Multivariable Calculus

Michael Taylor

Math. Dept., UNC
E-mail address: met@math.unc.edu

2010 *Mathematics Subject Classification.* 26B05, 26B10, 26B12, 26B15, 26B20

Key words and phrases. real numbers, complex numbers, Euclidean space, compact spaces, Cauchy sequences, continuous function, power series, derivative, mean value theorem, Riemann integral, fundamental theorem of calculus, arclength, exponential function, logarithm, trigonometric functions, Euler's formula, multiple integrals, surfaces, surface area, differential forms, Stokes theorem, Gauss theorem, Green theorem

Contents

Preface

This is a text for students with a background in one-variable calculus, who are ready to tackle calculus in several variables. It is designed for the honors section of Math 233 at the University of North Carolina.

Chapter 1 presents a brisk review of the basics of calculus in one variable: definitions and elementary properties of the derivative and integral, the fundamental theorem of calculus, and power series. One might skim over this introductory chapter to see if a refresher is needed for some of this material.

Multivariable calculus is done on multidimensional spaces. Chapter 2 introduces algebraic tools useful for this study. We start with a section on n-dimensional Euclidean space $\mathbb{R}^n$, which has a linear structure, and also a geometric structure, coming from a dot product. We then take up more general vector spaces, linear transforms between them, matrix representations of such transformations, and determinants of square matrices. This chapter concludes with a treatment of the cross product on $\mathbb{R}^3$.

Chapter 3 studies curves in Euclidean spaces, i.e., functions $\gamma : I \to \mathbb{R}^n$, where I is an interval in the real line. We derive a formula for the arclength of a C^1 curve, and discuss parametrizing the curve by arclength. Applying these considerations to the unit circle centered at the origin in $\mathbb{R}^2$ gives rise to the trigonometric functions $\cos t$ and $\sin t$. In §3.2 we define the exponential function, first for real arguments ($e^t,\ t \in \mathbb{R}$) and then for complex arguments ($e^z,\ z \in \mathbb{C}$). An examination of the planar curve $\gamma(t) = e^{it}$ shows that this is a unit speed parametrization of the unit circle, leading to the Euler identity,

$$e^{it} = \cos t + i \sin t.$$

Sections 3.3–3.4 present results on curvature, first for planar curves and then for curves in $\mathbb{R}^3$, where also the notion of torsion arises. Calculations of curves with given curvature (and, in 3D, torsion) lead to an extension of the exponential function, the matrix exponential.

Chapter 4 studies the derivative of functions of several variables. We define the derivative of a function $F : \mathcal{O} \to \mathbb{R}^m$, at a point x in an open set $\mathcal{O} \subset \mathbb{R}^n$, as a linear map from $\mathbb{R}^n$ to $\mathbb{R}^m$, relate it to partial derivatives, and establish basic properties, such as the chain rule, for the derivative of a composite map $G \circ F$. We consider higher-order derivatives, and study power series for functions on a set $\mathcal{O} \subset \mathbb{R}^n$. We also establish the inverse function theorem, stating that (when $m = n$) the map F has a smooth inverse on a neighborhood of x provided its derivative $DF(x)$ is an invertible linear transformation on $\mathbb{R}^n$ (i.e., its determinant is not 0).

Chapter 5 develops integral calculus on domains $S \subset \mathbb{R}^n$. We start with $S = R$, an n-dimensional rectangle, and give a definition parallel to that of the one-dimensional integral in §1.2. However, it is important to be able to integrate over other sets, such as balls and other regions with curvy boundaries. We can take a rectangle R containing S, and extend our function f from S to R by zero. This operation makes it crucial that we be able to integrate discontinuous functions, and the n-dimensional Riemann integral is up to the task. We show that a bounded function on R is Riemann integrable provided its set of points of discontinuity has negligible size, in an appropriate sense. Other important results covered in this chapter include a change of variable formula for multiple integrals and the reduction of multiple integrals to iterated integrals. We also treat integrals over all of $\mathbb{R}^n$ and integrals of a class of unbounded functions.

Chapter 6 extends the calculus developed in the previous two chapters from open sets in Euclidean space to smooth surfaces in $\mathbb{R}^n$. These surfaces have coordinate charts, used to perform differential and integral calculus. The inverse function theorem from Chapter 4 and the change of variable formula for integrals established in Chapter 5 play a crucial role in doing this analysis on surfaces. Applications include computation of areas of n-dimensional spheres, and integration over groups of rotations in $\mathbb{R}^n$ (averaging over rotations). In §6.3 we derive important integral identities known as theorems of Gauss, Green, and Stokes. In §6.4 we introduce a class of objects more general than surfaces, called manifolds, on which to develop differential and integral calculus.

This text concludes with several appendices, providing supplementary material that the reader might find useful. Appendix A develops the real numbers, as ideal limits of Cauchy sequences of rational numbers. It establishes key properties, such as completeness of the real number line $\mathbb{R}$ and compactness of nonempty, closed, bounded subsets of $\mathbb{R}$, which lie behind many phenomena important for calculus, such as existence of maxima and minima, and the intermediate value theorem. It also presents some basic results on the set $\mathbb{C}$ of complex numbers.

Appendix B has some basic results on continuous functions, and on sequences and series of such functions, including a sufficient condition for uniform convergence of such a series, known as the Weierstrass M-test, useful for our treatment of power series.

Appendix C has material on linear algebra, supplementing that presented in Chapter 2. This includes a treatment of inner product spaces, of which $\mathbb{R}^n$ with the dot product is the example in Chapter 2. It discusses eigenvalues and eigenvectors of linear transformations on finite-dimensional vector spaces, of particular use in the characterization of various types of critical points of a smooth, real-valued function on a region of $\mathbb{R}^n$, in terms of its matrix of second-order partial derivatives. There is a treatment of matrix norms, including the operator norm and the Hilbert-Schmidt norm. This appendix also treats the matrix exponential, extending the treatment of exponentials of complex numbers given in Chapter 3. This exponential is given as an infinite series, and material on matrix norms plays a role in showing the series converges.

Appendix D discusses functions $f : \mathcal{O} \to \mathbb{C}$ (with $\mathcal{O}$ open in $\mathbb{C}$) that are C^1 and complex differentiable, using Green's theorem to establish results known as the Cauchy integral theorem and the Cauchy integral formula, and a corollary known as Liouville's theorem. This appendix provides an introduction to the area of complex analysis, which the reader can pursue further in other texts, such as [**17**]. One application appears in the following appendix.

Appendix E treats the fundamental theorem of algebra, which says that each nonconstant polynomial $p(z) = a_n z^n + \cdots + a_0$, with coefficients $a_j \in \mathbb{C}$, vanishes for some $z \in \mathbb{C}$. Two proofs are given, one elementary, and the other using Liouville's theorem, established in §D.3.

We follow this introduction with a record of some standard notation that will be used throughout this text.

Some basic notation

$\mathbb{R}$ is the set of real numbers.

$\mathbb{C}$ is the set of complex numbers.

$\mathbb{Z}$ is the set of integers.

$\mathbb{Z}^+$ is the set of integers ≥ 0.

$\mathbb{N}$ is the set of integers ≥ 1 (the "natural numbers").

$\mathbb{Q}$ is the set of rational numbers.

$x \in \mathbb{R}$ means x is an element of $\mathbb{R}$, i.e., x is a real number.

(a, b) denotes the set of $x \in \mathbb{R}$ such that $a < x < b$.

$[a, b]$ denotes the set of $x \in \mathbb{R}$ such that $a \leq x \leq b$.

$\{x \in \mathbb{R} : a \leq x \leq b\}$ denotes the set of x in $\mathbb{R}$ such that $a \leq x \leq b$.

$[a, b) = \{x \in \mathbb{R} : a \leq x < b\}$ and $(a, b] = \{x \in \mathbb{R} : a < x \leq b\}$.

$\overline{z} = x - iy$ if $z = x + iy \in \mathbb{C},\ x, y \in \mathbb{R}$.

$\overline{\Omega}$ denotes the closure of the set Ω.

$f : A \to B$ denotes that the function f takes points in the set A to points in B. One also says f maps A to B.

$x \to x_0$ means the variable x tends to the limit x_0.

$f(x) = O(x)$ means $f(x)/x$ is bounded. Similarly $g(\varepsilon) = O(\varepsilon^k)$ means $g(\varepsilon)/\varepsilon^k$ is bounded.

$f(x) = o(x)$ as $x \to 0$ (resp., $x \to \infty$) means $f(x)/x \to 0$ as x tends to the specified limit.

$S = \sup_n |a_n|$ means S is the smallest real number that satisfies $S \geq |a_n|$ for all n. If there is no such real number then we take $S = +\infty$.

$\limsup_{k\to\infty} |a_k| = \lim_{n\to\infty} \big(\sup_{k\geq n} |a_k|\big)$.

Chapter 1

Basic one variable calculus

This first chapter provides a review of calculus for functions of one real variable. Students with a solid background in one-variable calculus might skim this quickly, to make sure they are familiar with the basic concepts. If there are any gaps, this chapter is designed to fill them in.

Section 1.1 introduces the derivative, establishes basic identities like the product rule and the chain rule, and also obtains some important theoretical results, such as the Mean Value Theorem and the Inverse Function Theorem. One application of the latter is the study of $x^{1/n}$, for $x > 0$, which leads more generally to x^r, for $x > 0$ and $r \in \mathbb{Q}$. (Extension to $r \in \mathbb{R}$, and beyond, is given in §3.2.)

Section 1.2 brings in the integral, more precisely the Riemann integral. A major result is the Fundamental Theorem of Calculus, whose proof makes essential use of the Mean Value Theorem. Another topic is the change of variable formula for integrals (treated in some exercises).

In §1.3 we treat power series. Topics include term by term differentiation of power series, and formulas for the remainder when a power series is truncated. An application of such remainder formulas is made to the study of convergence of the power series about $x = 0$ of $(1 - x)^b$.

In §1.4 we give a natural extension of the Riemann integral from the class of bounded (Riemann integrable) functions to a class of unbounded "integrable" functions. The treatment here is perhaps a desirable alternative to discussions one sees of "improper integrals."

1.1. The derivative

Consider a function f, defined on an interval $(a,b) \subset \mathbb{R}$, taking values in $\mathbb{R}$ or $\mathbb{C}$. Given $x \in (a,b)$, we say f is differentiable at x, with derivative $f'(x)$, provided

$$\lim_{h\to 0} \frac{f(x+h)-f(x)}{h} = f'(x). \tag{1.1.1}$$

We also use the notation

$$\frac{df}{dx}(x) = f'(x). \tag{1.1.2}$$

A characterization equivalent to (1.1.1) is

$$f(x+h) = f(x) + f'(x)h + r(x,h), \quad r(x,h) = o(h), \tag{1.1.3}$$

where

$$r(x,h) = o(h) \ \text{ means } \ \frac{r(x,h)}{h} \to 0 \ \text{ as } \ h \to 0. \tag{1.1.4}$$

Clearly if f is differentiable at x then it is continuous at x. We say f is differentiable on (a,b) provided it is differentiable at each point of (a,b). If also g is defined on (a,b) and differentiable at x, we have

$$\frac{d}{dx}(f+g)(x) = f'(x) + g'(x). \tag{1.1.5}$$

We also have the following *product rule*:

$$\frac{d}{dx}(fg)(x) = f'(x)g(x) + f(x)g'(x). \tag{1.1.6}$$

To prove (1.1.6), note that

$$\begin{aligned}&\frac{f(x+h)g(x+h)-f(x)g(x)}{h}\\ &= \frac{f(x+h)-f(x)}{h} g(x) + f(x+h)\frac{g(x+h)-g(x)}{h}.\end{aligned}$$

We can use the product rule to show inductively that

$$\frac{d}{dx} x^n = n x^{n-1}, \tag{1.1.7}$$

for all $n \in \mathbb{N}$. In fact, this is immediate from (1.1.1) if $n = 1$. Given that it holds for $n = k$, we have

$$\begin{aligned}\frac{d}{dx} x^{k+1} = \frac{d}{dx}(x\, x^k) &= \frac{dx}{dx} x^k + x \frac{d}{dx} x^k \\ &= x^k + k x^k \\ &= (k+1) x^k,\end{aligned}$$

completing the induction. We also have

$$\frac{1}{h}\Bigl(\frac{1}{x+h} - \frac{1}{x}\Bigr) = -\frac{1}{x(x+h)} \to -\frac{1}{x^2}, \quad \text{as } \ h \to 0,$$

for $x \neq 0$, hence

$$\frac{d}{dx}\frac{1}{x} = -\frac{1}{x^2}, \quad \text{if } \ x \neq 0. \tag{1.1.8}$$

From here, we can extend (1.1.7) from $n \in \mathbb{N}$ to all $n \in \mathbb{Z}$ (requiring $x \neq 0$ if $n < 0$).

A similar inductive argument yields

$$\frac{d}{dx} f(x)^n = n f(x)^{n-1} f'(x), \tag{1.1.9}$$

for $n \in \mathbb{N}$, and more generally for $n \in \mathbb{Z}$ (requiring $f(x) \neq 0$ if $n < 0$).

Going further, we have the following *chain rule*. Suppose $f : (a, b) \to (\alpha, \beta)$ is differentiable at x and $g : (\alpha, \beta) \to \mathbb{R}$ (or $\mathbb{C}$) is differentiable at $y = f(x)$. Form $G = g \circ f$, i.e., $G(x) = g(f(x))$. We claim

$$G = g \circ f \Longrightarrow G'(x) = g'(f(x)) f'(x). \tag{1.1.10}$$

To see this, write

$$\begin{aligned} G(x+h) &= g(f(x+h)) \\ &= g(f(x) + f'(x)h + r_f(x,h)) \\ &= g(f(x)) + g'(f(x))(f'(x)h + r_f(x,h)) \\ &\quad + r_g(f(x), f'(x)h + r_f(x,h)). \end{aligned} \tag{1.1.11}$$

Here,

$$\frac{r_f(x,h)}{h} \longrightarrow 0 \quad \text{as} \quad h \to 0,$$

and also

$$\frac{r_g(f(x), f'(x)h + r_f(x,h))}{h} \longrightarrow 0, \quad \text{as } h \to 0,$$

so the analogue of (1.1.3) applies.

The derivative has the following important connection to maxima and minima.

Proposition 1.1.1. *Let $f : (a, b) \to \mathbb{R}$. Suppose $x \in (a, b)$ and*

$$f(x) \geq f(y), \quad \forall\, y \in (a, b). \tag{1.1.12}$$

If f is differentiable at x, then $f'(x) = 0$. The same conclusion holds if $f(x) \leq f(y)$ for all $y \in (a, b)$.

Proof. Given (1.1.12), we have

$$\frac{f(x+h) - f(x)}{h} \leq 0, \quad \forall\, h \in (0, b - x), \tag{1.1.13}$$

and

$$\frac{f(x+h) - f(x)}{h} \geq 0, \quad \forall\, h \in (a - x, 0). \tag{1.1.14}$$

If f is differentiable at x, both (1.1.13) and (1.1.14) must converge to $f'(x)$ as $h \to 0$, so we simultaneously have $f'(x) \leq 0$ and $f'(x) \geq 0$. □

We next establish a key result known as the *Mean Value Theorem*. See Figure 1.1.1 for an illustration.

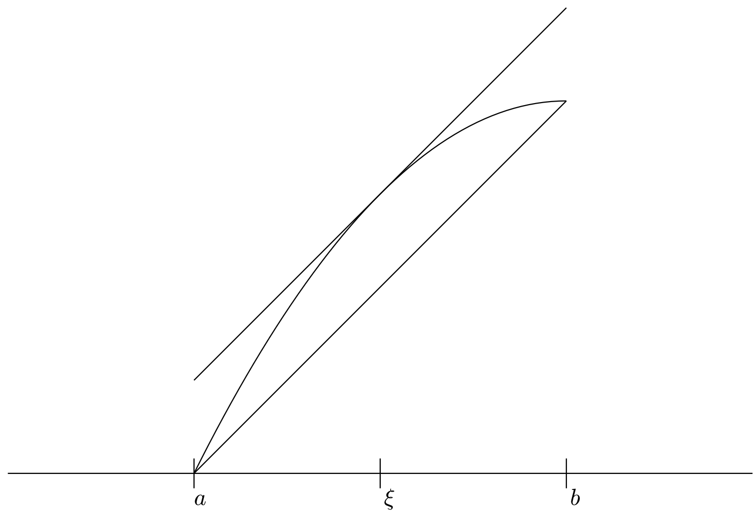

Figure 1.1.1. Illustration of the Mean Value Theorem

Theorem 1.1.2. *Let $f : [a, b] \to \mathbb{R}$. Assume f is continuous on $[a, b]$ and differentiable on (a, b). Then there exists $\xi \in (a, b)$ such that*

$$f'(\xi) = \frac{f(b) - f(a)}{b - a}. \tag{1.1.15}$$

Proof. Let $g(x) = f(x) - \kappa(x - a)$, where κ denotes the right side of (1.1.15). Then $g(a) = g(b)$. The result (1.1.15) is equivalent to the assertion that

$$g'(\xi) = 0 \tag{1.1.16}$$

for some $\xi \in (a, b)$. Now g is continuous on the compact set $[a, b]$, so it assumes both a maximum and a minimum on this set. If g has a maximum at a point $\xi \in (a, b)$, then (1.1.16) follows from Proposition 1.1.1. If not, the maximum must be $g(a) = g(b)$, and then g must assume a minimum at some point $\xi \in (a, b)$. Again Proposition 1.1.1 implies (1.1.16). □

We use the Mean Value Theorem to produce a criterion for constructing the inverse of a function. Let

$$f : [a, b] \longrightarrow \mathbb{R}, \quad f(a) = \alpha, \quad f(b) = \beta. \tag{1.1.17}$$

Assume f is continuous on $[a, b]$, differentiable on (a, b), and

$$0 < \gamma_0 \le f'(x) \le \gamma_1 < \infty, \quad \forall\, x \in (a, b). \tag{1.1.18}$$

We can apply Theorem 1.1.2 to f, restricted ot the interval $[x_1, x_2] \subset [a, b]$, to get

$$\gamma_0 \le \frac{f(x_2) - f(x_1)}{x_2 - x_1} \le \gamma_1, \quad \text{if } \ a \le x_1 < x_2 \le b, \tag{1.1.19}$$

or

$$\gamma_0(x_2 - x_1) \le f(x_2) - f(x_1) \le \gamma_1(x_2 - x_1). \tag{1.1.20}$$

It follows that

$$f : [a, b] \longrightarrow [\alpha, \beta] \ \text{ is one-to-one.} \tag{1.1.21}$$

The intermediate value theorem implies $f : [a, b] \to [\alpha, \beta]$ is onto. Consequently f has an inverse

$$g : [\alpha, \beta] \longrightarrow [a, b], \quad g(f(x)) = x, \quad f(g(y)) = y, \tag{1.1.22}$$

and (1.1.19) implies

$$\frac{1}{\gamma_1} \le \frac{g(y_2) - g(y_1)}{y_2 - y_1} \le \frac{1}{\gamma_0}, \quad \text{if } \ \alpha \le y_1 < y_2 \le \beta. \tag{1.1.23}$$

The following result is known as the *Inverse Function Theorem.*

Theorem 1.1.3. *If f is continuous on $[a, b]$ and differentiable on (a, b), and* (1.1.17)–(1.1.18) *hold, then its inverse $g : [\alpha, \beta] \to [a, b]$ is differentiable on (α, β), and*

$$g'(y) = \frac{1}{f'(x)}, \quad \textit{for } \ y = f(x) \in (\alpha, \beta). \tag{1.1.24}$$

The same conclusion holds if in place of (1.1.18) *we have*

$$-\gamma_1 \le f'(x) \le -\gamma_0 < 0, \quad \forall\, x \in (a, b), \tag{1.1.25}$$

except that then $\beta < \alpha$.

Proof. Fix $y \in (\alpha, \beta)$, and let $x = g(y)$, so $y = f(x)$. To say that f is differentiable at x is to say

$$\lim_{\xi \to x} \frac{f(x) - f(\xi)}{x - \xi} = f'(x). \tag{1.1.26}$$

Now take $\eta = f(\xi)$, so $\xi = g(\eta)$, and note from (1.1.19) that

$$\xi \to x \Longleftrightarrow \eta \to y. \tag{1.1.27}$$

Hence, by (1.1.18)–(1.1.19) and (1.1.23), we have

$$\lim_{\eta \to y} \frac{g(y) - g(\eta)}{y - \eta} = \frac{1}{f'(x)}, \tag{1.1.28}$$

which proves (1.1.24). □

Remark. If one knew that g were differentiable, as well as f, then the identity (1.1.24) would follow by differentiating $g(f(x)) = x$, applying the chain rule. However, an additional argument, such as given above, is necessary to guarantee that g is differentiable.

Theorem 1.1.3 applies to the functions

$$p_n(x) = x^n, \quad n \in \mathbb{N}. \tag{1.1.29}$$

By (1.1.7), $p_n'(x) > 0$ for $x > 0$, so (1.1.18) holds when $0 < a < b < \infty$. We can take $a \searrow 0$ and $b \nearrow \infty$ and see that

$$p_n : (0,\infty) \longrightarrow (0,\infty) \text{ is invertible,} \tag{1.1.30}$$

with differentiable inverse $q_n : (0,\infty) \to (0,\infty)$. We use the notation

$$x^{1/n} = q_n(x), \quad x > 0, \tag{1.1.31}$$

so, given $n \in \mathbb{N}$,

$$x > 0 \Longrightarrow x = x^{1/n} \cdots x^{1/n}, \quad (n \text{ factors}). \tag{1.1.32}$$

Given $m \in \mathbb{Z}, n \in \mathbb{N}$, we can set

$$x^{m/n} = (x^{1/n})^m, \quad x > 0, \tag{1.1.33}$$

and verify that $(x^{1/kn})^{km} = (x^{1/n})^m$ for $k \in \mathbb{N}$. Thus we have x^r defined for all $r \in \mathbb{Q}$, when $x > 0$. We have

$$x^{r+s} = x^r x^s, \quad \text{for } x > 0,\ r, s \in \mathbb{Q}. \tag{1.1.34}$$

Applying (1.1.24) to $f(x) = x^n$ and $g(y) = y^{1/n}$, we have

$$\frac{d}{dy} y^{1/n} = \frac{1}{nx^{n-1}}, \quad y = x^n,\ x > 0. \tag{1.1.35}$$

Now $x^{n-1} = y/x = y^{1-1/n}$, so we get

$$\frac{d}{dy} y^r = r y^{r-1}, \quad y > 0, \tag{1.1.36}$$

when $r = 1/n$. Putting this together with (1.1.9) (with m in place of n), we get (1.1.36) for all $r = m/n \in \mathbb{Q}$.

The definition of x^r for $x > 0$ and the identity (1.1.36) can be extended to all $r \in \mathbb{R}$, with some more work. We will find a neat way to do this in §3.2.

We recall another common notation, namely

$$\sqrt{x} = x^{1/2}, \quad x > 0. \tag{1.1.37}$$

Then (1.1.36) yields

$$\frac{d}{dx}\sqrt{x} = \frac{1}{2\sqrt{x}}. \tag{1.1.38}$$

In regard to this, note that, if we consider

$$\frac{\sqrt{x+h} - \sqrt{x}}{h}, \tag{1.1.39}$$

we can multiply numerator and denominator by $\sqrt{x+h} + \sqrt{x}$, to get

$$\frac{1}{\sqrt{x+h} + \sqrt{x}}, \tag{1.1.40}$$

whose convergence to the right side of (1.1.38) for $x > 0$ is equivalent to the statement that

$$\lim_{h \to 0} \sqrt{x+h} = \sqrt{x}, \tag{1.1.41}$$

i.e., to the continuity of $x \mapsto \sqrt{x}$ on $(0, \infty)$. Such continuity is a consequence of the fact that, for $0 < a < b < \infty$, $n = 2$,

$$p_n : [a, b] \longrightarrow [a^n, b^n] \tag{1.1.42}$$

is continuous, one-to-one, and onto, so, by the compactness of $[a, b]$, its inverse is continuous. Thus we have an alternative derivation of (1.1.38).

If $I \subset \mathbb{R}$ is an interval and $f : I \to \mathbb{R}$ (or $\mathbb{C}$), we say $f \in C^1(I)$ if f is differentiable on I and f' is continuous on I. If f' is in turn differentiable, we have the second derivative of f:

$$\frac{d^2 f}{dx^2}(x) = f''(x) = \frac{d}{dx} f'(x). \tag{1.1.43}$$

If f' is differentiable on I and f'' is continuous on I, we say $f \in C^2(I)$. Inductively, we can define higher order derivatives of f, $f^{(k)}$, also denoted $d^k f/dx^k$. Here, $f^{(1)} = f'$, $f^{(2)} = f''$, and if $f^{(k)}$ is differentiable,

$$f^{(k+1)}(x) = \frac{d}{dx} f^{(k)}(x). \tag{1.1.44}$$

If $f^{(k)}$ is continuous on I, we say $f \in C^k(I)$.

Sometimes we will run into functions of more than one variable, and will want to differentiate with respect to each one of them. For example, if $f(x, y)$ is defined for (x, y) in an open set in $\mathbb{R}^2$, we define partial derivatives,

$$\begin{aligned} \frac{\partial f}{\partial x}(x, y) &= \lim_{h \to 0} \frac{f(x + h, y) - f(x, y)}{h}, \\ \frac{\partial f}{\partial y}(x, y) &= \lim_{h \to 0} \frac{f(x, y + h) - f(x, y)}{h}. \end{aligned} \tag{1.1.45}$$

We will not need any more than the definition here. A serious study of the derivative of a function of several variables is given in Chapter 4.

We end this section with some results on the significance of the second derivative.

Proposition 1.1.4. *Assume f is differentiable on (a, b), $x_0 \in (a, b)$, and $f'(x_0) = 0$. Assume f' is differentiable at x_0 and $f''(x_0) > 0$. Then there exists $\delta > 0$ such that*

$$f(x_0) < f(x) \quad \text{for all} \quad x \in (x_0 - \delta, x_0 + \delta) \setminus \{x_0\}. \tag{1.1.46}$$

We say f has a local minimum at x_0.

Proof. Since

$$f''(x_0) = \lim_{h \to 0} \frac{f'(x_0 + h) - f'(x_0)}{h}, \tag{1.1.47}$$

the assertion that $f''(x_0) > 0$ implies that there exists $\delta > 0$ such that the right side of (1.1.47) is > 0 for all nonzero $h \in [-\delta, \delta]$. Hence

$$\begin{aligned} -\delta \le h < 0 &\Longrightarrow f'(x_0 + h) < 0, \\ 0 < h \le \delta &\Longrightarrow f'(x_0 + h) > 0. \end{aligned} \tag{1.1.48}$$

This plus the mean value theorem imply (1.1.46). ∎

Remark. Similarly,

$$f''(x_0) < 0 \Longrightarrow f \text{ has a local maximum at } x_0. \tag{1.1.49}$$

These two facts constitute the second derivative test for local maxima and local minima.

Let us now assume that f and f' are differentiable on (a, b), so f'' is defined at each point of (a, b). Let us further assume

$$f''(x) > 0, \quad \forall\, x \in (a, b). \tag{1.1.50}$$

The mean value theorem, applied to f', yields

$$a < x_0 < x_1 < b \Longrightarrow f'(x_0) < f'(x_1). \tag{1.1.51}$$

Here is another interesting property.

Proposition 1.1.5. *If* (1.1.50) *holds and* $a < x_0 < x_1 < b$, *then*

$$f(sx_0 + (1-s)x_1) < sf(x_0) + (1-s)f(x_1), \quad \forall\, s \in (0, 1). \tag{1.1.52}$$

Proof. For $s \in [0, 1]$, set

$$g(s) = sf(x_0) + (1-s)f(x_1) - f(sx_0 + (1-s)x_1). \tag{1.1.53}$$

The result (1.1.52) is equivalent to

$$g(s) > 0 \ \text{ for } \ 0 < s < 1. \tag{1.1.54}$$

Note that

$$g(0) = g(1) = 0. \tag{1.1.55}$$

If (1.1.54) fails, g must assume a minimum at some point $s_0 \in (0, 1)$. At such a point, $g'(s_0) = 0$. A computation gives

$$g'(s) = f(x_0) - f(x_0) - (x_0 - x_1)f'(sx_0 + (1-s)x_1),$$

and hence

$$g''(s) = -(x_0 - x_1)^2 f''(sx_0 + (1-s)x_1). \tag{1.1.56}$$

Thus (1.1.50) $\Rightarrow g''(s_0) < 0$. Then (1.1.49) $\Rightarrow g$ has a local *maximum* at s_0. This contradiction establishes (1.1.54), hence (1.1.52). □

Remark. The result (1.1.52) implies that, whenever $a < x_0 < x_1 < b$, the graph of $y = f(x)$ over $[x_0, x_1]$ lies *below* the chord, i.e., the line segment from $(x_0, f(x_0))$ to $(x_1, f(x_1))$ in $\mathbb{R}^2$. We say f is *convex*.

Exercises

For Exercises 1–3, compute the derivative of each of the following functions. Specify where each of these derivatives are defined.

1. $\sqrt{1+x^2}$,
2. $(x^2+x^3)^{-4}$,
3. $\sqrt{1+x^2}/(x^2+x^3)^4$.

4. Let $f : [0,\infty) \to \mathbb{R}$ be a C^2 function satisfying

$$f(x) > 0, \quad f'(x) > 0, \quad f''(x) < 0, \quad \text{for} \ \ x > 0. \tag{1.1.57}$$

Show that

$$x, y > 0 \Longrightarrow f(x+y) < f(x) + f(y). \tag{1.1.58}$$

5. Apply Exercise 4 to

$$f(x) = \frac{x}{1+x}. \tag{1.1.59}$$

Give a direct proof that (1.1.58) holds for f in (1.1.59), without using calculus.

6. If $f : I \to \mathbb{R}^n$, we define $f'(x)$ just as in (1.1.1). If $f(x) = (f_1(x), \dots, f_n(x))$, then f is differentiable at x if and only if each component f_j is, and

$$f'(x) = (f_1'(x), \dots, f_n'(x)).$$

Parallel to (1.1.6), show that if $g : I \to \mathbb{R}^n$, then the dot product satisfies

$$\frac{d}{dx} f(x) \cdot g(x) = f'(x) \cdot g(x) + f(x) \cdot g'(x).$$

7. Establish the following variant of Proposition 1.1.5. Suppose (1.1.50) is weakened to

$$f''(x) \geq 0, \quad \forall\, x \in (a,b). \tag{1.1.60}$$

Show that, in place of (1.1.52), one has

$$f(sx_0 + (1-s)x_1) \leq sf(x_0) + (1-s)f(x_1), \quad \forall\, s \in (0,1). \tag{1.1.61}$$

Hint. Consider $f_\varepsilon(x) = f(x) + \varepsilon x^2$.

8. The following is called the generalized mean value theorem. Let f and g be continuous on $[a,b]$ and differentiable on (a,b). Then there exists $\xi \in (a,b)$ such that

$$[f(b) - f(a)]g'(\xi) = [g(b) - g(a)]f'(\xi).$$

Show that this follows from the mean value theorem, applied to

$$h(x) = [f(b) - f(a)]g(x) - [g(b) - g(a)]f(x).$$

9. Take $f : [a, b] \to [\alpha, \beta]$ and $g : [\alpha, \beta] \to [a, b]$ as in the setting of the Inverse Function Theorem, Theorem 1.3. Write (1.1.24) as

$$g'(y) = \frac{1}{f'(g(y))}, \quad y \in (\alpha, \beta). \tag{1.1.62}$$

Show that

$$f \in C^1((a, b)) \Longrightarrow g \in C^1((\alpha, \beta)),$$

i.e., the right side of (1.1.62) is continuous on (α, β). Show inductively that, for $k \in \mathbb{N}$,

$$f \in C^k((a, b)) \Longrightarrow g \in C^k((\alpha, \beta)).$$

Example. Show that if $f \in C^2((a, b))$, then (having shown that $g \in C^1$) the right side of (1.1.62) is C^1 and hence

$$g''(y) = -\frac{1}{f'(g(y))^2} f''(g(y))g'(y).$$

10. Let $I \subset \mathbb{R}$ be an open interval and $f : I \to \mathbb{R}$ differentiable. (Do *not* assume f' is continuous.) Assume $a, b \in I,\ a < b$, and

$$f'(a) < u < f'(b).$$

Show that there exists $\xi \in (a, b)$ such that $f'(\xi) = u$.
Hint. Reduce to the case $u = 0$, so $f'(a) < 0 < f'(b)$. Show that then $f|_{[a,b]}$ has a minimum at a point $\xi \in (a, b)$.

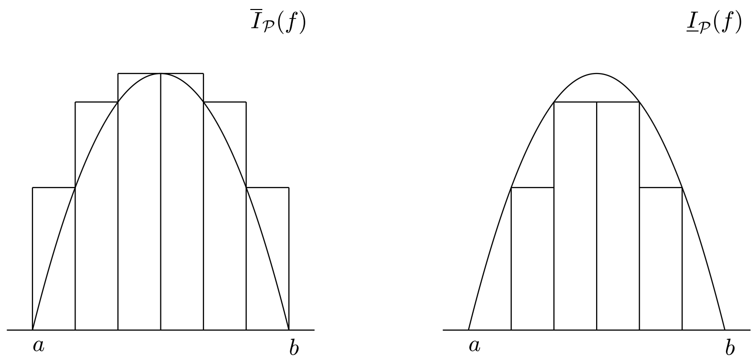

Figure 1.2.1. Upper and lower sums associated to a partition

1.2. The integral

In this section, we introduce the Riemann version of the integral, and relate it to the derivative. We will define the Riemann integral of a bounded function over an interval $I = [a, b]$ on the real line. For now, we assume f is real valued. To start, we partition I into smaller intervals. A *partition* $\mathcal{P}$ of I is a finite collection of subintervals $\{J_k : 0 \le k \le N\}$, disjoint except for their endpoints, whose union is I. We can order the J_k so that $J_k = [x_k, x_{k+1}]$, where

$$x_0 < x_1 < \cdots < x_N < x_{N+1}, \quad x_0 = a, \ x_{N+1} = b. \tag{1.2.1}$$

We call the points x_k the *endpoints* of $\mathcal{P}$. We set

$$\ell(J_k) = x_{k+1} - x_k, \quad \text{maxsize}\,(\mathcal{P}) = \max_{0 \le k \le N} \ \ell(J_k) \tag{1.2.2}$$

We then set

$$\begin{aligned} \overline{I}_{\mathcal{P}}(f) &= \sum_k \sup_{J_k} f(x)\, \ell(J_k), \\ \underline{I}_{\mathcal{P}}(f) &= \sum_k \inf_{J_k} f(x)\, \ell(J_k). \end{aligned} \tag{1.2.3}$$

Here,

$$\sup_{J_k} \ f(x) = \sup \ f(J_k), \quad \inf_{J_k} \ f(x) = \inf \ f(J_k),$$

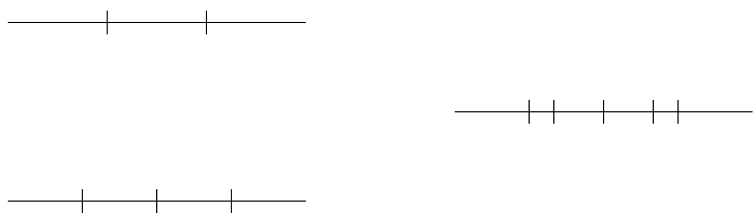

Figure 1.2.2. Two partitions, $\mathcal{P}_j$, of I and a common refinement, $\mathcal{Q} \succ \mathcal{P}_j$

and we note that if $S \subset \mathbb{R}$ is bounded, $\sup S$ and $\inf S$ are defined in §A.2; cf. (A.2.38) and (A.2.51). We call $\overline{I}_{\mathcal{P}}(f)$ and $\underline{I}_{\mathcal{P}}(f)$ respectively the upper sum and lower sum of f, associated to the partition $\mathcal{P}$. See Figure 1.2.1 for an illustration. Note that $\underline{I}_{\mathcal{P}}(f) \le \overline{I}_{\mathcal{P}}(f)$. These quantities should approximate the Riemann integral of f, if the partition $\mathcal{P}$ is sufficiently "fine."

To be more precise, if $\mathcal{P}$ and $\mathcal{Q}$ are two partitions of I, we say $\mathcal{Q}$ *refines* $\mathcal{P}$, and write $\mathcal{Q} \succ \mathcal{P}$, if $\mathcal{Q}$ is formed by partitioning each interval in $\mathcal{P}$. Equivalently, $\mathcal{Q} \succ \mathcal{P}$ if and only if all the endpoints of $\mathcal{P}$ are also endpoints of $\mathcal{Q}$. It is easy to see that any two partitions have a common refinement; just take the union of their endpoints, to form a new partition. See Figure 1.2.2. Note also that refining a partition lowers the upper sum of f and raises its lower sum:

$$\mathcal{Q} \succ \mathcal{P} \Longrightarrow \overline{I}_{\mathcal{Q}}(f) \le \overline{I}_{\mathcal{P}}(f), \text{ and } \underline{I}_{\mathcal{Q}}(f) \ge \underline{I}_{\mathcal{P}}(f). \tag{1.2.4}$$

Consequently, if $\mathcal{P}_j$ are any two partitions and $\mathcal{Q}$ is a common refinement, we have

$$\underline{I}_{\mathcal{P}_1}(f) \le \underline{I}_{\mathcal{Q}}(f) \le \overline{I}_{\mathcal{Q}}(f) \le \overline{I}_{\mathcal{P}_2}(f). \tag{1.2.5}$$

Now, whenever $f : I \to \mathbb{R}$ is bounded, the following quantities are well defined:

$$\overline{I}(f) = \inf_{\mathcal{P} \in \Pi(I)} \overline{I}_{\mathcal{P}}(f), \quad \underline{I}(f) = \sup_{\mathcal{P} \in \Pi(I)} \underline{I}_{\mathcal{P}}(f), \tag{1.2.6}$$

where $\Pi(I)$ is the set of all partitions of I. We call $\underline{I}(f)$ the lower integral of f and $\overline{I}(f)$ its upper integral. Clearly, by (1.2.5), $\underline{I}(f) \le \overline{I}(f)$. We then say that f is

Riemann integrable provided $\overline{I}(f) = \underline{I}(f)$, and in such a case, we set

$$\int_a^b f(x)\,dx = \int_I f(x)\,dx = \overline{I}(f) = \underline{I}(f). \tag{1.2.7}$$

We will denote the set of Riemann integrable functions on I by $\mathcal{R}(I)$.

We derive some basic properties of the Riemann integral.

Proposition 1.2.1. *If $f,\ g \in \mathcal{R}(I)$, then $f + g \in \mathcal{R}(I)$, and*

$$\int_I (f+g)\,dx = \int_I f\,dx + \int_I g\,dx. \tag{1.2.8}$$

Proof. If J_k is any subinterval of I, then

$$\sup_{J_k}(f+g) \le \sup_{J_k} f + \sup_{J_k} g, \quad \text{and} \quad \inf_{J_k}(f+g) \ge \inf_{J_k} f + \inf_{J_k} g,$$

so, for any partition $\mathcal{P}$, we have $\overline{I}_{\mathcal{P}}(f+g) \le \overline{I}_{\mathcal{P}}(f) + \overline{I}_{\mathcal{P}}(g)$. Also, using common refinements, we can *simultaneously* approximate $\overline{I}(f)$ and $\overline{I}(g)$ by $\overline{I}_{\mathcal{P}}(f)$ and $\overline{I}_{\mathcal{P}}(g)$, and ditto for $\overline{I}(f+g)$. Thus the characterization (1.2.6) implies $\overline{I}(f+g) \le \overline{I}(f) + \overline{I}(g)$. A parallel argument implies $\underline{I}(f+g) \ge \underline{I}(f) + \underline{I}(g)$, and the proposition follows. □

Next, there is a fair supply of Riemann integrable functions.

Proposition 1.2.2. *If f is continuous on I, then f is Riemann integrable.*

Proof. Any continuous function on a compact interval is bounded and uniformly continuous (see Propositions A.3.5 and B.1.3). Let $\omega(\delta)$ be a modulus of continuity for f, so

$$|x-y| \le \delta \Longrightarrow |f(x)-f(y)| \le \omega(\delta), \quad \omega(\delta) \to 0 \text{ as } \delta \to 0. \tag{1.2.9}$$

Then

$$\operatorname{maxsize}(\mathcal{P}) \le \delta \Longrightarrow \overline{I}_{\mathcal{P}}(f) - \underline{I}_{\mathcal{P}}(f) \le \omega(\delta)\cdot \ell(I), \tag{1.2.10}$$

which yields the proposition. □

We denote the set of continuous functions on I by $C(I)$. Thus Proposition 1.2.2 says

$$C(I) \subset \mathcal{R}(I).$$

The proof of Proposition 1.2.2 provides a criterion on a partition guaranteeing that $\overline{I}_{\mathcal{P}}(f)$ and $\underline{I}_{\mathcal{P}}(f)$ are close to $\int_I f\,dx$ when f is continuous. We produce an extension, giving a condition under which $\overline{I}_{\mathcal{P}}(f)$ and $\overline{I}(f)$ are close, and $\underline{I}_{\mathcal{P}}(f)$ and $\underline{I}(f)$ are close, given f bounded on I. Given a partition $\mathcal{P}_0$ of I, set

$$\operatorname{minsize}(\mathcal{P}_0) = \min\{\ell(J_k) : J_k \in \mathcal{P}_0\}. \tag{1.2.11}$$

Lemma 1.2.3. *Let $\mathcal{P}$ and $\mathcal{Q}$ be two partitions of I. Assume*

$$\text{maxsize}(\mathcal{P}) \le \frac{1}{k}\ \text{minsize}(\mathcal{Q}). \tag{1.2.12}$$

Let $|f| \le M$ on I. Then

$$\begin{aligned} \overline{I}_{\mathcal{P}}(f) &\le \overline{I}_{\mathcal{Q}}(f) + \frac{2M}{k}\ell(I), \\ \underline{I}_{\mathcal{P}}(f) &\ge \underline{I}_{\mathcal{Q}}(f) - \frac{2M}{k}\ell(I). \end{aligned} \tag{1.2.13}$$

Proof. Let $\mathcal{P}_1$ denote the minimal common refinement of $\mathcal{P}$ and $\mathcal{Q}$. Consider on the one hand those intervals in $\mathcal{P}$ that are contained in intervals in $\mathcal{Q}$ and on the other hand those intervals in $\mathcal{P}$ that are *not* contained in intervals in $\mathcal{Q}$. Each interval of the first type is also an interval in $\mathcal{P}_1$. Each interval of the second type gets partitioned, to yield two intervals in $\mathcal{P}_1$. Denote by $\mathcal{P}_1^b$ the collection of such divided intervals. By (1.2.12), the lengths of the intervals in $\mathcal{P}_1^b$ sum to $\le \ell(I)/k$. It follows that

$$|\overline{I}_{\mathcal{P}}(f) - \overline{I}_{\mathcal{P}_1}(f)| \le \sum_{J\in\mathcal{P}_1^b} 2M\ell(J) \le 2M\frac{\ell(I)}{k},$$

and similarly $|\underline{I}_{\mathcal{P}}(f) - \underline{I}_{\mathcal{P}_1}(f)| \le 2M\ell(I)/k$. Therefore

$$\overline{I}_{\mathcal{P}}(f) \le \overline{I}_{\mathcal{P}_1}(f) + \frac{2M}{k}\ell(I), \quad \underline{I}_{\mathcal{P}}(f) \ge \underline{I}_{\mathcal{P}_1}(f) - \frac{2M}{k}\ell(I).$$

Since also $\overline{I}_{\mathcal{P}_1}(f) \le \overline{I}_{\mathcal{Q}}(f)$ and $\underline{I}_{\mathcal{P}_1}(f) \ge \underline{I}_{\mathcal{Q}}(f)$, we obtain (1.2.13). □

The following consequence is sometimes called Darboux's Theorem.

Theorem 1.2.4. *Let $\mathcal{P}_\nu$ be a sequence of partitions of I into ν intervals $J_{\nu k}$, $1 \le k \le \nu$, such that*

$$\text{maxsize}(\mathcal{P}_\nu) \longrightarrow 0.$$

If $f : I \to \mathbb{R}$ is bounded, then

$$\overline{I}_{\mathcal{P}_\nu}(f) \to \overline{I}(f) \quad \text{and} \quad \underline{I}_{\mathcal{P}_\nu}(f) \to \underline{I}(f). \tag{1.2.14}$$

Consequently,

$$f \in \mathcal{R}(I) \Longleftrightarrow \overline{I}(f) = \lim_{\nu\to\infty} \sum_{k=1}^{\nu} f(\xi_{\nu k})\ell(J_{\nu k}), \tag{1.2.15}$$

for arbitrary $\xi_{\nu k} \in J_{\nu k}$, in which case the limit is $\int_I f\,dx$.

Proof. As before, assume $|f| \le M$. Pick $\varepsilon > 0$. Let $\mathcal{Q}$ be a partition such that

$$\begin{aligned} \overline{I}(f) &\le \overline{I}_{\mathcal{Q}}(f) \le \overline{I}(f) + \varepsilon, \\ \underline{I}(f) &\ge \underline{I}_{\mathcal{Q}}(f) \ge \underline{I}(f) - \varepsilon. \end{aligned}$$

Now pick N such that

$$\nu \ge N \Longrightarrow \text{maxsize}\,\mathcal{P}_\nu \le \varepsilon\,\text{minsize}\,\mathcal{Q}.$$

Lemma 2.3 yields, for $\nu \ge N$,

$$\begin{aligned} \overline{I}_{\mathcal{P}_\nu}(f) &\le \overline{I}_{\mathcal{Q}}(f) + 2M\ell(I)\varepsilon, \\ \underline{I}_{\mathcal{P}_\nu}(f) &\ge \underline{I}_{\mathcal{Q}}(f) - 2M\ell(I)\varepsilon. \end{aligned}$$

Hence, for $\nu \geq N$,

$$\overline{I}(f) \leq \overline{I}_{\mathcal{P}_\nu}(f) \leq \overline{I}(f) + [2M\ell(I) + 1]\varepsilon,$$
$$\underline{I}(f) \geq \underline{I}_{\mathcal{P}_\nu}(f) \geq \underline{I}(f) - [2M\ell(I) + 1]\varepsilon.$$

This proves (1.2.14). □

Remark. The sums on the right side of (1.2.15) are called Riemann sums, approximating $\int_I f\,dx$ (when f is Riemann integrable).

Remark. A second proof of Proposition 1.2.1 can readily be deduced from Theorem 1.2.4.

One should be warned that, once such a specific choice of $\mathcal{P}_\nu$ and $\xi_{\nu k}$ has been made, the limit on the right side of (1.2.15) might exist for a bounded function f that is *not* Riemann integrable. This and other phenomena are illustrated by the following example of a function which is not Riemann integrable. For $x \in I$, set

$$\vartheta(x) = 1 \ \text{ if } \ x \in \mathbb{Q}, \quad \vartheta(x) = 0 \ \text{ if } \ x \notin \mathbb{Q}, \tag{1.2.16}$$

where $\mathbb{Q}$ is the set of *rational* numbers. Now every interval $J \subset I$ of positive length contains points in $\mathbb{Q}$ and points not in $\mathbb{Q}$, so for any partition $\mathcal{P}$ of I we have $\overline{I}_{\mathcal{P}}(\vartheta) = \ell(I)$ and $\underline{I}_{\mathcal{P}}(\vartheta) = 0$, hence

$$\overline{I}(\vartheta) = \ell(I), \quad \underline{I}(\vartheta) = 0. \tag{1.2.17}$$

Note that, if $\mathcal{P}_\nu$ is a partition of I into ν equal subintervals, then we could pick each $\xi_{\nu k}$ to be rational, in which case the limit on the right side of (1.2.15) would be $\ell(I)$, or we could pick each $\xi_{\nu k}$ to be irrational, in which case this limit would be zero. Alternatively, we could pick half of them to be rational and half to be irrational, and the limit would be $\frac{1}{2}\ell(I)$.

Associated to the Riemann integral is a notion of size of a set S, called *content*. If S is a subset of I, define the "characteristic function"

$$\chi_S(x) = 1 \ \text{ if } \ x \in S, \ 0 \ \text{ if } \ x \notin S. \tag{1.2.18}$$

We define "upper content" cont^+ and "lower content" cont^- by

$$\text{cont}^+(S) = \overline{I}(\chi_S), \quad \text{cont}^-(S) = \underline{I}(\chi_S). \tag{1.2.19}$$

We say S "has content," or "is contented" if these quantities are equal, which happens if and only if $\chi_S \in \mathcal{R}(I)$, in which case the common value of $\text{cont}^+(S)$ and $\text{cont}^-(S)$ is

$$m(S) = \int_I \chi_S(x)\,dx. \tag{1.2.20}$$

It is easy to see that

$$\text{cont}^+(S) = \inf\Big\{\sum_{k=1}^N \ell(J_k) : S \subset J_1 \cup \cdots \cup J_N\Big\}, \tag{1.2.21}$$

where J_k are intervals. Here, we require S to be in the union of a *finite* collection of intervals.

There is a more sophisticated notion of the size of a subset of I, called Lebesgue measure. The key to the construction of Lebesgue measure is to cover a set S by a *countable* (either finite or *infinite*) set of intervals. The *outer measure* of $S \subset I$ is defined by

$$m^*(S) = \inf\Big\{\sum_{k\ge 1} \ell(J_k) : S \subset \bigcup_{k\ge 1} J_k\Big\}. \tag{1.2.22}$$

Here $\{J_k\}$ is a finite or countably infinite collection of intervals. Clearly

$$m^*(S) \le \text{cont}^+(S). \tag{1.2.23}$$

Note that, if $S = I \cap \mathbb{Q}$, then $\chi_S = \vartheta$, defined by (1.2.16). In this case it is easy to see that $\text{cont}^+(S) = \ell(I)$, but $m^*(S) = 0$. In fact, (1.2.22) readily yields the following:

$$S \ \text{ countable } \Longrightarrow m^*(S) = 0. \tag{1.2.24}$$

We point out that we can require the intervals J_k in (1.2.22) to be open. Consequently, since each open cover of a compact set has a finite subcover,

$$S \ \text{ compact } \Longrightarrow m^*(S) = \text{cont}^+(S). \tag{1.2.25}$$

See the material at the end of this section for a generalization of Proposition 1.2.2, giving a sufficient condition for a bounded function to be Riemann integrable on I, in terms of the upper content of its set of discontinuities, in Proposition 1.2.11, and then, in Proposition 1.2.12, a refinement, replacing upper content by outer measure.

It is useful to note that $\int_I f\,dx$ is additive in I, in the following sense.

Proposition 1.2.5. *If* $a < b < c,\ f : [a,c] \to \mathbb{R},\ f_1 = f\big|_{[a,b]},\ f_2 = f\big|_{[b,c]}$, *then*

$$f \in \mathcal{R}\big([a,c]\big) \Longleftrightarrow f_1 \in \mathcal{R}\big([a,b]\big) \quad \text{and} \quad f_2 \in \mathcal{R}\big([b,c]\big), \tag{1.2.26}$$

and, if this holds,

$$\int_a^c f\,dx = \int_a^b f_1\ dx + \int_b^c f_2\,dx. \tag{1.2.27}$$

Proof. Since any partition of $[a,c]$ has a refinement for which b is an endpoint, we may as well consider a partition $\mathcal{P} = \mathcal{P}_1 \cup \mathcal{P}_2$, where $\mathcal{P}_1$ is a partition of $[a,b]$ and $\mathcal{P}_2$ is a partition of $[b,c]$. Then

$$\overline{I}_{\mathcal{P}}(f) = \overline{I}_{\mathcal{P}_1}(f_1) + \overline{I}_{\mathcal{P}_2}(f_2), \quad \underline{I}_{\mathcal{P}}(f) = \underline{I}_{\mathcal{P}_1}(f_1) + \underline{I}_{\mathcal{P}_2}(f_2), \tag{1.2.28}$$

so

$$\overline{I}_{\mathcal{P}}(f) - \underline{I}_{\mathcal{P}}(f) = \big\{\overline{I}_{\mathcal{P}_1}(f_1) - \underline{I}_{\mathcal{P}_1}(f_1)\big\} + \big\{\overline{I}_{\mathcal{P}_2}(f_2) - \underline{I}_{\mathcal{P}_2}(f_2)\big\}. \tag{1.2.29}$$

Since both terms in braces in (1.2.29) are ≥ 0, we have equivalence in (1.2.26). Then (1.2.27) follows from (1.2.28) upon taking finer and finer partitions, and passing to the limit. ☐

Let $I = [a, b]$. If $f \in \mathcal{R}(I)$, then $f \in \mathcal{R}([a, x])$ for all $x \in [a, b]$, and we can consider the function

$$g(x) = \int_a^x f(t)\,dt. \tag{1.2.30}$$

If $a \le x_0 \le x_1 \le b$, then

$$g(x_1) - g(x_0) = \int_{x_0}^{x_1} f(t)\,dt, \tag{1.2.31}$$

so, if $|f| \le M$,

$$|g(x_1) - g(x_0)| \le M|x_1 - x_0|. \tag{1.2.32}$$

In other words, if $f \in \mathcal{R}(I)$, then g is Lipschitz continuous on I.

Recall from §1.1 that a function $g : (a, b) \to \mathbb{R}$ is said to be differentiable at $x \in (a, b)$ provided there exists the limit

$$\lim_{h \to 0} \frac{1}{h}\big[g(x + h) - g(x)\big] = g'(x). \tag{1.2.33}$$

When such a limit exists, $g'(x)$, also denoted dg/dx, is called the derivative of g at x. Clearly g is continuous wherever it is differentiable.

The next result is part of the Fundamental Theorem of Calculus.

Theorem 1.2.6. *If $f \in C([a, b])$, then the function g, defined by* (1.2.30)*, is differentiable at each point $x \in (a, b)$, and*

$$g'(x) = f(x). \tag{1.2.34}$$

Proof. Parallel to (1.2.31), we have, for $h > 0$,

$$\frac{1}{h}\big[g(x + h) - g(x)\big] = \frac{1}{h}\int_x^{x+h} f(t)\,dt. \tag{1.2.35}$$

If f is continuous at x, then, for any $\varepsilon > 0$, there exists $\delta > 0$ such that $|f(t) - f(x)| \le \varepsilon$ whenever $|t - x| \le \delta$. Thus the right side of (1.2.35) is within ε of $f(x)$ whenever $h \in (0, \delta]$. Thus the desired limit exists as $h \searrow 0$. A similar argument treats $h \nearrow 0$. □

The next result is the rest of the Fundamental Theorem of Calculus.

Theorem 1.2.7. *If G is differentiable and $G'(x)$ is continuous on $[a, b]$, then*

$$\int_a^b G'(t)\,dt = G(b) - G(a). \tag{1.2.36}$$

Proof. Consider the function

$$g(x) = \int_a^x G'(t)\,dt. \tag{1.2.37}$$

We have $g \in C([a, b])$, $g(a) = 0$, and, by Theorem 1.2.6,

$$g'(x) = G'(x), \quad \forall\, x \in (a, b). \tag{1.2.38}$$

Thus $f(x) = g(x) - G(x)$ is continuous on $[a, b]$, and

$$f'(x) = 0, \quad \forall\, x \in (a, b). \tag{1.2.39}$$

We claim that (1.2.39) implies f is *constant* on $[a,b]$. Granted this, since $f(a) = g(a) - G(a) = -G(a)$, we have $f(x) = -G(a)$ for all $x \in [a,b]$, so the integral (1.2.37) is equal to $G(x) - G(a)$ for all $x \in [a,b]$. Taking $x = b$ yields (1.2.36). $\square$

The fact that (1.2.39) implies f is constant on $[a,b]$ is a consequence of the Mean Value Theorem. This was established in §1.1; see Theorem 1.1.2. We repeat the statement here.

Theorem 1.2.8. *Let $f : [a,\beta] \to \mathbb{R}$ be continuous, and assume f is differentiable on (a,β). Then $\exists\ \xi \in (a,\beta)$ such that*

$$f'(\xi) = \frac{f(\beta) - f(a)}{\beta - a}. \tag{1.2.40}$$

Now, to see that (1.2.39) implies f is constant on $[a,b]$, if not, $\exists\ \beta \in (a,b]$ such that $f(\beta) \neq f(a)$. Then just apply Theorem 1.2.8 to f on $[a,\beta]$. This completes the proof of Theorem 1.2.7.

We now extend Theorems 1.2.6–1.2.7 to the setting of Riemann integrable functions.

Proposition 1.2.9. *Let $f \in \mathcal{R}([a,b])$, and define g by (1.2.28). If $x \in [a,b]$ and f is continuous at x, then g is differentiable at x, and $g'(x) = f(x)$.*

The proof is identical to that of Theorem 1.2.6.

Proposition 1.2.10. *Assume G is differentiable on $[a,b]$ and $G' \in \mathcal{R}([a,b])$. Then (1.2.36) holds.*

Proof. We have

$$\begin{aligned} G(b) - G(a) &= \sum_{k=0}^{n-1} \Bigl[G\Bigl(a + (b-a)\frac{k+1}{n}\Bigr) - G\Bigl(a + (b-a)\frac{k}{n}\Bigr)\Bigr] \\ &= \frac{b-a}{n} \sum_{k=0}^{n-1} G'(\xi_{kn}), \end{aligned} \tag{1.2.41}$$

for some ξ_{kn} satisfying

$$a + (b-a)\frac{k}{n} < \xi_{kn} < a + (b-a)\frac{k+1}{n}, \tag{1.2.42}$$

as a consequence of the Mean Value Theorem. Given $G' \in \mathcal{R}([a,b])$, Darboux's theorem (Theorem 1.2.4) implies that as $n \to \infty$ one gets $G(b) - G(a) = \int_a^b G'(t)\,dt$. $\square$

Note that the beautiful symmetry in Theorems 1.2.6–1.2.7 is not preserved in Propositions 1.2.9–1.2.10. The hypothesis of Proposition 1.2.10 requires G to be differentiable at each $x \in [a,b]$, but the conclusion of Proposition 1.2.9 does not yield differentiability at all points. For this reason, we regard Propositions 1.2.9–1.2.10 as less "fundamental" than Theorems 1.2.6–1.2.7. There are more satisfactory extensions of the fundamental theorem of calculus, involving the Lebesgue integral, and a more subtle notion of the "derivative" of a non-smooth function. For this, we can point the reader to Chapters 10-11 of the text [**14**], *Measure Theory and Integration*.

So far, we have dealt with integration of real valued functions. If $f : I \to \mathbb{C}$, we set $f = f_1 + if_2$ with $f_j : I \to \mathbb{R}$ and say $f \in \mathcal{R}(I)$ if and only if f_1 and f_2 are in $\mathcal{R}(I)$. Then

$$\int_I f\,dx = \int_I f_1\,dx + i\int_I f_2\,dx. \tag{1.2.43}$$

There are straightforward extensions of Propositions 1.2.5–1.2.10 to complex valued functions. Similar comments apply to functions $f : I \to \mathbb{R}^n$.

Complementary results on Riemann integrability

Here we provide a condition, more general then Proposition 1.2.2, which guarantees Riemann integrability.

Proposition 1.2.11. *Let $f : I \to \mathbb{R}$ be a bounded function, with $I = [a,b]$. Suppose that the set S of points of discontinuity of f has the property*

$$\text{cont}^+(S) = 0. \tag{1.2.44}$$

Then $f \in \mathcal{R}(I)$.

Proof. Say $|f(x)| \le M$. Take $\varepsilon > 0$. As in (1.2.21), take intervals $J_1, \dots, J_N$ such that $S \subset J_1 \cup \cdots \cup J_N$ and $\sum_{k=1}^N \ell(J_k) < \varepsilon$. In fact, fatten each J_k such that S is contained in the interior of this collection of intervals. Consider a partition $\mathcal{P}_0$ of I, whose intervals include $J_1, \dots, J_N$, amongst others, which we label $I_1, \dots, I_K$. Now f is continuous on each interval I_ν, so, subdividing each I_ν as necessary, hence refining $\mathcal{P}_0$ to a partition $\mathcal{P}_1$, we arrange that $\sup f - \inf f < \varepsilon$ on each such subdivided interval. Denote these subdivided intervals $I'_1, \dots, I'_L$. It readily follows that

$$\begin{aligned} 0 \le \overline{I}_{\mathcal{P}_1}(f) - \underline{I}_{\mathcal{P}_1}(f) &< \sum_{k=1}^N 2M\ell(J_k) + \sum_{k=1}^L \varepsilon\ell(I'_k) \\ &< 2\varepsilon M + \varepsilon\ell(I). \end{aligned} \tag{1.2.45}$$

Since ε can be taken arbitrarily small, this establishes that $f \in \mathcal{R}(I)$. □

With a little more effort, we can establish the following result, which, in light of (1.2.23), is a bit sharper than Proposition 1.2.11.

Proposition 1.2.12. *In the setting of Proposition 1.2.11, if we replace* (1.2.44) *by*

$$m^*(S) = 0, \tag{1.2.46}$$

we still conclude that $f \in \mathcal{R}(I)$.

Proof. As before, we assume $|f(x)| \le M$ and pick $\varepsilon > 0$. This time, take a countable collection of open intervals $\{J_k\}$ such that $S \subset \cup_{k\ge1} J_k$ and $\sum_{k\ge1} \ell(J_k) < \varepsilon$. Now f is continuous at each $p \in I \setminus S$, so there exists an interval K_p, open (in I), containing p, such that $\sup_{K_p} f - \inf_{K_p} f < \varepsilon$. Now $\{J_k : k \in \mathbb{N}\} \cup \{K_p : p \in I \setminus S\}$ is an open cover of I, so it has a finite subcover, which we denote $\{J_1, \dots, J_N, K_1, \dots, K_M\}$. We have

$$\sum_{k=1}^{N} \ell(J_k) < \varepsilon, \quad \text{and} \quad \sup_{K_j} f - \inf_{K_j} f < \varepsilon, \quad \forall\, j \in \{1, \dots, M\}. \tag{1.2.47}$$

Let $\mathcal{P}$ be the partition of I obtained by taking the union of all the endpoints of J_k and K_j in (1.2.47). Let us write

$$\begin{aligned} \mathcal{P} &= \{L_k : 0 \le k \le \mu\} \\ &= \Big(\bigcup_{k\in\mathcal{A}} L_k\Big) \cup \Big(\bigcup_{k\in\mathcal{B}} L_k\Big), \end{aligned}$$

where we say $k \in \mathcal{A}$ provided L_k is contained in an interval of the form K_j for some $j \in \{1, \dots, M\}$, as in (1.2.47). Consequently, if $k \in \mathcal{B}$, then $L_k \subset J_\ell$ for some $\ell \in \{1, \dots, N\}$, so

$$\bigcup_{k\in\mathcal{B}} L_k \subset \bigcup_{\ell=1}^{N} J_\ell. \tag{1.2.48}$$

We therefore have

$$\sum_{k\in\mathcal{B}} \ell(L_k) < \varepsilon, \quad \text{and} \quad \sup_{L_j} f - \inf_{L_j} f < \varepsilon, \quad \forall\, j \in \mathcal{A}. \tag{1.2.49}$$

It follows that

$$\begin{aligned} 0 \le \overline{I}_{\mathcal{P}}(f) - \underline{I}_{\mathcal{P}}(f) &< \sum_{k\in\mathcal{B}} 2M\ell(L_k) + \sum_{j\in\mathcal{A}} \varepsilon\ell(L_j) \\ &< 2\varepsilon M + \varepsilon\ell(I). \end{aligned} \tag{1.2.50}$$

Since ε can be taken arbitrarily small, this establishes that $f \in \mathcal{R}(I)$. □

Remark. Proposition 1.2.12 is part of the sharp result that a bounded function f on $I = [a, b]$ is Riemann integrable if and *only if* its set S of points of discontinuity satisfies (1.2.46). Standard books on measure theory, including [**7**] and [**14**], establish this.

We give an example of a function to which Proposition 1.2.11 applies, and then an example for which Proposition 1.2.11 fails to apply, but Proposition 1.2.12 applies.

Example 1. Let $I = [0, 1]$. Define $f : I \to \mathbb{R}$ by

$$\begin{aligned} &f(0) = 0, \\ &f(x) = (-1)^j \ \text{ for } x \in (2^{-(j+1)}, 2^{-j}], \ j \ge 0. \end{aligned} \tag{1.2.51}$$

Then $|f| \le 1$ and the set of points of discontinuity of f is

$$S = \{0\} \cup \{2^{-j} : j \ge 1\}. \tag{1.2.52}$$

It is easy to see that $\text{cont}^+ S = 0$. Hence $f \in \mathcal{R}(I)$.

See Exercises 16-17 below for a more elaborate example to which Proposition 1.2.11 applies.

Example 2. Again $I = [0,1]$. Define $f : I \to \mathbb{R}$ by

$$\begin{aligned} f(x) = 0 \quad &\text{if } \ x \notin \mathbb{Q}, \\ \frac{1}{n} \quad &\text{if } \ x = \frac{m}{n}, \text{ in lowest terms.} \end{aligned} \tag{1.2.53}$$

Then $|f| \le 1$ and the set of points of discontinuity of f is

$$S = I \cap \mathbb{Q}. \tag{1.2.54}$$

As we have seen below (1.2.23), $\text{cont}^+ S = 1$, so Proposition 1.2.11 does not apply. Nevertheless, it is fairly easy to see directly that

$$\overline{I}(f) = \underline{I}(f) = 0, \quad \text{so} \ \ f \in \mathcal{R}(I). \tag{1.2.55}$$

In fact, given $\varepsilon > 0$, $f \ge \varepsilon$ only on a finite set, hence

$$\overline{I}(f) \le \varepsilon, \quad \forall\, \varepsilon > 0. \tag{1.2.56}$$

As indicated below (1.2.23), (1.2.46) does apply to this function, so Proposition 1.2.12 applies. Example 2 is illustrative of the following general phenomenon, which is worth recording.

Corollary 1.2.13. *If $f : I \to \mathbb{R}$ is bounded and its set S of points of discontinuity is countable, then $f \in \mathcal{R}(I)$.*

Proof. By virtue of (1.2.24), Proposition 1.2.12 applies. □

Here is another useful sufficient condition condition for Riemann integrability.

Proposition 1.2.14. *If $f : I \to \mathbb{R}$ is bounded and monotone, then $f \in \mathcal{R}(I)$.*

Proof. It suffices to consider the case that f is monotone increasing. Let $\mathcal{P}_N = \{J_k : 1 \le k \le N\}$ be the partition of I into N intervals of equal length. Note that $\sup_{J_k} f \le \inf_{J_{k+1}} f$. Hence

$$\begin{aligned} \overline{I}_{\mathcal{P}_N}(f) &\le \sum_{k=1}^{N-1} (\inf_{J_{k+1}} f)\ell(J_k) + (\sup_{J_N} f)\ell(J_N) \\ &\le \underline{I}_{\mathcal{P}_N}(f) + 2M\frac{\ell(I)}{N}, \end{aligned} \tag{1.2.57}$$

if $|f| \le M$. Taking $N \to \infty$, we deduce from Theorem 1.2.4 that $\overline{I}(f) \le \underline{I}(f)$, which proves $f \in \mathcal{R}(I)$. □

Remark. It can be shown that if f is monotone, then its set of points of discontinuity is countable. Given this, Proposition 1.2.14 is also a consequence of Corollary 1.2.13.

By contrast, the function ϑ in (1.2.16) is discontinuous at each point of I.

We mention some alternative characterizations of $\overline{I}(f)$ and $\underline{I}(f)$, which can be useful. Given $I = [a, b]$, we say $g : I \to \mathbb{R}$ is *piecewise constant* on I (and write $g \in \mathrm{PK}(I)$) provided there exists a partition $\mathcal{P} = \{J_k\}$ of I such that g is constant on the interior of each interval J_k. Clearly $\mathrm{PK}(I) \subset \mathcal{R}(I)$. It is easy to see that, if $f : I \to \mathbb{R}$ is bounded,

$$\begin{aligned} \overline{I}(f) &= \inf\Big\{ \int_I f_1\, dx : f_1 \in \mathrm{PK}(I),\ f_1 \ge f \Big\}, \\ \underline{I}(f) &= \sup\Big\{ \int_I f_0\, dx : f_0 \in \mathrm{PK}(I),\ f_0 \le f \Big\}. \end{aligned} \tag{1.2.58}$$

Hence, given $f : I \to \mathbb{R}$ bounded,

$$\begin{aligned} f \in \mathcal{R}(I) \Leftrightarrow &\text{ for each } \varepsilon > 0,\ \exists f_0, f_1 \in \mathrm{PK}(I) \text{ such that} \\ &f_0 \le f \le f_1 \ \text{ and } \int_I (f_1 - f_0)\, dx < \varepsilon. \end{aligned} \tag{1.2.59}$$

This can be used to prove

$$f, g \in \mathcal{R}(I) \Longrightarrow fg \in \mathcal{R}(I), \tag{1.2.60}$$

via the fact that

$$f_j, g_j \in \mathrm{PK}(I) \Longrightarrow f_j g_j \in \mathrm{PK}(I). \tag{1.2.61}$$

In fact, we have the following, which can be used to prove (1.2.60), based on the identity

$$2fg = (f+g)^2 - f^2 - g^2.$$

Proposition 1.2.15. *Let $f \in \mathcal{R}(I)$, and assume $|f| \le M$. Let*

$$\varphi : [-M, M] \to \mathbb{R}$$

be continuous. Then $\varphi \circ f \in \mathcal{R}(I)$.

Proof. We proceed in steps.

Step 1. We can obtain φ as a uniform limit on $[-M, M]$ of a sequence φ_ν of continuous, piecewise linear functions. Then $\varphi_\nu \circ f \to \varphi \circ f$ uniformly on I. A uniform limit g of functions $g_\nu \in \mathcal{R}(I)$ is in $\mathcal{R}(I)$ (see Exercise 9). So it suffices to prove Proposition 1.2.15 when φ is continuous and piecewise linear.

Step 2. Given $\varphi : [-M, M] \to \mathbb{R}$ continuous and piecewise linear, it is an exercise to write $\varphi = \varphi_1 - \varphi_2$, with $\varphi_j : [-M, M] \to \mathbb{R}$ monotone, continuous, and piecewise linear. Now $\varphi_1 \circ f, \varphi_2 \circ f \in \mathcal{R}(I) \Rightarrow \varphi \circ f \in \mathcal{R}(I)$.

STEP 3. We now demonstrate Proposition 1.2.15 when $\varphi : [-M, M] \to \mathbb{R}$ is monotone and Lipschitz. By Step 2, this will suffice. So we assume

$$-M \le x_1 < x_2 \le M \Longrightarrow \varphi(x_1) \le \varphi(x_2) \ \text{ and } \ \varphi(x_2) - \varphi(x_1) \le L(x_2 - x_1),$$

for some $L < \infty$. Given $\varepsilon > 0$, pick $f_0, f_1 \in \mathrm{PK}(I)$, as in (2.59). Then

$$\varphi \circ f_0,\ \varphi \circ f_1 \in \mathrm{PK}(I), \quad \varphi \circ f_0 \le \varphi \circ f \le \varphi \circ f_1,$$

and

$$\int_I (\varphi \circ f_1 - \varphi \circ f_0)\, dx \le L \int_I (f_1 - f_0)\, dx \le L\varepsilon.$$

This proves $\varphi \circ f \in \mathcal{R}(I)$. □

For another characterization of $\mathcal{R}(I)$, we can deduce from (1.2.58) that, if $f : I \to \mathbb{R}$ is bounded,

$$\overline{I}(f) = \inf\Big\{ \int_I \varphi_1\, dx : \varphi_1 \in C(I),\ \varphi_1 \ge f \Big\},$$
$$\underline{I}(f) = \sup\Big\{ \int_I \varphi_0\, dx : \varphi_0 \in C(I),\ \varphi_0 \le f \Big\}, \tag{1.2.62}$$

and this leads to the following variant of (1.2.59).

Proposition 1.2.16. *Given $f : I \to \mathbb{R}$ bounded, $f \in \mathcal{R}(I)$ if and only if for each $\varepsilon > 0$, there exist $\varphi_0, \varphi_1 \in C(I)$ such that*

$$\varphi_0 \le f \le \varphi_1, \quad \textit{and} \quad \int_I (\varphi_1 - \varphi_0)\, dx < \varepsilon. \tag{1.2.63}$$

Exercises

1. Let $c > 0$ and let $f : [ac, bc] \to \mathbb{R}$ be Riemann integrable. Working directly with the definition of integral, show that

$$\int_a^b f(cx)\, dx = \frac{1}{c} \int_{ac}^{bc} f(x)\, dx. \tag{1.2.64}$$

More generally, show that

$$\int_{a-d/c}^{b-d/c} f(cx + d)\, dx = \frac{1}{c} \int_{ac}^{bc} f(x)\, dx. \tag{1.2.65}$$

2. Let $f : I \times S \to \mathbb{R}$ be continuous, where $I = [a, b]$ and $S \subset \mathbb{R}^n$. Take $\varphi(y) =$

$\int_I f(x,y)\,dx$. Show that φ is continuous on S.
Hint. If $f_j : I \to \mathbb{R}$ are continuous and $|f_1(x) - f_2(x)| \le \delta$ on I, then

$$\Bigl| \int\limits_I f_1\,dx - \int\limits_I f_2\;dx \Bigr| \le \ell(I)\delta. \tag{1.2.66}$$

3. With f as in Exercise 2, suppose $g_j : S \to \mathbb{R}$ are continuous and $a \le g_0(y) < g_1(y) \le b$. Take $\varphi(y) = \int_{g_0(y)}^{g_1(y)} f(x,y)\,dx$. Show that φ is continuous on S.
Hint. Make a change of variables, linear in x, to reduce this to Exercise 2.

4. Let $\varphi : [a,b] \to [A,B]$ be C^1 on a neighborhood J of $[a,b]$, with $\varphi'(x) > 0$ for all $x \in [a,b]$. Assume $\varphi(a) = A,\ \varphi(b) = B$. Show that the identity

$$\int_A^B f(y)\,dy = \int_a^b f\bigl(\varphi(t)\bigr)\varphi'(t)\,dt, \tag{1.2.67}$$

for any $f \in C([A,B])$, follows from the chain rule and the Fundamental Theorem of Calculus. The identity (1.2.67) is called the *change of variable formula* for the integral.
Hint. Replace b by x, B by $\varphi(x)$, and differentiate.
Going further, using (1.2.62)–(1.2.63), show that $f \in \mathcal{R}([A,B]) \Rightarrow f \circ \varphi \in \mathcal{R}([a,b])$ and (1.2.67) holds. (This result contains that of Exercise 1.)

5. Show that, if f and g are C^1 on a neighborhood of $[a,b]$, then

$$\int_a^b f(s)g'(s)\,ds = -\int_a^b f'(s)g(s)\,ds + \bigl[f(b)g(b) - f(a)g(a)\bigr]. \tag{1.2.68}$$

This transformation of integrals is called "integration by parts."

6. Let $f : (-a,a) \to \mathbb{R}$ be a C^{j+1} function. Show that, for $x \in (-a,a)$,

$$f(x) = f(0) + f'(0)x + \frac{f''(0)}{2}x^2 + \cdots + \frac{f^{(j)}(0)}{j!}x^j + R_j(x), \tag{1.2.69}$$

where

$$R_j(x) = \int_0^x \frac{(x-s)^j}{j!} f^{(j+1)}(s)\,ds \tag{1.2.70}$$

This is Taylor's formula with remainder.
Hint. Use induction. If (1.2.69)–(1.2.70) holds for $0 \le j \le k$, show that it holds for $j = k+1$, by showing that

$$\int_0^x \frac{(x-s)^k}{k!} f^{(k+1)}(s)\,ds = \frac{f^{(k+1)}(0)}{(k+1)!}x^{k+1} + \int_0^x \frac{(x-s)^{k+1}}{(k+1)!} f^{(k+2)}(s)\,ds. \tag{1.2.71}$$

To establish this, use the integration by parts formula (1.2.68), with $f(s)$ replaced by $f^{(k+1)}(s)$, and with appropriate $g(s)$. Note that another presentation of (1.2.70) is

$$R_j(x) = \frac{x^{j+1}}{(j+1)!} \int_0^1 f^{(j+1)}\Bigl(\bigl(1 - t^{1/(j+1)}\bigr)x\Bigr)\,dt. \tag{1.2.72}$$

For another demonstration of (1.2.70), see the proof of Proposition 1.3.4.

7. Assume $f : (-a, a) \to \mathbb{R}$ is a C^j function. Show that, for $x \in (-a, a)$, (1.2.69) holds, with

$$R_j(x) = \frac{1}{(j-1)!} \int_0^x (x-s)^{j-1} \big[f^{(j)}(s) - f^{(j)}(0) \big]\, ds. \tag{1.2.73}$$

Hint. Apply (1.2.70) with j replaced by $j-1$. Add and subtract $f^{(j)}(0)$ to the factor $f^{(j)}(s)$ in the resulting integrand.

8. Given $I = [a, b]$, show that

$$f, g \in \mathcal{R}(I) \Longrightarrow fg \in \mathcal{R}(I), \tag{1.2.74}$$

as advertised in (1.2.60).

9. Assume $f_k \in \mathcal{R}(I)$ and $f_k \to f$ uniformly on I. Prove that $f \in \mathcal{R}(I)$ and

$$\int_I f_k \, dx \longrightarrow \int_I f \, dx. \tag{1.2.75}$$

10. Given $I = [a, b]$, $I_\varepsilon = [a + \varepsilon, b - \varepsilon]$, assume $f_k \in \mathcal{R}(I)$, $|f_k| \le M$ on I for all k, and

$$f_k \longrightarrow f \quad \text{uniformly on} \quad I_\varepsilon, \tag{1.2.76}$$

for all $\varepsilon \in (0, (b-a)/2)$. Prove that $f \in \mathcal{R}(I)$ and (1.2.75) holds.

11. Use the fundamental theorem of calculus and results of §1.1 to compute

$$\int_a^b x^r \, dx, \quad r \in \mathbb{Q} \setminus \{-1\}, \tag{1.2.77}$$

where $-\infty < a < b < \infty$ if $r \in \mathbb{N}$ and $0 < a < b < \infty$ if $r \notin \mathbb{N}$. See §3.2 for (1.2.77) with $r = -1$ (and also for general $r \in \mathbb{R}$, even $r \in \mathbb{C}$).

12. Use the change of variable result of Exercise 4 to compute

$$\int_0^1 x\sqrt{1 + x^2}\, dx. \tag{1.2.78}$$

13. We say $f \in \mathcal{R}(\mathbb{R})$ provided $f|_{[k,k+1]} \in \mathcal{R}([k, k+1])$ for each $k \in \mathbb{Z}$, and

$$\sum_{k=-\infty}^{\infty} \int_k^{k+1} |f(x)|\, dx < \infty. \tag{1.2.79}$$

If $f \in \mathcal{R}(\mathbb{R})$, we set

$$\int_{-\infty}^{\infty} f(x)\, dx = \lim_{k \to \infty} \int_{-k}^{k} f(x)\, dx. \tag{1.2.80}$$

Formulate and demonstrate basic properties of the integral over $\mathbb{R}$ of elements of $\mathcal{R}(\mathbb{R})$.

14. This exercise discusses the integral test for absolute convergence of an infinite series, which goes as follows. Let f be a positive, monotonically decreasing, continuous function on $[0,\infty)$, and suppose $|a_k| = f(k)$. Then

$$\sum_{k=0}^{\infty} |a_k| < \infty \Longleftrightarrow \int_0^\infty f(x)\,dx < \infty.$$

Prove this.
Hint. Use

$$\sum_{k=1}^{N} |a_k| \le \int_0^N f(x)\,dx \le \sum_{k=0}^{N-1} |a_k|.$$

15. Use the integral test to show that, if $p > 0$,

$$\sum_{k=1}^{\infty} \frac{1}{k^p} < \infty \Longleftrightarrow p > 1.$$

NOTE. Compare Exercise 7 in §A.2. (For now, $p \in \mathbb{Q}^+$. Results of §3.2 allow one to take $p \in \mathbb{R}^+$.) *Hint.* Use Exercise 11 to evaluate $I_N(p) = \int_1^N x^{-p}\,dx$, for $p \neq -1$, and let $N \to \infty$. See if you can show $\int_1^\infty x^{-1}\,dx = \infty$ without knowing about $\log N$. *Subhint.* Show that $\int_1^2 x^{-1}\,dx = \int_N^{2N} x^{-1}\,dx$.

In Exercises 16–17, $\mathcal{C} \subset [a,b]$ is the Cantor set introduced in the exercises for §A.3. As in (A.3.24), $\mathcal{C} = \cap_{j\ge 0}\mathcal{C}_j$.

16. Show that $\operatorname{cont}^+ \mathcal{C}_j = (2/3)^j(b-a)$, and conclude that

$$\operatorname{cont}^+ \mathcal{C} = 0.$$

17. Define $f : [a,b] \to \mathbb{R}$ as follows. We call an interval of length $3^{-j}(b-a)$, omitted in passing from $\mathcal{C}_{j-1}$ to $\mathcal{C}_j$, a "j-interval." Set

$$\begin{aligned} f(x) = 0, &\quad \text{if } \ x \in \mathcal{C}, \\ (-1)^j, &\quad \text{if } x \text{ belongs to a } j\text{-interval}. \end{aligned}$$

Show that the set of discontinuities of f is $\mathcal{C}$. Hence Proposition 1.2.11 implies $f \in \mathcal{R}([a,b])$.

18. Let $f_k \in \mathcal{R}([a,b])$ and $f : [a,b] \to \mathbb{R}$ satisfy the following conditions.
(a) $|f_k| \le M < \infty, \quad \forall\, k,$
(b) $f_k(x) \longrightarrow f(x), \quad \forall\, x \in [a,b],$
(c) Given $\varepsilon > 0$, there exists $S_\varepsilon \subset [a,b]$ such that
$\operatorname{cont}^+ S_\varepsilon < \varepsilon, \quad$ and $f_k \to f$ uniformly on $[a,b] \setminus S_\varepsilon$.

Show that $f \in \mathcal{R}([a,b])$ and

$$\int_a^b f_k(x)\,dx \longrightarrow \int_a^b f(x)\,dx, \quad \text{as} \ \ k \to \infty.$$

REMARK. In the Lebesgue theory of integration, there is a stronger result, known as the Lebesgue dominated convergence theorem. See Exercises 12–14 in §1.4 for more on this.

19. Recall that one ingredient in the proof of Theorem 1.2.7 was that if $f : (a,b) \to \mathbb{R}$, then

$$f'(x) = 0 \text{ for all } x \in (a,b) \Longrightarrow f \text{ is constant on } (a,b). \tag{1.2.81}$$

Consider the following approach to proving (1.2.81), which avoids use of the Mean Value Theorem.
(a) Assume $a < x_0 < y_0 < b$ and $f(x_0) \neq f(y_0)$. Say $f(y_0) = f(x_0) + A(y_0 - x_0)$, and we may as well assume $A > 0$.
(b) Divide $I_0 = [x_0, y_0]$ into two equal intervals, $I_{0\ell}$ and I_{0r}, meeting at the midpoint $\xi_0 = (x_0 + y_0)/2$. Show that either

$$f(\xi_0) \geq f(x_0) + A(\xi_0 - x_0) \ \text{ or } \ f(y_0) \geq f(\xi_0) + A(y_0 - \xi_0).$$

Set $I_1 = I_{0\ell}$ if the former holds; otherwise, set $I_1 = I_{0r}$. Say $I_1 = [x_1, y_1]$.
(c) Inductively, having $I_k = [x_k, y_k]$, of length $2^{-k}(y_0 - x_0)$, divide it into two equal intervals, $I_{k\ell}$ and I_{kr}, meeting at the midpoint $\xi_k = (x_k + y_k)/2$. Show that either

$$f(\xi_k) \geq f(x_k) + A(\xi_k - x_k) \ \text{ or } \ f(y_k) \geq f(\xi_k) + A(y_k - \xi_k).$$

Set $I_{k+1} = I_{k\ell}$ if the former holds; otherwise set $I_{k+1} = I_{kr}$.
(d) Show that

$$x_k \nearrow x, \quad y_k \searrow x, \quad x \in [x_0, y_0],$$

and that, if f is differentiable at x, then $f'(x) \geq A$. Note that this contradicts the hypothesis that $f'(x) = 0$ for all $x \in (a,b)$.

1.3. Power series

We consider power series, of the form

$$f(z) = \sum_{k=0}^{\infty} a_k z^k, \tag{1.3.1}$$

with $a_k \in \mathbb{C}$. We begin with the following result.

Proposition 1.3.1. *If the series* (1.3.1) *converges for some $z_1 \neq 0$, then either this series is absolutely convergent for all $z \in \mathbb{C}$ or there is some $R \in (0, \infty)$ such that the series is absolutely convergent for $|z| < R$ and divergent for $|z| > R$. The series converges uniformly on*

$$D_S = \{z \in \mathbb{C} : |z| < S\}, \tag{1.3.2}$$

for each $S < R$, and f is continuous on D_R.

Proof. If (1.3.1) converges for $z = z_1 \neq 0$, then there exists $C < \infty$ such that

$$|a_k z_1^k| \leq C, \quad \forall\, k. \tag{1.3.3}$$

Hence, if $|z| = r|z_1|,\ r < 1$, we have

$$\sum_{k=0}^{\infty} |a_k z^k| \leq C \sum_{k=0}^{\infty} r^k = \frac{C}{1-r} < \infty, \tag{1.3.4}$$

the last identity being the classical geometric series computation. This yields the first part of Proposition 1.3.1.

To proceed, say the series (1.3.1) converges for all $|z| < R$, defining $f : D_R \to \mathbb{C}$. Take $S \in (0, R)$ and then pick $T \in (S, R)$. We know there exists $C < \infty$ such that $|a_k T^k| \leq C$ for all k. Hence

$$z \in D_S \Longrightarrow |a_k z^k| \leq C\Big(\frac{S}{T}\Big)^k. \tag{1.3.5}$$

Since

$$\sum_{k=0}^{\infty} \Big(\frac{S}{T}\Big)^k < \infty, \tag{1.3.6}$$

the Weierstrass M-test (Proposition B.3.1) applies, to yield uniform convergence on D_S. This yields continuity of f on D_S, for all $S < R$, hence continuity on D_R. ☐

The quantity R described above is called the radius of convergence of the power series (1.3.1). We now restrict attention to cases where $z = t \in \mathbb{R}$, and apply calculus to the study of such power series. We emphasize that we still allow the coefficients a_k to be complex numbers.

Proposition 1.3.2. *Assume $a_k \in \mathbb{C}$ and*

$$f(t) = \sum_{k=0}^{\infty} a_k t^k \tag{1.3.7}$$

converges for real t satisfying $|t| < R$. Then f is differentiable on the interval $-R < t < R$, and its derivative is given by

$$f'(t) = \sum_{k=1}^{\infty} k a_k t^{k-1}, \tag{1.3.8}$$

the latter series being absolutely convergent for $|t| < R$.

We first check absolute convergence of the series (1.3.8). Let $S < T < R$. Convergence of (1.3.7) implies there exists $C < \infty$ such that

$$|a_k| T^k \le C, \quad \forall\, k. \tag{1.3.9}$$

Hence, if $|t| \le S$,

$$|k a_k t^{k-1}| \le \frac{C}{S}\, k \left(\frac{S}{T}\right)^k, \tag{1.3.10}$$

which readily yields absolute convergence. (See Exercise 1 below.) Hence

$$g(t) = \sum_{k=1}^{\infty} k a_k t^{k-1} \tag{1.3.11}$$

is continuous on $(-R, R)$. To show that $f'(t) = g(t)$, by the fundamental theorem of calculus, it is equivalent to show

$$\int_0^t g(s)\, ds = f(t) - f(0). \tag{1.3.12}$$

The following result implies this.

Proposition 1.3.3. *Assume $b_k \in \mathbb{C}$ and*

$$g(t) = \sum_{k=0}^{\infty} b_k t^k \tag{1.3.13}$$

converges for real t, satisfying $|t| < R$. Then, for $|t| < R$,

$$\int_0^t g(s)\, ds = \sum_{k=0}^{\infty} \frac{b_k}{k+1} t^{k+1}, \tag{1.3.14}$$

the series being absolutely convergent for $|t| < R$.

Proof. Since, for $|t| < R$,

$$\left| \frac{b_k}{k+1} t^{k+1} \right| \le R |b_k t^k|, \tag{1.3.15}$$

convergence of the series in (1.3.14) is clear. Next, write

$$\begin{aligned} g(t) &= S_N(t) + R_N(t), \\ S_N(t) &= \sum_{k=0}^{N} b_k t^k, \quad R_N(t) = \sum_{k=N+1}^{\infty} b_k t^k. \end{aligned} \tag{1.3.16}$$

To continue, as in the proof of Proposition 1.3.1, pick $S < T < R$. There exists $C < \infty$ such that $|b_k T^k| \le C$ for all k. Hence

$$|t| \le S \Rightarrow |R_N(t)| \le C \sum_{k=N+1}^{\infty} \Bigl(\frac{S}{T}\Bigr)^k = C\varepsilon_N \to 0, \quad \text{as} \quad N \to \infty. \tag{1.3.17}$$

so

$$\int_0^t g(s)\,ds = \sum_{k=0}^{N} \frac{b_k}{k+1} t^{k+1} + \int_0^t R_N(s)\,ds, \tag{1.3.18}$$

and, for $|t| \le S$,

$$\Bigl|\int_0^t R_N(s)\,ds\Bigr| \le \int_0^t |R_N(s)|\,ds \le CR\varepsilon_N. \tag{1.3.19}$$

This gives (1.3.14). □

Remark. The definition of (1.3.14) for $t < 0$ follows standard convention. More generally, if $a < b$ and $g \in \mathcal{R}([a,b])$, then

$$\int_b^a g(s)\,ds = -\int_a^b g(s)\,ds.$$

More generally, if we have a power series about t_0,

$$f(t) = \sum_{k=0}^{\infty} a_k (t-t_0)^k, \quad \text{for} \quad |t-t_0| < R, \tag{1.3.20}$$

then f is differentiable for $|t-t_0| < R$ and

$$f'(t) = \sum_{k=1}^{\infty} k a_k (t-t_0)^{k-1}. \tag{1.3.21}$$

We can then differentiate this power series, and inductively obtain

$$f^{(n)}(t) = \sum_{k=n}^{\infty} k(k-1)\cdots(k-n+1) a_k (t-t_0)^{k-n}. \tag{1.3.22}$$

In particular,

$$f^{(n)}(t_0) = n!\, a_n. \tag{1.3.23}$$

We can turn (1.3.23) around and write

$$a_n = \frac{f^{(n)}(t_0)}{n!}. \tag{1.3.24}$$

This suggests the following method of taking a given function and deriving a power series representation. Namely, if we can, we compute $f^{(k)}(t_0)$ and propose that

$$f(t) = \sum_{k=0}^{\infty} \frac{f^{(k)}(t_0)}{k!} (t-t_0)^k, \tag{1.3.25}$$

at least on some interval about t_0.

To take an example, consider

$$f(t) = (1-t)^{-r}, \tag{1.3.26}$$

with $r \in \mathbb{Q}$ (but $-r \notin \mathbb{N}$), and take $t_0 = 0$. (Results of §3.2 will allow us to extend this analysis to $r \in \mathbb{R}$.) Using (1.1.36), we get

$$f'(t) = r(1-t)^{-(r+1)}, \tag{1.3.27}$$

for $t < 1$. Inductively, for $k \in \mathbb{N}$,

$$f^{(k)}(t) = r(r+1)\cdots(r+k-1)\,(1-t)^{-(r+k)}. \tag{1.3.28}$$

Hence, for $k \geq 1$,

$$f^{(k)}(0) = r(r+1)\cdots(r+k-1) = \prod_{\ell=0}^{k-1}(r+\ell). \tag{1.3.29}$$

Consequently, we propose that

$$(1-t)^{-r} = \sum_{k=0}^{\infty} \frac{a_k}{k!} t^k, \quad |t| < 1, \tag{1.3.30}$$

with

$$a_0 = 1, \quad a_k = \prod_{\ell=0}^{k-1}(r+\ell), \quad \text{for} \quad k \geq 1. \tag{1.3.31}$$

We can verify convergence of the right side of (1.3.30) by using the ratio test:

$$\Bigl|\frac{a_{k+1}t^{k+1}/(k+1)!}{a_k t^k/k!}\Bigr| = \frac{k+r}{k+1}|t|. \tag{1.3.32}$$

This computation implies that the power series on the right side of (1.3.30) is absolutely convergent for $|t| < 1$, yielding a function

$$g(t) = \sum_{k=0}^{\infty} \frac{a_k}{k!} t^k, \quad |t| < 1. \tag{1.3.33}$$

It remains to establish that $g(t) = (1-t)^{-r}$.

We take up this task, on a more general level. Establishing that the series

$$\sum_{k=0}^{\infty} \frac{f^{(k)}(t_0)}{k!}(t-t_0)^k \tag{1.3.34}$$

converges to $f(t)$ is equivalent to examining the remainder $R_n(t, t_0)$ in the finite expansion

$$f(t) = \sum_{k=0}^{n} \frac{f^{(k)}(t_0)}{k!}(t-t_0)^k + R_n(t, t_0). \tag{1.3.35}$$

The series (1.3.34) converges to $f(t)$ if and only if $R_n(t, t_0) \to 0$ as $n \to \infty$. To see when this happens, we need a compact formula for the remainder R_n, which we proceed to derive.

It seems to clarify matters if we switch notation a bit, and write

$$f(x) = f(y) + f'(y)(x-y) + \cdots + \frac{f^{(n)}(y)}{n!}(x-y)^n + R_n(x,y). \tag{1.3.36}$$

We now take the y-derivative of each side of (1.3.36). The y-derivative of the left side is 0, and when we apply $\partial/\partial y$ to the right side, we observe an enormous amount of cancellation. There results the identity

$$\frac{\partial R_n}{\partial y}(x,y) = -\frac{1}{n!} f^{(n+1)}(y)(x-y)^n. \tag{1.3.37}$$

Also,

$$R_n(x,x) = 0. \tag{1.3.38}$$

If we concentrate on $R_n(x,y)$ as a function of y and look at the difference quotient $[R_n(x,y) - R_n(x,x)]/(y-x)$, an immediate consequence of the mean value theorem is that, if f is real valued,

$$R_n(x,y) = \frac{1}{n!}(x-y)(x-\xi_n)^n f^{(n+1)}(\xi_n), \tag{1.3.39}$$

for some ξ_n betweeen x and y. This is known as Cauchy's formula for the remainder. If $f^{(n+1)}$ is continuous, we can apply the fundamental theorem of calculus to (1.3.37)–(1.3.38), and obtain the following integral formula for the remainder in the power series.

Proposition 1.3.4. *If $I \subset \mathbb{R}$ is an interval, $x, y \in I$, and $f \in C^{n+1}(I)$, then the remainder $R_n(x,y)$ in* (1.3.36) *is given by*

$$R_n(x,y) = \frac{1}{n!}\int_y^x (x-s)^n f^{(n+1)}(s)\, ds. \tag{1.3.40}$$

This works regardless of whether f is real valued. Another derivation of (1.3.40) arose in the exercise set for §1.2. The change of variable $x - s = t(x-y)$ gives the integral formula

$$R_n(x,y) = \frac{1}{n!}(x-y)^{n+1}\int_0^1 t^n f^{(n+1)}(ty + (1-t)x)\, dt. \tag{1.3.41}$$

If we think of this integral as $1/(n+1)$ times a weighted mean of $f^{(n+1)}$, we get the Lagrange formula for the remainder,

$$R_n(x,y) = \frac{1}{(n+1)!}(x-y)^{n+1} f^{(n+1)}(\zeta_n), \tag{1.3.42}$$

for some ζ_n between x and y, provided f is real valued. The Lagrange formula is shorter and neater than the Cauchy formula, but the Cauchy formula is actually more powerful. The calculations in (1.3.45)–(1.3.56) below will illustrate this.

Note that, if $I(x,y)$ denotes the interval with endpoints x and y (e.g., (x,y) if $x < y$), then (1.3.40) implies

$$|R_n(x,y)| \le \frac{|x-y|}{n!} \sup_{\xi \in I(x,y)} |(x-\xi)^n f^{(n+1)}(\xi)|, \tag{1.3.43}$$

while (1.3.41) implies

$$|R_n(x,y)| \le \frac{|x-y|^{n+1}}{(n+1)!} \sup_{\xi \in I(x,y)} |f^{(n+1)}(\xi)|. \tag{1.3.44}$$

In case f is real valued, (1.3.43) also follows from the Cauchy formula (1.3.39) and (1.3.44) follows from the Lagrange formula (1.3.42).

Let us apply these estimates with f as in (1.3.26), i.e.,

$$f(x) = (1-x)^{-r}, \tag{1.3.45}$$

and $y = 0$. By (1.3.28),

$$f^{(n+1)}(\xi) = a_{n+1}(1-\xi)^{-(r+n+1)}, \quad a_{n+1} = \prod_{\ell=0}^{n}(r+\ell). \tag{1.3.46}$$

Consequently,

$$\frac{f^{(n+1)}(\xi)}{n!} = b_n(1-\xi)^{-(r+n+1)}, \quad b_n = \frac{a_{n+1}}{n!}. \tag{1.3.47}$$

Note that

$$\frac{b_{n+1}}{b_n} = \frac{n+1+r}{n+1} \to 1, \quad \text{as } n \to \infty. \tag{1.3.48}$$

Let us first investigate the estimate of $R_n(x,0)$ given by (1.3.44) (as in the Lagrange formula), and see how it leads to a suboptimal conclusion. (The impatient reader might skip (1.3.49)–(1.3.52) and go to (1.3.53).) By (1.3.47), if n is sufficiently large that $r+n+1>0$,

$$\begin{aligned} \sup_{\xi \in I(x,0)} \frac{|f^{(n+1)}(\xi)|}{(n+1)!} &= \frac{|b_n|}{n+1} \quad \text{if } \ -1 \le x \le 0, \\ &\ \frac{|b_n|}{n+1}(1-x)^{-(r+n+1)} \quad \text{if } \ 0 \le x < 1. \end{aligned} \tag{1.3.49}$$

Thus (1.3.44) implies

$$\begin{aligned} |R_n(x,0)| \le &\frac{|b_n|}{n+1}|x|^{n+1} \quad \text{if } \ -1 \le x \le 0, \\ &\frac{|b_n|}{n+1}\,\frac{1}{(1-x)^r}\Big(\frac{x}{1-x}\Big)^{n+1} \quad \text{if } \ 0 \le x < 1. \end{aligned} \tag{1.3.50}$$

Note that, by (1.3.48),

$$c_n = \frac{|b_n|}{n+1} \Longrightarrow \frac{c_{n+1}}{c_n} = \frac{|b_{n+1}|}{|b_n|}\,\frac{n+1}{n+2} \to 1 \ \text{ as } \ n \to \infty,$$

so we conclude from the first part of (1.3.50) that

$$R_n(x,0) \longrightarrow 0 \ \text{ as } \ n \to \infty, \ \text{if } -1 < x \le 0. \tag{1.3.51}$$

On the other hand, $x/(1-x)$ is < 1 for $0 \le x < 1/2$, but not for $1/2 \le x < 1$. Hence the factor $(x/(1-x))^{n+1}$ decreases geometrically for $0 \le x < 1/2$, but not for $1/2 \le x < 1$. Thus the second part of (1.3.50) yields only

$$R_n(x,0) \longrightarrow 0 \ \text{ as } \ n \to \infty, \ \text{if } 0 \le x < \frac{1}{2}. \tag{1.3.52}$$

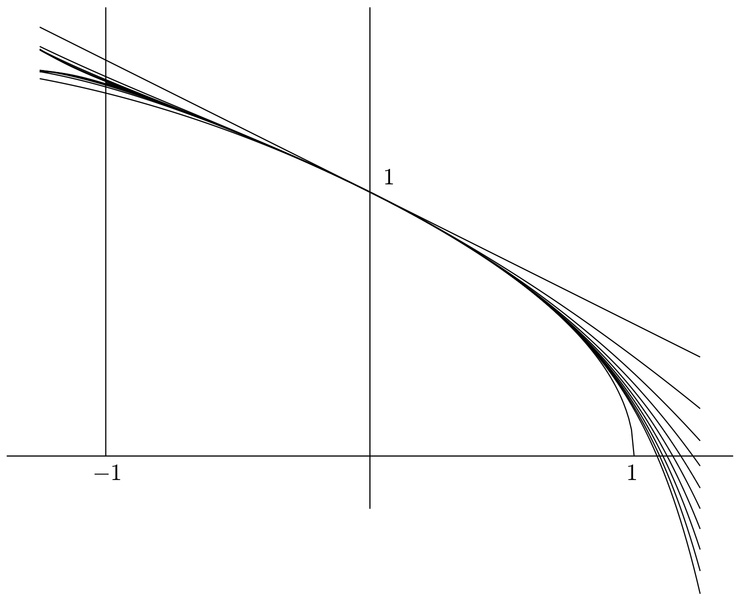

Figure 1.3.1. Power series approximations $\mathcal{S}_n(t)$ to $\sqrt{1-t}$, $1 \le n \le 10$

This is what the remainder estimate (1.3.44) yields.

To get the stronger result

$$R_n(x,0) \longrightarrow 0 \quad \text{as} \quad n \to \infty, \ \text{for } |x| < 1, \tag{1.3.53}$$

we use the remainder estimate (1.3.43) (as in the Cauchy formula). This gives

$$|R_n(x,0)| \le |b_n| \cdot |x| \sup_{\xi \in I(x,0)} \frac{|x-\xi|^n}{|1-\xi|^{n+1+r}}, \tag{1.3.54}$$

with b_n as in (1.3.47). Now

$$\begin{aligned} 0 \le \xi \le x < 1 &\Longrightarrow \frac{x-\xi}{1-\xi} \le x, \\ -1 < x \le \xi \le 0 &\Longrightarrow \left|\frac{x-\xi}{1-\xi}\right| \le |x-\xi| \le |x|. \end{aligned} \tag{1.3.55}$$

The first conclusion holds since it is equivalent to $x - \xi \le x(1-\xi) = x - x\xi$, hence to $x\xi \le \xi$. The second conclusion in (1.3.55) holds since $\xi \le 0 \Rightarrow 1 - \xi \ge 1$. We deduce from (1.3.54)–(1.3.55) that

$$|x| < 1 \Longrightarrow |R_n(x,0)| \le |b_n| \cdot |x|^{n+1}. \tag{1.3.56}$$

Using (1.3.48) then gives the desired conclusion (1.3.53).

We can now conclude that (1.3.30) holds, with a_k given by (1.3.31). For another proof of (1.3.30), see Exercise 14.

We illustrate this result in Figure 1.3.1, with $r = -1/2$. This figure shows the graphs of the approximations

$$\mathcal{S}_n(t) = \sum_{k=0}^{n} a_k t^k \tag{1.3.57}$$

to

$$(1-t)^{1/2} = \sum_{k=0}^{\infty} a_k t^k, \quad a_0 = 1, \quad a_{k+1} = \frac{2k-1}{2k+2} a_k, \tag{1.3.58}$$

for $1 \le n \le 10$. Results established above imply that

$$\mathcal{S}_n(t) \longrightarrow (1-t)^{1/2}, \quad \text{as} \quad n \to \infty, \tag{1.3.59}$$

for $|t| < 1$. In this case, this can be sharpened to yield uniform convergence for $t \in [-1, 1]$. For $t > 0$, the sequence $\mathcal{S}_n(t)$ is monotonically decreasing. It decreases to $(1-t)^{1/2}$ for $t \in [0, 1]$, and to $-\infty$ for $t > 1$. For $t < 0$, the terms in the series (1.3.57) alternate signs, for $n \ge 1$. One again has divergence for $t < 1$, as can be seen via the ratio test.

Often it is useful to use a substitution of variables in power series. For example, one can take $t = x^2$ in the power series for $f(t) = (1-t)^{-r}$ to get the power series for $(1-x^2)^{-r}$. Just replace t^k by x^{2k} in (1.3.30). To take a specific example, we have

$$(1-x^2)^{1/2} = \sum_{k=0}^{\infty} a_k x^{2k}, \quad a_0 = 1, \quad a_{k+1} = \frac{2k-1}{2k+2} a_k. \tag{1.3.60}$$

Figure 1.3.2 shows the graphs of the approximations

$$S_{2n}(x) = \sum_{k=0}^{n} a_k x^{2k} \quad \text{to} \quad \sqrt{1-x^2}, \tag{1.3.61}$$

for $1 \le n \le 10$. As indicated in the graph, this series *diverges* for $|x| > 1$. Results established above for $\mathcal{S}_n(t)$ imply

$$S_{2n}(x) \longrightarrow (1-x^2)^{1/2}, \quad \text{as} \quad n \to \infty, \tag{1.3.62}$$

for $|x| < 1$. Again, this can be sharpened to yield uniform convergence for $x \in [-1, 1]$.

There are some important examples of power series representations for which one does not need to use remainder estimates like (1.3.43) or (1.3.44). For example, we have

$$\sum_{k=0}^{n} x^k = \frac{1-x^{n+1}}{1-x}, \tag{1.3.63}$$

if $x \ne 1$. The right side tends to $1/(1-x)$ as $n \to \infty$, if $|x| < 1$, so we get

$$\frac{1}{1-x} = \sum_{k=0}^{\infty} x^k, \quad |x| < 1, \tag{1.3.64}$$

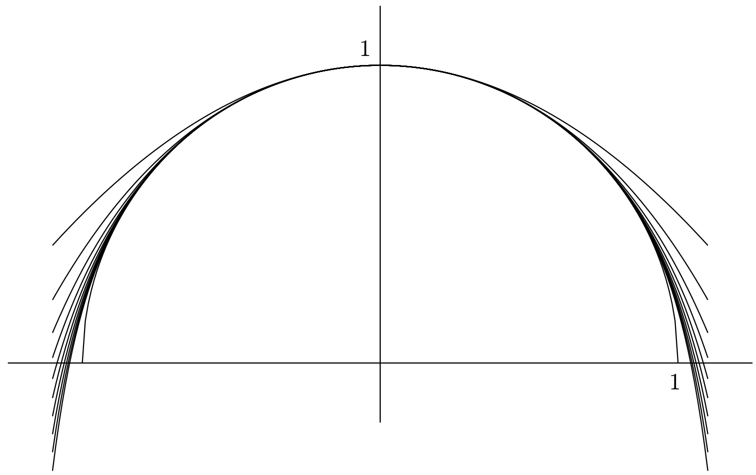

Figure 1.3.2. Power series approximations $S_{2n}(x)$ to $\sqrt{1-x^2}$, $1 \le n \le 10$

without further ado, which is the case $r = 1$ of (1.3.30)–(1.3.31). We can differentiate (1.3.64) repeatedly to get

$$(1-x)^{-n} = \sum_{k=0}^{\infty} c_k(n) x^k, \quad |x| < 1, \quad n \in \mathbb{N}, \tag{1.3.65}$$

and verify that (1.3.65) agrees with (1.3.30)–(1.3.31) with $r = n$. However, when $r \notin \mathbb{Z}$, such an analysis of $R_n(x, 0)$ as made above seems necessary. (But see Exercise 14 below.)

Let us also note that we can apply Proposition 1.3.3 to (1.3.64), obtaining

$$\sum_{k=0}^{\infty} \frac{x^{k+1}}{k+1} = \int_0^x \frac{dy}{1-y}, \quad |x| < 1. \tag{1.3.66}$$

Material covered in §3.2 will produce another formula for the right side of (1.3.66).

Returning to the integral formula for the remainder $R_n(x, y)$ in (1.3.36), we record the following variant of Proposition 1.3.4.

Proposition 1.3.5. *If $I \in \mathbb{R}$ is an interval, $x, y \in I$, and $f \in C^n(I)$, then*

$$R_n(x, y) = \frac{1}{(n-1)!} \int_y^x (x-s)^{n-1} [f^{(n)}(s) - f^{(n)}(y)]\, ds. \tag{1.3.67}$$

Proof. Do (1.3.36)–(1.3.40) with n replaced by $n-1$, and then write

$$R_{n-1}(x,y) = \frac{f^{(n)}(y)}{n!} + R_n(x,y). \tag{1.3.68}$$

□

Remark. An advantage of (1.3.67) over (1.3.40) is that for (1.3.67), we need only $f \in C^n(I)$, rather than $f \in C^{n+1}(I)$.

Exercises

1. Show that (1.3.10) yields the absolute convergence asserted in the proof of Proposition 1.3.2. More generally, show that, for any $n \in \mathbb{N}$, $r \in (0,1)$,

$$\sum_{k=1}^{\infty} k^n r^k < \infty.$$

Hint. Use the ratio test.

2. A special case of (1.3.20)–(1.3.23) is that, given a polynomial $p(t) = a_n t^n + \cdots + a_1 t + a_0$, we have $p^{(k)}(0) = k!\, a_k$. Apply this to

$$P_n(t) = (1+t)^n.$$

Compute $P_n^{(k)}(t)$ using (1.1.7) repeatedly, then compute $P_n^{(k)}(0)$, and use this to establish the binomial formula:

$$(1+t)^n = \sum_{k=0}^{n} \binom{n}{k} t^k, \quad \binom{n}{k} = \frac{n!}{k!(n-k)!}. \tag{1.3.69}$$

3. Going further, and building on the analysis in (1.3.26)–(1.3.56), show that, for $|t| < 1$,

$$(1+t)^r = \sum_{k=0}^{\infty} \binom{r}{k} t^k, \tag{1.3.70}$$

with

$$\binom{r}{0} = 1, \quad \binom{r}{k} = \frac{r(r-1)\cdots(r-k+1)}{k!}, \quad k \in \mathbb{N}. \tag{1.3.71}$$

The coefficients of t^k in (1.3.70), extending those that arise in (1.3.69), are also called binomial coefficients. Here, we take $r \in \mathbb{Q}$, but results of §3.2 will allow us to extend this result to $r \in \mathbb{R}$, and further, to $r \in \mathbb{C}$.

4. Find the coefficients in the power series

$$\frac{1}{\sqrt{1-x^4}} = \sum_{k=0}^{\infty} b_k x^k.$$

Show that this series converges to the left side for $|x| < 1$.
Hint. Take $r = 1/2$ in (1.3.30)–(1.3.31) and set $t = x^4$.

Now expand

$$\int_0^x \frac{dy}{\sqrt{1-y^4}}$$

in a power series in x. Show this holds for $|x| < 1$.

5. Expand

$$\int_0^x \frac{dy}{\sqrt{1+y^4}}$$

as a power series in x. Show that this holds for $|x| < 1$.

6. Expand

$$\int_0^1 \frac{dt}{\sqrt{1+xt^4}}$$

as a power series in x. Show that this holds for $|x| < 1$.

7. Let $I \subset \mathbb{R}$ be an open interval, $x_0 \in I$, and assume $f \in C^2(I)$ and $f'(x_0) = 0$. Use Proposition 1.3.4 to show that

$$\begin{aligned} f''(x_0) > 0 &\Rightarrow f \text{ has a local mimimum at } x_0, \\ f''(x_0) < 0 &\Rightarrow f \text{ has a local maximum at } x_0. \end{aligned}$$

Compare the proof of Proposition 1.1.4.

8. Note that

$$\sqrt{2} = 2\sqrt{1 - \frac{1}{2}}.$$

Expand the right side in a power series, using (1.3.30)–(1.3.31). How many terms suffice to approximate $\sqrt{2}$ to 12 digits?

9. In the setting of Exercise 8, investigate series that converge faster, such as series obtained from

$$\begin{aligned} \sqrt{2} &= \frac{3}{2}\sqrt{1 - \frac{1}{9}} \\ &= \frac{10}{7}\sqrt{1 - \frac{1}{50}}. \end{aligned}$$

10. Apply variants of the methods of Exercises 8–9 to approximate $\sqrt{3}$, $\sqrt{5}$, $\sqrt{7}$, and $\sqrt{1001}$.

11. Given a rational approximation x_n to $\sqrt{2}$, write

$$\sqrt{2} = x_n\sqrt{1+\delta_n}, \quad \text{so} \quad 1 + \delta_n = 2/x_n^2.$$

Assume $|\delta_n| \le 1/2$. Then set

$$x_{n+1} = x_n\Bigl(1 + \frac{1}{2}\delta_n\Bigr), \quad 2 = x_{n+1}^2(1 + \delta_{n+1}).$$

Estimate δ_{n+1}. Does the sequence (x_n) approach $\sqrt{2}$ faster than a power series? Apply this method to the last approximation in Exercise 9.

12. Assume $F \in C([a, b]),\ g \in \mathcal{R}([a, b]),\ F$ real valued, and $g \ge 0$ on $[a, b]$. Show that

$$\int_a^b g(t)F(t)\,dt = \Bigl(\int_a^b g(t)\,dt\Bigr)F(\zeta),$$

for some $\zeta \in (a, b)$. Show how this result justifies passing from (1.3.41) to (1.3.42). *Hint.* If $A = \min F,\ B = \max F$, and $M = \int_a^b g(t)\,dt$, show that

$$AM \le \int_a^b g(t)F(t)\,dt \le BM.$$

13. Recall that the Cauchy formula (1.3.39) for the remainder $R_n(x, y)$ was obtained by applying the Mean Value Theorem to the difference quotient

$$\frac{R_n(x, y) - R_n(x, x)}{y - x}.$$

Now apply the generalized mean value theorem, described in Exercise 8 of §1.1, with

$$f(y) = R(x, y), \quad g(y) = (x - y)^{n+1},$$

to obtain the Lagrange formula (1.3.42).

14. Here is an approach to the proof of (1.3.30) that avoids formulas for the remainder $R_n(x, 0)$. Set

$$f_r(t) = (1 - t)^{-r}, \quad g_r(t) = \sum_{k=0}^{\infty} \frac{a_k}{k!} t^k, \quad \text{for} \quad |t| < 1,$$

with a_k given by (1.3.31). Show that, for $|t| < 1$,

$$f_r'(t) = \frac{r}{1 - t} f_r(t), \quad \text{and} \quad (1 - t)g_r'(t) = r g_r(t).$$

Then show that

$$\frac{d}{dt}(1 - t)^r g_r(t) = 0,$$

and deduce that $f_r(t) = g_r(t)$.

15. Assume $f, g \in C^k(I),\ 0 \in I$, and write

$$f(x) = \sum_{i=0}^{k} f_i x^i + o(x^k), \quad g(x) = \sum_{j=0}^{k} g_j x^j + o(x^k),$$

with

$$f_i = \frac{f^{(i)}(0)}{i!}, \quad g_j = \frac{g^{(j)}(0)}{j!}.$$

Show that $h(x) = f(x)g(x)$ satisfies

$$h(x) = \sum_{i,j=0}^{k} f_i g_j x^{i+j} + o(x^k),$$

and deduce that

$$\frac{h^{(k)}(0)}{k!} = \sum_{i+j=k} f_i g_j = \sum_{i+j=k} \frac{1}{i!j!} f^{(i)}(0) g^{(j)}(0).$$

From this deduce that

$$\frac{d^k}{dx^k}(fg)(0) = \sum_{i+j=k} \frac{k!}{i!j!} f^{(i)}(0) g^{(j)}(0).$$

Pass from this to the identity

$$\frac{d^k}{dx^k}(fg)(x) = \sum_{i+j=k} \frac{k!}{i!j!} f^{(i)}(x) g^{(j)}(x), \tag{1.3.72}$$

for $x \in I$. This identity is called the *Leibniz identity*.

1.4. Unbounded integrable functions

There are lots of unbounded functions we would like to be able to integrate. For example, consider $f(x) = x^{-1/2}$ on $(0, 1]$ (defined any way you like at $x = 0$). Since, for $\varepsilon \in (0, 1)$,

$$\int_{\varepsilon}^{1} x^{-1/2}\,dx = 2 - 2\sqrt{\varepsilon}, \tag{1.4.1}$$

this has a limit as $\varepsilon \searrow 0$, and it is natural to set

$$\int_{0}^{1} x^{-1/2}\,dx = 2. \tag{1.4.2}$$

Sometimes (1.4.2) is called an "improper integral," but we do not consider that to be a proper designation. Here, we define a class $\mathcal{R}^{\#}(I)$ of not necessarily bounded "integrable" functions on an interval $I = [a, b]$, as follows.

First, assume $f \geq 0$ on I, and for $A \in (0, \infty)$, set

$$\begin{aligned} f_A(x) = \ & f(x) \quad \text{if} \ \ f(x) \leq A, \\ & A, \quad \text{if} \ \ f(x) > A. \end{aligned} \tag{1.4.3}$$

We say $f \in \mathcal{R}^{\#}(I)$ provided

$$\begin{aligned} & f_A \in \mathcal{R}(I), \quad \forall\, A < \infty, \quad \text{and} \\ & \exists \ \text{uniform bound} \ \int_I f_A\,dx \leq M. \end{aligned} \tag{1.4.4}$$

If $f \geq 0$ satisfies (1.4.4), then $\int_I f_A\,dx$ increases monotonically to a finite limit as $A \nearrow +\infty$, and we call the limit $\int_I f\,dx$:

$$\int_I f_A\,dx \nearrow \int_I f\,dx, \quad \text{for} \ \ f \in \mathcal{R}^{\#}(I), \ f \geq 0. \tag{1.4.5}$$

We also use the notation $\int_a^b f\,dx$, if $I = [a, b]$. If I is understood, we might just write $\int f\,dx$. It is valuable to have the following.

Proposition 1.4.1. *If $f, g : I \to \mathbb{R}^+$ are in $\mathcal{R}^{\#}(I)$, then $f + g \in \mathcal{R}^{\#}(I)$, and*

$$\int_I (f + g)\,dx = \int_I f\,dx + \int_I g\,dx. \tag{1.4.6}$$

Proof. To start, note that $(f + g)_A \leq f_A + g_A$. In fact,

$$(f + g)_A = (f_A + g_A)_A. \tag{1.4.7}$$

Hence $(f + g)_A \in \mathcal{R}(I)$ and $\int (f + g)_A\,dx \leq \int f_A\,dx + \int g_A\,dx \leq \int f\,dx + \int g\,dx$, so we have $f + g \in \mathcal{R}^{\#}(I)$ and

$$\int (f + g)\,dx \leq \int f\,dx + \int g\,dx. \tag{1.4.8}$$

On the other hand, if $B > 2A$, then $(f + g)_B \geq f_A + g_A$, so

$$\int (f + g)\,dx \geq \int f_A\,dx + \int g_A\,dx, \tag{1.4.9}$$

for all $A < \infty$, and hence

$$\int (f+g)\,dx \geq \int f\,dx + \int g\,dx. \tag{1.4.10}$$

Together, (1.4.8) and (1.4.10) yield (1.4.6). □

Next, we take $f : I \to \mathbb{R}$ and set

$$f = f^+ - f^-, \quad f^+(x) = \begin{cases} f(x) & \text{if } f(x) \geq 0, \\ 0 & \text{if } f(x) < 0. \end{cases} \tag{1.4.11}$$

Then we say

$$f \in \mathcal{R}^\#(I) \Longleftrightarrow f^+, f^- \in \mathcal{R}^\#(I), \tag{1.4.12}$$

and set

$$\int_I f\,dx = \int_I f^+\,dx - \int_I f^-\,dx, \tag{1.4.13}$$

where the two terms on the right are defined as in (1.4.5). To extend the additivity, we begin as follows

Proposition 1.4.2. *Assume that $g \in \mathcal{R}^\#(I)$ and that $g_j \geq 0$, $g_j \in \mathcal{R}^\#(I)$, and*

$$g = g_0 - g_1. \tag{1.4.14}$$

Then

$$\int g\,dx = \int g_0\,dx - \int g_1\,dx. \tag{1.4.15}$$

Proof. Take $g = g^+ - g^-$ as in (1.4.11). Then (1.4.14) implies

$$g^+ + g_1 = g_0 + g^-, \tag{1.4.16}$$

which by Proposition 1.4.1 yields

$$\int g^+\,dx + \int g_1\,dx = \int g_0\,dx + \int g^-\,dx. \tag{1.4.17}$$

This implies

$$\int g^+\,dx - \int g^-\,dx = \int g_0\,dx - \int g_1\,dx, \tag{1.4.18}$$

which yields (1.4.15) □

We now extend additivity.

Proposition 1.4.3. *Assume $f_1, f_2 \in \mathcal{R}^\#(I)$. Then $f_1 + f_2 \in \mathcal{R}^\#(I)$ and*

$$\int_I (f_1+f_2)\,dx = \int_I f_1\,dx + \int_I f_2\,dx. \tag{1.4.19}$$

Proof. If $g = f_1 + f_2 = (f_1^+ - f_1^-) + (f_2^+ - f_2^-)$, then

$$g = g_0 - g_1, \quad g_0 = f_1^+ + f_2^+, \quad g_1 = f_1^- + f_2^-. \tag{1.4.20}$$

We have $g_j \in \mathcal{R}^\#(I)$, and then

$$\begin{aligned} \int (f_1 + f_2)\, dx &= \int g_0\, dx - \int g_1\, dx \\ &= \int (f_1^+ + f_2^+)\, dx - \int (f_1^- + f_2^-)\, dx \\ &= \int f_1^+\, dx + \int f_2^+\, dx - \int f_1^-\, dx - \int f_2^-\, dx, \end{aligned} \tag{1.4.21}$$

the first equality by Proposition 1.4.2, the second tautologically, and the third by Proposition 1.4.1. Since

$$\int f_j\, dx = \int f_j^+\, dx - \int f_j^-\, dx, \tag{1.4.22}$$

this gives (1.4.19). □

If $f : I \to \mathbb{C}$, we set $f = f_1 + i f_2,\ f_j : I \to \mathbb{R}$, and say $f \in \mathcal{R}^\#(I)$ if and only if f_1 and f_2 belong to $\mathcal{R}^\#(I)$. Then we set

$$\int f\, dx = \int f_1\, dx + i \int f_2\, dx. \tag{1.4.23}$$

Similar comments apply to $f : I \to \mathbb{R}^n$.

Given $f \in \mathcal{R}^\#(I)$, we set

$$\|f\|_{L^1(I)} = \int_I |f(x)|\, dx. \tag{1.4.24}$$

We have, for $f, g \in \mathcal{R}^\#(I),\ a \in \mathbb{C}$,

$$\|af\|_{L^1(I)} = |a|\, \|f\|_{L^1(I)}, \tag{1.4.25}$$

and

$$\begin{aligned} \|f + g\|_{L^1(I)} &= \int_I |f + g|\, dx \\ &\le \int_I (|f| + |g|)\, dx \\ &= \|f\|_{L^1(I)} + \|g\|_{L^1(I)}. \end{aligned} \tag{1.4.26}$$

Note that, if $S \subset I$,

$$\mathrm{cont}^+(S) = 0 \Longrightarrow \int_I |\chi_S|\, dx = 0, \tag{1.4.27}$$

where $\mathrm{cont}^+(S)$ is defined by (1.2.21). Thus, to get a metric, we need to form equivalence classes. The set of equivalence classes $[f]$ of elements of $\mathcal{R}^\#(I)$, where

$$f \sim \tilde{f} \Longleftrightarrow \int_I |f - \tilde{f}|\, dx = 0, \tag{1.4.28}$$

forms a metric space, with distance function

$$D([f],[g]) = \|f-g\|_{L^1(I)}. \tag{1.4.29}$$

However, this metric space is not complete. One needs the Lebesgue integral to obtain a complete metric space. One can see [**7**] or [**14**].

We next show that each $f \in \mathcal{R}^\#(I)$ can be approximated in L^1 by a sequence of bounded, Riemann integrable functions.

Proposition 1.4.4. *If $f \in \mathcal{R}^\#(I)$, then there exist $f_k \in \mathcal{R}(I)$ such that*

$$\|f-f_k\|_{L^1(I)} \longrightarrow 0, \quad \text{as} \quad k \to \infty. \tag{1.4.30}$$

Proof. If we separately approximate $\operatorname{Re} f$ and $\operatorname{Im} f$ by such sequences, then we approximate f, so it suffices to treat the case where f is real. Similarly, writing $f = f^+ - f^-$, we see that it suffices to treat the case where $f \ge 0$ on I. For such f, simply take

$$f_k = f_A, \quad A = k, \tag{1.4.31}$$

with f_A as in (1.4.3). Then (1.4.5) implies

$$\int\limits_I f_k\,dx \nearrow \int\limits_I f\,dx, \tag{1.4.32}$$

and Proposition 1.4.3 gives

$$\begin{aligned} \int\limits_I |f-f_k|\,dx &= \int\limits_I (f-f_k)\,dx \\ &= \int\limits_I f\,dx - \int\limits_I f_k\,dx \\ &\to 0 \ \text{ as } \ k \to \infty. \end{aligned} \tag{1.4.33}$$

□

So far, we have dealt with integrable functions on a bounded interval. Now, we say $f : \mathbb{R} \to \mathbb{R}$ (or $\mathbb{C}$, or $\mathbb{R}^n$) belongs to $\mathcal{R}^\#(\mathbb{R})$ provided $f|_I \in \mathcal{R}^\#(I)$ for each closed, bounded interval $I \subset \mathbb{R}$ and

$$\exists A < \infty \ \text{ such that } \int_{-R}^{R} |f|\,dx \le A, \quad \forall\, R < \infty. \tag{1.4.34}$$

In such a case, we set

$$\int_{-\infty}^{\infty} f\,dx = \lim_{R\to\infty} \int_{-R}^{R} f\,dx. \tag{1.4.35}$$

One can similarly define $\mathcal{R}^\#(\mathbb{R}^+)$.

Exercises

1. Let $f : [0,1] \to \mathbb{R}^+$ and assume f is continuous on $(0,1]$. Show that

$$f \in \mathcal{R}^\#([0,1]) \Longleftrightarrow \int_\varepsilon^1 f\,dx \text{ is bounded as } \varepsilon \searrow 0.$$

In such a case, show that

$$\int_0^1 f\,dx = \lim_{\varepsilon\to 0} \int_\varepsilon^1 f\,dx.$$

2. Let $a > 0$. Define $p_a : [0,1] \to \mathbb{R}$ by $p_a = x^{-a}$ if $0 < x \le 1$ Set $p_a(0) = 0$. Show that

$$p_a \in \mathcal{R}^\#([0,1]) \Longleftrightarrow a < 1.$$

3. (See §3.2 for a development of $\log x$.) Let $b > 0$. Define $q_b : [0,1/2] \to \mathbb{R}$ by

$$q_b(x) = \frac{1}{x|\log x|^b},$$

if $0 < x \le 1/2$. Set $q_b(0) = 0$. Show that

$$q_b \in \mathcal{R}^\#([0,1/2]) \Longleftrightarrow b > 1.$$

4. Show that if $a \in \mathbb{C}$ and if $f \in \mathcal{R}^\#(I)$, then

$$af \in \mathcal{R}^\#(I), \quad \text{and} \quad \int af\,dx = a\int f\,dx.$$

Hint. Check this for $a > 0$, $a = -1$, and $a = i$.

5. Show that

$$f \in \mathcal{R}(I),\ g \in \mathcal{R}^\#(I) \Longrightarrow fg \in \mathcal{R}^\#(I).$$

Hint. Use (1.2.53). First treat the case $f, g \ge 1$, $f \le M$. Show that in such a case,

$$(fg)_A = (f_A g_A)_A, \quad \text{and} \quad (fg)_A \le M g_A.$$

6. Peek ahead to §3.2 and compute

$$\int_0^1 \log t\,dt.$$

Hint. To compute $\int_\varepsilon^1 \log t\,dt$, first compute

$$\frac{d}{dt}(t\log t).$$

7. Given $g \in \mathcal{R}(I)$, show that there exist $g_k \in \mathrm{PK}(I)$ such that

$$\|g - g_k\|_{L^1(I)} \longrightarrow 0.$$

Given $h \in \mathrm{PK}(I)$, show that there exist $h_k \in C(I)$ such that

$$\|h - h_k\|_{L^1(I)} \longrightarrow 0.$$

8. Using Exercise 7 and Proposition 1.4.4, prove the following: given $f \in \mathcal{R}^{\#}(I)$, there exist $f_k \in C(I)$ such that

$$\|f - f_k\|_{L^1(I)} \longrightarrow 0.$$

9. Recall Exercise 4 of §1.2. If $\varphi : [a,b] \to [A,B]$ is C^1, with $\varphi'(x) > 0$ for all $x \in [a,b]$, then

$$\int_A^B f(y)\,dy = \int_a^b f(\varphi(t))\varphi'(t)\,dt, \tag{1.4.36}$$

for each $f \in C([a,b])$, where $A = \varphi(a)$, $B = \varphi(b)$. Using Exercise 8, show that (1.4.36) holds for each $f \in \mathcal{R}^{\#}([a,b])$.

10. If $f \in \mathcal{R}^{\#}(\mathbb{R})$, so (1.4.34) holds, prove that the limit exists in (1.4.35).

11. Given $f(x) = x^{-1/2}(1+x^2)^{-1}$ for $x > 0$, show that $f \in \mathcal{R}^{\#}(\mathbb{R}^+)$. Show that

$$\int_0^\infty \frac{1}{1+x^2}\frac{dx}{\sqrt{x}} = 2\int_0^\infty \frac{dy}{1+y^4}.$$

12. Let $f_k \in \mathcal{R}^{\#}([a,b])$, $f : [a,b] \to \mathbb{R}$ satisfy

(a) $|f_k| \le g, \quad \forall\, k,$ for some $g \in \mathcal{R}^{\#}([a,b])$,

(b) Given $\varepsilon > 0$, $\exists$ contented $S_\varepsilon \subset [a,b]$ such that
$\int\limits_{S_\varepsilon} g\,dx < \varepsilon$, and $f_k \to f$ uniformly on $[a,b] \setminus S_\varepsilon$.

Show that $f \in \mathcal{R}^{\#}([a,b])$ and

$$\int_a^b f_k(x)\,dx \longrightarrow \int_a^b f(x)\,dx, \quad \text{as } k \to \infty.$$

13. Let $g \in \mathcal{R}^{\#}([a,b])$ be ≥ 0. Show that for each $\varepsilon > 0$, there exists $\delta > 0$ such that

$$S \subset [a,b] \text{ contented, } \operatorname{cont} S < \delta \Longrightarrow \int\limits_S g\,dx < \varepsilon.$$

Hint. With g_A defined as in (1.4.3), pick A such that $\int g_A\,dx \ge \int g\,dx - \varepsilon/2$. Then pick $\delta < \varepsilon/2A$.

14. Deduce from Exercises 12–13 the following. Let $f_k \in \mathcal{R}^{\#}([a,b])$, $f : [a,b] \to \mathbb{R}$ satisfy

(a) $|f_k| \le g, \quad \forall\, k,$ for some $g \in \mathcal{R}^{\#}([a,b])$,

(b) Given $\delta > 0$, $\exists$ contented $S_\delta \subset [a,b]$ such that
$\operatorname{cont} S_\delta < \delta$, and $f_k \to f$ uniformly on $[a,b] \setminus S_\delta$.

Show that $f \in \mathcal{R}^{\#}([a,b])$ and

$$\int_a^b f_k(x)\,dx \longrightarrow \int_a^b f(x)\,dx, \quad \text{as} \quad k \to \infty.$$

REMARK. Compare Exercise 18 of §1.2. As mentioned there, the Lebesgue theory of integration has a stronger result, known as the Lebesgue dominated convergence theorem.

Chapter 2

Multidimensional spaces

Multivariable calculus is set in multidimensional spaces. The paradigmatic case is n-dimensional Euclidean space $\mathbb{R}^n$. We present basic material on this in §2.1. The space $\mathbb{R}^n$ has both a linear structure and a geometric structure, coming from the dot product, which gives rise to the notion of distance and of convergence of sequences.

While $\mathbb{R}^n$ is the paradigm, it is convenient to consider more general vector spaces, and we do this in §2.2. We also study linear transformations $T : V \to W$ between two vector spaces. We define the class of finite-dimensional vector spaces, and show that the dimension of such a vector space is well defined. If V is a real vector space and $\dim V = n$, then V is isomorphic to $\mathbb{R}^n$. Linear transformations from $\mathbb{R}^n$ to $\mathbb{R}^m$ are given by $m \times n$ matrices. In Chapter 4, such linear transformations arise as derivatives of nonlinear maps, and understanding the behavior of these derivatives is basic to many key results in multivariable calculus, both in Chapter 4 and in subsequent chapters.

In §2.3 we define the determinant, $\det A$, of an $n \times n$ matrix A, and show that A is invertible if and only if $\det A \neq 0$. In Chapter 5 we will see the determinant of the derivative $DF(x)$ of a map $F : \mathcal{O} \to \Omega$ between regions of $\mathbb{R}^n$ entering into the change of variable formula for the integral.

In §2.4 we define the trace of a matrix $A \in M(n, \mathbb{R})$ and explore some of its basic properties, including the Euclidean space structure on $M(n, \mathbb{R})$ that arises from $\langle A, B\rangle = \operatorname{Tr} AB^t$. Some exercises relate the trace and the determinant.

Section 2.5 treats the cross product of vectors in $\mathbb{R}^3$. Results derived here will be useful for the study of curves in $\mathbb{R}^3$ in §3.4, and for the study of surface area in §6.1.

2.1. Euclidean spaces

The space $\mathbb{R}^n$, n-dimensional Euclidean space, consists of n-tuples of real numbers:

$$x = (x_1, \dots, x_n) \in \mathbb{R}^n, \quad x_j \in \mathbb{R},\ 1 \le j \le n. \tag{2.1.1}$$

The number x_j is called the jth component of x. Here we discuss some important algebraic and metric structures on $\mathbb{R}^n$. First, there is addition. If x is as in (2.1.1) and also $y = (y_1, \dots, y_n) \in \mathbb{R}^n$, we have

$$x + y = (x_1 + y_1, \dots, x_n + y_n) \in \mathbb{R}^n. \tag{2.1.2}$$

Addition is done componentwise. Also, given $a \in \mathbb{R}$, we have

$$ax = (ax_1, \dots, ax_n) \in \mathbb{R}^n. \tag{2.1.3}$$

This is scalar multiplication. In (2.1.1), we represent x as a row vector. Sometimes we want to represent x by a column vector,

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}. \tag{2.1.4}$$

Then (2.1.2)–(2.1.3) are converted to

$$x + y = \begin{pmatrix} x_1 + y_1 \\ \vdots \\ x_n + y_n \end{pmatrix}, \quad ax = \begin{pmatrix} ax_1 \\ \vdots \\ ax_n \end{pmatrix}. \tag{2.1.5}$$

We also have the dot product,

$$x \cdot y = \sum_{j=1}^n x_j y_j = x_1 y_1 + \cdots + x_n y_n \in \mathbb{R}, \tag{2.1.6}$$

given $x, y \in \mathbb{R}^n$. The dot product has the properties

$$\begin{aligned} &x \cdot y = y \cdot x, \\ &x \cdot (ay + bz) = a(x \cdot y) + b(x \cdot z), \\ &x \cdot x > 0 \ \text{ unless } \ x = 0. \end{aligned} \tag{2.1.7}$$

Note that

$$x \cdot x = x_1^2 + \cdots + x_n^2. \tag{2.1.8}$$

We set

$$|x| = \sqrt{x \cdot x}, \tag{2.1.9}$$

which we call the norm of x. Note that (2.1.7) implies

$$(ax) \cdot (ax) = a^2 (x \cdot x), \tag{2.1.10}$$

hence

$$|ax| = |a| \cdot |x|, \quad \text{for } \ a \in \mathbb{R},\ x \in \mathbb{R}^n. \tag{2.1.11}$$

Taking a cue from the Pythagorean theorem, we say that the *distance* from x to y in $\mathbb{R}^n$ is

$$d(x, y) = |x - y|. \tag{2.1.12}$$

For us, (2.1.9) and (2.1.12) are simply definitions. We do not need to depend on a derivation of the Pythagorean theorem via classical Euclidean geometry. Significant properties will be derived below, without recourse to a prior theory of Euclidean geometry.

A set X equipped with a distance function is called a metric space. One can find a discussion of metric spaces in general in [**15**]. Here, we want to show that the Euclidean distance, defined by (2.1.12), satisfies the "triangle inequality,"

$$d(x, y) \le d(x, z) + d(z, y), \quad \forall\, x, y, z \in \mathbb{R}^n. \tag{2.1.13}$$

This in turn is a consequence of the following, also called the triangle inequality.

Proposition 2.1.1. *The norm* (2.1.9) *on* $\mathbb{R}^n$ *has the property*

$$|x + y| \le |x| + |y|, \quad \forall\, x, y \in \mathbb{R}^n. \tag{2.1.14}$$

Proof. We compare the squares of the two sides of (2.1.14). First,

$$\begin{aligned} |x + y|^2 &= (x + y) \cdot (x + y) \\ &= x \cdot x + y \cdot x + y \cdot x + y \cdot y \\ &= |x|^2 + 2x \cdot y + |y|^2. \end{aligned} \tag{2.1.15}$$

Next,

$$(|x| + |y|)^2 = |x|^2 + 2|x| \cdot |y| + |y|^2. \tag{2.1.16}$$

We see that (2.1.14) holds if and only if $x \cdot y \le |x| \cdot |y|$. Thus the proof of Proposition 2.1.1 is finished off by the following result, known as Cauchy's inequality. ☐

Proposition 2.1.2. *For all* $x, y \in \mathbb{R}^n$,

$$|x \cdot y| \le |x| \cdot |y|. \tag{2.1.17}$$

Proof. We start with the chain

$$0 \le |x - y|^2 = (x - y) \cdot (x - y) = |x|^2 + |y|^2 - 2x \cdot y, \tag{2.1.18}$$

which implies

$$2x \cdot y \le |x|^2 + |y|^2, \quad \forall\, x, y \in \mathbb{R}^n. \tag{2.1.19}$$

If we replace x by tx and y by $t^{-1}y$, with $t > 0$, the left side of (2.1.19) is unchanged, so we have

$$2x \cdot y \le t^2|x|^2 + t^{-2}|y|^2, \quad \forall\, t > 0. \tag{2.1.20}$$

Now we pick t so that the two terms on the right side of (2.1.20) are equal, namely

$$t^2 = \frac{|y|}{|x|}, \quad t^{-2} = \frac{|x|}{|y|}. \tag{2.1.21}$$

(At this point, note that (2.1.17) is obvious if $x = 0$ or $y = 0$, so we will assume that $x \ne 0$ and $y \ne 0$.) Plugging (2.1.21) into (2.1.20) gives

$$x \cdot y \le |x| \cdot |y|, \quad \forall\, x, y \in \mathbb{R}^n. \tag{2.1.22}$$

This is almost (2.1.17). To finish, we can replace x in (2.1.22) by $-x = (-1)x$, getting

$$-(x \cdot y) \leq |x| \cdot |y|, \tag{2.1.23}$$

and together (2.1.22) and (2.1.23) give (2.1.17). $\square$

We now discuss a number of notions and results related to convergence in $\mathbb{R}^n$. First, a sequence of points (p_j) in $\mathbb{R}^n$ converges to a limit $p \in \mathbb{R}^n$ (we write $p_j \to p$) if and only if

$$|p_j - p| \longrightarrow 0, \tag{2.1.24}$$

where $|\ \cdot\ |$ is the Euclidean norm on $\mathbb{R}^n$, defined by (2.1.9), and the meaning of (2.1.24) is that for every $\varepsilon > 0$ there exists N such that

$$j \geq N \Longrightarrow |p_j - p| < \varepsilon. \tag{2.1.25}$$

If we write $p_j = (p_{1j}, \dots, p_{nj})$ and $p = (p_1, \dots, p_n)$, then (2.1.24) is equivalent to

$$(p_{1j} - p_1)^2 + \cdots + (p_{nj} - p_n)^2 \longrightarrow 0, \quad \text{as} \quad j \to \infty,$$

which holds if and only if

$$|p_{\ell j} - p_\ell| \longrightarrow 0 \quad \text{as} \quad j \to \infty, \quad \text{for each} \quad \ell \in \{1, \dots, n\}.$$

That is to say, convergence $p_j \to p$ in $\mathbb{R}^n$ is eqivalent to convergence of each component.

A set $S \subset \mathbb{R}^n$ is said to be *closed* if and only if

$$p_j \in S, \ p_j \to p \Longrightarrow p \in S. \tag{2.1.26}$$

The complement $\mathbb{R}^n \setminus S$ of a closed set S is *open*. Alternatively, $\Omega \subset \mathbb{R}^n$ is open if and only if, given $q \in \Omega$, there exists $\varepsilon > 0$ such that $B_\varepsilon(q) \subset \Omega$, where

$$B_\varepsilon(q) = \{p \in \mathbb{R}^n : |p - q| < \varepsilon\}, \tag{2.1.27}$$

so q cannot be a limit of a sequence of points in $\mathbb{R}^n \setminus \Omega$.

An important property of $\mathbb{R}^n$ is *completeness*, a property defined as follows. A sequence (p_j) of points in $\mathbb{R}^n$ is called a Cauchy sequence if and only if

$$|p_j - p_k| \longrightarrow 0, \quad \text{as} \quad j, k \to \infty. \tag{2.1.28}$$

Again we see that (p_j) is Cauchy in $\mathbb{R}^n$ if and only if each component is Cauchy in $\mathbb{R}$. It is easy to see that if $p_j \to p$ for some $p \in \mathbb{R}^n$, then (2.1.28) holds. The completeness property is the converse.

Theorem 2.1.3. *If (p_j) is a Cauchy sequence in $\mathbb{R}^n$, then it has a limit, i.e.,* (2.1.24) *holds for some $p \in \mathbb{R}^n$.*

Proof. Since convergence $p_j \to p$ in $\mathbb{R}^n$ is equivalent to convergence in $\mathbb{R}$ of each component, the result is a consequence of the completeness of $\mathbb{R}$. This is proved in §A.2. $\square$

Completeness provides a path to the following key notion of *compactness*. A nonempty set $K \subset \mathbb{R}^n$ is said to be compact if and only if the following property holds.

$$\text{Each infinite sequence } (p_j) \text{ in } K \text{ has a subsequence that converges to a point in } K. \tag{2.1.29}$$

It is clear that if K is compact, then it must be closed. It must also be bounded, i.e., there exists $R < \infty$ such that $K \subset B_R(0)$. Indeed, if K is not bounded, there exist $p_j \in K$ such that $|p_{j+1}| \ge |p_j| + 1$. In such a case, $|p_j - p_k| \ge 1$ whenever $j \ne k$, so (p_j) cannot have a convergent subsequence. The following converse result is the n-dimensional Bolzano-Weierstrass theorem.

Theorem 2.1.4. *If a nonempty $K \subset \mathbb{R}^n$ is closed and bounded, then it is compact.*

Proof. If $K \subset \mathbb{R}^n$ is closed and bounded, it is a closed subset of some box

$$\mathcal{B} = \{(x_1, \dots, x_n) \in \mathbb{R}^n : a \le x_k \le b, \ \forall\, k\}. \tag{2.1.30}$$

Clearly every closed subset of a compact set is compact, so it suffices to show that $\mathcal{B}$ is compact. Now, each closed bounded interval $[a, b]$ in $\mathbb{R}$ is compact, as shown in Appendix A.3, and (by reasoning similar to the proof of Theorem 2.1.3) the compactness of $\mathcal{B}$ follows readily from this. □

We establish some further properties of compact sets $K \subset \mathbb{R}^n$, leading to the important result, Proposition 2.1.8 below.

Proposition 2.1.5. *Let $K \subset \mathbb{R}^n$ be compact. Assume $X_1 \supset X_2 \supset X_3 \supset \cdots$ form a decreasing sequence of closed subsets of K. If each $X_m \ne \emptyset$, then $\cap_m X_m \ne \emptyset$.*

Proof. Pick $x_m \in X_m$. If K is compact, (x_m) has a convergent subsequence, $x_{m_k} \to y$. Since $\{x_{m_k} : k \ge \ell\} \subset X_{m_\ell}$, which is closed, we have $y \in \cap_m X_m$. □

Corollary 2.1.6. *Let $K \subset \mathbb{R}^n$ be compact. Assume $U_1 \subset U_2 \subset U_3 \subset \cdots$ form an increasing sequence of open sets in $\mathbb{R}^n$. If $\cup_m U_m \supset K$, then $U_M \supset K$ for some M.*

Proof. Consider $X_m = K \setminus U_m$. □

Before getting to Proposition 2.1.8, we bring in the following. Let $\mathbb{Q}$ denote the set of rational numbers, and let $\mathbb{Q}^n$ denote the set of points in $\mathbb{R}^n$ all of whose components are rational. The set $\mathbb{Q}^n \subset \mathbb{R}^n$ has the following "denseness" property: given $p \in \mathbb{R}^n$ and $\varepsilon > 0$, there exists $q \in \mathbb{Q}^n$ such that $|p - q| < \varepsilon$. Let

$$\mathcal{R} = \{B_r(q) : q \in \mathbb{Q}^n, \ r \in \mathbb{Q} \cap (0, \infty)\}. \tag{2.1.31}$$

Note that $\mathbb{Q}$ and $\mathbb{Q}^n$ are *countable*, i.e., they can be put in one-to-one correspondence with $\mathbb{N}$. Hence $\mathcal{R}$ is a countable collection of balls. The following lemma is left as an exercise for the reader.

Lemma 2.1.7. *Let $\Omega \subset \mathbb{R}^n$ be a nonempty open set. Then*

$$\Omega = \bigcup \{B : B \in \mathcal{R}, \ B \subset \Omega\}. \tag{2.1.32}$$

To state the next result, we say that a collection $\{U_\alpha : \alpha \in \mathcal{A}\}$ covers K if $K \subset \cup_{\alpha\in\mathcal{A}} U_\alpha$. If each $U_\alpha \subset \mathbb{R}^n$ is open, it is called an open cover of K. If $\mathcal{B} \subset \mathcal{A}$ and $K \subset \cup_{\beta\in\mathcal{B}} U_\beta$, we say $\{U_\beta : \beta \in \mathcal{B}\}$ is a subcover. The following is part of the n-dimensional Heine-Borel theorem.

Proposition 2.1.8. *If $K \subset \mathbb{R}^n$ is compact, then it has the following property.*

$$\text{Every open cover } \{U_\alpha : \alpha \in \mathcal{A}\} \text{ of } K \text{ has a finite subcover.} \tag{2.1.33}$$

Proof. By Lemma 2.1.7, it suffices to prove the following.

$$\begin{gathered}\text{Every countable cover } \{B_j : j \in \mathbb{N}\} \text{ of } K \text{ by open balls} \\ \text{has a finite subcover.}\end{gathered} \tag{2.1.34}$$

To see this, write $\mathcal{R} = \{B_j : j \in \mathbb{N}\}$. Given the cover $\{U_\alpha\}$, pass to $\{B_j : j \in J\}$, where $j \in J$ if and only of B_j is contained in some U_α. By (2.1.32), $\{B_j : j \in J\}$ covers K. If (2.1.34) holds, we have a subcover $\{B_\ell : \ell \in L\}$ for some finite $L \subset J$. Pick $\alpha_\ell \in \mathcal{A}$ such that $B_\ell \subset U_{\alpha_\ell}$. The $\{U_{\alpha_\ell} : \ell \in L\}$ is the desired finite subcover advertised in (2.1.33).

Finally, to prove (2.1.34), we set

$$U_m = B_1 \cup \cdots \cup B_m \tag{2.1.35}$$

and apply Corollary 2.1.6. □

Exercises

1. Identifying $z = x + iy \in \mathbb{C}$ with $(x, y) \in \mathbb{R}^2$ and $w = u + iv \in \mathbb{C}$ with $(u, v) \in \mathbb{R}^2$, show that the dot product satisfies

$$z \cdot w = \operatorname{Re} z\overline{w}.$$

2. Take $x, y \in \mathbb{R}^n$. We write

$$x \perp y \Longleftrightarrow x \cdot y = 0,$$

and say x and y are *orthogonal*. Show that

$$x \perp y \Longleftrightarrow |x + y|^2 = |x|^2 + |y|^2.$$

3. Given $x_\nu \in \mathbb{R}^n$, we say $\{x_\nu : 1 \le \nu \le m\}$ is an orthonormal set provided

$$\begin{aligned} x_\nu \cdot x_\mu = \delta_{\mu\nu} = 1 \quad &\text{if} \quad \mu = \nu, \\ 0 \quad &\text{if} \quad \mu \neq \nu. \end{aligned}$$

Show that, if $\{x_\nu : 1 \le \nu \le m\}$ is an orthonormal set, then, for $a_\nu \in \mathbb{R}$,

$$|a_1 x_1 + \cdots + a_m x_m|^2 = a_1^2 + \cdots + a_m^2.$$

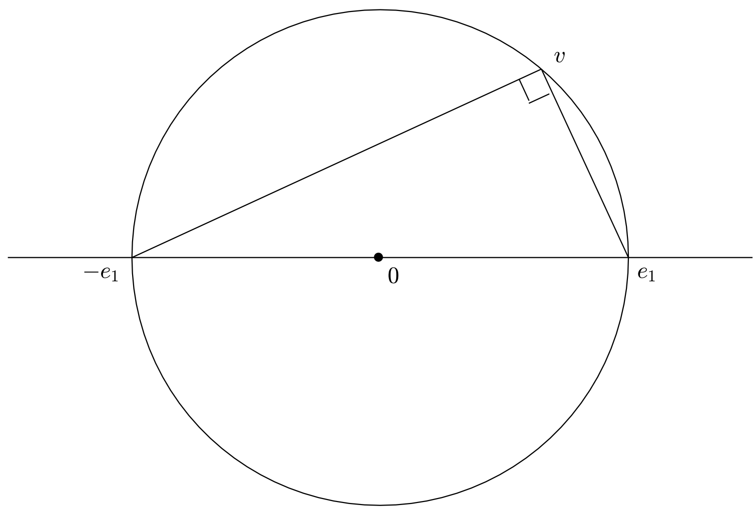

Figure 2.1.1. Right triangle in a circle

4. Assume $\{u_j : 1 \le j \le m\}$ is an orthonormal set in $\mathbb{R}^n$. Take $x \in \mathbb{R}^n$ and set

$$a_j = x \cdot u_j, \quad y = a_1 u_1 + \cdots + a_m u_m.$$

Show that

$$x - y \perp u_j, \quad \forall\, j \in \{1, \dots, m\}.$$

See Exercise 9 of §2.2 for a complementary result.

5. Show that the inequality (2.1.14) implies (2.1.13).

6. Let $e_1, v \in \mathbb{R}^n$ and assume $|e_1| = |v| = 1$. Show that

$$e_1 - v \perp e_1 + v.$$

Hint. Expand $(e_1 - v) \cdot (e_1 + v)$.
See Figure 2.1.1 for the geometrical significance of this, when $n = 2$.

7. Let $S^1 = \{x \in \mathbb{R}^2 : |x| = 1\}$ denote the unit circle in $\mathbb{R}^2$, and set $e_1 = (1, 0) \in S^1$. Pick $a \in \mathbb{R}$ such that $0 < a < 1$, and set $u = (1 - a)e_1$. See Figure 2.1.2. Then pick

$$v \in S^1 \text{ such that } v - u \perp e_1, \text{ and set } b = |v - e_1|.$$

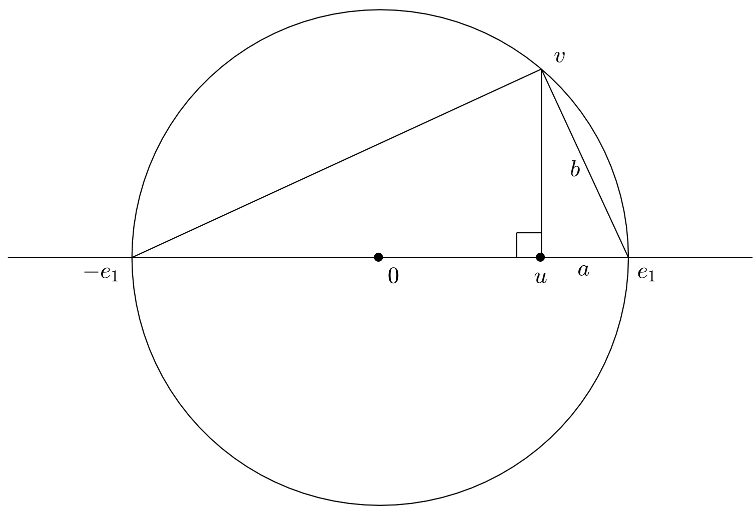

Figure 2.1.2. Geometric construction of $b = \sqrt{2a}$

Show that

$$b = \sqrt{2a}. \tag{2.1.36}$$

Hint. Note that $1 - a = u \cdot e_1 = v \cdot e_1$, hence $a = 1 - v \cdot e_1$.
Then expand $b^2 = (v - e_1) \cdot (v - e_1)$.

8. Recall the approach to (2.1.36) in classical Euclidean geometry, using similarity of triangles, leading to

$$\frac{a}{b} = \frac{b}{2}.$$

What is the relevance of Exercise 6 to this?

9. Prove Lemma 2.1.7.

10. Use Proposition 2.1.8 to prove the following extension of Proposition 2.1.5.

Proposition 2.1.9. *Let $K \subset \mathbb{R}^n$ be compact. Assume $\{X_\alpha : \alpha \in \mathcal{A}\}$ is a collection of closed subsets of K. Assume that for each finite set $\mathcal{B} \subset \mathcal{A}$, $\cap_{\alpha \in \mathcal{B}} X_\alpha \neq \emptyset$. Then*

$$\bigcap_{\alpha \in \mathcal{A}} X_\alpha \neq \emptyset.$$

Hint. Consider $U_\alpha = \mathbb{R}^n \setminus X_\alpha$.

11. Let $K \subset \mathbb{R}^n$ be compact. Show that there exist $x_0, x_1 \in K$ such that

$$|x_0| \le |x|, \quad \forall\, x \in K,$$
$$|x_1| \ge |x|, \quad \forall\, x \in K.$$

We say

$$|x_0| = \min_{x \in K} |x|, \quad |x_1| = \max_{x \in K} |x|.$$

2.2. Vector spaces and linear transformations

We have seen in §2.1 how $\mathbb{R}^n$ is a vector space, with vector operations given by (2.1.2)–(2.1.3), for row vectors, and by (2.1.4)–(2.1.5) for column vectors. We could also use complex numbers, replacing $\mathbb{R}^n$ by $\mathbb{C}^n$, and allowing $a \in \mathbb{C}$ in (2.1.3) and (2.1.5). We will use $\mathbb{F}$ to denote $\mathbb{R}$ or $\mathbb{C}$.

Many other vector spaces arise naturally. We define this general notion now. A vector space over $\mathbb{F}$ is a set V, endowed with two operations, that of vector addition and multiplication by scalars. That is, given $v, w \in V$ and $a \in \mathbb{F}$, then $v + w$ and av are defined in V. Furthermore, the following properties are to hold, for all $u, v, w \in V,\ a, b \in \mathbb{F}$. First there are laws for vector addition:

$$\begin{aligned} \text{Commutative law}: &\quad u + v = v + u, \\ \text{Associative law}: &\quad (u + v) + w = u + (v + w), \\ \text{Zero vector}: &\quad \exists\, 0 \in V,\ v + 0 = v, \\ \text{Negative}: &\quad \exists - v,\ v + (-v) = 0. \end{aligned} \tag{2.2.1}$$

Next there are laws for multiplication by scalars:

$$\begin{aligned} \text{Associative law}: &\quad a(bv) = (ab)v, \\ \text{Unit}: &\quad 1 \cdot v = v. \end{aligned} \tag{2.2.2}$$

Finally there are two distributive laws:

$$\begin{aligned} a(u + v) &= au + av, \\ (a + b)u &= au + bu. \end{aligned} \tag{2.2.3}$$

It is easy to see that $\mathbb{R}^n$ and $\mathbb{C}^n$ satisfy all these rules. We will present a number of other examples below. Let us also note that a number of other simple identities are automatic consequences of the rules given above. Here are some, which the reader is invited to verify:

$$\begin{aligned} &v + w = v \Rightarrow w = 0, \\ &v + 0 \cdot v = (1 + 0)v = v, \\ &0 \cdot v = 0, \\ &v + w = 0 \Rightarrow w = -v, \\ &v + (-1)v = 0 \cdot v = 0, \\ &(-1)v = -v. \end{aligned} \tag{2.2.4}$$

Here are some other examples of vector spaces. Let $I = [a, b]$ denote an interval in $\mathbb{R}$, and take a non-negative integer k. Then $C^k(I)$ denotes the set of functions $f : I \to \mathbb{F}$ whose derivatives up to order k are continuous. We denote by $\mathcal{P}$ the set of polynomials in x, with coefficients in $\mathbb{F}$. We denote by $\mathcal{P}_k$ the set of polynomials in x of degree $\le k$. In these various cases,

$$(f + g)(x) = f(x) + g(x), \quad (af)(x) = af(x). \tag{2.2.5}$$

Such vector spaces and certain of their linear subspaces play a major role in the material developed in these notes.

Regarding the notion just mentioned, we say a subset W of a vector space V is a linear subspace provided

$$w_j \in W,\ a_j \in \mathbb{F} \Longrightarrow a_1 w_1 + a_2 w_2 \in W. \tag{2.2.6}$$

Then W inherits the structure of a vector space.

Linear transformations and matrices

If V and W are vector spaces over $\mathbb{F}$ ($\mathbb{R}$ or $\mathbb{C}$), a map

$$T : V \longrightarrow W \tag{2.2.7}$$

is said to be a *linear transformation* provided

$$T(a_1 v_1 + a_2 v_2) = a_1 T v_1 + a_2 T v_2, \quad \forall\ a_j \in \mathbb{F},\ v_j \in V. \tag{2.2.8}$$

We also write $T \in \mathcal{L}(V, W)$. In case $V = W$, we also use the notation $\mathcal{L}(V) = \mathcal{L}(V, V)$.

Linear transformations arise in a number of ways. For example, an $m \times n$ matrix A with entries in $\mathbb{F}$ defines a linear transformation

$$A : \mathbb{F}^n \longrightarrow \mathbb{F}^m \tag{2.2.9}$$

by

$$\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} = \begin{pmatrix} \Sigma a_{1\ell} b_\ell \\ \vdots \\ \Sigma a_{m\ell} b_\ell \end{pmatrix}. \tag{2.2.10}$$

We also have linear transformations on function spaces, such as multiplication operators

$$M_f : C^k(I) \longrightarrow C^k(I), \quad M_f g(x) = f(x) g(x), \tag{2.2.11}$$

given $f \in C^k(I),\ I = [a, b]$, and the operation of differentiation:

$$D : C^{k+1}(I) \longrightarrow C^k(I), \quad D f(x) = f'(x). \tag{2.2.12}$$

We also have integration:

$$\mathcal{I} : C^k(I) \longrightarrow C^{k+1}(I), \quad \mathcal{I} f(x) = \int_a^x f(y)\, dy. \tag{2.2.13}$$

Note also that

$$D : \mathcal{P}_{k+1} \longrightarrow \mathcal{P}_k, \quad \mathcal{I} : \mathcal{P}_k \longrightarrow \mathcal{P}_{k+1}, \tag{2.2.14}$$

where $\mathcal{P}_k$ denotes the space of polynomials in x of degree $\le k$.

Two linear transformations $T_j \in \mathcal{L}(V, W)$ can be added:

$$T_1 + T_2 : V \longrightarrow W, \quad (T_1 + T_2) v = T_1 v + T_2 v. \tag{2.2.15}$$

Also $T \in \mathcal{L}(V, W)$ can be multiplied by a scalar:

$$aT : V \longrightarrow W, \quad (aT) v = a(Tv). \tag{2.2.16}$$

This makes $\mathcal{L}(V, W)$ a vector space.

We can also compose linear transformations $S \in \mathcal{L}(W, X),\ T \in \mathcal{L}(V, W)$:

$$ST : V \longrightarrow X, \quad (ST)v = S(Tv). \tag{2.2.17}$$

For example, we have

$$M_f D : C^{k+1}(I) \longrightarrow C^k(I), \quad M_f Dg(x) = f(x)g'(x), \tag{2.2.18}$$

given $f \in C^k(I)$. When two transformations

$$A : \mathbb{F}^n \longrightarrow \mathbb{F}^m, \quad B : \mathbb{F}^k \longrightarrow \mathbb{F}^n \tag{2.2.19}$$

are represented by matrices, e.g., A as in (2.2.10) and

$$B = \begin{pmatrix} b_{11} & \cdots & b_{1k} \\ \vdots & & \vdots \\ b_{n1} & \cdots & b_{nk} \end{pmatrix}, \tag{2.2.20}$$

then

$$AB : \mathbb{F}^k \longrightarrow \mathbb{F}^m \tag{2.2.21}$$

is given by matrix multiplication:

$$AB = \begin{pmatrix} \Sigma a_{1\ell} b_{\ell 1} & \cdots & \Sigma a_{1\ell} b_{\ell k} \\ \vdots & & \vdots \\ \Sigma a_{m\ell} b_{\ell 1} & \cdots & \Sigma a_{m\ell} b_{\ell k} \end{pmatrix}. \tag{2.2.22}$$

For example,

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} = \begin{pmatrix} a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} \\ a_{21}b_{11} + a_{22}b_{21} & a_{21}b_{12} + a_{22}b_{22} \end{pmatrix}.$$

Another way of writing (2.2.22) is to represent A and B as

$$A = (a_{ij}), \quad B = (b_{ij}), \tag{2.2.23}$$

and then we have

$$AB = (d_{ij}), \quad d_{ij} = \sum_{\ell=1}^{n} a_{i\ell} b_{\ell j}. \tag{2.2.24}$$

To establish the identity (2.2.22), we note that it suffices to show the two sides have the same effect on each $e_j \in \mathbb{F}^k,\ 1 \le j \le k$, where e_j is the column vector in $\mathbb{F}^k$ whose jth entry is 1 and whose other entries are 0. First note that

$$Be_j = \begin{pmatrix} b_{1j} \\ \vdots \\ b_{nj} \end{pmatrix}, \tag{2.2.25}$$

the jth column in B, as one can see via (2.2.10). Similarly, if D denotes the right side of (2.2.22), De_j is the jth column of this matrix, i.e.,

$$De_j = \begin{pmatrix} \Sigma a_{1\ell} b_{\ell j} \\ \vdots \\ \Sigma a_{m\ell} b_{\ell j} \end{pmatrix}. \tag{2.2.26}$$

On the other hand, applying A to (2.2.25), via (2.2.10), gives the same result, so (2.2.25) holds.

Associated with a linear transformation as in (2.2.7) there are two special linear spaces, the *null space* of T and the *range* of T. The null space of T is

$$\mathcal{N}(T) = \{v \in V : Tv = 0\}, \tag{2.2.27}$$

and the range of T is

$$\mathcal{R}(T) = \{Tv : v \in V\}. \tag{2.2.28}$$

Note that $\mathcal{N}(T)$ is a linear subspace of V and $\mathcal{R}(T)$ is a linear subspace of W. If $\mathcal{N}(T) = 0$ we say T is injective; if $\mathcal{R}(T) = W$ we say T is surjective. Note that T is injective if and only if T is one-to-one, i.e.,

$$Tv_1 = Tv_2 \Longrightarrow v_1 = v_2. \tag{2.2.29}$$

If T is surjective, we also say T is *onto*. If T is one-to-one and onto, we say it is an *isomorphism*. In such a case the *inverse*

$$T^{-1} : W \longrightarrow V \tag{2.2.30}$$

is well defined, and it is a linear transformation. We also say T is invertible, in such a case.

Basis and dimension

Given a finite set $S = \{v_1, \dots, v_k\}$ in a vector space V, the *span* of S is the set of vectors in V of the form

$$c_1 v_1 + \cdots + c_k v_k, \tag{2.2.31}$$

with c_j arbitrary scalars, ranging over $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$. This set, denoted Span(S) is a linear subspace of V. The set S is said to be *linearly dependent* if and only if there exist scalars $c_1, \dots, c_k$, not all zero, such that (2.2.31) vanishes. Otherwise we say S is *linearly independent*.

If $\{v_1, \dots, v_k\}$ is linearly independent, we say S is a *basis* of Span(S), and that k is the *dimension* of Span(S). In particular, if this holds and Span(S) $= V$, we say $k = \dim V$. We also say V has a finite basis, and that V is finite dimensional.

By convention, if V has only one element, the zero element, we say $V = 0$ and $\dim V = 0$.

It is easy to see that any finite set $S = \{v_1, \dots, v_k\} \subset V$ has a maximal subset that is linearly independent, and such a subset has the same span as S, so Span(S) has a basis. To take a complementary perspective, S will have a minimal subset S_0 with the same span, and any such minimal subset will be a basis of Span(S). Soon we will show that any two bases of a finite-dimensional vector space V have the same number of elements (so $\dim V$ is well defined). First, let us relate V to $\mathbb{F}^k$.

So say V has a basis $S = \{v_1, \dots, v_k\}$. We define a linear transformation

$$\mathcal{J}_S : \mathbb{F}^k \longrightarrow V \tag{2.2.32}$$

by

$$\mathcal{J}_S(c_1 e_1 + \cdots + c_k e_k) = c_1 v_1 + \cdots + c_k v_k, \tag{2.2.33}$$

where

$$e_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \dots\dots, e_k = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}. \tag{2.2.34}$$

We say $\{e_1, \dots, e_k\}$ is the standard basis of $\mathbb{F}^k$. The linear independence of S is equivalent to the injectivity of $\mathcal{J}_S$ and the statement that S spans V is equivalent to the surjectivity of $\mathcal{J}_S$. Hence the statement that S is a basis of V is equivalent to the statement that $\mathcal{J}_S$ is an isomorphism, with inverse uniquely specified by

$$\mathcal{J}_S^{-1}(c_1 v_1 + \cdots + c_k v_k) = c_1 e_1 + \cdots + c_k e_k. \tag{2.2.35}$$

We begin our demonstration that $\dim V$ is well defined, with the following concrete result.

Lemma 2.2.1. *If $v_1, \dots, v_{k+1}$ are vectors in $\mathbb{F}^k$, then they are linearly dependent.*

Proof. We use induction on k. The result is obvious if $k = 1$. We can suppose the last component of some v_j is nonzero, since otherwise we can regard these vectors as elements of $\mathbb{F}^{k-1}$ and use the inductive hypothesis. Reordering these vectors, we can assume the last component of v_{k+1} is nonzero, and it can be assumed to be 1. Form

$$w_j = v_j - v_{kj} v_{k+1}, \quad 1 \le j \le k,$$

where $v_j = (v_{1j}, \dots, v_{kj})^t$. Then the last component of each of the vectors $w_1, \dots, w_k$ is 0, so we can regard these as k vectors in $\mathbb{F}^{k-1}$. By induction, there exist scalars $a_1, \dots, a_k$, not all zero, such that

$$a_1 w_1 + \cdots + a_k w_k = 0,$$

so we have

$$a_1 v_1 + \cdots + a_k v_k = (a_1 v_{k1} + \cdots + a_k v_{kk}) v_{k+1},$$

the desired linear dependence relation on $\{v_1, \dots, v_{k+1}\}$. □

With this result in hand, we proceed.

Proposition 2.2.2. *If V has a basis $S = \{v_1, \dots, v_k\}$ with k elements and if the set $\{w_1, \dots, w_\ell\} \subset V$ is linearly independent, then $\ell \le k$.*

Proof. Take the isomorphism $\mathcal{J}_S : \mathbb{F}^k \to V$ described in (2.2.32)–(2.2.33). The hypotheses imply that $\{\mathcal{J}_S^{-1} w_1, \dots, \mathcal{J}_S^{-1} w_\ell\}$ is linearly independent in $\mathbb{F}^k$, so Lemma 2.2.1 implies $\ell \le k$. □

Corollary 2.2.3. *If V is finite-dimensional, any two bases of V have the same number of elements. If V is isomorphic to W, these spaces have the same dimension.*

Proof. If S (with $\#S$ elements) and T are bases of V, we have $\#S \le \#T$ and $\#T \le \#S$, hence $\#S = \#T$. For the latter part, an isomorphism of V onto W takes a basis of V to a basis of W. □

The following is an easy but useful consequence.

Proposition 2.2.4. *If V is finite dimensional and $W \subset V$ a linear subspace, then W has a finite basis, and* $\dim W \le \dim V$.

Proof. Suppose $\{w_1, \dots, w_\ell\}$ is a linearly independent subset of W. Proposition 2.2.2 implies $\ell \le \dim V$. If this set spans W, we are done. If not, there is an element $w_{\ell+1} \in W$ not in this span, and $\{w_1, \dots, w_{\ell+1}\}$ is a linearly independent subset of W. Again $\ell + 1 \le \dim V$. Continuing this process a finite number of times must produce a basis of W. □

A similar argument establishes:

Proposition 2.2.5. *Suppose V is finite dimensional, $W \subset V$ a linear subspace, and $\{w_1, \dots, w_\ell\}$ a basis of W. Then V has a basis of the form $\{w_1, \dots, w_\ell, u_1, \dots, u_m\}$, and $\ell + m = \dim V$.*

Having this, we can establish the following result, sometimes called the fundamental theorem of linear algebra.

Proposition 2.2.6. *Assume V and W are vector spaces, V finite dimensional, and*

$$A : V \longrightarrow W \tag{2.2.36}$$

a linear map. Then

$$\dim \mathcal{N}(A) + \dim \mathcal{R}(A) = \dim V. \tag{2.2.37}$$

Proof. Let $\{w_1, \dots, w_\ell\}$ be a basis of $\mathcal{N}(A) \subset V$, and complete it to a basis

$$\{w_1, \dots, w_\ell, u_1, \dots, u_m\}$$

of V. Set $L = \text{Span}\{u_1, \dots, u_m\}$, and consider

$$A_0 : L \longrightarrow W, \quad A_0 = A\big|_L. \tag{2.2.38}$$

Clearly $w \in \mathcal{R}(A) \Rightarrow w = A(a_1 w_1 + \cdots + a_\ell w_\ell + b_1 u_1 + \cdots + b_m u_m) = A_0(b_1 u_1 + \cdots + b_m u_m)$, so

$$\mathcal{R}(A_0) = \mathcal{R}(A). \tag{2.2.39}$$

Furthermore,

$$\mathcal{N}(A_0) = \mathcal{N}(A) \cap L = 0. \tag{2.2.40}$$

Hence $A_0 : L \to \mathcal{R}(A_0)$ is an isomorphism. Thus $\dim \mathcal{R}(A) = \dim \mathcal{R}(A_0) = \dim L = m$, and we have (2.2.37). □

The following is a significant special case.

Corollary 2.2.7. *Let V be finite dimensional, and let $A : V \to V$ be linear. Then*

$$A \ \text{ injective} \Longleftrightarrow A \ \text{ surjective} \Longleftrightarrow A \ \text{ isomorphism.}$$

We mention that these equivalences can fail for infinite dimensional spaces. For example, if $\mathcal{P}$ denotes the space of polynomials in x, then $M_x : \mathcal{P} \to \mathcal{P}$ ($M_x f(x) = x f(x)$) is injective but not surjective, while $D : \mathcal{P} \to \mathcal{P}$ ($Df(x) = f'(x)$) is surjective but not injective.

Next we have the following important characterization of injectivity and surjectivity.

Proposition 2.2.8. *Assume V and W are finite dimensional and $A : V \to W$ is linear. Then*

$$A \;\; \text{surjective} \Longleftrightarrow AB = I_W, \;\; \text{for some } B \in \mathcal{L}(W, V), \tag{2.2.41}$$

and

$$A \;\text{injective} \Longleftrightarrow CA = I_V, \;\; \text{for some } C \in \mathcal{L}(W, V). \tag{2.2.42}$$

Proof. Clearly $AB = I \Rightarrow A$ surjective and $CA = I \Rightarrow A$ injective. We establish the converses.

First assume $A : V \to W$ is surjective. Let $\{w_1, \dots, w_\ell\}$ be a basis of W. Pick $v_j \in V$ such that $Av_j = w_j$. Set

$$B(a_1 w_1 + \cdots + a_\ell w_\ell) = a_1 v_1 + \cdots + a_\ell v_\ell. \tag{2.2.43}$$

This works in (2.2.41).

Next assume $A : V \to W$ is injective. Let $\{v_1, \dots, v_k\}$ be a basis of V. Set $w_j = Av_j$. Then $\{w_1, \dots, w_k\}$ is linearly independent, hence a basis of $\mathcal{R}(A)$, and we then can produce a basis $\{w_1, \dots, w_k, u_1, \dots, u_m\}$ of W. Set

$$C(a_1 w_1 + \cdots + a_k w_k + b_1 u_1 + \cdots + b_m u_m) = a_1 v_1 + \cdots + a_k v_k. \tag{2.2.44}$$

This works in (2.2.42). □

An $m \times n$ matrix A defines a linear transformation $A : \mathbb{F}^n \to \mathbb{F}^m$, as in (2.2.9)–(2.2.10). The columns of A are

$$a_j = \begin{pmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{pmatrix}. \tag{2.2.45}$$

As seen in (2.2.25),

$$Ae_j = a_j, \tag{2.2.46}$$

where $e_1, \dots, e_n$ is the standard basis of $\mathbb{F}^n$. Hence

$$\mathcal{R}(A) = \text{ linear span of the columns of } A, \tag{2.2.47}$$

so

$$\mathcal{R}(A) = \mathbb{F}^m \Longleftrightarrow a_1, \dots, a_n \text{ span } \mathbb{F}^m. \tag{2.2.48}$$

Furthermore,

$$A\Bigl(\sum_{j=1}^n c_j e_j\Bigr) = 0 \Longleftrightarrow \sum_{j=1}^n c_j a_j = 0, \tag{2.2.49}$$

so

$$\mathcal{N}(A) = 0 \Longleftrightarrow \{a_1, \dots, a_n\} \text{ is linearly independent.} \tag{2.2.50}$$

We have the following conclusion, in case $m = n$.

Proposition 2.2.9. *Let A be an $n \times n$ matrix, defining $A : \mathbb{F}^n \to \mathbb{F}^n$. Then the following are equivalent:*

$$\begin{aligned} &A \text{ is invertible}, \\ &\text{The columns of } A \text{ are linearly independent}, \\ &\text{The columns of } A \text{ span } \mathbb{F}^n. \end{aligned} \tag{2.2.51}$$

Exercises

1. Show that the results in (2.2.4) follow from the basic rules (2.2.1)–(2.2.3).
Hint. To start, add $-v$ to both sides of the identity $v + w = v$, and take account first of the associative law in (2.2.1), and then of the rest of (2.2.1). For the second line of (2.2.4), use the rules (2.2.2) and (2.2.3). Then use the first two lines of (2.2.4) to justify the third line...

2. Demonstrate the following results for any vector space. Take $a \in \mathbb{F},\ v \in V$.

$$\begin{aligned} a \cdot 0 &= 0 \in V, \\ a(-v) &= -av. \end{aligned}$$

Hint. Feel free to use the results of (2.2.4).

Let V be a vector space (over $\mathbb{F}$) and $W, X \subset V$ linear subspaces. We say

$$V = W + X \tag{2.2.52}$$

provided each $v \in V$ can be written

$$v = w + x, \quad w \in W,\ x \in X. \tag{2.2.53}$$

We say

$$V = W \oplus X \tag{2.2.54}$$

provided each $v \in V$ has a unique representation (2.2.53).

3. Show that

$$V = W \oplus X \Longleftrightarrow V = W + X \text{ and } W \cap X = 0.$$

4. Let $A : \mathbb{F}^n \to \mathbb{F}^m$ be defined by an $m \times n$ matrix, as in (2.2.9)–(2.2.10).
(a) Show that $\mathcal{R}(A)$ is the span of the columns of A.
Hint. See (2.2.25).
(b) Show that $\mathcal{N}(A) = 0$ if and only if the columns of A are linearly independent.

5. Define the transpose of an $m \times n$ matrix $A = (a_{jk})$ to be the $n \times m$ matrix

$A^t = (a_{kj})$. Thus, if A is as in (2.2.9)–(2.2.10),

$$A^t = \begin{pmatrix} a_{11} & \cdots & a_{m1} \\ \vdots & & \vdots \\ a_{1n} & \cdots & a_{mn} \end{pmatrix}. \tag{2.2.55}$$

For example,

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix} \Longrightarrow A^t = \begin{pmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{pmatrix}.$$

Suppose also B is an $n \times k$ matrix, as in (2.2.20), so AB is defined, as in (2.2.21). Show that

$$(AB)^t = B^t A^t. \tag{2.2.56}$$

6. Let

$$A = \begin{pmatrix} 1 & 2 & 3 \end{pmatrix}, \quad B = \begin{pmatrix} 2 \\ 0 \\ 2 \end{pmatrix}.$$

Compute AB and BA. Then compute $A^t B^t$ and $B^t A^t$.

7. Let $\mathcal{P}_5$ be the space of real polynomials in x of degree ≤ 5 and set

$$T : \mathcal{P}_5 \longrightarrow \mathbb{R}^3, \quad Tp = \bigl(p(-1), p(0), p(1)\bigr).$$

Specify $\mathcal{R}(T)$ and $\mathcal{N}(T)$, and verify (2.2.37) for $V = \mathcal{P}_5,\ W = \mathbb{R}^3,\ A = T$.

8. Denote the space of $m \times n$ matrices with entries in $\mathbb{F}$ (as in (2.2.10)) by

$$M(m \times n, \mathbb{F}). \tag{2.2.57}$$

If $m = n$, denote it by

$$M(n, \mathbb{F}). \tag{2.2.58}$$

Show that

$$\dim M(m \times n, \mathbb{F}) = mn,$$

especially

$$\dim M(n, \mathbb{F}) = n^2.$$

9. Assume $\{u_j : 1 \le j \le n\}$ is an orthonormal set in $\mathbb{R}^n$. Pick $x \in \mathbb{R}^n$ and set $a_j = x \cdot u_j$. Show that

$$x = a_1 u_1 + \cdot + a_n u_n.$$

We say $\{u_j : 1 \le j \le n\}$ is an orthonormal basis of $\mathbb{R}^n$.
Hint. Show that this orthonormal set is linearly independent, and deduce that it spans $\mathbb{R}^n$. Then see Exercise 4 of §2.1.

Given $T \in M(n, \mathbb{R})$, we say

$$T \in O(n) \Longleftrightarrow T^t T = I,$$

or equivalently, if and only if

$$Tx \cdot Ty = x \cdot y, \quad \forall\, x, y \in \mathbb{R}^n.$$

10. Show that, given $T \in M(n, \mathbb{R})$,

$$T \in O(n) \Longleftrightarrow \text{the columns of } T \text{ form an orthonormal basis of } \mathbb{R}^n.$$

11. Given $T \in M(n, \mathbb{R})$, show that

$$T \in O(n) \Longleftrightarrow |Tx| = |x|, \quad \forall\, x \in \mathbb{R}^n.$$

Hint. Expand $|T(x + y)|^2 = (Tx + Ty) \cdot (Tx + Ty)$.

12. If $\{u_j : 1 \le j \le n\}$ and $\{v_j : 1 \le j \le n\}$ are both orthonormal bases of $\mathbb{R}^n$, show that there is a unique $T \in O(n)$ such that

$$Tu_j = v_j, \quad \forall j \in \{1, \dots, n\}.$$

13. Take a peek at §C.1 and show that if $V \subset \mathbb{R}^n$ is a linear subspace, then V has an orthonormal basis. Going further, show that $\mathbb{R}^n$ has an orthonormal basis $\{u_1, \dots, u_n\}$ such that $\{u_1, \dots u_d\}$ is a basis of V, where $d = \dim V$.

14. Let $\{e_j : 1 \le j \le n\}$ denote the standard basis of $\mathbb{R}^n$. Assume $n \ge 3$. Let $V \subset \mathbb{R}^n$ be a 2-dimensional subspace, with basis $\{u, v\}$, $|u| = 1$. Show that there exists $T \in O(n)$ such that

$$Tu = e_1, \quad Tv \in \operatorname{Span}\{e_1, e_2\}.$$

2.3. Determinants

Determinants arise in the study of inverting a matrix. To take the 2×2 case, solving for x and y the system

$$\begin{aligned} ax + by &= u, \\ cx + dy &= v \end{aligned} \tag{2.3.1}$$

can be done by multiplying these equations by d and b, respectively, and subtracting, and by multiplying them by c and a, respectively, and subtracting, yielding

$$\begin{aligned} (ad - bc)x &= du - bv, \\ (ad - bc)y &= av - cu. \end{aligned} \tag{2.3.2}$$

The factor on the left is

$$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc, \tag{2.3.3}$$

and solving (2.3.2) for x and y leads to

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \Longrightarrow A^{-1} = \frac{1}{\det A} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}, \tag{2.3.4}$$

provided $\det A \neq 0$.

We now consider determinants of $n \times n$ matrices. Let $M(n, \mathbb{F})$ denote the set of $n \times n$ matrices with entries in $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$. We write

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix} = (a_1, \dots, a_n), \tag{2.3.5}$$

where

$$a_j = \begin{pmatrix} a_{1j} \\ \vdots \\ a_{nj} \end{pmatrix} \tag{2.3.6}$$

is the jth column of A. The determinant is defined as follows.

Proposition 2.3.1. *There is a unique function*

$$\vartheta : M(n, \mathbb{F}) \longrightarrow \mathbb{F}, \tag{2.3.7}$$

satisfying the following three properties:

(a) ϑ *is linear in each column* a_j *of* A*,*
(b) $\vartheta(\widetilde{A}) = -\vartheta(A)$ *if* $\widetilde{A}$ *is obtained from* A *by interchanging two columns,*
(c) $\vartheta(I) = 1$.

This defines the determinant:

$$\vartheta(A) = \det A. \tag{2.3.8}$$

If (c) is replaced by

(c′) $\vartheta(I) = r$,

then

$$\vartheta(A) = r \det A. \tag{2.3.9}$$

The proof will involve constructing an explicit formula for $\det A$ by following the rules (a)–(c). We start with the case $n = 3$. We have

$$\det A = \sum_{j=1}^{3} a_{j1} \det(e_j, a_2, a_3), \tag{2.3.10}$$

by applying (a) to the first column of A, $a_1 = \sum_j a_{j1} e_j$. Here and below, $\{e_j : 1 \le j \le n\}$ denotes the standard basis of $\mathbb{F}^n$, so e_j has a 1 in the jth slot and 0s elsewhere. Applying (a) to the second and third columns gives

$$\begin{aligned} \det A &= \sum_{j,k=1}^{3} a_{j1} a_{k2} \det(e_j, e_k, a_3) \\ &= \sum_{j,k,\ell=1}^{3} a_{j1} a_{k2} a_{\ell 3} \det(e_j, e_k, e_\ell). \end{aligned} \tag{2.3.11}$$

This is a sum of 27 terms, but most of them are 0. Note that rule (b) implies

$$\det B = 0 \quad \text{whenever} \quad B \quad \text{has two identical columns.} \tag{2.3.12}$$

Hence $\det(e_j, e_k, e_\ell) = 0$ unless j, k, and ℓ are distinct, that is, unless (j, k, ℓ) is a *permutation* of $(1, 2, 3)$. Now rule (c) says

$$\det(e_1, e_2, e_3) = 1, \tag{2.3.13}$$

and we see from rule (b) that $\det(e_j, e_k, e_\ell) = 1$ if one can convert (e_j, e_k, e_ℓ) to (e_1, e_2, e_3) by an even number of column interchanges, and $\det(e_j, e_k, e_\ell) = -1$ if it takes an odd number of interchanges. Explicitly,

$$\begin{aligned} &\det(e_1, e_2, e_3) = 1, \quad \det(e_1, e_3, e_2) = -1, \\ &\det(e_2, e_3, e_1) = 1, \quad \det(e_2, e_1, e_3) = -1, \\ &\det(e_3, e_1, e_2) = 1, \quad \det(e_3, e_2, e_1) = -1. \end{aligned} \tag{2.3.14}$$

Consequently (2.3.11) yields

$$\begin{aligned} \det A = a_{11} a_{22} a_{33} &- a_{11} a_{32} a_{23} \\ + a_{21} a_{32} a_{13} &- a_{21} a_{12} a_{33} \\ + a_{31} a_{12} a_{23} &- a_{31} a_{22} a_{13}. \end{aligned} \tag{2.3.15}$$

Note that the second indices occur in $(1, 2, 3)$ order in each product. We can rearrange these products so that the *first* indices occur in $(1, 2, 3)$ order:

$$\begin{aligned} \det A = a_{11} a_{22} a_{33} &- a_{11} a_{23} a_{32} \\ + a_{13} a_{21} a_{32} &- a_{12} a_{21} a_{33} \\ + a_{12} a_{23} a_{31} &- a_{13} a_{22} a_{31}. \end{aligned} \tag{2.3.16}$$

Now we tackle the case of general n. Parallel to (2.3.10)–(2.3.11), we have

$$\begin{aligned}\det A &= \sum_j a_{j1} \det(e_j, a_2, \dots, a_n) = \cdots \\ &= \sum_{j_1,\dots,j_n} a_{j_1 1} \cdots a_{j_n n} \det(e_{j_1}, \dots e_{j_n}),\end{aligned} \tag{2.3.17}$$

by applying rule (a) to each of the n columns of A. As before, (2.3.12) implies $\det(e_{j_1}, \dots, e_{j_n}) = 0$ unless $(j_1, \dots, j_n)$ are all distinct, that is, unless $(j_1, \dots, j_n)$ is a permutation of the set $(1, 2, \dots, n)$. We set

$$S_n = \text{ set of permutations of } (1, 2, \dots, n). \tag{2.3.18}$$

That is, S_n consists of elements σ, mapping the set $\{1, \dots, n\}$ to itself,

$$\sigma : \{1, 2, \dots, n\} \longrightarrow \{1, 2, \dots, n\}, \tag{2.3.19}$$

that are one-to-one and onto. We can compose two such permutations, obtaining the product $\sigma\tau \in S_n$, given σ and τ in S_n. A permutation that interchanges just two elements of $\{1, \dots, n\}$, say j and k ($j \neq k$), is called a *transposition*, and labeled (jk). It is easy to see that each permutation of $\{1, \dots, n\}$ can be achieved by successively transposing pairs of elements of this set. That is, each element $\sigma \in S_n$ is a product of transpositions. We claim that

$$\det(e_{\sigma(1)1}, \dots, e_{\sigma(n)n}) = (\operatorname{sgn} \sigma) \det(e_1, \dots, e_n) = \operatorname{sgn} \sigma, \tag{2.3.20}$$

where

$$\begin{aligned}\operatorname{sgn} \sigma = \ & 1 \quad \text{if } \sigma \text{ is a product of an even number of transpositions,} \\ & -1 \quad \text{if } \sigma \text{ is a product of an odd number of transpositions.}\end{aligned} \tag{2.3.21}$$

In fact, the first identity in (2.3.20) follows from rule (b) and the second identity from rule (c).

There is one point to be checked here. Namely, we claim that a given $\sigma \in S_n$ cannot simultaneously be written as a product of an even number of transpositions and an odd number of transpositions. If σ could be so written, $\operatorname{sgn} \sigma$ would not be well defined, and it would be impossible to satisfy condition (b), so Proposition 2.3.1 would fail. One neat way to see that $\operatorname{sgn} \sigma$ is well defined is the following. Let $\sigma \in S_n$ act on functions of n variables by

$$(\sigma f)(x_1, \dots, x_n) = f(x_{\sigma(1)}, \dots, x_{\sigma(n)}). \tag{2.3.22}$$

It is readily verified that if also $\tau \in S_n$,

$$g = \sigma f \Longrightarrow \tau g = (\tau\sigma) f. \tag{2.3.23}$$

Now, let P be the polynomial

$$P(x_1, \dots, x_n) = \prod_{1 \le j < k \le n} (x_j - x_k). \tag{2.3.24}$$

One readily has

$$(\sigma P)(x) = -P(x), \quad \text{whenever } \ \sigma \ \text{ is a transposition,} \tag{2.3.25}$$

and hence, by (2.3.23),

$$(\sigma P)(x) = (\operatorname{sgn} \sigma) P(x), \quad \forall \, \sigma \in S_n, \tag{2.3.26}$$

and $\operatorname{sgn} \sigma$ is well defined.

The proof of (2.3.20) is complete, and substitution into (2.3.17) yields the formula

$$\det A = \sum_{\sigma\in S_n} (\operatorname{sgn}\sigma) a_{\sigma(1)1}\cdots a_{\sigma(n)n}. \tag{2.3.27}$$

It is routine to check that this satisfies the properties (a)–(c). Regarding (b), note that if $\vartheta(A)$ denotes the right side of (2.3.27) and $\widetilde{A}$ is obtained from A by applying a permutation τ to the columns of A, so $\widetilde{A} = (a_{\tau(1)},\dots,a_{\tau(n)})$, then

$$\begin{aligned}\vartheta(\widetilde{A}) &= \sum_{\sigma\in S_n} (\operatorname{sgn}\sigma) a_{\sigma(1)\tau(1)}\cdots a_{\sigma(n)\tau(n)} \\ &= \sum_{\sigma\in S_n} (\operatorname{sgn}\sigma) a_{\sigma\tau^{-1}(1)1}\cdots a_{\sigma\tau^{-1}(n)n} \\ &= \sum_{\omega\in S_n} (\operatorname{sgn}\omega\tau) a_{\omega(1)1}\cdots a_{\omega(n)n} \\ &= (\operatorname{sgn}\tau)\vartheta(A),\end{aligned}$$

the last identity because

$$\operatorname{sgn}\omega\tau = (\operatorname{sgn}\omega)(\operatorname{sgn}\tau), \quad \forall\, \omega,\tau\in S_n.$$

As for the final part of Proposition 2.3.1, if (c) is replaced by (c′), then (2.3.20) is replaced by

$$\vartheta(e_{\sigma(1)},\dots,e_{\sigma(n)}) = r(\operatorname{sgn}\sigma), \tag{2.3.28}$$

and (2.3.9) follows.

Remark. (2.3.27) is taken as a definition of the determinant by some authors. While it is a useful *formula* for the determinant, it is a bad *definition*, which has perhaps led to a bit of fear and loathing among math students.

Remark. Here is another formula for $\operatorname{sgn}\sigma$, which the reader is invited to verify. If $\sigma\in S_n$,

$$\operatorname{sgn}\sigma = (-1)^{\kappa(\sigma)},$$

where

$$\begin{aligned}\kappa(\sigma) = {}&\text{number of pairs } (j,k) \text{ such that } 1\le j<k\le n,\\ &\text{but } \sigma(j)>\sigma(k).\end{aligned}$$

Note that

$$a_{\sigma(1)1}\cdots a_{\sigma(n)n} = a_{1\tau(1)}\cdots a_{n\tau(n)}, \quad \text{with } \ \tau=\sigma^{-1}, \tag{2.3.29}$$

and $\operatorname{sgn}\sigma = \operatorname{sgn}\sigma^{-1}$, so, parallel to (2.3.16), we also have

$$\det A = \sum_{\sigma\in S_n} (\operatorname{sgn}\sigma) a_{1\sigma(1)}\cdots a_{n\sigma(n)}. \tag{2.3.30}$$

Comparison with (2.3.27) gives

$$\det A = \det A^t, \tag{2.3.31}$$

where $A = (a_{jk}) \Rightarrow A^t = (a_{kj})$. Note that the jth column of A^t has the same entries as the jth row of A. In light of this, we have:

Corollary 2.3.2. *In Proposition 2.3.1, one can replace "columns" by "rows."*

The following is a key property of the determinant, called multiplicativity:

Proposition 2.3.3. *Given A and B in $M(n, \mathbb{F})$,*

$$\det(AB) = (\det A)(\det B). \tag{2.3.32}$$

Proof. For fixed A, apply Proposition 2.3.1 to

$$\vartheta_1(B) = \det(AB). \tag{2.3.33}$$

If $B = (b_1, \dots, b_n)$, with jth column b_j, then

$$AB = (Ab_1, \dots, Ab_n). \tag{2.3.34}$$

Clearly rule (a) holds for ϑ_1. Also, if $\widetilde{B} = (b_{\sigma(1)}, \dots, b_{\sigma(n)})$ is obtained from B by permuting its columns, then $A\widetilde{B}$ has columns $(Ab_{\sigma(1)}, \dots, Ab_{\sigma(n)})$, obtained by permuting the columns of AB in the same fashion. Hence rule (b) holds for ϑ_1. Finally, rule (c′) holds for ϑ_1, with $r = \det A$, and (2.3.32) follows, □

Corollary 2.3.4. *If $A \in M(n, \mathbb{F})$ is invertible, then* $\det A \neq 0$.

Proof. If A is invertible, there exists $B \in M(n, \mathbb{F})$ such that $AB = I$. Then, by (2.3.32), $(\det A)(\det B) = 1$, so $\det A \neq 0$. □

The converse of Corollary 2.3.4 also holds. Before proving it, it is convenient to show that the determinant is invariant under a certain class of column operations, given as follows.

Proposition 2.3.5. *If $\widetilde{A}$ is obtained from $A = (a_1, \dots, a_n) \in M(n, \mathbb{F})$ by adding ca_ℓ to a_k for some $c \in \mathbb{F}$, $\ell \neq k$, then*

$$\det \widetilde{A} = \det A. \tag{2.3.35}$$

Proof. By rule (a), $\det \widetilde{A} = \det A + c \det A^b$, where A^b is obtained from A by replacing the column a_k by a_ℓ. Hence A^b has two identical columns, so $\det A^b = 0$, and (2.3.35) holds. □

We now extend Corollary 2.3.4.

Proposition 2.3.6. *If $A \in M(n, \mathbb{F})$, then A is invertible if and only if* $\det A \neq 0$.

Proof. We have half of this from Corollary 2.3.4. To finish, assume A is not invertible. As seen in §2.2, this implies the columns $a_1, \dots, a_n$ of A are linearly dependent. Hence, for some k,

$$a_k + \sum_{\ell \neq k} c_\ell a_\ell = 0, \tag{2.3.36}$$

with $c_\ell \in \mathbb{F}$. Now we can apply Proposition 2.3.5 to obtain $\det A = \det \widetilde{A}$, where $\widetilde{A}$ is obtained by adding $\sum c_\ell a_\ell$ to a_k. But then the kth column of $\widetilde{A}$ is 0, so $\det A = \det \widetilde{A} = 0$. This finishes the proof of Proposition 2.3.6. □

Further useful facts about determinants arise in the following exercises.

Exercises

1. Show that

$$\det \begin{pmatrix} 1 & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ 0 & a_{n2} & \cdots & a_{nn} \end{pmatrix} = \det \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ 0 & a_{n2} & \cdots & a_{nn} \end{pmatrix} = \det A_{11} \tag{2.3.37}$$

where $A_{11} = (a_{jk})_{2\le j,k\le n}$.

Hint. Do the first identity using Proposition 2.3.5. Then exploit uniqueness for det on $M(n-1, \mathbb{F})$.

2. Deduce that $\det(e_j, a_2, \dots, a_n) = (-1)^{j-1} \det A_{1j}$ where A_{kj} is formed by deleting the kth column and the jth row from A.

3. Deduce from the first sum in (2.3.17) that

$$\det A = \sum_{j=1}^{n} (-1)^{j-1} a_{j1} \det A_{1j}. \tag{2.3.38}$$

More generally, for any $k \in \{1, \dots, n\}$,

$$\det A = \sum_{j=1}^{n} (-1)^{j-k} a_{jk} \det A_{kj}. \tag{2.3.39}$$

This is called an expansion of $\det A$ by minors, down the kth column.

4. By definition, the cofactor matrix of A is given by

$$\mathrm{Cof}(A)_{jk} = c_{kj} = (-1)^{j-k} \det A_{kj}.$$

Show that

$$\sum_{j=1}^{n} a_{j\ell} c_{kj} = 0, \quad \text{if } \ell \neq k. \tag{2.3.40}$$

Deduce from this and (5.39) that

$$\mathrm{Cof}(A)^t A = (\det A) I. \tag{2.3.41}$$

Hint. Reason as in Exercises 1–3 that the left side of (2.3.40) is equal to

$$\det(a_1, \dots, a_\ell, \dots, a_\ell, \dots, a_n),$$

with a_ℓ in the kth column as well as in the ℓth column. The identity (2.3.41) is known as Cramer's formula. Note how this generalizes (2.3.4).

5. Show that

$$\det \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ & a_{22} & \cdots & a_{2n} \\ & & \ddots & \vdots \\ & & & a_{nn} \end{pmatrix} = a_{11}a_{22}\cdots a_{nn}. \tag{2.3.42}$$

Hint. Use (2.3.37) and induction. *Alternative*: Use (2.3.27). Show that $\sigma \in S_n,\ \sigma(k) \le k \,\forall\, k \Rightarrow \sigma(k) \equiv k$.

6. Recall that $O(n) = \{T \in M(n, \mathbb{R}) : T^t T = I\}$. Show that

$$T \in O(n) \Longrightarrow \det T = \pm 1.$$

We say

$$SO(n) = \{T \in O(n) : \det T = 1\}.$$

Show that

$$S, T \in SO(n) \Longrightarrow ST,\ T^{-1} \in SO(n).$$

2.4. The trace of a matrix, and the Euclidean structure of $M(n,\mathbb{R})$

Let $A \in M(n,\mathbb{R})$ be as in (2.3.5). We define the *trace* of A as

$$\operatorname{Tr} A = \sum_{j=1}^{n} a_{jj}. \tag{2.4.1}$$

If also $B = (b_{jk}) \in M(n,\mathbb{R})$, we have

$$AB = C, \quad c_{jk} = \sum_{\ell} a_{j\ell} b_{\ell k}, \tag{2.4.2}$$

so

$$\operatorname{Tr} AB = \sum_{j,\ell} a_{j\ell} b_{\ell j}, \tag{2.4.3}$$

from which we deduce that

$$\operatorname{Tr} AB = \operatorname{Tr} BA. \tag{2.4.4}$$

Replacing B by B^t in (2.4.2)–(2.4.3) gives

$$\operatorname{Tr} AB^t = \sum_{j,\ell} a_{j\ell} b_{j\ell}, \tag{2.4.5}$$

which is just the Euclidean dot product on

$$M(n,\mathbb{R}) \approx \mathbb{R}^{n^2}. \tag{2.4.6}$$

We denote it by

$$\langle A, B\rangle = \operatorname{Tr} AB^t. \tag{2.4.7}$$

Note also that

$$\begin{aligned} \langle A, B\rangle &= \langle B, A\rangle \\ &= \operatorname{Tr} BA^t \\ &= \operatorname{Tr} A^t B. \end{aligned} \tag{2.4.8}$$

We denote the corresponding Euclidean norm on $A \in M(n,\mathbb{R})$ (called the *Hilbert-Schmidt norm*) by $\|A\|_{\mathrm{HS}}$:

$$\|A\|_{\mathrm{HS}}^2 = \langle A, A\rangle = \operatorname{Tr} AA^t. \tag{2.4.9}$$

In other words,

$$\|A\|_{\mathrm{HS}}^2 = \sum_{j,k} a_{jk}^2. \tag{2.4.10}$$

Note that the Cauchy inequality (2.1.17), applied to $\mathbb{R}^{n^2}$, yields

$$|\langle A, B\rangle| \le \|A\|_{\mathrm{HS}} \|B\|_{\mathrm{HS}}. \tag{2.4.11}$$

Exercises

1. Deduce from (2.4.4) that if $B \in M(n, \mathbb{R})$ is invertible,
$$\operatorname{Tr} B^{-1}AB = \operatorname{Tr} A.$$

2. Recall that
$$O(n) = \{T \in M(n, \mathbb{R}) : T^t T = I\}.$$
Show that, if $A, B \in M(n, \mathbb{R})$ and $T \in O(n)$, then
$$\begin{aligned} \langle A, B \rangle &= \langle TA, TB \rangle \\ &= \langle AT, BT \rangle. \end{aligned}$$

3. Take $A \in M(n, \mathbb{R})$. Show that, as $t \to 0$,
$$\det(I + tA) = 1 + t \operatorname{Tr} A + O(t^2).$$

4. Deduce from the previous exercise that, if $B \in M(n, \mathbb{R})$ is invertible,
$$\det(B + tA) = (\det B)(1 + t \operatorname{Tr} B^{-1}A) + O(t^2).$$

2.5. The cross product on $\mathbb{R}^3$

If $u, v \in \mathbb{R}^3$, we define the cross product $u \times v = \Pi(u, v)$ to be the unique bilinear map $\Pi : \mathbb{R}^3 \times \mathbb{R}^3$ satisfying

$$\begin{gathered} u \times v = -v \times u, \quad \text{and} \\ i \times j = k, \quad j \times k = i, \quad k \times i = j, \end{gathered} \tag{2.5.1}$$

where $\{i, j, k\}$ is the standard basis of $\mathbb{R}^3$. Here, to say Π is bilinear is to say $\Pi(u, v)$ is linear in both u and v.

The following result relates the cross product on $\mathbb{R}^3$ to the 3×3 determinant. The proof is a straightforward consequence of results of §2.3.

Proposition 2.5.1. *If $u, v, w \in \mathbb{R}^3$, then*

$$w \cdot (u \times v) = \det \begin{pmatrix} w_1 & u_1 & v_1 \\ w_2 & u_2 & v_2 \\ w_3 & u_3 & v_3 \end{pmatrix}. \tag{2.5.2}$$

Note (by Proposition 2.3.1) that the right side of (2.5.2) is linear in u and in v, and it changes sign when u and v are switched. It remains to check the identity for $\{u, v\} = \{i, j\}, \{j, k\}$, and $\{k, i\}$, which the reader can do.

We mention that (2.5.2) can be rewritten (symbolically) as

$$u \times v = \det \begin{pmatrix} i & u_1 & v_1 \\ j & u_2 & v_2 \\ k & u_3 & v_3 \end{pmatrix} = \begin{pmatrix} u_2 v_3 - u_3 v_2 \\ u_3 v_1 - u_1 v_3 \\ u_1 v_2 - u_2 v_1 \end{pmatrix}. \tag{2.5.3}$$

One can also readily check this against the multiplication table in (2.5.1).

It is an important geometrical fact that the cross product is preserved by rotations on $\mathbb{R}^3$. To state the result, we say

$$T \in SO(3) \tag{2.5.4}$$

provided

$$T \in M(3, \mathbb{R}), \quad T^t T = I, \quad \text{and} \quad \det T > 0. \tag{2.5.5}$$

Recall §2.2, Exercises 9–14, §2.3, Exercise 6, and §2.4, Exercise 2. We note that these conditions actually imply

$$\det T = 1. \tag{2.5.6}$$

Here is the result

Proposition 2.5.2.

$$T \in SO(3) \Longrightarrow Tu \times Tv = T(u \times v). \tag{2.5.7}$$

Proof. Multiply the 3×3 matrix in Proposition 2.5.1 on the left by T. The resulting determinant is unchanged, since $\det T = 1$ On the other hand, the quantity one gets is

$$Tw \cdot (Tu \times Tv), \tag{2.5.8}$$

but the fact that $T^t T = I$ implies that

$$w \cdot (u \times v) = Tw \cdot T(u \times v). \tag{2.5.9}$$

The desired identity follows. □

We can apply Proposition 2.5.2 to establish the following useful identity.

Proposition 2.5.3. *For all* $u, v, w, x \in \mathbb{R}^3$,

$$(u \times v) \cdot (w \times x) = \det \begin{pmatrix} u \cdot w & v \cdot w \\ u \cdot x & v \cdot x \end{pmatrix}. \tag{2.5.10}$$

Proof. By Proposition 2.5.2 (in concert with Exercise 3 below), it suffices to check this for

$$w = i, \quad x = ai + bj,$$

in which case $w \times x = bk$. Then the left side of (2.5.10) is

$$(u \times v) \cdot bk = \det \begin{pmatrix} 0 & u \cdot i & v \cdot i \\ 0 & u \cdot j & v \cdot j \\ b & u \cdot k & v \cdot k \end{pmatrix}.$$

Meanwhile, the right side of (2.5.10) is

$$\begin{aligned} &\det \begin{pmatrix} u \cdot i & v \cdot i \\ au \cdot i + bu \cdot j & av \cdot i + bv \cdot j \end{pmatrix} \\ &= b \det \begin{pmatrix} u \cdot i & v \cdot i \\ u \cdot j & v \cdot j \end{pmatrix}. \end{aligned}$$

But one sees that the last two right sides are equal. □

In case $u = w$ and $v = x$, this specializes to the following.

Corollary 2.5.4. *If* θ *is the angle between* u *and* v *in* $\mathbb{R}^3$, *then*

$$|u \times v| = |u| \; |v| \; |\sin \theta|. \tag{2.5.11}$$

Proof. From (2.5.10), we have

$$\begin{aligned} |u \times v|^2 &= \det \begin{pmatrix} u \cdot u & v \cdot u \\ u \cdot v & v \cdot v \end{pmatrix} \\ &= |u|^2 |v|^2 - (u \cdot v)^2, \end{aligned} \tag{2.5.12}$$

a result known as *Lagrange's identity*. Since

$$u \cdot v = |u| \, |v| \cos \theta, \tag{2.5.13}$$

this gives (2.5.11). □

Remark. See §3.2 for a self-contained treatment of the trigonometric functions $\sin \theta$ and $\cos \theta$.

Exercises

1. Show that if $u, v \in \mathbb{R}^3$, then $u \times v$ is orthogonal to u and v, i.e.,

$$u \cdot (u \times v) = v \cdot (u \times v) = 0.$$

2. Show that

$$w \cdot (u \times v) = u \cdot (v \times w), \quad \forall\, u, v, w \in \mathbb{R}^3.$$

3. Suppose $w, x \in \mathbb{R}^3,\ |w| = 1$. Show that there exists $T \in SO(3)$ such that

$$Tw = i, \quad Tx \in \operatorname{Span}\{i, j\},$$

where $\{i, j, k\}$ denotes the standard basis of $\mathbb{R}^3$. Discuss how this figures in the proof of Proposition 2.5.3.
Hint. See Exercise 14 in §2.2. Show that, in that exercise, you can actually take $T \in SO(n)$.

4. Show that $\kappa : \mathbb{R}^3 \to$ Skew(3), the set of antisymmetric real 3×3 matrices, given by

$$\kappa(y_1, y_2, y_3) = \begin{pmatrix} 0 & -y_3 & y_2 \\ y_3 & 0 & -y_1 \\ -y_2 & y_1 & 0 \end{pmatrix} \tag{2.5.14}$$

satisfies

$$Kx = y \times x, \quad K = \kappa(y). \tag{2.5.15}$$

Show that, with $[A, B] = AB - BA$,

$$\begin{aligned} \kappa(x \times y) &= \bigl[\kappa(x), \kappa(y)\bigr], \\ \operatorname{Tr}\bigl(\kappa(x)\kappa(y)^t\bigr) &= 2x \cdot y. \end{aligned} \tag{2.5.16}$$

5. Assume $\{u_1, u_2, u_3\}$ is an orthonormal basis of $\mathbb{R}^3$, and form

$$U = (u_1, u_2, u_3) \in O(3).$$

Show that

$$U \in SO(3) \Longleftrightarrow u_1 = u_2 \times u_3.$$

Show also that

$$u_1 = u_2 \times u_3 \Leftrightarrow u_2 = u_3 \times u_1 \Leftrightarrow u_3 = u_1 \times u_2.$$

Chapter 3

Curves in Euclidean space

Our transition from one variable to multivariable calculus starts with the consideration of n dependent variables, as functions of one independent variable, that is, a function $\gamma(t)$ of $t \in I \subset \mathbb{R}$, having n components, or as we say a curve.

Section 3.1 starts the study of curves in Euclidean space $\mathbb{R}^n$, with particular attention to arc length. We derive an integral formula for arc length. We show that a smooth curve can be reparametrized by arc length, as an application of the Inverse Function Theorem. We then take a look at the unit circle S^1 in $\mathbb{R}^2$. Using the parametrization of part of S^1 as $(t, \sqrt{1-t^2})$, we obtain a power series for arc lengths, as an application of material of §1.3 on power series of $(1-x)^b$, with $b = -1/2$, and x replaced by t^2. We also bring in the trigonometric functions, having the property that $(\cos t, \sin t)$ provides a parametrization of S^1 by arc length.

Section 3.2 goes much further into the study of the trigonometric functions. Actually, it begins with a treatment of the exponential function e^t, observes that such treatment extends readily to e^{at}, given $a \in \mathbb{C}$, and then establishes that e^{it} provides a unit speed parametrization of S^1. This directly gives Euler's formula

$$e^{it} = \cos t + i \sin t, \tag{3.0.1}$$

and provides for a unified treatment of the exponential and trigonometric functions. We also bring in log as the inverse function to the exponential, and we use the formula $x^r = e^{r \log x}$ to generalize results of §1.1 on x^r from $r \in \mathbb{Q}$ to $r \in \mathbb{R}$, and further, to $r \in \mathbb{C}$.

We next examine curvature, which is a measure of how far a curve is from being a straight line. If $\gamma : (a, b) \to \mathbb{R}^n$ is a smooth curve, parametrized by arc length, with unit tangent vector $T(s) = \gamma'(s)$, then γ is a straight line if and only if $T'(s) \equiv 0$, so $T'(s)$ serves as the "curvature vector." The case $n = 2$, treated in §3.3, leads to

$$T'(s) = \kappa(s) J T(s), \quad J = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \tag{3.0.2}$$

defining the curvature $\kappa(s)$. We show that the solution can be represented using the matrix exponential, e^{tJ}. General material on the matrix exponential e^{tA}, for $A \in M(n, \mathbb{R})$, or $A \in M(n, \mathbb{C})$, is given in §C.4. In case $A = J$, we have the following variant of the Euler identity:

$$e^{tJ} = (\cos t)I + (\sin t)J, \tag{3.0.3}$$

which leads to an explicit solution to (3.0.2), yielding an analysis of planar curves with constant curvature, as circles.

In §3.4 we look at smooth curves in $\mathbb{R}^3$. When parametrized by arc length, these curves have curvature characterized by the norm of $T'(s)$. In addition, there is torsion, τ, measuring whether such a curve is actually contained in some plane. In this setting, the 2×2 system (3.0.2) is replaced by a 9×9 system, involving both κ and τ, known as the Frenet-Serret equations. In case κ and τ are constant, this system is also amenable to solution via the matrix exponential, leading to curves that are helices.

3.1. Curves and arc length

The term "curve" is commonly used to refer to a couple of different, but closely related, objects. In one meaning, a curve is a continuous function from an interval $I \subset \mathbb{R}$ to n-dimensional Euclidean space:

$$\gamma : I \longrightarrow \mathbb{R}^n, \quad \gamma(t) = (\gamma_1(t), \dots, \gamma_n(t)). \tag{3.1.1}$$

We say γ is differentiable provided each component γ_j is, in which case

$$\gamma'(t) = (\gamma_1'(t), \dots, \gamma_n'(t)). \tag{3.1.2}$$

$\gamma'(t)$ is the *velocity* of γ, at "time" t, and its *speed* is the magnitude of $\gamma'(t)$:

$$|\gamma'(t)| = \sqrt{\gamma_1'(t)^2 + \cdots + \gamma_n'(t)^2}. \tag{3.1.3}$$

We say γ is smooth of class C^k provided each component $\gamma_j(t)$ has this property.

One also calls the image of I under the map γ a curve in $\mathbb{R}^n$. If $u : J \to I$ is continuous, one-to-one, and onto, the map

$$\sigma : J \longrightarrow \mathbb{R}^n, \quad \sigma(t) = \gamma(u(t)) \tag{3.1.4}$$

has the same image as γ. We say σ is a reparametrization of γ. We usually require that u be C^1, with C^1 inverse. If γ is C^k and u is also C^k, so is σ, and the chain rule gives

$$\sigma'(t) = u'(t)\gamma'(u(t)). \tag{3.1.5}$$

Let us assume $I = [a, b]$ is a closed, bounded interval, and γ is C^1. We want to define the *length* of this curve. To get started, we take a partition $\mathcal{P}$ of $[a, b]$, given by

$$a = t_0 < t_1 < \cdots < t_N = b, \tag{3.1.6}$$

and set

$$\ell_{\mathcal{P}}(\gamma) = \sum_{j=1}^{N} |\gamma(t_j) - \gamma(t_{j-1})|. \tag{3.1.7}$$

See Figure 3.1.1.

We will massage the right side of (3.1.7) into something that looks like a Riemann sum for $\int_a^b |\gamma'(t)|\, dt$. We have

$$\begin{aligned} \gamma(t_j) - \gamma(t_{j-1}) &= \int_{t_{j-1}}^{t_j} \gamma'(t)\, dt \\ &= \int_{t_{j-1}}^{t_j} \big[\gamma'(t_j) + \gamma'(t) - \gamma'(t_j)\big]\, dt \\ &= (t_j - t_{j-1})\gamma'(t_j) + \int_{t_{j-1}}^{t_j} \big[\gamma'(t) - \gamma'(t_j)\big]\, dt. \end{aligned} \tag{3.1.8}$$

We get

$$|\gamma(t_j) - \gamma(t_{j-1})| = (t_j - t_{j-1})|\gamma'(t_j)| + r_j, \tag{3.1.9}$$

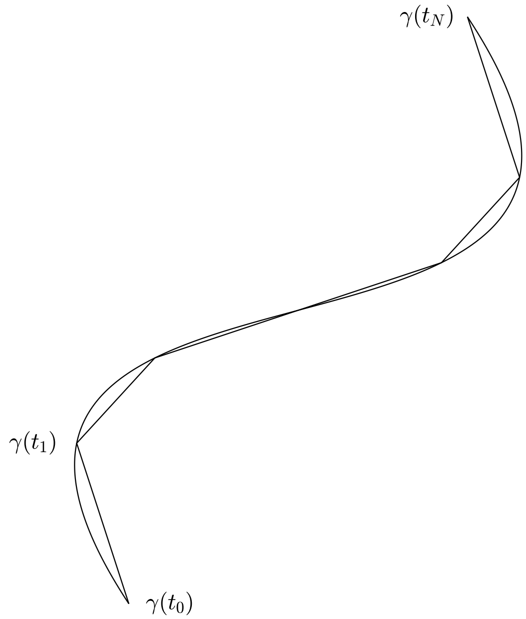

Figure 3.1.1. Approximating $\ell(\gamma)$ by $\ell_{\mathcal{P}}(\gamma)$

with

$$|r_j| \le \int_{t_{j-1}}^{t_j} |\gamma'(t) - \gamma'(t_j)|\, dt. \tag{3.1.10}$$

Now if γ' is continuous on $[a, b]$, so is $|\gamma'|$, and hence both are uniformly continuous on $[a, b]$. We have

$$s, t \in [a, b],\ |s - t| \le h \Longrightarrow |\gamma'(t) - \gamma'(s)| \le \omega(h), \tag{3.1.11}$$

where $\omega(h) \to 0$ as $h \to 0$. Summing (3.1.9) over j, we get

$$\ell_{\mathcal{P}}(\gamma) = \sum_{j=1}^{N} |\gamma'(t_j)|(t_j - t_{j-1}) + R_{\mathcal{P}}, \tag{3.1.12}$$

with

$$|R_{\mathcal{P}}| \le (b - a)\omega(h), \quad \text{if each } \ t_j - t_{j-1} \le h. \tag{3.1.13}$$

Since the sum on the right side of (3.1.12) is a Riemann sum, we can apply Theorem 1.2.4 to get the following.

Proposition 3.1.1. *Assume $\gamma : [a, b] \to \mathbb{R}^n$ is a C^1 curve. Then*

$$\ell_{\mathcal{P}}(\gamma) \longrightarrow \int_a^b |\gamma'(t)|\, dt \quad \textit{as} \quad \textit{maxsize}\ \mathcal{P} \to 0. \tag{3.1.14}$$

We call this limit the length of the curve γ, and write

$$\ell(\gamma) = \int_a^b |\gamma'(t)|\, dt. \tag{3.1.15}$$

Note that if $u : [\alpha, \beta] \to [a, b]$ is a C^1 map with C^1 inverse, and $\sigma = \gamma \circ u$, as in (3.1.4), we have from (3.1.5) that $|\sigma'(t)| = |u'(t)| \cdot |\gamma'(u(t))|$, and the change of variable formula (1.2.67) for the integral gives

$$\int_\alpha^\beta |\sigma'(t)|\, dt = \int_a^b |\gamma'(t)|\, dt, \tag{3.1.16}$$

hence we have the geometrically natural result

$$\ell(\sigma) = \ell(\gamma). \tag{3.1.17}$$

Given such a C^1 curve γ, it is natural to consider the length function

$$\ell_\gamma(t) = \int_a^t |\gamma'(s)|\, ds, \quad \ell'_\gamma(t) = |\gamma'(t)|. \tag{3.1.18}$$

If we assume also that γ' is nowhere vanishing on $[a, b]$, Theorem 1.1.3, the inverse function theorem, implies that $\ell_\gamma : [a, b] \to [0, \ell(\gamma)]$ has a C^1 inverse

$$u : [0, \ell(\gamma)] \longrightarrow [a, b], \tag{3.1.19}$$

and then $\sigma = \gamma \circ u : [0, \ell(\gamma)] \to \mathbb{R}^n$ satisfies

$$\begin{aligned} \sigma'(t) &= u'(t)\gamma'(u(t)) \\ &= \frac{1}{\ell'_\gamma(s)}\gamma'(u(t)), \quad \text{for } \ t = \ell_\gamma(s),\ s = u(t), \end{aligned} \tag{3.1.20}$$

since the chain rule applied to $u(\ell_\gamma(t)) = t$ yields $u'(\ell_\gamma(t))\ell'_\gamma(t) = 1$. Also, by (4.18), $\ell'_\gamma(s) = |\gamma'(s)| = |\gamma'(u(t))|$, so

$$|\sigma'(t)| \equiv 1. \tag{3.1.21}$$

Then σ is a reparametrization of γ, and σ has unit speed. We say σ is a reparametrization by arc length.

We now focus on that most classical example of a curve in the plane $\mathbb{R}^2$, the unit circle

$$S^1 = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}. \tag{3.1.22}$$

We can parametrize S^1 away from $(x, y) = (\pm 1, 0)$ by

$$\gamma_+(t) = (t, \sqrt{1 - t^2}), \quad \gamma_-(t) = (t, -\sqrt{1 - t^2}), \tag{3.1.23}$$

on the intersection of S^1 with $\{(x, y) : y > 0\}$ and $\{(x, y) : y < 0\}$, respectively. Here $\gamma_\pm : (-1, 1) \to \mathbb{R}^2$, and both maps are smooth. In fact, we can take $\gamma_\pm : [-1, 1] \to \mathbb{R}^2$, but these functions are not differentiable at ± 1. We can also parametrize S^1 away from $(x, y) = (0, \pm 1)$, by

$$\gamma_\ell(t) = (-\sqrt{1 - t^2}, t), \quad \gamma_r(t) = (\sqrt{1 - t^2}, t), \tag{3.1.24}$$

again with $t \in (-1, 1)$. Note that

$$\gamma'_+(t) = (1, -t(1 - t^2)^{-1/2}), \tag{3.1.25}$$

so

$$|\gamma_+'(t)|^2 = 1 + \frac{t^2}{1-t^2} = \frac{1}{1-t^2}. \tag{3.1.26}$$

Hence, if $\ell(t)$ is the length of the image $\gamma_+([0,t])$, we have

$$\ell(t) = \int_0^t \frac{1}{\sqrt{1-s^2}}\,ds, \quad \text{for} \quad 0 < t < 1. \tag{3.1.27}$$

The same formula holds with γ_+ replaced by γ_-, γ_ℓ, or γ_r.

We can evaluate the integral (3.1.27) as a power series in t, as follows. As seen in §1.3,

$$(1-r)^{-1/2} = \sum_{k=0}^{\infty} \frac{a_k}{k!} r^k, \quad \text{for} \quad |r| < 1, \tag{3.1.28}$$

where

$$a_0 = 1, \quad a_1 = \frac{1}{2}, \quad a_k = \Bigl(\frac{1}{2}\Bigr)\Bigl(\frac{3}{2}\Bigr)\cdots\Bigl(k-\frac{1}{2}\Bigr). \tag{3.1.29}$$

The power series converges uniformly on $[-\rho,\rho]$, for each $\rho \in (0,1)$. It follows that

$$(1-s^2)^{-1/2} = \sum_{k=0}^{\infty} \frac{a_k}{k!} s^{2k}, \quad |s| < 1, \tag{3.1.30}$$

uniformly convergent on $[-a,a]$ for each $a \in (0,1)$. Hence we can integrate (3.1.30) term by term to get

$$\ell(t) = \sum_{k=0}^{\infty} \frac{a_k}{k!} \frac{t^{2k+1}}{2k+1}, \quad 0 \le t < 1. \tag{3.1.31}$$

One can use (3.1.27)–(3.1.31) to get a rapidly convergent infinite series for the number π, defined as

$$\pi \text{ is half the length of } S^1. \tag{3.1.32}$$

See Exercise 7 in §3.2.

Since S^1 is a smooth curve, it can be parametrized by arc length. We will let $C : \mathbb{R} \to S^1$ be such a parametrization, satisfying

$$C(0) = (1,0), \quad C'(0) = (0,1), \tag{3.1.33}$$

so $C(t)$ traverses S^1 counter-clockwise, as t increases. For t moderately bigger than 0, the rays from $(0,0)$ to $(1,0)$ and from $(0,0)$ to $C(t)$ make an angle that, measured in radians, is t. This leads to the standard trigonometrical functions $\cos t$ and $\sin t$, defined by

$$C(t) = (\cos t, \sin t), \tag{3.1.34}$$

when C is such a unit-speed parametrization of S^1. See Figure 3.1.2.

We can evaluate the derivative of $C(t)$ by the following device. Applying d/dt to the identity

$$C(t) \cdot C(t) = 1 \tag{3.1.35}$$

and using the product formula gives

$$C'(t) \cdot C(t) = 0. \tag{3.1.36}$$

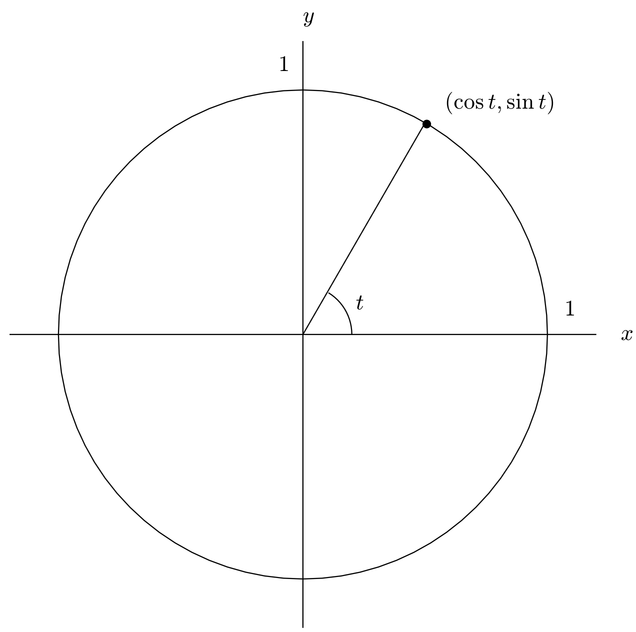

Figure 3.1.2. The circle $C(t) = (\cos t, \sin t)$

since both $|C(t)| \equiv 1$ and $|C'(t)| \equiv 1$, (3.1.36) allows only two possibilities. Either

$$C'(t) = (\sin t, -\cos t). \tag{3.1.37}$$

or

$$C'(t) = (-\sin t, \cos t). \tag{3.1.38}$$

Since $C'(0) = (0, 1)$, (3.1.37) is not a possibility. This implies

$$\frac{d}{dt}\cos t = -\sin t, \quad \frac{d}{dt}\sin t = \cos t. \tag{3.1.39}$$

We will derive further important results on $\cos t$ and $\sin t$ in §3.2.

One can think of $\cos t$ and $\sin t$ as special functions arising to analyze the length of arcs in the circle. Related special functions arise to analyze the length of portions of a parabola in $\mathbb{R}^2$, say the graph of

$$y = \frac{1}{2}x^2. \tag{3.1.40}$$

This curve is parametrized by

$$\gamma(t) = \Big(t, \frac{1}{2}t^2\Big), \tag{3.1.41}$$

so

$$\gamma'(t) = (1, t). \tag{3.1.42}$$

In such a case, the length of $\gamma([0,t])$ is

$$\ell_\gamma(t) = \int_0^t \sqrt{1+s^2}\, ds. \tag{3.1.43}$$

Methods to evaluate the integral in (4.42) are provided in §3.2. See Exercise 10 of §3.2.

The study of lengths of other curves has stimulated much work in analysis. Another example is the ellipse

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1, \tag{3.1.44}$$

given $a, b \in (0, \infty)$. This curve is parametrized by

$$\gamma(t) = (a \cos t, b \sin t). \tag{3.1.45}$$

In such a case, by (4.38), $\gamma'(t) = (-a \sin t, b \cos t)$, so

$$\begin{aligned} |\gamma'(t)|^2 &= a^2 \sin^2 t + b^2 \cos^2 t \\ &= b^2 + \eta \sin^2 t, \quad \eta = a^2 - b^2, \end{aligned} \tag{3.1.46}$$

and hence the length of $\gamma([0,t])$ is

$$\ell_\gamma(t) = b \int_0^t \sqrt{1 + \sigma \sin^2 s}\, ds, \quad \sigma = \frac{\eta}{b^2}. \tag{3.1.47}$$

If $a \neq b$, this is called an elliptic integral, and it gives rise to a more subtle family of special functions, called elliptic functions. Material on this can be found in Chapter 6 of [**17**], *Introduction to Complex Analysis*.

We end this section with a brief discussion of curves in *polar coordinates*. We define a map

$$\Pi : \mathbb{R}^2 \longrightarrow \mathbb{R}^2, \quad \Pi(r, \theta) = (r \cos \theta, r \sin \theta). \tag{3.1.48}$$

We say (r, θ) are polar coordinates of $(x, y) \in \mathbb{R}^2$ if $\Pi(r, \theta) = (x, y)$. See Figure 3.1.3.

Now, Π in (3.1.48) is not bijective, since

$$\Pi(r, \theta + 2\pi) = \Pi(r, \theta), \quad \Pi(r, \theta + \pi) = \Pi(-r, \theta), \tag{3.1.49}$$

and $\Pi(0, \theta)$ is independent of θ. So polar coordinates are not unique, but we will not belabor this point. The point we make is that an equation

$$r = \rho(\theta), \quad \rho : [a, b] \to \mathbb{R}, \tag{3.1.50}$$

yields a curve in $\mathbb{R}^2$, namely (with $\theta = t$)

$$\gamma(t) = (\rho(t) \cos t, \rho(t) \sin t), \quad a \le t \le b. \tag{3.1.51}$$

The circle (3.1.34) corresponds to $\rho(\theta) \equiv 1$. Other cases include

$$\rho(\theta) = a \cos \theta, \quad -\frac{\pi}{2} \le \theta \le \frac{\pi}{2}, \tag{3.1.52}$$

yielding a circle of diameter $a/2$ centered at $(a/2, 0)$ (see Exercise 6 below), and

$$\rho(\theta) = a \cos 3\theta, \tag{3.1.53}$$

yielding a figure called a three-leaved rose. See Figure 3.1.4.

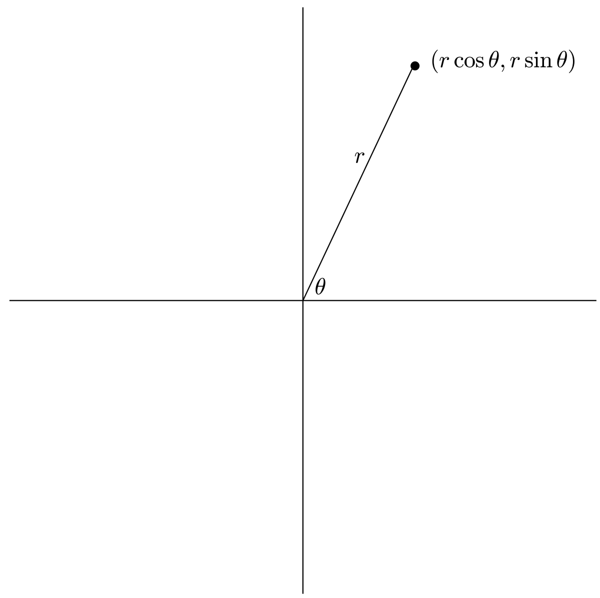

Figure 3.1.3. Polar coordinates on $\mathbb{R}^2$

To compute the arc length of (3.1.51), we note that, by (3.1.39),

$$\begin{aligned} &x(t) = \rho(t)\cos t, \quad y(t) = \rho(t)\sin t \\ &\Rightarrow x'(t) = \rho'(t)\cos t - \rho(t)\sin t,\ y'(t) = \rho'(t)\sin t + \rho(t)\cos t, \end{aligned} \tag{3.1.54}$$

hence

$$\begin{aligned} x'(t)^2 + y'(t)^2 &= \rho'(t)^2\cos^2 t - 2\rho(t)\rho'(t)\cos t\sin t + \rho(t)^2\sin^2 t \\ &\qquad + \rho'(t)^2\sin^2 t + 2\rho(t)\rho'(t)\sin t\cos t + \rho(t)^2\cos^2 t \\ &= \rho'(t)^2 + \rho(t)^2. \end{aligned} \tag{3.1.55}$$

Therefore

$$\ell(\gamma) = \int_a^b |\gamma'(t)|\,dt = \int_a^b \sqrt{\rho(t)^2 + \rho'(t)^2}\,dt. \tag{3.1.56}$$

A more systematic treatment of polar coordinates is given in §4.3.

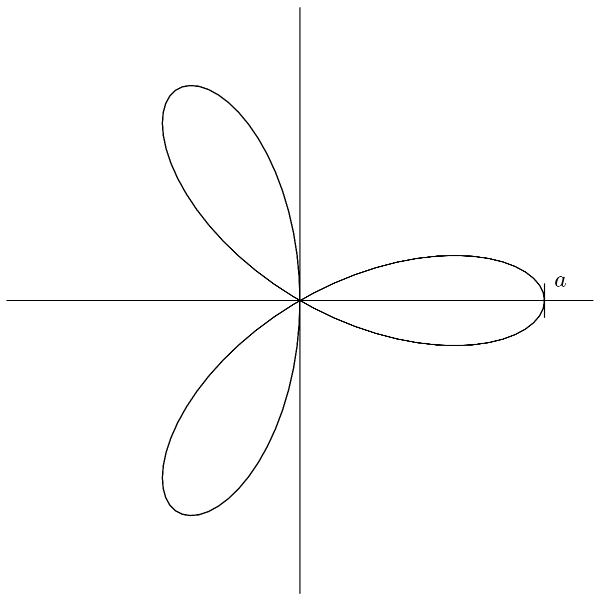

Figure 3.1.4. Three-leafed rose: $r = a\cos 3\theta$

Exercises

1. Let $\gamma(t) = (t^2, t^3)$. Compute the length of $\gamma([0,t])$.

2. With $a, b > 0$, the curve

$$\gamma(t) = (a\cos t, a\sin t, bt)$$

is a helix. Compute the length of $\gamma([0,t])$.

3. Let

$$\gamma(t) = \Big(t, \frac{2\sqrt{2}}{3}t^{3/2}, \frac{1}{2}t^2\Big).$$

Compute the length of $\gamma([0,t])$.

4. In case $b > a$ for the ellipse (3.1.45), the length formula (3.1.47) becomes

$$\ell_\gamma(t) = b\int_0^t \sqrt{1-\beta^2\sin^2 s}\,ds, \quad \beta^2 = \frac{b^2-a^2}{b^2} \in (0,1).$$

Apply the change of variable $x = \sin s$ to this integral (cf. (1.2.46)), and write out the resulting integral.

5. The second half of (3.1.49) is equivalent to the identity

$$(\cos(\theta + \pi), \sin(\theta + \pi)) = -(\cos\theta, \sin\theta).$$

Deduce this from the definition (3.1.32) of π, together with the characterization of $C(t)$ in (3.1.34) as the unit speed parametrization of S^1, satisfying (3.1.33). For a more general identity, see (3.2.44).

6. The curve defined by (3.1.52) can be written

$$\gamma(t) = (a\cos^2 t, a\cos t\sin t), \quad -\frac{\pi}{2} \le t \le \frac{\pi}{2}.$$

Peek ahead at (3.2.44) and show that

$$\gamma(t) = \Bigl(\frac{a}{2} + \frac{a}{2}\cos 2t, \frac{a}{2}\sin 2t\Bigr).$$

Verify that this traces out a circle of radius $a/2$, centered at $(a/2, 0)$.

7. Use (3.1.56) to write the arc length of the curve given by (3.1.53) as an integral. Show this integral has the same general form as (3.1.46)–(3.1.47).

8. Let $\gamma : [a, b] \to \mathbb{R}^n$ be a C^1 curve. Show that

$$\ell(\gamma) \ge |\gamma(b) - \gamma(a)|,$$

with strict inequality if there exists $t \in (a, b)$ such that $\gamma(t)$ does not lie on the line segment from $\gamma(a)$ to $\gamma(b)$.
Hint. To get started, show that, in (3.1.7), $\ell_{\mathcal{P}}(\gamma) \ge |\gamma(b) - \gamma(a)|$.

9. Consider the curve $C(t) = (\cos t, \sin t)$, discussed in (3.1.33)–(3.1.38). Note that the length $\ell_C(t)$ of $C([0, t])$ is t, for $t > 0$. Show that

$$C\Bigl(\frac{\pi}{2}\Bigr) = (0, 1), \quad C(\pi) = (-1, 0), \quad C(2\pi) = (1, 0).$$

10. In the setting of Exercise 9, compute $|C(t) - (1, 0)|$. Then deduce from Exercise 8 that, for $0 < t \le \pi/2$,

$$1 - \cos t < \frac{t^2}{2},$$

hence (multiplying by $1 + \cos t$),

$$\sin^2 t < t^2\,\frac{1 + \cos t}{2}. \tag{3.1.57}$$

Hint. $\sin^2 t = 1 - \cos^2 t$.

11. Let $\gamma : [a, b] \to \mathbb{R}^n$ be a C^1 curve, and assume that $|\gamma(t)| \ge 1$ for all $t \in [a, b]$. Set

$$\sigma(t) = \frac{1}{|\gamma(t)|}\gamma(t).$$

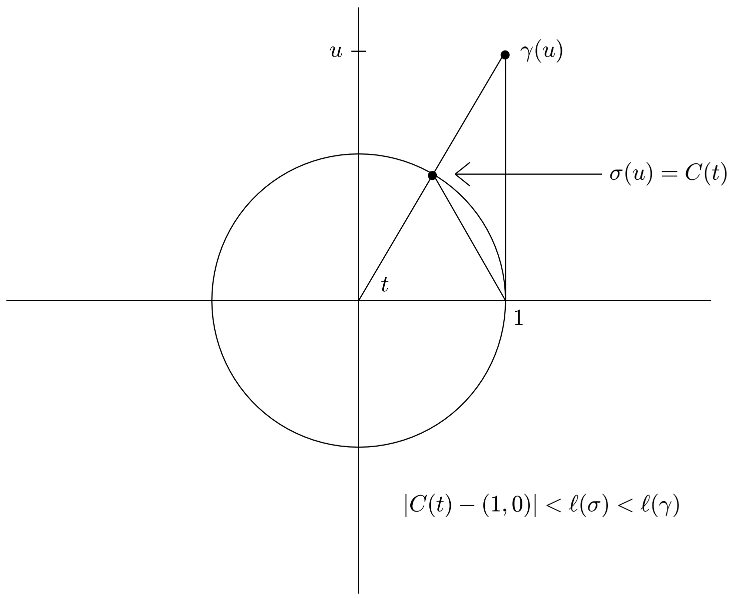

Figure 3.1.5. $\tan t = u$, and key to estimates (3.1.57) and (3.1.58)

Show that

$$\ell(\sigma) \leq \ell(\gamma).$$

Hint. Show that

$$x, y \in \mathbb{R}^n, \ |x| \geq 1, |y| \geq 1 \Longrightarrow \left| \frac{x}{|x|} - \frac{y}{|y|} \right| \leq |x - y|,$$

and deduce that $\ell_{\mathcal{P}}(\sigma) \leq \ell_{\mathcal{P}}(\gamma)$.

12. Consider curves $\gamma, \sigma : \mathbb{R} \to \mathbb{R}^2$ given by

$$\gamma(u) = (1, u), \quad \sigma(u) = \frac{1}{|\gamma(u)|}\gamma(u),$$

so $\sigma(u)$ lies on the unit circle centered at the origin. Show that

$$\sigma(\tan t) = C(t),$$

where $C(t)$ is as in (3.1.34) and

$$\tan t = \frac{\sin t}{\cos t}.$$

See Figure 3.1.5.

13. With $\ell_\gamma(u)$ defined to be the length of $\gamma([0, u])$ and $\ell_\sigma(u)$ and $\ell_C(t)$ similarly

defined (cf. Exercise 9), deduce from Exercises 11–12 that, for $0 \le t < \pi/2$,

$$t \le \tan t. \tag{3.1.58}$$

14. Deduce from Exercises 10 and 13 that, for $0 \le t < \pi/2$,

$$\sin t \le t \le \tan t,$$

and hence

$$\cos t \le \frac{\sin t}{t} \le 1.$$

Use this to give a demonstration that

$$\lim_{t\to 0} \frac{\sin t}{t} = 1, \tag{3.1.59}$$

independent of the use of (3.1.39).

15. Use the conclusion of Exercise 14, together with the identity

$$(1+\cos t)(1-\cos t) = \sin^2 t,$$

to show that

$$\lim_{t\to 0} \frac{1-\cos t}{t^2} = \frac{1}{2}, \tag{3.1.60}$$

independent of the use of (3.1.39).

16. A derivation of the formula for $(d/dt)\sin t$ in (3.1.39) often found in calculus texts goes as follows. One starts with the addition formula

$$\sin(t+s) = (\cos t)(\sin s) + (\sin t)(\cos s), \tag{3.1.61}$$

and writes

$$\frac{1}{h}\big(\sin(t+h) - \sin t\big) = \cos t\,\frac{\sin h}{h} - \frac{1-\cos h}{h}\,\sin t.$$

Use the results of Exercises 14 and 15 to conclude that

$$\lim_{h\to 0} \frac{\sin(t+h) - \sin t}{h} = \cos t.$$

Remark. See §3.2 for a derivation of (3.1.61) of a different nature than typically seen in trigonometry texts.

17. Using the formulas (3.1.39) for the derivatives of $\cos t$ and $\sin t$, in conjunction with the formulas (1.3.35)–(1.3.42) for power series, write

$$\begin{aligned}\cos t &= \sum_{k=0}^{n} \frac{(-1)^k}{(2k)!} t^{2k} + C_{2n}^b(t) = C_{2n}(t) + C_{2n}^b(t),\\ \sin t &= \sum_{k=0}^{n} \frac{(-1)^k}{(2k+1)!} t^{2k+1} + S_{2n+1}^b(t) = S_{2n+1}(t) + S_{2n+1}^b(t),\end{aligned} \tag{3.1.62}$$

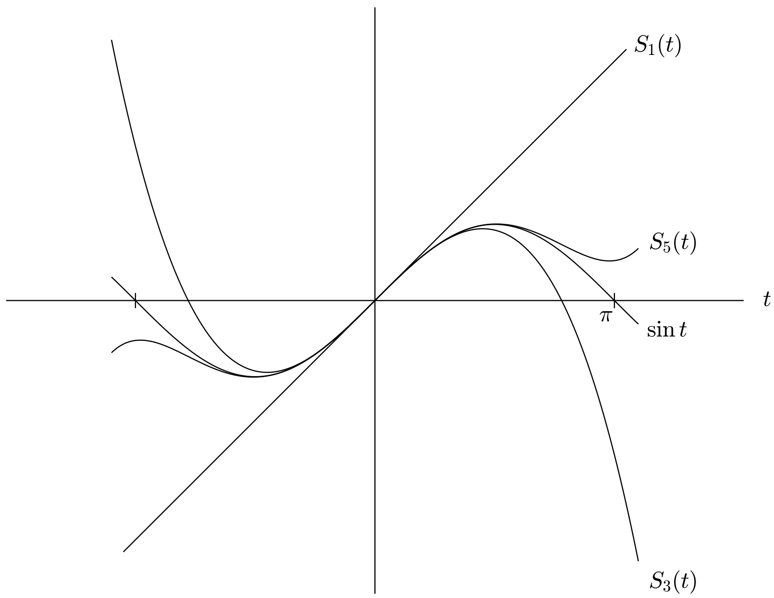

Figure 3.1.6. Power series approximations to $\sin t$

and show that

$$C^b_{2n}(t) = \pm\frac{t^{2n+1}}{(2n+1)!}\sin\xi_n,$$
$$S^b_{2n+1}(t) = \pm\frac{t^{2n+2}}{(2n+2)!}\sin\zeta_n,$$

for some $\xi_n, \zeta_n \in [-|t|, |t|]$. Deduce that

$$C^b_{2n}(t),\ S^b_{2n+1}(t) \longrightarrow 0, \quad \text{as} \ \ n \to \infty,$$

uniformly for t in a bounded set. See Figure 3.1.6 for graphs of $\sin t$ and the power series approximations $S_1(t)$, $S_3(t)$, and $S_5(t)$.

The formula $u \cdot v = |u|\,|v|\,\cos\theta$

18. Show that

$$\cos : [0,\pi] \longrightarrow [-1,1],$$

and that this function is monotone decreasing, one-to-one, and onto.

19. Given nonzero vectors u and v in $\mathbb{R}^n$, we define the angle between them, $\theta(u,v)$, to be the unique number $\theta \in [0,\pi]$ such that

$$u \cdot v = |u|\,|v|\,\cos\theta, \tag{3.1.63}$$

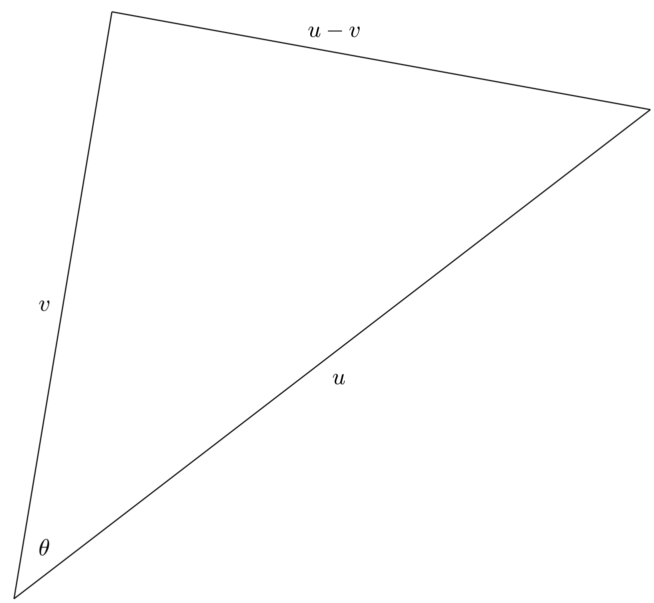

Figure 3.1.7. Setting for law of cosines

i.e., such that $\cos\theta = u \cdot v/(|u|\,|v|)$. Deduce from Cauchy's inequality plus Exercise 18 that $\theta(u, v)$ is well defined.

20. Show that, if u, v are nonzero vectors in $\mathbb{R}^n$, then

$$\theta(u, v) = \theta(Tu, Tv), \quad \forall\, T \in O(n),$$

and

$$\theta(u, v) = \theta(au, bv), \quad \forall\, a, b \in \mathbb{R} \setminus 0,\ \frac{b}{a} > 0.$$

21. Say$\{e_1, \dots, e_n\}$ is the standard orthonormal basis of $\mathbb{R}^n$. Take u, v as above. Then take $T \in O(n)$ such that

$$\begin{aligned} Tu &= |u| e_1, \\ Tv &= b_1 e_1 + b_2 e_2, \quad b_j \in \mathbb{R}. \end{aligned}$$

(Cf. Exercise 14 of §2.2.) Deduce that

$$\theta(u, v) = \theta(e_1, b_1 e_1 + b_2 e_2).$$

22. Assume

$$y = (\cos t) e_1 + (\sin t) e_2, \quad t \in [-\pi, \pi].$$

Show that

$$\theta(e_1, y) = |t|.$$

23. Take $u, v \in \mathbb{R}^n$ as above. The *law of cosines* states that

$$|u - v|^2 = |u|^2 + |v|^2 - 2|u|\,|v|\,\cos\theta. \tag{3.1.64}$$

See Figure 3.1.7. Show that (3.1.64) is equivalent to (3.1.63).
Hint. Expand $(u - v) \cdot (u - v)$.

3.2. The exponential and trigonometric functions

The exponential function is one of the central objects of analysis. In this section we define the exponential function, both for real and complex arguments, and establish a number of basic properties, including fundamental connections to the trigonometric functions.

We construct the exponential function to solve the differential equation

$$\frac{dx}{dt} = x, \quad x(0) = 1. \tag{3.2.1}$$

We seek a solution as a power series

$$x(t) = \sum_{k=0}^{\infty} a_k t^k. \tag{3.2.2}$$

In such a case, if this series converges for $|t| < R$, then, by Proposition 1.3.2,

$$\begin{aligned} x'(t) &= \sum_{k=1}^{\infty} k a_k t^{k-1} \\ &= \sum_{\ell=0}^{\infty} (\ell+1) a_{\ell+1} t^{\ell}, \end{aligned} \tag{3.2.3}$$

so for (3.2.1) to hold we need

$$a_0 = 1, \quad a_{k+1} = \frac{a_k}{k+1}, \tag{3.2.4}$$

i.e., $a_k = 1/k!$, where $k! = k(k-1)\cdots 2 \cdot 1$. Thus (3.2.1) is solved by

$$x(t) = e^t = \sum_{k=0}^{\infty} \frac{1}{k!} t^k, \quad t \in \mathbb{R}. \tag{3.2.5}$$

This defines the exponential function e^t.

More generally, we can define

$$e^z = \sum_{k=0}^{\infty} \frac{1}{k!} z^k, \quad z \in \mathbb{C}. \tag{3.2.6}$$

The ratio test then shows that the series (3.2.6) is absolutely convergent for all $z \in \mathbb{C}$, and uniformly convergent for $|z| \le R$, for each $R < \infty$. Note that, again by Proposition 1.3.2,

$$e^{at} = \sum_{k=0}^{\infty} \frac{a^k}{k!} t^k \tag{3.2.7}$$

solves

$$\frac{d}{dt} e^{at} = a e^{at}, \tag{3.2.8}$$

and this works for each $a \in \mathbb{C}$.

We claim that e^{at} is the *unique* solution to

$$\frac{dy}{dt} = ay, \quad y(0) = 1. \tag{3.2.9}$$

To see this, compute the derivative of $e^{-at}y(t)$:

$$\frac{d}{dt}\big(e^{-at}y(t)\big) = -ae^{-at}y(t) + e^{-at}ay(t) = 0, \tag{3.2.10}$$

where we use the product rule, (3.2.8) (with a replaced by $-a$) and (3.2.9). Thus $e^{-at}y(t)$ is independent of t. Evaluating at $t = 0$ gives

$$e^{-at}y(t) = 1, \quad \forall\, t \in \mathbb{R}, \tag{3.2.11}$$

whenever $y(t)$ solves (3.2.9). Since e^{at} solves (3.2.9), we have $e^{-at}e^{at} = 1$, hence

$$e^{-at} = \frac{1}{e^{at}}, \quad \forall\, t \in \mathbb{R},\ a \in \mathbb{C}. \tag{3.2.12}$$

Thus multiplying both sides of (3.2.11) by e^{at} gives the asserted uniqueness:

$$y(t) = e^{at}, \quad \forall\, t \in \mathbb{R}. \tag{3.2.13}$$

We can draw further useful conclusions from applying d/dt to products of exponential functions. In fact, let $a, b \in \mathbb{C}$; then

$$\begin{aligned} &\frac{d}{dt}\Big(e^{-at}e^{-bt}e^{(a+b)t}\Big) \\ &\quad = -ae^{-at}e^{-bt}e^{(a+b)t} - be^{-at}e^{-bt}e^{(a+b)t} + (a+b)e^{-at}e^{-bt}e^{(a+b)t} \\ &\quad = 0, \end{aligned} \tag{3.2.14}$$

so again we are differentiating a function that is independent of t. Evaluation at $t = 0$ gives

$$e^{-at}e^{-bt}e^{(a+b)t} = 1, \quad \forall\, t \in \mathbb{R}. \tag{3.2.15}$$

Again using (3.2.12), we get

$$e^{(a+b)t} = e^{at}e^{bt}, \quad \forall\, t \in \mathbb{R},\ a, b \in \mathbb{C}, \tag{3.2.16}$$

or, setting $t = 1$,

$$e^{a+b} = e^a e^b, \quad \forall\, a, b \in \mathbb{C}. \tag{3.2.17}$$

We next record some properties of $\exp(t) = e^t$ for real t. The power series (3.2.5) clearly gives $e^t > 0$ for $t \ge 0$. Since $e^{-t} = 1/e^t$, we see that $e^t > 0$ for all $t \in \mathbb{R}$. Since $de^t/dt = e^t > 0$, the function is monotone increasing in t, and since $d^2e^t/dt^2 = e^t > 0$, this function is convex. (See Proposition 1.1.5 and the remark that follows it.) Note that, for $t > 0$,

$$e^t = 1 + t + \frac{t^2}{2} + \cdots > 1 + t \nearrow +\infty, \tag{3.2.18}$$

as $t \nearrow \infty$. Hence

$$\lim_{t \to +\infty} e^t = +\infty. \tag{3.2.19}$$

Since $e^{-t} = 1/e^t$,

$$\lim_{t \to -\infty} e^t = 0. \tag{3.2.20}$$

As a consequence,

$$\exp : \mathbb{R} \longrightarrow (0, \infty) \tag{3.2.21}$$

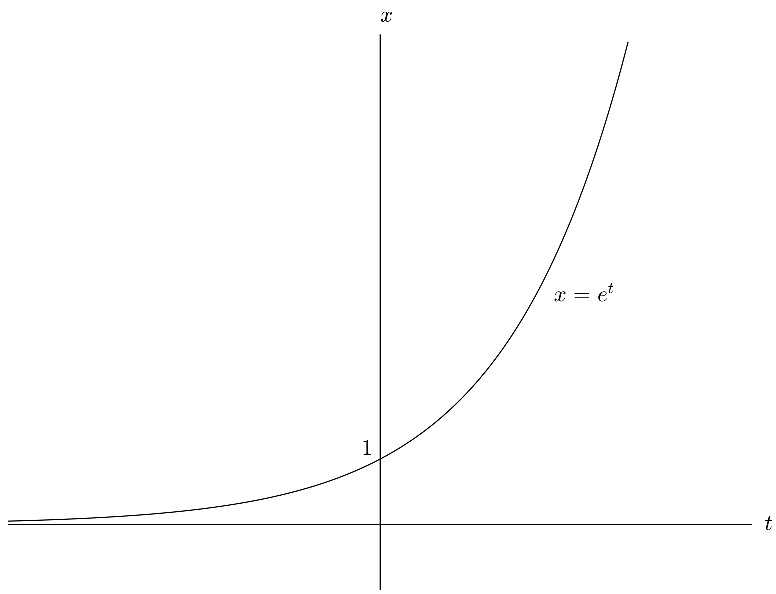

Figure 3.2.1. Exponential function

is one-to-one and onto, with positive derivative, so there is a smooth inverse

$$L : (0, \infty) \longrightarrow \mathbb{R}. \tag{3.2.22}$$

We call this inverse the natural logarithm:

$$\log x = L(x). \tag{3.2.23}$$

See Figures 3.2.1 and 3.2.2 for graphs of $x = e^t$ and $t = \log x$.

Applying d/dt to

$$L(e^t) = t \tag{3.2.24}$$

gives

$$L'(e^t)e^t = 1, \quad \text{hence} \quad L'(e^t) = \frac{1}{e^t}, \tag{3.2.25}$$

i.e.,

$$\frac{d}{dx} \log x = \frac{1}{x}. \tag{3.2.26}$$

Since $\log 1 = 0$, we get

$$\log x = \int_1^x \frac{dy}{y}. \tag{3.2.27}$$

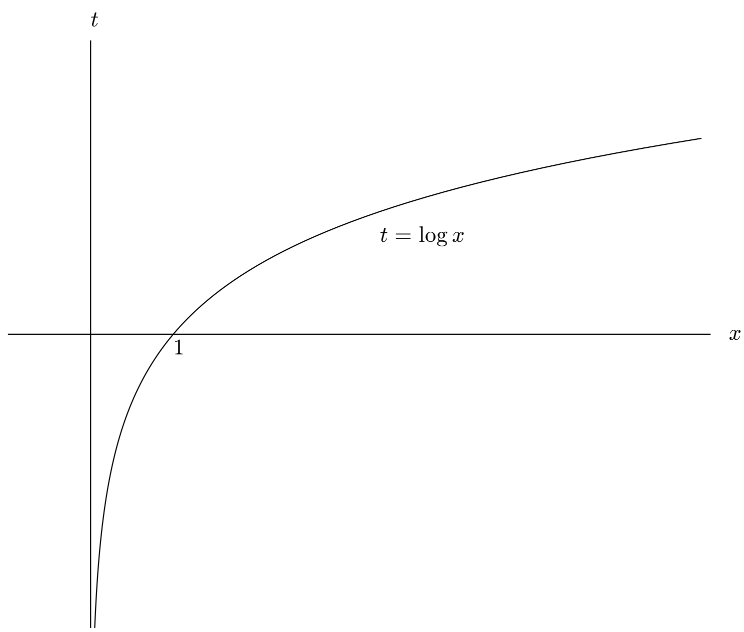

Figure 3.2.2. Logarithm

An immediate consequence of (3.2.17) (for $a, b \in \mathbb{R}$) is the identity

$$\log xy = \log x + \log y, \quad x, y \in (0, \infty). \tag{3.2.28}$$

We move on to a study of e^z for purely imaginary z, i.e., of

$$\gamma(t) = e^{it}, \quad t \in \mathbb{R}. \tag{3.2.29}$$

This traces out a curve in the complex plane, and we want to understand which curve it is. Let us set

$$e^{it} = c(t) + is(t), \tag{3.2.30}$$

with $c(t)$ and $s(t)$ real valued. First we calculate $|e^{it}|^2 = c(t)^2 + s(t)^2$. For $x, y \in \mathbb{R}$,

$$z = x + iy \Longrightarrow \overline{z} = x - iy \Longrightarrow z\overline{z} = x^2 + y^2 = |z|^2. \tag{3.2.31}$$

It is elementary that

$$\begin{gathered} z, w \in \mathbb{C} \Longrightarrow \overline{zw} = \overline{z}\,\overline{w} \Longrightarrow \overline{z^n} = \overline{z}^n, \\ \text{and} \quad \overline{z + w} = \overline{z} + \overline{w}. \end{gathered} \tag{3.2.32}$$

Hence

$$\overline{e^z} = \sum_{k=0}^{\infty} \frac{\overline{z}^k}{k!} = e^{\overline{z}}. \tag{3.2.33}$$

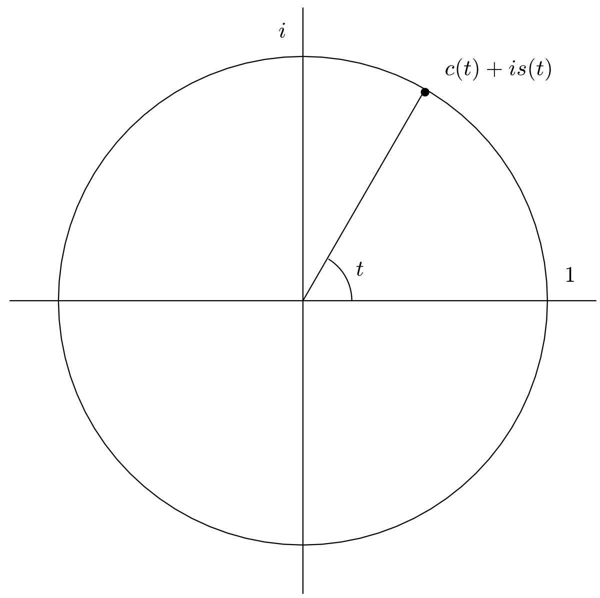

Figure 3.2.3. The circle $e^{it} = c(t) + is(t)$

In particular,

$$t \in \mathbb{R} \Longrightarrow |e^{it}|^2 = e^{it}e^{-it} = 1. \tag{3.2.34}$$

Hence $t \mapsto \gamma(t) = e^{it}$ traces out the unit circle centered at the origin in $\mathbb{C}$. Also

$$\gamma'(t) = ie^{it} \Longrightarrow |\gamma'(t)| \equiv 1, \tag{3.2.35}$$

so $\gamma(t)$ moves at unit speed on the unit circle. We have

$$\gamma(0) = 1, \quad \gamma'(0) = i. \tag{3.2.36}$$

Thus, for moderate $t > 0$, the arc from $\gamma(0)$ to $\gamma(t)$ is an arc on the unit circle, pictured in Figure 3.2.3, of length

$$\ell(t) = \int_0^t |\gamma'(s)|\, ds = t. \tag{3.2.37}$$

In other words, $\gamma(t) = e^{it}$ is the parametrization of the unit circle by arc length, introduced in (3.1.33). As in (3.1.34), standard definitions from trigonometry give

$$\cos t = c(t), \quad \sin t = s(t). \tag{3.2.38}$$

Thus (3.2.30) becomes

$$e^{it} = \cos t + i \sin t, \tag{3.2.39}$$

which is Euler's formula. The identity

$$\frac{d}{dt} e^{it} = i e^{it}, \tag{3.2.40}$$

applied to (3.2.39), yields

$$\frac{d}{dt} \cos t = -\sin t, \quad \frac{d}{dt} \sin t = \cos t. \tag{3.2.41}$$

Compare the derivation of (3.1.39). We can use (3.2.17) to derive formulas for sin and cos of the sum of two angles. Indeed, comparing

$$e^{i(s+t)} = \cos(s+t) + i \sin(s+t) \tag{3.2.42}$$

with

$$e^{is} e^{it} = (\cos s + i \sin s)(\cos t + i \sin t) \tag{3.2.43}$$

gives

$$\begin{aligned} \cos(s+t) &= (\cos s)(\cos t) - (\sin s)(\sin t), \\ \sin(s+t) &= (\sin s)(\cos t) + (\cos s)(\sin t). \end{aligned} \tag{3.2.44}$$

Further material on the trigonometric functions is developed in the exercises below.

Remark. An alternative approach to Euler's formula (3.2.39) is to take the power series for e^{it}, via (3.2.7), and compare it to the power series for $\cos t$ and $\sin t$, given in (3.1.62). This author regards the demonstration via (3.2.33)–(3.2.37), which yields a direct geometrical description of the curve $\gamma(t) = e^{it}$, to be more natural and fundamental than one via the observation of coincident power series.

For yet another derivation of Euler's formula, we can set

$$\text{cis}(t) = \cos t + i \sin t, \tag{3.2.45}$$

and use (3.2.41) (relying on the proof in (3.1.39)) to get

$$\frac{d}{dt} \text{cis}(t) = i \, \text{cis}(t), \quad \text{cis}(0) = 1. \tag{3.2.46}$$

Then the uniqueness result (3.2.9)–(3.2.13) implies that $\text{cis}(t) = e^{it}$.

Exercises

1. Show that

$$|t| < 1 \Rightarrow \log(1+t) = \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k} t^k = t - \frac{t^2}{2} + \frac{t^3}{3} - \cdots . \tag{3.2.47}$$

Hint. Rewrite (3.2.27) as

$$\log(1+t) = \int_0^t \frac{ds}{1+s},$$

expand

$$\frac{1}{1+s} = 1 - s + s^2 - s^3 + \cdots , \quad |s| < 1,$$

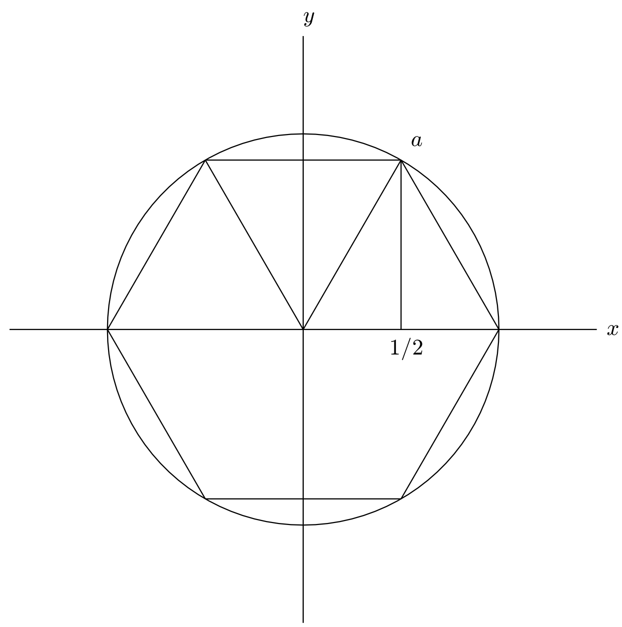

Figure 3.2.4. Regular hexagon, $a = e^{\pi i/3}$

and integrate term by term.

2. In §3.1, π was defined to be half the length of the unit circle S^1. Equivalently, π is the smallest positive number such that $e^{\pi i} = -1$. Show that

$$e^{\pi i/2} = i, \quad e^{\pi i/3} = \frac{1}{2} + \frac{\sqrt{3}}{2} i.$$

Hint. See Figure 3.2.4.

3. Show that

$$\cos^2 t + \sin^2 t = 1,$$

and

$$1 + \tan^2 t = \sec^2 t,$$

where

$$\tan t = \frac{\sin t}{\cos t}, \quad \sec t = \frac{1}{\cos t}.$$

4. Show that

$$\frac{d}{dt}\tan t = \sec^2 t = 1 + \tan^2 t,$$
$$\frac{d}{dt}\sec t = \sec t\ \tan t.$$

5. Evaluate

$$\int_0^y \frac{dx}{1+x^2}.$$

Hint. Set $x = \tan t$.

6. Evaluate

$$\int_0^y \frac{dx}{\sqrt{1-x^2}}.$$

Hint. Set $x = \sin t$.

7. Show that

$$\frac{\pi}{6} = \int_0^{1/2} \frac{dx}{\sqrt{1-x^2}}.$$

Use (3.1.27)–(3.1.31) to obtain a rapidly convergent infinite series for π.
Hint. Show that $\sin \pi/6 = 1/2$. Use Exercise 2 and the identity $e^{\pi i/6} = e^{\pi i/2}e^{-\pi i/3}$. Note that a_k in (3.1.29)-(3.1.31) satisfies $a_{k+1} = (k+1/2)a_k$. Deduce that

$$\pi = \sum_{k=0}^{\infty} \frac{b_k}{2k+1}, \quad b_0 = 3, \quad b_{k+1} = \frac{1}{4}\,\frac{2k+1}{2k+2}\,b_k. \tag{3.2.48}$$

Note that $b_k \le 3 \cdot 4^{-k}$. Deduce that

$$\mathrm{pi}(n) = \sum_{k=0}^{n} \frac{b_k}{2k+1} \Longrightarrow 0 < \pi - \mathrm{pi}(n) < \frac{1}{n+1}\,2^{-2n-1}. \tag{3.2.49}$$

In particular,

$$\pi - \mathrm{pi}(20) < 10^{-13}. \tag{3.2.50}$$

8. Set

$$\cosh t = \frac{1}{2}(e^t + e^{-t}), \quad \sinh t = \frac{1}{2}(e^t - e^{-t}).$$

Show that

$$\frac{d}{dt}\cosh t = \sinh t, \quad \frac{d}{dt}\sinh t = \cosh t,$$

and

$$\cosh^2 t - \sinh^2 t = 1.$$

9. Evaluate

$$\int_0^y \frac{dx}{\sqrt{1+x^2}}.$$

Hint. Set $x = \sinh t$.

10. Evaluate

$$\int_0^y \sqrt{1+x^2}\,dx.$$

11. Using Exercise 4, verify that

$$\begin{aligned}\frac{d}{dt}(\sec t+\tan t) &= \sec t(\sec t+\tan t),\\ \frac{d}{dt}(\sec t\ \ \tan t) &= \sec^3 t+\sec t\ \ \tan^2 t,\\ &= 2\sec^3 t-\sec t.\end{aligned}$$

12. Next verify that

$$\begin{aligned}\frac{d}{dt}\log|\sec t| &= \tan t,\\ \frac{d}{dt}\log|\sec t+\tan t| &= \sec t.\end{aligned}$$

13. Now verify that

$$\begin{aligned}\int \tan t\,dt &= \log|\sec t|,\\ \int \sec t\,dt &= \log|\sec t+\tan t|,\\ 2\int \sec^3 t\,dt &= \sec t\ \ \tan t+\int \sec t\,dt.\end{aligned}$$

(Here and below, we omit the arbitrary additive constants in indefinite integrals.) See the next exercise, and also Exercises 40–43 for other approaches to evaluating these and related integrals.

14. Here is another approach to the evaluation of $\int \sec t\,dt$. We evaluate

$$I(u)=\int_0^u \frac{dv}{\sqrt{1+v^2}}$$

in two ways.
(a) Using $v=\sinh y$, show that

$$I(u)=\int_0^{\sinh^{-1} u} dy=\sinh^{-1} u.$$

(b) Using $v=\tan t$, show that

$$I(u)=\int_0^{\tan^{-1} u} \sec t\,dt.$$

Deduce that

$$\int_0^x \sec t\,dt=\sinh^{-1}(\tan x),\quad \text{for}\ \ |x|<\frac{\pi}{2}.$$

Deduce from this that

$$\cosh\Bigl(\int_0^x \sec t\,dt\Bigr) = \sec x,$$

and hence that

$$\exp\Bigl(\int_0^x \sec t\,dt\Bigr) = \sec x + \tan x.$$

Compare these formulas with the analogue in Exercise 13.

15. Show that

$$E_n^a(t) = \sum_{k=0}^n \frac{a^k}{k!} t^k \quad \text{satisfies} \quad \frac{d}{dt} E_n^a(t) = a E_{n-1}^a(t).$$

From this, show that

$$\frac{d}{dt}\Bigl(e^{-at} E_n^a(t)\Bigr) = -\frac{a^{n+1}}{n!} t^n e^{-at}.$$

16. Use Exercise 15 and the fundamental theorem of calculus to show that

$$\begin{aligned}\int t^n e^{-at}\,dt &= -\frac{n!}{a^{n+1}} E_n^a(t) e^{-at} \\ &= -\frac{n!}{a^{n+1}}\Bigl(1 + at + \frac{a^2t^2}{2!} + \cdots + \frac{a^n t^n}{n!}\Bigr) e^{-at}.\end{aligned}$$

17. Take $a = -i$ in Exercise 16 to produce formulas for

$$\int t^n \cos t\,dt \quad \text{and} \quad \int t^n \sin t\,dt.$$

Exercises on x^r

We define x^r for $x > 0$ and $r \in \mathbb{C}$, as follows:

(3.2.51) $$x^r = e^{r \log x}.$$

18. Show that if $r = n \in \mathbb{N}$, (3.2.51) yields $x^n = x \cdots x$ (n factors).

19. Show that if $r = 1/n$, $x^{1/n}$ defined by (3.2.51) satisfies

$$x = x^{1/n} \cdots x^{1/n} \quad (n \text{ factors}).$$

21. Show that, for $x > 0$,

$$x^{r+s} = x^r x^s, \quad \text{and} \quad (x^r)^s = x^{rs}, \quad \forall\, r, s \in \mathbb{C}.$$

22. Show that, given $r \in \mathbb{C}$,

$$\frac{d}{dx} x^r = r x^{r-1}, \quad \forall\, x > 0.$$

22A. For $y > 0$, evaluate $\int_0^y \cos(\log x)\,dx$ and $\int_0^x \sin(\log x)\,dx$.
Hint. Deduce from (3.2.51) and Euler's formula that

$$\cos(\log x) + i\sin(\log x) = x^i.$$

Use the result of Exercise 22 to integrate x^i.

23. Show that, given $r, r_j \in \mathbb{C},\ x > 0$,

$$r_j \to r \Longrightarrow x^{r_j} \to x^r.$$

24. Given $a > 0$, compute

$$\frac{d}{dx}a^x, \quad x \in \mathbb{R}.$$

25. Compute

$$\frac{d}{dx}x^x, \quad x > 0.$$

26. Prove that

$$x^{1/x} \longrightarrow 1, \quad \text{as} \;\; x \to \infty.$$

Hint. Show that

$$\frac{\log x}{x} \longrightarrow 0, \quad \text{as} \;\; x \to \infty.$$

Some unbounded integrable functions

27. Given $g(s) = 1/\sqrt{1-s^2}$, show that $g \in \mathcal{R}^{\#}([-1,1])$, and that

$$\int_{-1}^{1} \frac{ds}{\sqrt{1-s^2}} = \pi.$$

28. Given $f(t) = 1/\sqrt{t(1-t)}$, show that $f \in \mathcal{R}^{\#}([0,1])$, and that

$$\int_0^1 \frac{dt}{\sqrt{t(1-t)}} = \pi.$$

Hint. Set $t = s^2$.

The arctangent

29. Show that

$$\tan : \Bigl(-\frac{\pi}{2}, \frac{\pi}{2}\Bigr) \longrightarrow \mathbb{R}$$

is one-to-one and onto, with inverse

$$\tan^{-1} : \mathbb{R} \longrightarrow \Bigl(-\frac{\pi}{2}, \frac{\pi}{2}\Bigr),$$

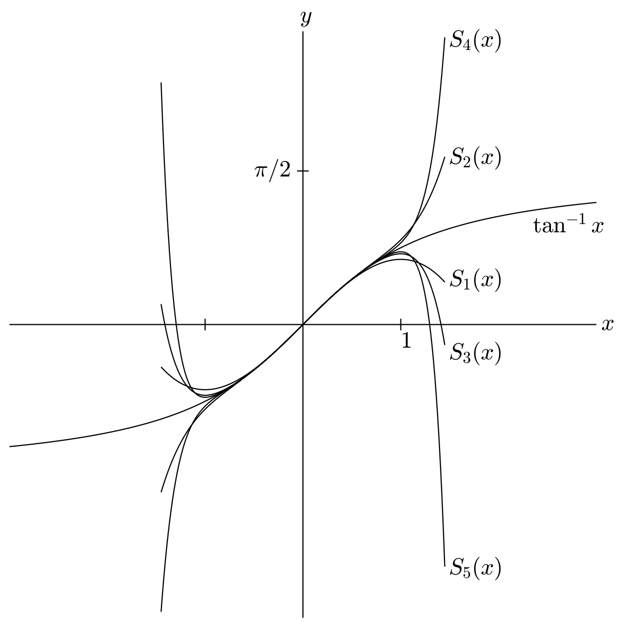

Figure 3.2.5. Power series approximations $S_n(x)$ to $\tan^{-1} x$

given, via Exercise 5, by

$$\tan^{-1} x = \int_0^x \frac{dt}{1+t^2}.$$

30. Use the integral formula above to show that $\tan^{-1} x$ is given by the power series

$$\tan^{-1} x = \sum_{k=0}^{\infty} (-1)^k \frac{x^{2k+1}}{2k+1}, \quad \text{for} \quad |x| \leq 1.$$

Show that this series diverges for $|x| > 1$.

See Figure 3.2.5 for the graphs of $\tan^{-1} x$ over $|x| \leq 3$ and of $S_n(x)$ over $|x| \leq 1.45$, where

$$S_n(x) = \sum_{k=0}^{n} (-1)^k \frac{x^{2k+1}}{2k+1},$$

and $1 \leq n \leq 5$.

Making a trig table

These exercises guide the reader who can use a computer or calculator for numerical work through the following task:

31. Make a table of $\cos \ell^\circ$ and $\sin \ell^\circ$, for the integers ℓ. Achieve at least 10 digits of accuracy.

Here ℓ° is the same as $(\pi/180)\ell$ radians, so, by Euler's identity,

$$\cos \ell^\circ + i \sin \ell^\circ = e^{i\theta_\ell}, \quad \theta_\ell = \frac{\pi \ell}{180}. \tag{3.2.52}$$

It suffices to compute $\cos \ell^\circ$ and $\sin \ell^\circ$ for $0 \le \ell \le 45$, since trigonometric identities then lead to the computation for other integer values of ℓ.

One approach would be to use the power series for e^z, with $z = i\theta_\ell$. To implement this requires having an accurate numerical evaluation of π. A method of obtaining this was presented in Exercise 7. Here we want to explore an alternative approach to the computation of (3.2.52), which does not require a previously computed evaluation of π. It starts with the following identities:

$$e^{\pi i/3} = \frac{1}{2}(1 + i\sqrt{3}), \quad e^{\pi i/4} = \frac{\sqrt{2}}{2}(1 + i), \tag{3.2.53}$$

cf. Exercise 2, supplemented by

$$e^{2\pi i/5} = c_5 + i s_5, \quad c_5 = \frac{1}{4}(\sqrt{5} - 1), \tag{3.2.54}$$

obtained in Exercise 4 of Appendix A.4, which in turn yields

$$s_5 = \sqrt{1 - c_5^2} = \frac{1}{4}\sqrt{10 + 2\sqrt{5}}. \tag{3.2.55}$$

In light of this, we formulate the following exercise:

32. Numerically evaluate the real and imaginary parts of

$$e^{\pi i/3}, \quad e^{\pi i/4}, \quad e^{2\pi i/5}.$$

Equivalently, numerically evaluate

$$\sqrt{3}, \quad \sqrt{3}, \quad \sqrt{5}, \quad \sqrt{10 + 2\sqrt{5}}.$$

The following expands the scope of Exercise 32.

33. Here is a way to approximate $\sqrt{a}$, given $a \in \mathbb{R}^+$. Suppose you have an approximation x_k to $\sqrt{a}$,

$$x_k - \sqrt{a} = \delta_k.$$

Square this to obtain $x_k^2 + a - 2x_k\sqrt{a} = \delta_k^2$, hence

$$\sqrt{a} = x_{k+1} - \frac{\delta_k^2}{2x_k}, \quad x_{k+1} = \frac{a + x_k^2}{2x_k}.$$

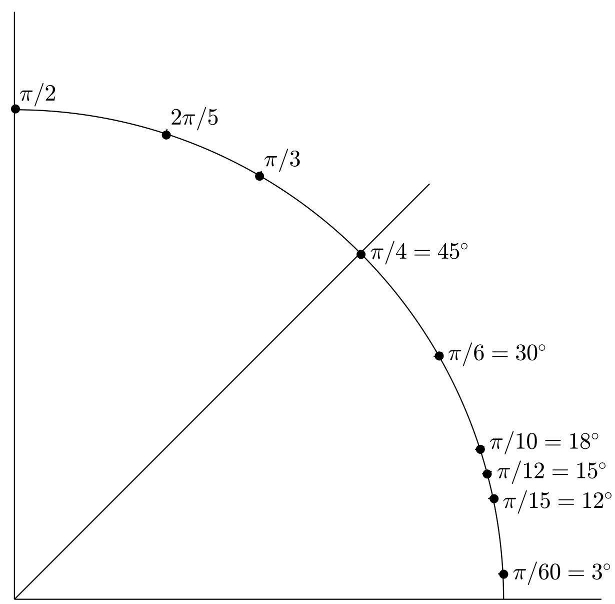

Figure 3.2.6. Special angles θ, at which to evaluate $\sin\theta$ and $\cos\theta$

Then x_{k+1} is an improved approximation, as long as $|\delta_k| < 2x_k$. One can iterate this. Try it on

$$\sqrt{2} \approx \frac{7}{5}, \quad \sqrt{3} \approx \frac{7}{4}, \quad \sqrt{5} \approx \frac{9}{4}.$$

How many iterations does it take to approximate these quantities to 12 digits of accuracy? Going further, take $10 + 2\sqrt{5} \approx 14.5$, and hence

$$\sqrt{10 + 2\sqrt{5}} \approx 3.8.$$

34. Verify the following identities:

$$\begin{aligned} e^{\pi i/2} e^{-\pi i/3} &= e^{\pi i/6}, \\ e^{\pi i/2} e^{-2\pi i/5} &= e^{\pi i/10}, \\ e^{\pi i/3} e^{-\pi i/4} &= e^{\pi i/12}, \\ e^{2\pi i/5} e^{-\pi i/3} &= e^{\pi i/15}. \end{aligned}$$

See Figure 3.2.6 for representations of the relevant angles.

35. Verify also that

$$e^{\pi i/12} e^{-\pi i/15} = e^{\pi i/60},$$

and deduce from results of Exercise 34 that

$$\begin{aligned}e^{\pi i/60} &= e^{2\pi i/3}e^{-\pi i/4}e^{-2\pi i/5} \\ &= \frac{\sqrt{2}}{4}(-1+i\sqrt{3})(1-i)(c_5 - is_5).\end{aligned}$$

36. Deduce from Exercise 35 that

$$2\sqrt{2}\,\sin\frac{\pi}{60} = (\sqrt{3}+1)c_5 - (\sqrt{3}-1)s_5.$$

Use this to produce a numerical evaluation of $\sin(\pi/60)$. Similarly, numerically evaluate $\cos(\pi/60)$.

37. Use the results of Exercises 32–36 to fill in the trig table for $\sin \ell^\circ$ and $\cos \ell^\circ$, when ℓ is an integer in $\{0, \dots, 45\}$ that is divisible by 3.

38. (Application to the evaluation of π.) Use the result of Exercise 36 to produce a numerical evaluation of

$$\tan\frac{\pi}{60} = \alpha.$$

Having this result, apply the power series in Exercise 30 to $\tan^{-1}\alpha$, to evaluate π to 10 digits of accuracy. How many terms in the power series are needed for this task?

39. Think about ways to proceed from results of the exercises above to numerically evaluate

$$e^{\pi i/180} = \cos 1^\circ + i \sin 1^\circ,$$

and from there to complete Exercise 31.

One approach. Note that $(e^{\pi i/180})^3 = e^{\pi i/60}$, which is evaluated in Exercise 36. To evaluate

$$(1+a)^{1/3}, \quad \text{given } a \in \mathbb{C}, \text{ small},$$

take $1 + a/3$ as a first approximation. Then evaluate

$$(1+a)(1+a/3)^{-3} = 1 + a_1,$$

with $a_1 \in \mathbb{C}$, substantially *smaller*, and iterate, obtaining

$$(1+a)^{1/3} = \Big(1+\frac{a}{3}\Big)\Big(1+\frac{a_1}{3}\Big)\cdots.$$

Alternative. Having evaluated π in Exercise 38, plug this into the power series for $e^{\pi i/180}$.

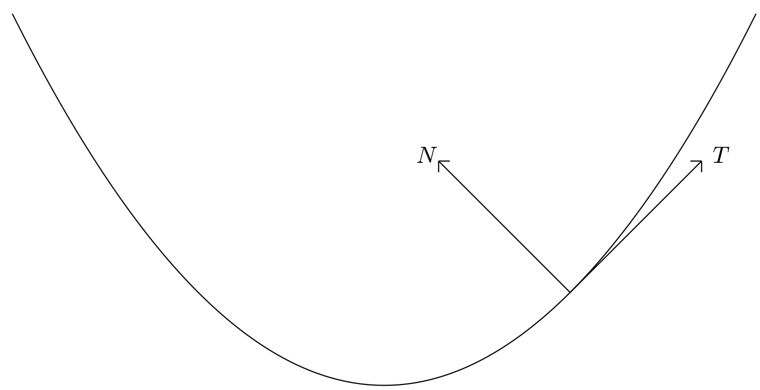

Figure 3.3.1. Unit tangent and normal vectors to a parabolic curve

3.3. Curvature of planar curves

The *curvature* of a curve $\gamma : (a, b) \to \mathbb{R}^n$ is a measure of how it is not straight. Assume γ has non-vanishing velocity. We can parametrize γ by arclength, so we have the unit tangent vector

$$\gamma'(s) = T(s), \quad \|T(s)\| \equiv 1. \tag{3.3.1}$$

Then γ is a straight line if and only if $T(s)$ is constant. Thus a measure of how γ curves is given by

$$T'(s). \tag{3.3.2}$$

We call this the curvature vector of γ. Note that

$$T \cdot T \equiv 1 \Longrightarrow T'(s) \cdot T(s) = 0, \tag{3.3.3}$$

so $T'(s)$ is orthogonal to $T(s)$.

Let us now specialize to planar curves, so $\gamma : (a, b) \to \mathbb{R}^2$. In such a case, we apply counterclockwise rotation by 90° to $T(s)$ to get a unit normal to γ:

$$N(s) = JT(s), \quad J = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}. \tag{3.3.4}$$

(Here we represent vectors in $\mathbb{R}^2$ as column vectors.) See Figure 3.3.1 for an illustration of the unit tangent and normal vectors to a parabolic curve at a point.

In this situation, (3.3.3) implies that $T'(s)$ is parallel to $N(s)$, say

$$T'(s) = \kappa(s)N(s), \tag{3.3.5}$$

and we call $\kappa(s)$ the *curvature* of γ. Note that, by (3.3.4),

$$\begin{aligned} N'(s) &= \kappa(s)JN(s) \\ &= \kappa(s)J^2T(s) \\ &= -\kappa(s)T(s). \end{aligned} \tag{3.3.6}$$

We set up the pair of equations

$$\begin{aligned} T'(s) &= \kappa(s)N(s), \\ N'(s) &= -\kappa(s)T(s), \end{aligned} \tag{3.3.7}$$

as a precursor to the more elaborate Frenet-Serret equations for curves in $\mathbb{R}^3$, given in (3.4.19). However, in the planar case, we can make do with

$$T'(s) = \kappa(s)JT(s) \tag{3.3.8}$$

as the defining equation for curvature of γ.

This sets us up to consider the following problem. Given a smooth function

$$\kappa : (a,b) \longrightarrow \mathbb{R}, \tag{3.3.9}$$

see if it determines a unit-speed curve $\gamma : (a,b) \to \mathbb{R}^2$ with curvature κ. We should impose initial conditions: take $s_0 \in (a,b)$ and specify

$$\gamma(s_0) = p_0, \quad T(s_0) = T_0, \quad p_0, T_0 \in \mathbb{R}^2, \ \|T_0\| = 1. \tag{3.3.10}$$

As we will show below, (3.3.8)–(3.3.10) has a unique solution

$$T : (a,b) \longrightarrow \mathbb{R}^2. \tag{3.3.11}$$

Furthermore,

$$\begin{aligned} \frac{d}{ds}\|T(s)\|^2 &= 2T'(s) \cdot T(s) \\ &= 2\kappa(s)JT(s) \cdot T(s) \\ &= 0, \end{aligned} \tag{3.3.12}$$

so

$$\|T(s)\| = 1, \quad \forall\, s \in (a,b). \tag{3.3.13}$$

Then we take

$$\gamma(s) = p_0 + \int_{s_0}^{s} T(\tau)\, d\tau, \tag{3.3.14}$$

to obtain the desired curve.

We will produce a specific formula for the solution to (3.3.8)–(3.3.10). We start with the case

$$\kappa(s) \equiv \kappa, \quad \text{real constant}. \tag{3.3.15}$$

In this case, the differential equation (3.3.8) becomes

$$\frac{dT}{ds} = \kappa JT(s). \tag{3.3.16}$$

Say $s_0 = 0$, so the initial condition is

$$T(0) = T_0. \tag{3.3.17}$$

This equation is formally similar to the equation (3.2.9), with $a \in \mathbb{C}$ replaced by $\kappa J \in M(2, \mathbb{R})$. The solution is given in terms of the *matrix exponential*. In general, for $A \in M(n, \mathbb{R})$, or $A \in M(n, \mathbb{C})$, we set

$$e^{tA} = \sum_{k=0}^{\infty} \frac{t^k}{k!} A^k. \tag{3.3.18}$$

A development of the matrix exponential, parallel to that of the exponential of complex numbers, is presented in §C.2. It follows that the solution to (3.3.16)–(3.3.17) is

$$T(s) = e^{s\kappa J} T_0. \tag{3.3.19}$$

This leaves us with the task of evaluating the matrix exponential e^{tJ}, for $t \in \mathbb{R}$. In view of the similarity

$$J^2 = -I, \quad i^2 = -1, \tag{3.3.20}$$

it is natural to guess that e^{tJ} satisfies the following variant of the Euler identity:

$$e^{tJ} = (\cos t)I + (\sin t)J, \quad t \in \mathbb{R}. \tag{3.3.21}$$

This is the case. One way to prove it is the following. Denote the right side of (3.3.21) by $X(t)$. Then, thanks to (3.1.39), or (3.2.41),

$$\begin{aligned} X'(t) &= -(\sin t)I + (\cos t)J \\ &= JX(t), \end{aligned} \tag{3.3.22}$$

and $X(0) = I$. This has the same form as (3.3.16)–(3.3.17), and §C.2 shows that this leads to $X(t) = e^{tJ}$, hence to (3.3.21).

Returning to (3.3.18), we see that

$$T(s) = (\cos s\kappa)T_0 + (\sin s\kappa)JT_0, \tag{3.3.23}$$

and

$$\gamma(s) = p_0 + \int_0^s T(\tau)\, d\tau. \tag{3.3.24}$$

For example, if

$$\gamma(0) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad T(0) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad \kappa = 1, \tag{3.3.25}$$

we have

$$T(s) = \begin{pmatrix} -\sin s \\ \cos s \end{pmatrix}, \tag{3.3.26}$$

hence

$$\gamma(s) = \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \int_0^s \begin{pmatrix} -\sin \tau \\ \cos \tau \end{pmatrix} d\tau = \begin{pmatrix} \cos s \\ \sin s \end{pmatrix}, \tag{3.3.27}$$

revealing γ as the unit circle.

We return to (3.3.8), for general smooth functions $\kappa(s)$. Using the formula

$$\frac{d}{ds} e^{\alpha(s)J} = \alpha'(s) J e^{\alpha(s)J}, \tag{3.3.28}$$

we see that (3.3.8)–(3.3.10) is solved by

$$T(s) = e^{\alpha(s)J} T_0, \quad \alpha(s) = \int_{s_0}^{s} \kappa(\sigma)\, d\sigma. \tag{3.3.29}$$

Exercises

1. Consider a curve $c(t)$ in $\mathbb{R}^2$ (not necessarily unit speed), with velocity $v(t)$ and acceleration $a(t)$, given by

$$v(t) = c'(t), \quad a(t) = v'(t).$$

Assume $v(t) \neq 0$. Take

$$T(t) = \frac{v(t)}{\|v(t)\|}, \quad N(t) = JT(t), \quad s(t) = \int_{t_0}^{t} \|v(\tau)\|\, d\tau,$$

so $s(t)$ is the arclength parameter. Show that

$$a(t) = \frac{d^2 s}{dt^2} T(t) + \kappa(t) \Bigl(\frac{ds}{dt}\Bigr)^2 N(t). \tag{3.3.30}$$

Hint. Differentiate $v(t) = (ds/dt)T(t)$ and use the chain rule $dT/dt = (ds/dt)(dT/ds)$.

2. Deduce from Exercise 1 that

$$\kappa \Bigl(\frac{ds}{dt}\Bigr)^2 = a \cdot N = \frac{a \cdot Jv}{\|v\|},$$

hence

$$\kappa = \frac{a \cdot Jv}{\|v\|^3}. \tag{3.3.31}$$

3. Consider the ellipse

$$\gamma(t) = (a \cos t, b \sin t).$$

Use the results of Exercise 2 to compute its curvature. Verify that, for such an ellipse,

$$\kappa(t) = \frac{ab}{|\gamma(t + \pi/2)|^3}.$$

Similarly, compute the curvature of the following curves:

4. Parabola

$$\gamma(t) = (t, t^2).$$

5. Sine curve

$$\gamma(t) = (t, \sin t).$$

6. Spiral

$$\gamma(t) = (e^t \cos t, e^t \sin t).$$

7. Find the unit-speed curve $\gamma : (0, \infty) \to \mathbb{R}^2$ satisfying

$$\gamma(1) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \gamma'(1) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \kappa(s) = \frac{1}{s}.$$

Hint. To compute

$$\int_1^s \begin{pmatrix} \cos \log \sigma \\ \sin \log \sigma \end{pmatrix} d\sigma,$$

use

$$\begin{aligned} &\int_1^s (\cos \log \sigma + i \sin \log \sigma)\, d\sigma \\ &= \int_1^s e^{i \log \sigma}\, d\sigma = \int_1^s \sigma^i\, d\sigma = \frac{\sigma^{i+1}}{i+1}\bigg|_1^s. \end{aligned}$$

3.4. Curvature and torsion of curves in $\mathbb{R}^3$

Given a curve $c(t) = (x(t), y(t), z(t))$ in 3-space, we define its velocity and acceleration by

$$v(t) = c'(t), \quad a(t) = v'(t) = c''(t). \tag{3.4.1}$$

We also define its speed $s'(t)$ and arclength by

$$s'(t) = \|v(t)\|, \quad s(t) = \int_{t_0}^{t} s'(\tau)\, d\tau, \tag{3.4.2}$$

assuming we start at $t = t_0$. We define the unit tangent vector to the curve as

$$T(t) = \frac{v(t)}{\|v(t)\|}. \tag{3.4.3}$$

Henceforth we assume the curve is parametrized by arclength.

We define the *curvature* $\kappa(s)$ of the curve and the normal $N(s)$ by

$$\kappa(s) = \left\| \frac{dT}{ds} \right\|, \quad \frac{dT}{ds} = \kappa(s)N(s). \tag{3.4.4}$$

Note that

$$T(s) \cdot T(s) = 1 \Longrightarrow T'(s) \cdot T(s) = 0, \tag{3.4.5}$$

so indeed $N(s)$ is orthogonal to $T(s)$. We then define the binormal $B(s)$ by

$$B(s) = T(s) \times N(s). \tag{3.4.6}$$

For each s, the vectors $T(s)$, $N(s)$ and $B(s)$ are mutually orthogonal unit vectors, known as the Frenet frame for the curve $c(s)$. Rules governing the cross product yield

$$T(s) = N(s) \times B(s), \quad N(s) = B(s) \times T(s). \tag{3.4.7}$$

For material on the cross product, see §2.5. The result (3.4.7) follows from (2.5.7); see Exercise 5 in §2.5. See Figure 3.4.1 for an illustration of a Frenet frame at a point.

The *torsion* of a curve measures the change in the plane generated by $T(s)$ and $N(s)$, or equivalently it measures the rate of change of $B(s)$. Note that, parallel to (3.4.5),

$$B(s) \cdot B(s) = 1 \Longrightarrow B'(s) \cdot B(s) = 0.$$

Also, differentiating (3.4.6) and using (3.4.4), we have

$$B'(s) = T'(s) \times N(s) + T(s) \times N'(s) = T(s) \times N'(s) \Longrightarrow B'(s) \cdot T(s) = 0. \tag{3.4.8}$$

We deduce that $B'(s)$ is parallel to $N(s)$. We define the torsion by

$$\frac{dB}{ds} = -\tau(s)N(s). \tag{3.4.9}$$

We complement the formulas (3.4.4) and (3.4.9) for dT/ds and dB/ds with one for dN/ds. Since $N(s) = B(s) \times T(s)$, we have

$$\frac{dN}{ds} = \frac{dB}{ds} \times T + B \times \frac{dT}{ds} = \tau N \times T + \kappa B \times N, \tag{3.4.10}$$

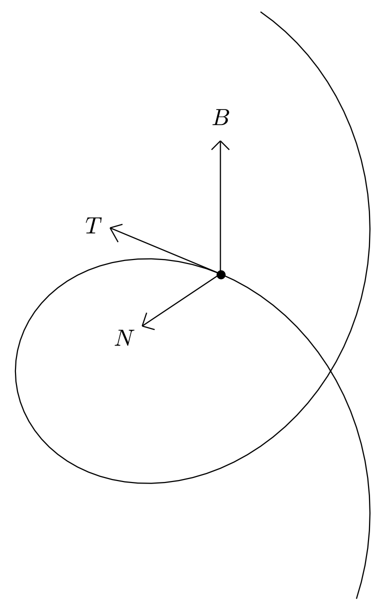

Figure 3.4.1. Frenet frame at a point on a 3D curve

or

$$\frac{dN}{ds} = -\kappa(s)T(s) + \tau(s)B(s). \tag{3.4.11}$$

Together, (3.4.4), (3.4.9) and (3.4.11) are known as the Frenet-Serret formulas.

Example. Pick $a, b > 0$ and consider the helix

$$c(t) = (a\cos t, a\sin t, bt). \tag{3.4.12}$$

Then $v(t) = (-a\sin t, a\cos t, b)$ and $\|v(t)\| = \sqrt{a^2+b^2}$, so we can pick $s = t\sqrt{a^2+b^2}$ to parametrize by arc length. We have

$$T(s) = \frac{1}{\sqrt{a^2+b^2}}(-a\sin t, a\cos t, b), \tag{3.4.13}$$

hence

$$\frac{dT}{ds} = \frac{1}{a^2+b^2}(-a\cos t, -a\sin t, 0). \tag{3.4.14}$$

By (7.4), this gives

$$\kappa(s) = \frac{a}{a^2+b^2}, \quad N(s) = (-\cos t, -\sin t, 0). \tag{3.4.15}$$

Hence

$$B(s) = T(s) \times N(s) = \frac{1}{\sqrt{a^2+b^2}}(b\sin t, -b\cos t, a). \tag{3.4.16}$$

Then

$$\frac{dB}{ds} = \frac{1}{a^2+b^2}(b\cos t, b\sin t, 0), \tag{3.4.17}$$

so, by (3.4.9),

$$\tau(s) = \frac{b}{a^2+b^2}. \tag{3.4.18}$$

In particular, for the helix (3.4.12), we see that the curvature and torsion are *constant*.

Let us collect the Frenet-Serret equations

$$\begin{aligned} \frac{dT}{ds} &= \qquad \kappa N \\ \frac{dN}{ds} &= -\kappa T \qquad\qquad + \tau B \\ \frac{dB}{ds} &= \qquad -\tau N \end{aligned} \tag{3.4.19}$$

for a smooth curve $c(s)$ in $\mathbb{R}^3$, parametrized by arclength, with unit tangent $T(s)$, normal $N(s)$, and binormal $B(s)$, given by

$$N(s) = \frac{1}{\kappa(s)} T'(s), \quad B(s) = T(s) \times N(s), \tag{3.4.20}$$

assuming $\kappa(s) = \|T'(s)\| > 0$.

The differential equation (3.4.19) is treated in texts on differential equations. A treatment can be found in [**19**]. If $\kappa(s)$ and $\tau(s)$ are given smooth functions on an interval $I = (a, b)$ and $s_0 \in I$, then, given $T_0, N_0, B_0 \in \mathbb{R}^3$, (3.4.19) has a unique solution on $s \in I$ satisfying

$$T(s_0) = T_0, \quad N(s_0) = N_0, \quad B(s_0) = B_0. \tag{3.4.21}$$

We now establish the following.

Proposition 3.4.1. *Assume κ and τ are given smooth functions on I, with $\kappa > 0$ on I. Assume $\{T_0, N_0, B_0\}$ is an orthonormal basis of $\mathbb{R}^3$, such that $B_0 = T_0 \times N_0$. Then there exists a smooth, unit-speed curve $c(s)$, $s \in I$, for which the solution to* (3.4.19) *and* (3.4.21) *is the Frenet frame.*

To construct the curve, take $T(s), N(s)$, and $B(s)$ to solve (3.4.19) and (3.4.21), pick $p \in \mathbb{R}^3$ and set

$$c(s) = p + \int_{s_0}^{s} T(\sigma)\, d\sigma, \tag{3.4.22}$$

so $T(s) = c'(s)$ is the velocity of this curve. To deduce that $\{T(s), N(s), B(s)\}$ is the Frenet frame for $c(s)$, for all $s \in I$, we need to know:

$$\{T(s), N(s), B(s)\} \text{ orthonormal, with } B(s) = T(s) \times N(s), \quad \forall\, s \in I. \tag{3.4.23}$$

In order to pursue the analysis further, it is convenient to form the 3×3 matrix-valued function

$$F(s) = (T(s), N(s), B(s)), \tag{3.4.24}$$

whose *columns* consist respectively of $T(s)$, $N(s)$, and $B(s)$. Then (3.4.23) is equivalent to

$$F(s) \in SO(3), \quad \forall\, s \in I, \tag{3.4.25}$$

with $SO(3)$ defined as above (2.5.7). The hypothesis on $\{T_0, N_0, B_0\}$ stated in Proposition 3.4.1 is equivalent to $F_0 = (T_0, N_0, B_0) \in SO(3)$. Now $F(s)$ satisfies the differential equation

$$F'(s) = F(s)A(s), \quad F(s_0) = F_0, \tag{3.4.26}$$

where

$$A(s) = \begin{pmatrix} 0 & -\kappa(s) & 0 \\ \kappa(s) & 0 & -\tau(s) \\ 0 & \tau(s) & 0 \end{pmatrix}. \tag{3.4.27}$$

Note that

$$\frac{dF^*}{ds} = A(s)^* F(s)^* = -A(s)F(s)^*, \tag{3.4.28}$$

since $A(s)$ in (3.4.27) is skew-adjoint. Hence

$$\begin{aligned} \frac{d}{ds}\, F(s)F(s)^* &= \frac{dF}{ds} F(s)^* + F(s)\frac{dF^*}{ds} \\ &= F(s)A(s)F(s)^* - F(s)A(s)F(s)^* \\ &= 0. \end{aligned} \tag{3.4.29}$$

Thus, whenever (3.4.26)–(3.4.27) hold,

$$F_0 F_0^* = I \Longrightarrow F(s)F(s)^* \equiv I, \tag{3.4.30}$$

and we have (3.4.23).

Let us specialize the system (3.4.19), or equivalently (3.4.26), to the case where κ and τ are *constant*, i.e.,

$$F'(s) = F(s)A, \quad A = \begin{pmatrix} 0 & -\kappa & 0 \\ \kappa & 0 & -\tau \\ 0 & \tau & 0 \end{pmatrix}, \tag{3.4.31}$$

with solution

$$F(s) = F_0\, e^{(s-s_0)A}. \tag{3.4.32}$$

We have already seen in that a helix of the form (3.4.12) has curvature κ and torsion τ, with

$$\kappa = \frac{a}{a^2 + b^2}, \quad \tau = \frac{b}{a^2 + b^2}, \tag{3.4.33}$$

and hence

$$a = \frac{\kappa}{\kappa^2 + \tau^2}, \quad b = \frac{\tau}{\kappa^2 + \tau^2}. \tag{3.4.34}$$

In (3.4.12), s and t are related by $t = s\sqrt{\kappa^2 + \tau^2}$.

We can also see such a helix arise via a direct calculation of e^{sA}, which we now produce. First, a straightforward calculation gives, for A as in (3.4.31),

$$\det(\lambda I - A) = \lambda(\lambda^2 + \kappa^2 + \tau^2), \tag{3.4.35}$$

hence

$$\operatorname{Spec}(A) = \{0, \pm i\sqrt{\kappa^2 + \tau^2}\}. \tag{3.4.36}$$

An inspection shows that we can take

$$v_1 = \frac{1}{\sqrt{\kappa^2+\tau^2}} \begin{pmatrix} \tau \\ 0 \\ \kappa \end{pmatrix}, \quad v_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad v_3 = \frac{1}{\sqrt{\kappa^2+\tau^2}} \begin{pmatrix} -\kappa \\ 0 \\ \tau \end{pmatrix}, \tag{3.4.37}$$

and then

$$Av_1 = 0, \quad Av_2 = \sqrt{\kappa^2+\tau^2}\, v_3, \quad Av_3 = -\sqrt{\kappa^2+\tau^2}\, v_2. \tag{3.4.38}$$

In particular, with respect to the basis $\{v_2, v_3\}$ of $V = \operatorname{Span}\{v_2, v_3\}$, $A|_V$ has the matrix representation

$$B = \sqrt{\kappa^2+\tau^2} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}. \tag{3.4.39}$$

We see that

$$e^{sA} v_1 = v_1, \tag{3.4.40}$$

while, in light of the calculations giving (3.3.21),

$$\begin{aligned} e^{sA} v_2 &= (\cos s\sqrt{\kappa^2+\tau^2}) v_2 \quad + (\sin s\sqrt{\kappa^2+\tau^2}) v_3, \\ e^{sA} v_3 &= -(\sin s\sqrt{\kappa^2+\tau^2}) v_2 + (\cos s\sqrt{\kappa^2+\tau^2}) v_3. \end{aligned} \tag{3.4.41}$$

For general variable $\kappa(s)$ and $\tau(s)$, the system (3.4.26) does not have a neat closed-form solution. However, here is a class, more general than the case of constant curvature and torsion, where it does. Assume there exist $\kappa_0, \tau_0 \in \mathbb{R}$ and a smooth function $\beta : I \to \mathbb{R}$ such that

$$\kappa(s) = \beta(s)\kappa_0, \quad \tau(s) = \beta(s)\tau_0. \tag{3.4.42}$$

Then, in (3.4.27),

$$A(s) = \beta(s) A_0, \quad A_0 = \begin{pmatrix} 0 & -\kappa_0 & 0 \\ \kappa_0 & 0 & -\tau_0 \\ 0 & \tau_0 & 0 \end{pmatrix}, \tag{3.4.43}$$

and, parallel to (3.3.29), a solution to (3.4.26) is given by

$$F(s) = F_0 e^{\alpha(s) A_0}, \quad \alpha(s) = \int_{s_0}^{s} \beta(\tau)\, d\tau. \tag{3.4.44}$$

Exercises

1. Consider a curve $c(t)$ in $\mathbb{R}^3$, not necessarily parametrized by arclength. Show that the acceleration $a(t)$ is given by

$$a(t) = \frac{d^2s}{dt^2} T + \kappa \Big(\frac{ds}{dt}\Big)^2 N. \tag{3.4.45}$$

Hint. Differentiate $v(t) = (ds/dt)T(t)$ and use the chain rule $dT/dt = (ds/dt)(dT/ds)$, plus (3.4.4).

2. Show that

$$\kappa B = \frac{v \times a}{\|v\|^3}. \tag{3.4.46}$$

Hint. Take the cross product of both sides of (3.4.45) with T, and use (3.4.6).

3. In the setting of Exercises 1–2, show that

$$\kappa^2 \tau \|v\|^6 = -a \cdot (v \times a'). \tag{3.4.47}$$

Deduce from (3.4.46)–(3.4.47) that

$$\tau = \frac{(v \times a) \cdot a'}{\|v \times a\|^2}. \tag{3.4.48}$$

Hint. Proceed from (3.4.46) to

$$\frac{d}{dt}(\kappa \|v\|^3)\, B + \kappa \|v\|^3 \frac{dB}{dt} = \frac{d}{dt}(v \times a) = v \times a',$$

and use $dB/dt = -\tau(ds/dt)N$, as a consequence of (3.4.9). Then dot with a, and use $a \cdot N = \kappa \|v\|^2$, from (3.4.45), to get (3.4.47).

4. Consider the curve $c(t)$ in $\mathbb{R}^3$ given by

$$c(t) = (a\, \cos t, b\, \sin t, t),$$

where a and b are given positive constants. Compute the curvature, torsion, and Frenet frame.
Hint. Use (3.4.46) to compute κ and B. Then use $N = B \times T$. Use (3.4.48) to compute τ.

5. Suppose c and $\tilde{c}$ are two curves, both parametrized by arc length over $0 \le s \le L$, and both having the same curvature $\kappa(s) > 0$ and the same torsion $\tau(s)$. Show that there exit $x_0 \in \mathbb{R}^3$ and $A \in O(3)$ such that

$$\tilde{c}(s) = Ac(s) + x_0, \quad \forall\ s \in [0, L].$$

Hint. To begin, show that if their Frenet frames coincide at $s = 0$, i.e., $\widetilde{T}(0) = T(0),\ \widetilde{N}(0) = N(0),\ \widetilde{B}(0) = B(0)$, then $\widetilde{T} \equiv T,\ \widetilde{N} \equiv N,\ \widetilde{B} \equiv B$.

6. Suppose c is a curve in $\mathbb{R}^3$ with curvature $\kappa > 0$. Show that there exists a plane in which $c(t)$ lies for all t if and only if $\tau \equiv 0$.
Hint. When $\tau \equiv 0$, the plane should be parallel to the orthogonal complement of B.

7. Let $\gamma : I \to \mathbb{R}^3$ be a smooth, unit-speed curve, with curvature and torsion κ, τ. Assume $\kappa > 0$. Take $A \in O(3)$, and set

$$\sigma(s) = A\gamma(s).$$

Show that the curvature κ_σ and torsion τ_σ of σ satisfy $\kappa_\sigma(s) = \kappa(s)$, and

$$\tau_\sigma(s) = \pm\tau(s), \quad \text{if} \quad \det A = \pm 1.$$

8. Let $\gamma : I \to \mathbb{R}^3$ be a unit-speed curve, with Frenet frame (T, N, B). Assume $\kappa, \tau > 0$. Set

$$\sigma(s) = \int_0^s B(t)\, dt,$$

also a unit-speed curve. Show that it has the Frenet frame

$$T_\sigma = B, \quad N_\sigma = -N, \quad B_\sigma = T.$$

Compute its curvature and torsion.

Curves on the unit sphere in $\mathbb{R}^3$

In Exercises 9–12, let $\gamma : I \to \mathbb{R}^3$ be a unit-speed curve satisfying

$$\gamma(s) \cdot \gamma(s) \equiv 1, \quad \text{i.e.,} \quad \gamma : I \longrightarrow S^2.$$

9. Show that

$$\gamma(s) \cdot T'(s) \equiv -1, \quad \text{i.e.,} \quad \kappa(s)\,\gamma(s) \cdot N(s) \equiv -1,$$

and hence the curvature satisfies

$$\kappa(s) \geq 1.$$

Hint. First show $\gamma \cdot T \equiv 0$. Apply d/ds to this.

10. For $s \in I$, set

$$\nu(s) = \gamma(s) \times T(s).$$

Show that $(\gamma(s), T(s), \nu(s))$ is an orthonormal basis of $\mathbb{R}^3$. Show that, for each $s \in I$,

$$N(s) = a(s)\gamma(s) + b(s)\nu(s), \tag{3.4.49}$$

with

$$a(s) = N(s) \cdot \gamma(s) = -\frac{1}{\kappa(s)}.$$

Hint. Use $N \cdot T \equiv 0$ and $\nu \cdot \gamma \equiv 0$.

11. Deduce that

$$\kappa \equiv 1 \Longleftrightarrow b \equiv 0 \Longleftrightarrow N \equiv -\gamma,$$

hence

$$\kappa(s) \equiv 1 \Longleftrightarrow \gamma''(s) \equiv -\gamma(s).$$

12. In (3.4.49), show that

$$b(s) = \gamma(s) \cdot B(s).$$

Deduce that

$$b'(s) = \frac{\tau(s)}{\kappa(s)}.$$

Hint. Show that $b = N \cdot (\gamma \times T) = \gamma \cdot (T \times N)$.

Chapter 4

Multivariable differential calculus

This chapter develops differential calculus on domains in n-dimensional Euclidean space $\mathbb{R}^n$.

In §4.1 we define the derivative of a function $F : \mathcal{O} \to \mathbb{R}^m$, where $\mathcal{O}$ is an open subset of $\mathbb{R}^n$, as a linear map from $\mathbb{R}^n$ to $\mathbb{R}^m$. We establish some basic properties, such as the chain rule. We use the one-dimensional integral as a tool to show that, if the matrix of first order partial derivatives of F is continuous on $\mathcal{O}$, then F is differentiable on $\mathcal{O}$.

In §4.2 we consider higher derivatives of functions with additional smoothness. We discuss two convenient multi-index notations for higher derivatives, and an alternative multi-linear notation. We derive the Taylor formula with remainder for the power series of a smooth function F on $\mathcal{O} \subset \mathbb{R}^n$, producing expressions of this formula in each of these three notations.

We also look at critical points of a real-valued, smooth function F on $\mathcal{O} \subset \mathbb{R}^n$, and give conditions that such a critical point gives a local maximum, a local minimum, or a saddle, in terms of the behavior of the $n \times n$ matrix of second-order partial derivatives of F.

In §4.3 we establish the Inverse Function Theorem, stating that a smooth map $F : \mathcal{O} \to \mathbb{R}^n$ with an invertible derivative $DF(p)$ has a smooth inverse defined near $q = F(p)$. We derive the Implicit Function Theorem as a consequence of this. As a tool in proving the Inverse Function Theorem, we use a fixed point theorem known as the Contraction Mapping Principle. The inverse and implicit function theorems will be essential tools in our study of surfaces, in Chapter 6.

4.1. The derivative

Let $\mathcal{O}$ be an open subset of $\mathbb{R}^n$, and $F : \mathcal{O} \to \mathbb{R}^m$ a continuous function. We say F is differentiable at a point $x \in \mathcal{O}$, with derivative L, if $L : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation such that, for $y \in \mathbb{R}^n$, small,

$$F(x+y) = F(x) + Ly + R(x,y) \tag{4.1.1}$$

with

$$\frac{\|R(x,y)\|}{\|y\|} \to 0 \quad \text{as} \quad y \to 0. \tag{4.1.2}$$

We write (4.1.2) as

$$R(x,y) = o(\|y\|). \tag{4.1.3}$$

In (4.1.2), we use the *Euclidean* norm on $\mathbb{R}^n$ and $\mathbb{R}^m$. As seen in §2.1, this norm is defined by

$$\|x\| = \bigl(x_1^2 + \cdots + x_n^2\bigr)^{1/2} \tag{4.1.4}$$

for $x = (x_1, \dots, x_n) \in \mathbb{R}^n$. We denote the derivative at x by $DF(x) = L$, and rewrite (4.1.1) as

$$F(x+y) = F(x) + DF(x)y + R(x,y). \tag{4.1.5}$$

In particular, if $\{e_1, \dots, e_n\}$ denotes the standard basis of $\mathbb{R}^n$, and if $DF(x)$ exists, we have, for $h \in \mathbb{R}$ small,

$$F(x + he_j) = F(x) + hDF(x)e_j + o(h), \tag{4.1.6}$$

or equivalently

$$DF(x)e_j = \lim_{h\to 0} \frac{1}{h}\bigl[F(x+he_j) - F(x)\bigr]. \tag{4.1.7}$$

This last limit, when it exists, is the *partial derivative* (compare (1.1.45)):

$$\frac{\partial F}{\partial x_j}(x) = \lim_{h\to 0} \frac{1}{h}\bigl[F(x+he_j) - F(x)\bigr]. \tag{4.1.8}$$

Thus, if F is differentiable at x, we have

$$DF(x)e_j = \frac{\partial F}{\partial x_j}(x) = \begin{pmatrix} \partial F_1/\partial x_j \\ \vdots \\ \partial F_m/\partial x_j \end{pmatrix}. \tag{4.1.9}$$

Consequently, with respect to the standard bases of $\mathbb{R}^n$ and $\mathbb{R}^m$, $DF(x)$ is simply the matrix of partial derivatives,

$$DF(x) = \left(\frac{\partial F_j}{\partial x_k}\right) = \begin{pmatrix} \partial F_1/\partial x_1 & \cdots & \partial F_1/\partial x_n \\ \vdots & & \vdots \\ \partial F_m/\partial x_1 & \cdots & \partial F_m/\partial x_n \end{pmatrix}, \tag{4.1.10}$$

so that, if $v = (v_1, \dots, v_n)^t$, (regarded as a column vector) then

$$DF(x)v = \begin{pmatrix} \sum_k (\partial F_1/\partial x_k) v_k \\ \vdots \\ \sum_k (\partial F_m/\partial x_k) v_k \end{pmatrix}. \tag{4.1.11}$$

Another handy notation is

$$\partial_k F_j = \frac{\partial F_j}{\partial x_k}. \tag{4.1.12}$$

In case $n = 1$, so $F : \mathcal{O} \to \mathbb{R}$, the matrix $DF(x)$ has one row,

$$F : \mathcal{O} \to \mathbb{R} \Longrightarrow DF(x) = (\partial_1 F \ \cdots \ \partial_n F). \tag{4.1.13}$$

We typically put in commas and write this as a row vector. It is also common to use the notation ∇F:

$$\nabla F(x) = (\partial_1 F(x), \dots, \partial_n F(x)). \tag{4.1.14}$$

It will be shown below that F is differentiable whenever all the partial derivatives exist and are *continuous* on $\mathcal{O}$. In such a case we say F is a C^1 function on $\mathcal{O}$. More generally, F is said to be C^k if all its partial derivatives of order $\le k$ exist and are continuous. If F is C^k for all k, we say F is C^∞.

An application of the Fundamental Theorem of Calculus, to functions of each variable x_j separately, yields the following. If we assume $F : \mathcal{O} \to \mathbb{R}^m$ is differentiable in each variable separately, and that each $\partial F/\partial x_j$ is continuous on $\mathcal{O}$, then

$$\begin{aligned} F(x+y) &= F(x) + \sum_{j=1}^n \bigl[F(x+z_j) - F(x+z_{j-1})\bigr] \\ &= F(x) + \sum_{j=1}^n A_j(x,y) y_j, \\ A_j(x,y) &= \int_0^1 \frac{\partial F}{\partial x_j}\bigl(x + z_{j-1} + t y_j e_j\bigr)\, dt, \end{aligned} \tag{4.1.15}$$

where

$$z_0 = 0, \quad z_j = (y_1, \dots, y_j, 0, \dots, 0) = z_{j-1} + y_j e_j,$$

and $\{e_j\}$ is the standard basis of $\mathbb{R}^n$. Here we have used

$$F(x+z_j) - F(x+z_{j-1}) = \int_0^1 \frac{d}{dt} F(x + z_{j-1} + t y_j e_j)\, dt,$$

and the one-variable chain rule, (1.1.10). Consequently,

$$\begin{aligned} F(x+y) &= F(x) + \sum_{j=1}^n \frac{\partial F}{\partial x_j}(x)\, y_j + R(x,y), \\ R(x,y) &= \sum_{j=1}^n y_j \int_0^1 \Bigl\{ \frac{\partial F}{\partial x_j}(x + z_{j-1} + t y_j e_j) - \frac{\partial F}{\partial x_j}(x) \Bigr\}\, dt. \end{aligned} \tag{4.1.16}$$

Now (4.1.16) implies F is differentiable on $\mathcal{O}$, as we stated below (4.1.11). Thus we have established the following.

Proposition 4.1.1. *If $\mathcal{O}$ is an open subset of $\mathbb{R}^n$ and $F : \mathcal{O} \to \mathbb{R}^m$ is of class C^1, then F is differentiable at each point $x \in \mathcal{O}$.*

One can use the Mean Value Theorem in place of the fundamental theorem of calculus and obtain a slightly more general result. See Exercise 2 below for prompts on how to accomplish this.

Let us give some examples of derivatives. First, take $n = 2,\ m = 1$, and set

$$F(x) = (\sin x_1)(\sin x_2). \tag{4.1.17}$$

Then

$$DF(x) = ((\cos x_1)(\sin x_2),\ (\sin x_1)(\cos x_2)). \tag{4.1.18}$$

Next, take $n = m = 2$ and

$$F(x) = \begin{pmatrix} x_1 x_2 \\ x_1^2 - x_2^2 \end{pmatrix}. \tag{4.1.19}$$

Then

$$DF(x) = \begin{pmatrix} x_2 & x_1 \\ 2x_1 & -2x_2 \end{pmatrix}. \tag{4.1.20}$$

We can replace $\mathbb{R}^n$ and $\mathbb{R}^m$ by more general finite-dimensional real vector spaces, isomorphic to Euclidean space. For example, the space $M(n, \mathbb{R})$ of real $n \times n$ matrices is isomorphic to $\mathbb{R}^{n^2}$. Consider the function

$$S : M(n, \mathbb{R}) \longrightarrow M(n, \mathbb{R}), \quad S(X) = X^2. \tag{4.1.21}$$

We have

$$\begin{aligned} (X + Y)^2 &= X^2 + XY + YX + Y^2 \\ &= X^2 + DS(X)Y + R(X, Y), \end{aligned} \tag{4.1.22}$$

with $R(X, Y) = Y^2$, and hence

$$DS(X)Y = XY + YX. \tag{4.1.23}$$

For our next example, we take

$$\mathcal{O} = G\ell(n, \mathbb{R}) = \{X \in M(n, \mathbb{R}) : \det X \neq 0\}, \tag{4.1.24}$$

which is open in $M(n, \mathbb{R})$, since $\det : M(n, \mathbb{R}) \to \mathbb{R}$, being a polynomial in the matrix entries of its argument, is continuous. We consider

$$\Phi : G\ell(n, \mathbb{R}) \longrightarrow M(n, \mathbb{R}), \quad \Phi(X) = X^{-1}, \tag{4.1.25}$$

and compute $D\Phi(I)$. We use the following. If, for $A \in M(n, \mathbb{R})$,

$$\|A\| = \sup\{\|Av\| : v \in \mathbb{R}^n,\ \|v\| \le 1\}, \tag{4.1.26}$$

then (cf. §C.3)

$$\begin{aligned} A, B \in M(n, \mathbb{R}) \Rightarrow \|A + B\| &\le \|A\| + \|B\| \\ \text{and}\ \ \|AB\| &\le \|A\| \cdot \|B\|, \\ \text{so } Y \in M(n, \mathbb{R}) \Rightarrow \|Y^k\| &\le \|Y\|^k. \end{aligned} \tag{4.1.27}$$

Also

$$S_k = I - Y + Y^2 - \cdots + (-1)^k Y^k$$
$$(4.1.28) \qquad \Rightarrow Y S_k = S_k Y = Y - Y^2 + Y^3 - \cdots + (-1)^k Y^{k+1}$$
$$\Rightarrow (I+Y)S_k = S_k(I+Y) = I + (-1)^k Y^{k+1},$$

hence

$$(4.1.29) \qquad \|Y\| < 1 \Longrightarrow (I+Y)^{-1} = \sum_{k=0}^{\infty} (-1)^k Y^k = I - Y + Y^2 - \cdots,$$

so

$$(4.1.30) \qquad D\Phi(I)Y = -Y.$$

Going further, we see that, given $X \in G\ell(n, \mathbb{R}),\ Y \in M(n, \mathbb{R})$,

$$(4.1.31) \qquad X + Y = X(I + X^{-1}Y),$$

which by (4.1.29) is invertible as long as

$$(4.1.32) \qquad \|X^{-1}Y\| < 1.$$

One can proceed from here to compute that, for $X \in G\ell(n, \mathbb{R})$,

$$(4.1.33) \qquad D\Phi(X)Y = -X^{-1}YX^{-1}.$$

See Exercise 7 below.

We return to general considerations, and derive the *chain rule* for the derivative. Let $F : \mathcal{O} \to \mathbb{R}^m$ be differentiable at $x \in \mathcal{O}$, as above, let U be a neighborhood of $z = F(x)$ in $\mathbb{R}^m$, and let $G : U \to \mathbb{R}^k$ be differentiable at z. Consider $H = G \circ F$. We have

$$(4.1.34) \qquad \begin{aligned} H(x+y) &= G(F(x+y)) \\ &= G\big(F(x) + DF(x)y + R(x,y)\big) \\ &= G(z) + DG(z)\big(DF(x)y + R(x,y)\big) + R_1(x,y) \\ &= G(z) + DG(z)DF(x)y + R_2(x,y) \end{aligned}$$

with

$$\frac{\|R_2(x,y)\|}{\|y\|} \to 0 \quad \text{as} \quad y \to 0.$$

This establishes the following.

Proposition 4.1.2. *Let $\mathcal{O} \subset \mathbb{R}^n$ and $U \subset \mathbb{R}^m$ be open. Assume $F : \mathcal{O} \to U$ is differentiable at $x \in \mathcal{O}$ and $G : U \to \mathbb{R}^k$ is differentiable at $z = F(x)$. Then $G \circ F$ is differentiable at x, and*

$$(4.1.35) \qquad D(G \circ F)(x) = DG(F(x)) \cdot DF(x).$$

Another useful remark is that, by the Fundamental Theorem of Calculus, applied to $\varphi(t) = F(x + ty)$,

$$(4.1.36) \qquad F(x+y) = F(x) + \int_0^1 DF(x+ty)y\, dt,$$

provided F is C^1. Compare (4.1.15).

Exercises

1. Compute $DF(x)$ in each of the following cases:

$$F(x) = x_1^2 + x_1 x_2,$$

$$F(x) = \begin{pmatrix} x_1^2 \\ x_1 x_2 \end{pmatrix},$$

$$F(x) = \begin{pmatrix} x_2 x_3 \\ x_1 x_3 \\ x_1 x_2 \end{pmatrix},$$

$$F(x) = \begin{pmatrix} x_1 e^{x_2 \cos x_3} \\ \cos(x_1 e^{x_2 x_3}) \end{pmatrix}.$$

2. Here we provide a path to a strengthening of Proposition 4.1.1. Let $\mathcal{O} \subset \mathbb{R}^n$ be open, $f : \mathcal{O} \to \mathbb{R}$. Assume $\partial f/\partial x_j$ exists on $\mathcal{O}$ for each j. Fix $x \in \mathcal{O}$ and assume that

$$\frac{\partial f}{\partial x_j} \text{ is continuous at } x, \text{ for each } j. \tag{4.1.37}$$

Task: prove that f is differentiable at x.
Hint. Start as in (4.1.15), with

$$f(x+y) = f(x) + \sum_{j=1}^{n} \Bigl\{ f(x+z_j) - f(x+z_{j-1}) \Bigr\},$$

where $z_0 = 0,\ z_j = (y_1, \dots, y_j, 0, \dots, 0) = z_{j-1} + y_j e_j,\ z_n = y$. Deduce from the mean value theorem that, for each j,

$$f(x+z_j) - f(x+z_{j-1}) = \frac{\partial f}{\partial x_j}(x + z_{j-1} + \theta_j y_j e_j) y_j,$$

for some $\theta_j \in (0,1)$. Deduce that

$$f(x+y) = f(x) + \sum_{j=1}^{n} \frac{\partial f}{\partial x_j}(x) y_j + R(x,y),$$

where

$$R(x,y) = \sum_{j=1}^{n} \Bigl\{ \frac{\partial f}{\partial x_j}(x + z_{j-1} + \theta_j y_j e_j) - \frac{\partial f}{\partial x_j}(x) \Bigr\} y_j.$$

Show that the hypothesis (4.1.37) implies that $R(x,y) = o(\|y\|)$.

3. Consider

$$f(x,y,z,w) = \det \begin{pmatrix} x & y \\ z & w \end{pmatrix}.$$

Compute $\nabla f(x,y,z,w)$.

4. Let $P_k : M(n,\mathbb{R}) \to M(n,\mathbb{R})$ be given by $P_k(X) = X^k$. Show that

$$DP_3(X)Y = YX^2 + XYX + X^2Y.$$

Hint. Expand $(X+Y)^3$ and isolate terms that vanish faster than $\|Y\|$ as $Y \to 0$.

5. In the setting of Exercise 4, show that, for $k \geq 2$,

$$DP_k(X)Y = \sum_{j=0}^{k-1} X^j Y X^{k-1-j}.$$

6. Let $M(n,\mathbb{R})$ denote the space of real $n \times n$ matrices, and let $\Omega \subset M(n,\mathbb{R})$ be open. Assume $F, G : \Omega \to M(n,\mathbb{R})$ are of class C^1. Show that $H(X) = F(X)G(X)$ defines a C^1 map $H : \Omega \to M(n,\mathbb{R})$, and

$$DH(X)Y = \Big(DF(X)Y\Big)G(X) + F(X)\Big(DG(X)Y\Big).$$

Use this to produce an inductive approach to Exercise 5.

6A. More generally, if $\Omega \subset \mathbb{R}^k$ is open $F, G : \Omega \to M(n,\mathbb{R})$, $H(x) = F(x)G(x)$, show that

$$DH(x)y = \Big(DF(x)y\Big)G(x) + F(x)\Big(DG(x)y\Big).$$

7. Let $Gl(n,\mathbb{R}) \subset M(n,\mathbb{R})$ denote the set of invertible matrices. Show that

$$\Phi : Gl(n,\mathbb{R}) \longrightarrow M(n,\mathbb{R}), \quad \Phi(X) = X^{-1}$$

is of class C^1 and that

$$D\Phi(X)Y = -X^{-1}YX^{-1}.$$

Hint. Start with (4.1.31), yielding

$$(X+Y)^{-1} = (I + X^{-1}Y)^{-1}X^{-1},$$

and apply the series expansion (4.1.29), with Y replaced by $X^{-1}Y$.

8. Define $S, \Phi, F : G\ell(n,\mathbb{R}) \to M(n,\mathbb{R})$ by

$$S(X) = X^2, \quad \Phi(X) = X^{-1}, \quad F(X) = X^{-2}.$$

Compute $DF(X)Y$ using each of the following approaches:
(a) Take $F(X) = \Phi(X)\Phi(X)$ and use the product rule (Exercise 6).
(b) Take $F(X) = \Phi(S(X))$ and use the chain rule.
(c) Take $F(X) = S(\Phi(X))$ and use the chain rule.

9. Identify $\mathbb{R}^2$ and $\mathbb{C}$ via $z = x + iy$. Then multiplication by i on $\mathbb{C}$ corresponds to applying

$$J = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

Let $\mathcal{O} \subset \mathbb{R}^2$ be open, $f : \mathcal{O} \to \mathbb{R}^2$ be C^1. Say $f = (u, v)$. Regard $Df(x, y)$ as a 2×2 real matrix. One says f is *holomorphic*, or complex-analytic, provided the Cauchy-Riemann equations hold:

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}, \quad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}.$$

Show that this is equivalent to the condition

$$Df(x, y)\, J = J\, Df(x, y).$$

10. Proceeding from the identity $e^{x+iy} = e^x \cos y + ie^x \sin y$, derived in §3.2, define $E : \mathbb{R}^2 \to \mathbb{R}^2$ by

$$E(x, y) = \begin{pmatrix} e^x \cos y \\ e^x \sin y \end{pmatrix}.$$

Compute $DE(x, y)$, as a 2×2 matrix, and verify that $DE(x, y)\, J = J\, DE(x, y)$.

11. Let f be C^1 on a region in $\mathbb{R}^2$ containing $[a, b] \times \{y\}$. Show that, as $h \to 0$,

$$\frac{1}{h}\bigl[f(x, y + h) - f(x, y)\bigr] \longrightarrow \frac{\partial f}{\partial y}(x, y), \quad \text{uniformly on} \quad [a, b] \times \{y\}.$$

Hint. Show that the left side is equal to

$$\frac{1}{h} \int_0^h \frac{\partial f}{\partial y}(x, y + s)\, ds,$$

and use the uniform continuity of $\partial f / \partial y$ on $[a, b] \times [y - \delta, y + \delta]$.

12. In the setting of Exercise 11, show that

$$\frac{d}{dy} \int_a^b f(x, y)\, dx = \int_a^b \frac{\partial f}{\partial y}(x, y)\, dx.$$

Exercises 13–15 deal with properties of the determinant, as a differentiable function on spaces of matrices.

13. Let $M(n, \mathbb{R})$ be the space of $n \times n$ matrices with real coefficients, $\det : M(n, \mathbb{R}) \to \mathbb{R}$ the determinant. Show that, if I is the identity matrix,

$$D \det(I) B = \operatorname{Tr} B,$$

i.e.,

$$\frac{d}{dt} \det(I + tB)|_{t=0} = \operatorname{Tr} B.$$

Hint. Brush up on the exercises in §2.3.

14. If $A(t) = (a_{jk}(t))$ is a smooth curve in $M(n, \mathbb{R})$, use the expansion of $(d/dt) \det A(t)$ as a sum of n determinants, in which the rows of $A(t)$ are successively differentiated, to show that

$$\frac{d}{dt} \det A(t) = \operatorname{Tr}\Bigl(\operatorname{Cof}(A(t))^t \cdot A'(t)\Bigr),$$

and deduce that, for $A, B \in M(n, \mathbb{R})$,

$$D \det(A)B = \mathrm{Tr}\big(\mathrm{Cof}(A)^t \cdot B\big).$$

Here $\mathrm{Cof}(A)$, the cofactor matrix, is defined in Exercise 4 of §2.3.

15. Suppose $A \in M(n, \mathbb{R})$ is invertible. Using

$$\det(A + tB) = (\det A) \det(I + tA^{-1}B),$$

show that

$$D \det(A)B = (\det A)\,\mathrm{Tr}(A^{-1}B).$$

Comparing this result with that of Exercise 14, establish Cramer's formula:

$$(\det A)A^{-1} = \mathrm{Cof}(A)^t.$$

Compare the derivation in Exercise 4 of §2.3.

16. Define $f(x, y)$ on $\mathbb{R}^2$ by

$$\begin{aligned} f(x, y) = \frac{xy}{\sqrt{x^2 + y^2}}, &\quad (x, y) \neq (0, 0), \\ 0, &\quad (x, y) = (0, 0). \end{aligned}$$

Show that f is continuous on $\mathbb{R}^2$ and smooth on $\mathbb{R}^2 \setminus (0, 0)$. Show that $\partial f/\partial x$ and $\partial f/\partial y$ exist at each point of $\mathbb{R}^2$, and are continuous on $\mathbb{R}^2 \setminus (0, 0)$, but not on $\mathbb{R}^2$. Show that

$$\frac{\partial f}{\partial x}(0, 0) = \frac{\partial f}{\partial y}(0, 0) = 0.$$

Show that f is not differentiable at $(0, 0)$.
Hint. Show that $f(x, y)$ is not $o(\|(x, y)\|)$ as $(x, y) \to (0, 0)$, by considering $f(x, x)$.

17. Let $f, g, h : \mathbb{R}^2 \to \mathbb{R}$ be of class C^1, and define $F : \mathbb{R}^2 \to \mathbb{R}$ by

$$F(x) = h(f(x_1, x_2), g(x_1, x_2)).$$

Show that the chain rule implies

$$\frac{\partial F}{\partial x_j} = \partial_1 h(f(x), g(x)) \frac{\partial f}{\partial x_j} + \partial_2 h(f(x), g(x)) \frac{\partial g}{\partial x_j}.$$

4.2. Higher derivatives and power series

For the study of higher order derivatives of a function, the following result is fundamental.

Proposition 4.2.1. *Assume $F : \mathcal{O} \to \mathbb{R}^m$ is of class C^2, with $\mathcal{O}$ open in $\mathbb{R}^n$. Then, for each $x \in \mathcal{O}$, $1 \le j, k \le n$,*

$$\frac{\partial}{\partial x_j}\frac{\partial F}{\partial x_k}(x) = \frac{\partial}{\partial x_k}\frac{\partial F}{\partial x_j}(x). \tag{4.2.1}$$

Proof. It suffices to take $m = 1$. We label our function $f : \mathcal{O} \to \mathbb{R}$. For $1 \le j \le n$, we set

$$\Delta_{j,h} f(x) = \frac{1}{h}\big(f(x + he_j) - f(x)\big), \tag{4.2.2}$$

where $\{e_1, \dots, e_n\}$ is the standard basis of $\mathbb{R}^n$. The mean value theorem (for functions of x_j alone) implies that if $\partial_j f = \partial f/\partial x_j$ exists on $\mathcal{O}$, then, for $x \in \mathcal{O}$, $h > 0$ sufficiently small,

$$\Delta_{j,h} f(x) = \partial_j f(x + \alpha_j h e_j), \tag{4.2.3}$$

for some $\alpha_j \in (0, 1)$, depending on x and h. Iterating this, if $\partial_j(\partial_k f)$ exists on $\mathcal{O}$, then, for $x \in \mathcal{O}$, $h > 0$ sufficiently small,

$$\begin{aligned} \Delta_{k,h}\Delta_{j,h} f(x) &= \partial_k(\Delta_{j,h} f)(x + \alpha_k h e_k) \\ &= \Delta_{j,h}(\partial_k f)(x + \alpha_k h e_k) \\ &= \partial_j \partial_k f(x + \alpha_k h e_k + \alpha_j h e_j), \end{aligned} \tag{4.2.4}$$

with $\alpha_j, \alpha_k \in (0, 1)$. See Figure 4.2.1 for an illustration, with $n = 2$, $j = 1, k = 2$. For the second identity in (4.2.4), we have used the elementary result

$$\partial_k \Delta_{j,h} f = \Delta_{j,h}(\partial_k f). \tag{4.2.5}$$

We deduce the following.

Proposition 4.2.2. *If $\partial_k f$ and $\partial_j \partial_k f$ exist on $\mathcal{O}$ and $\partial_j \partial_k f$ is continuous at $x_0 \in \mathcal{O}$, then*

$$\partial_j \partial_k f(x_0) = \lim_{h \to 0} \Delta_{k,h}\Delta_{j,h} f(x_0). \tag{4.2.6}$$

The following identity is also elementary (see Exercise 8):

$$\Delta_{k,h}\Delta_{j,h} f = \Delta_{j,h}\Delta_{k,h} f. \tag{4.2.7}$$

Hence we have the following, which readily implies Proposition 4.2.1. □

Corollary 4.2.3. *In the setting of Proposition 4.2.2, if also $\partial_j f$ and $\partial_k \partial_j f$ exist on $\mathcal{O}$ and $\partial_k \partial_j f$ is continuous at x_0, then*

$$\partial_j \partial_k f(x_0) = \partial_k \partial_j f(x_0). \tag{4.2.8}$$

We now describe two convenient notations to express higher order derivatives of a C^k function $f : \Omega \to \mathbb{R}$, where $\Omega \subset \mathbb{R}^n$ is open. In one, let J be a k-tuple of integers between 1 and n; $J = (j_1, \dots, j_k)$. We set

$$f^{(J)}(x) = \partial_{j_k} \cdots \partial_{j_1} f(x), \quad \partial_j = \frac{\partial}{\partial x_j}. \tag{4.2.9}$$

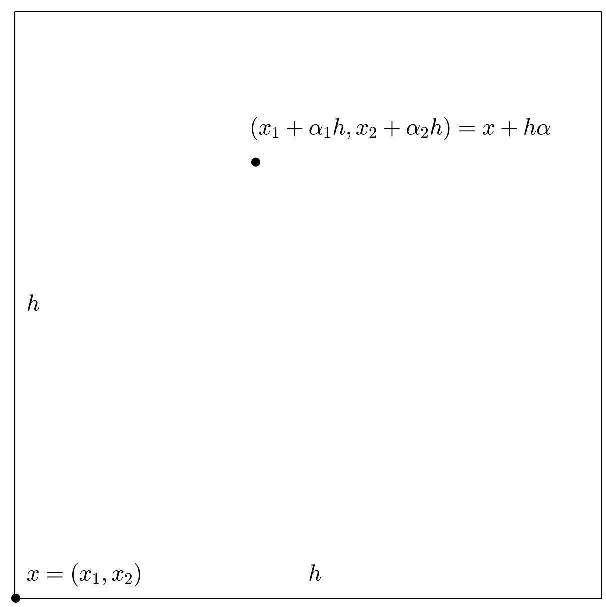

Figure 4.2.1. $\Delta_{2,h}\Delta_{1,h}f(x) = \partial_1\partial_2 f(x + h\alpha)$

We set $|J| = k$, the total order of differentiation. As we have seen in Proposition 4.2.1, $\partial_i\partial_j f = \partial_j\partial_i f$ provided $f \in C^2(\Omega)$. It follows that, if $f \in C^k(\Omega)$, then $\partial_{j_k} \cdots \partial_{j_1} f = \partial_{\ell_k} \cdots \partial_{\ell_1} f$ whenever $\{\ell_1, \dots, \ell_k\}$ is a permutation of $\{j_1, \dots, j_k\}$. Thus, another convenient notation to use is the following. Let α be an n-tuple of non-negative integers, $\alpha = (\alpha_1, \dots, \alpha_n)$. Then we set

$$f^{(\alpha)}(x) = \partial_1^{\alpha_1} \cdots \partial_n^{\alpha_n} f(x), \quad |\alpha| = \alpha_1 + \cdots + \alpha_n. \tag{4.2.10}$$

Note that, if $|J| = |\alpha| = k$ and $f \in C^k(\Omega)$,

$$f^{(J)}(x) = f^{(\alpha)}(x), \text{ with } \alpha_i = \#\{\ell : j_\ell = i\}. \tag{4.2.11}$$

Correspondingly, there are two expressions for monomials in $x = (x_1, \dots, x_n)$:

$$x^J = x_{j_1} \cdots x_{j_k}, \quad x^\alpha = x_1^{\alpha_1} \cdots x_n^{\alpha_n}, \tag{4.2.12}$$

and $x^J = x^\alpha$ provided J and α are related as in (4.2.11). Both these notations are called "multi-index" notations.

Multivariable power series

We now consider multivariable power series, and derive Taylor's formula with remainder for a smooth function $F : \Omega \to \mathbb{R}$, making use of the multi-index notations introduced above. We will apply the one variable formula derived in §1.3

(cf. Proposition 1.3.4),

$$\varphi(t) = \varphi(0) + \varphi'(0)t + \frac{1}{2}\varphi''(0)t^2 + \cdots + \frac{1}{k!}\varphi^{(k)}(0)t^k + r_k(t), \tag{4.2.13}$$

with

$$r_k(t) = \frac{1}{k!}\int_0^t (t-s)^k \varphi^{(k+1)}(s)\, ds, \tag{4.2.14}$$

given $\varphi \in C^{k+1}(I)$, $I = (-a, a)$. (See Exercise 1 below for a reminder.) Let us assume $0 \in \Omega$, and that the line segment from 0 to x is contained in Ω. We set $\varphi(t) = F(tx)$, and apply (4.2.13)–(4.2.14) with $t = 1$. Applying the chain rule, we have

$$\varphi'(t) = \sum_{j=1}^n \partial_j F(tx) x_j. \tag{4.2.15}$$

Differentiating again, we have

$$\varphi''(t) = \sum_{j_1, j_2} \partial_{j_2}\partial_{j_1} F(tx) x_{j_1} x_{j_2}. \tag{4.2.16}$$

Inductively, we have

$$\varphi^{(k)}(t) = \sum_{j_1,\dots,j_k} \partial_{j_k}\cdots\partial_{j_1} F(tx) x_{j_1}\cdots x_{j_k} = \sum_{|J|=k} F^{(J)}(tx) x^J. \tag{4.2.17}$$

Hence, from (4.2.13) with $t = 1$,

$$F(x) = F(0) + \sum_{|J|=1} F^{(J)}(0)x^J + \cdots + \frac{1}{k!}\sum_{|J|=k} F^{(J)}(0)x^J + R_k(x), \tag{4.2.18}$$

or, more briefly,

$$F(x) = \sum_{|J|\le k} \frac{1}{|J|!} F^{(J)}(0)x^J + R_k(x), \tag{4.2.19}$$

where

$$R_k(x) = \frac{1}{k!}\sum_{|J|=k+1}\Bigl(\int_0^1 (1-s)^k F^{(J)}(sx)\, ds\Bigr) x^J. \tag{4.2.20}$$

This gives Taylor's formula with remainder for $F \in C^{k+1}(\Omega)$, in the J-multi-index notation.

We also want to write the formula in the α-multi-index notation. We have

$$\sum_{|J|=k} F^{(J)}(tx)x^J = \sum_{|\alpha|=k} \nu(\alpha) F^{(\alpha)}(tx) x^\alpha, \tag{4.2.21}$$

where

$$\nu(\alpha) = \#\{J : \alpha = \alpha(J)\}, \tag{4.2.22}$$

and we define the relation $\alpha = \alpha(J)$ to hold provided the condition (4.2.11) holds, or equivalently provided $x^J = x^\alpha$. Thus $\nu(\alpha)$ is uniquely defined by

$$\sum_{|\alpha|=k} \nu(\alpha) x^\alpha = \sum_{|J|=k} x^J = (x_1 + \cdots + x_n)^k. \tag{4.2.23}$$

To evaluate $\nu(\alpha)$, we can expand $(x_1 + \cdots + x_n)^k$ in terms of x^α by a repeated application of the binomial formula (cf. §1.3, Exercise 2):

$$
\begin{aligned}
(x_1 + \cdots + x_n)^k &= \big(x_1 + (x_2 + \cdots + x_n)\big)^k \\
&= \sum_{\alpha_1 \le k} \binom{k}{\alpha_1} x_1^{\alpha_1} (x_2 + \cdots + x_n)^{k-\alpha_1} \\
&= \sum_{\alpha_1+\alpha_2 \le k} \binom{k}{\alpha_1}\binom{k-\alpha_1}{\alpha_2} x_1^{\alpha_1} x_2^{\alpha_2} (x_3 + \cdots + x_n)^{k-\alpha_1-\alpha_2} \\
&= \cdots \\
&= \sum_{|\alpha|=k} \binom{k}{\alpha_1}\binom{k-\alpha_1}{\alpha_2} \cdots \binom{k-\alpha_1-\cdots-\alpha_{n-1}}{\alpha_n} x_1^{\alpha_1} \cdots x_n^{\alpha_n} \\
&= \sum_{|\alpha|=k} \nu(\alpha) x^\alpha.
\end{aligned} \tag{4.2.24}
$$

We have $\nu(\alpha)$ equal to the product of binomial coefficients given above, i.e., to

$$
\begin{aligned}
&\frac{k!}{\alpha_1!(k-\alpha_1)!} \cdot \frac{(k-\alpha_1)!}{\alpha_2!(k-\alpha_1-\alpha_2)!} \cdots \frac{(k-\alpha_1-\cdots-\alpha_{n-1})!}{\alpha_n!(k-\alpha_1-\cdots-\alpha_n)!} \\
&= \frac{k!}{\alpha_1!\cdots\alpha_n!}.
\end{aligned}
$$

In other words, for $|\alpha| = k$,

$$\nu(\alpha) = \frac{k!}{\alpha!}, \text{ where } \alpha! = \alpha_1! \cdots \alpha_n! \tag{4.2.25}$$

Thus the Taylor formula (4.2.19) can be rewritten

$$F(x) = \sum_{|\alpha| \le k} \frac{1}{\alpha!} F^{(\alpha)}(0) x^\alpha + R_k(x), \tag{4.2.26}$$

where

$$R_k(x) = \sum_{|\alpha|=k+1} \frac{k+1}{\alpha!} \Big(\int_0^1 (1-s)^k F^{(\alpha)}(sx)\, ds \Big) x^\alpha. \tag{4.2.27}$$

The formula (4.2.26)–(4.2.27) holds for $F \in C^{k+1}$. It is significant that (4.2.26), with a variant of (4.2.27), holds for $F \in C^k$. In fact, for such F, we can apply (4.2.27) with k replaced by $k-1$, to get

$$F(x) = \sum_{|\alpha| \le k-1} \frac{1}{\alpha!} F^{(\alpha)}(0) x^\alpha + R_{k-1}(x), \tag{4.2.28}$$

with

$$R_{k-1}(x) = \sum_{|\alpha|=k} \frac{k}{\alpha!} \Big(\int_0^1 (1-s)^{k-1} F^{(\alpha)}(sx)\, ds \Big) x^\alpha. \tag{4.2.29}$$

We can add and subtract $F^{(\alpha)}(0)$ to $F^{(\alpha)}(sx)$ in the integrand above, and obtain the following.

Proposition 4.2.4. *If $F \in C^k$ on a ball $B_r(0)$, the formula (4.2.26) holds for $x \in B_r(0)$, with*

$$R_k(x) = \sum_{|\alpha|=k} \frac{k}{\alpha!} \Big(\int_0^1 (1-s)^{k-1} \big[F^{(\alpha)}(sx) - F^{(\alpha)}(0) \big] \, ds \Big) x^\alpha. \tag{4.2.30}$$

Remark. Note that (4.2.30) yields the estimate

$$|R_k(x)| \le \sum_{|\alpha|=k} \frac{|x^\alpha|}{\alpha!} \sup_{0\le s\le 1} |F^{(\alpha)}(sx) - F^{(\alpha)}(0)|. \tag{4.2.31}$$

The term corresponding to $|J| = 2$ in (4.2.19), or $|\alpha| = 2$ in (4.2.26), is of particular interest. It is

$$\frac{1}{2} \sum_{|J|=2} F^{(J)}(0)\, x^J = \frac{1}{2} \sum_{j,k=1}^{n} \frac{\partial^2 F}{\partial x_k \partial x_j}(0)\, x_j x_k. \tag{4.2.32}$$

We define the *Hessian* of a C^2 function $F : \mathcal{O} \to \mathbb{R}$ as an $n \times n$ matrix:

$$D^2 F(y) = \left(\frac{\partial^2 F}{\partial x_k \partial x_j}(y) \right). \tag{4.2.33}$$

Then the power series expansion of second order about 0 for F takes the form

$$F(x) = F(0) + DF(0)x + \frac{1}{2} x \cdot D^2 F(0) x + R_2(x), \tag{4.2.34}$$

where, by (4.2.31),

$$|R_2(x)| \le C_n |x|^2 \sup_{0\le s\le 1, |\alpha|=2} |F^{(\alpha)}(sx) - F^{(\alpha)}(0)|. \tag{4.2.35}$$

In all these formulas we can translate coordinates and expand about $y \in \mathcal{O}$. For example, (4.2.34) extends to

$$F(x) = F(y) + DF(y)(x-y) + \frac{1}{2}(x-y) \cdot D^2 F(y)(x-y) + R_2(x,y), \tag{4.2.36}$$

with

$$|R_2(x,y)| \le C_n |x-y|^2 \sup_{0\le s\le 1, |\alpha|=2} |F^{(\alpha)}(y + s(x-y)) - F^{(\alpha)}(y)|. \tag{4.2.37}$$

Example. If we take $F(x)$ as in (4.1.17), so $DF(x)$ is as in (4.1.18), then

$$D^2 F(x) = \begin{pmatrix} -\sin x_1 \ \sin x_2 & \cos x_1 \ \cos x_2 \\ \cos x_1 \ \cos x_2 & -\sin x_1 \ \sin x_2 \end{pmatrix}.$$

Extremal problems and critical points

The results (4.2.36)–(4.2.37) are useful for extremal problems, i.e., determining where a sufficiently smooth function $F : \mathcal{O} \to \mathbb{R}$ has local maxima and local minima. Clearly if $F \in C^1(\mathcal{O})$ and F has a local maximum or minimum at $x_0 \in \mathcal{O}$, then $DF(x_0) = 0$. (Compare Proposition 1.1.1.) In such a case, we say x_0 is a

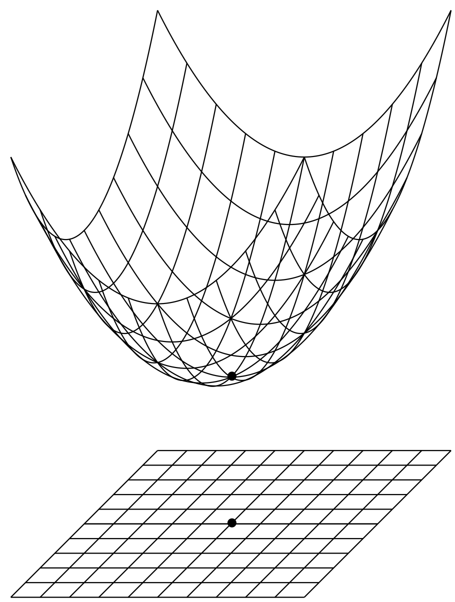

Figure 4.2.2. Critical point that is a minimum

critical point of F. To check what kind of critical point x_0 is, we look at the $n \times n$ matrix $A = D^2F(x_0)$, assuming $F \in C^2(\mathcal{O})$. By Proposition 4.2.1, A is a symmetric matrix. A basic result in linear algebra, treated in Appendix C.2, is that if A is a real, symmetric $n \times n$ matrix, then $\mathbb{R}^n$ has an orthonormal basis of eigenvectors, $\{v_1, \dots, v_n\}$, satisfying $Av_j = \lambda_j v_j$; the real numbers λ_j are the eigenvalues of A. We say A is positive definite if all $\lambda_j > 0$, and we say A is negative definite if all $\lambda_j < 0$. We say A is strongly indefinite if some $\lambda_j > 0$ and another $\lambda_k < 0$. Equivalently, given a real, symmetric matrix A,

$$\begin{aligned} A \ \text{ positive definite } &\Longleftrightarrow v \cdot Av \geq C|v|^2, \\ A \ \text{ negative definite } &\Longleftrightarrow v \cdot Av \leq -C|v|^2, \end{aligned} \tag{4.2.38}$$

for some $C > 0$, all $v \in \mathbb{R}^n$, and

$$\begin{aligned} A \text{ strongly indefinite } \Longleftrightarrow \exists\, v, w \in \mathbb{R}^n, &\text{ nonzero, such that} \\ v \cdot Av \geq C|v|^2,\ & w \cdot Aw \leq -C|w|^2, \end{aligned} \tag{4.2.39}$$

for some $C > 0$. In light of (4.2.19)–(4.2.20), we have the following result.

Proposition 4.2.5. *Assume $F \in C^2(\mathcal{O})$ is real valued, $\mathcal{O}$ open in $\mathbb{R}^n$. Let $x_0 \in \mathcal{O}$ be a critical point for F. Then*
(i) $D^2F(x_0)$ positive definite $\Rightarrow F$ has a local minimum at x_0,
(ii) $D^2F(x_0)$ negative definite $\Rightarrow F$ has a local maximum at x_0,

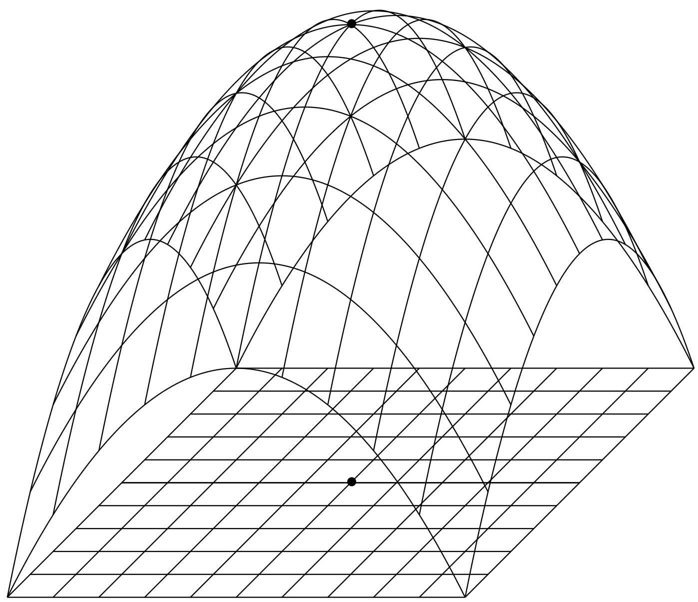

Figure 4.2.3. Critical point that is a maximum

(iii) $D^2F(x_0)$ *strongly indefinite* $\Rightarrow$ F *has neither a local maximum nor a local minimum at* x_0.

In case (iii), we say x_0 is a *saddle point* for F. See Figures 4.2.2–4.2.4 for illustrations.

The following is a test for positive definiteness.

Proposition 4.2.6. *Let* $A = (a_{ij})$ *be a real, symmetric,* $n \times n$ *matrix. For* $1 \le \ell \le n$, *form the* $\ell \times \ell$ *matrix* $A_\ell = (a_{ij})_{1 \le i,j \le \ell}$. *Then*

$$A \ \textit{positive definite} \iff \det A_\ell > 0, \quad \forall\, \ell \in \{1, \dots, n\}. \tag{4.2.40}$$

Regarding the implication $\Rightarrow$, note that if A is positive definite, then $\det A = \det A_n$ is the product of its eigenvalues, all > 0, hence is > 0. Also in this case, it follows from the hypothesis on the left side of (4.2.40) that each A_ℓ must be positive definite, hence have positive determinant, so we have $\Rightarrow$.

The implication $\Leftarrow$ is easy enough for 2×2 matrices. If A is symmetric and $\det A > 0$, then either both its eigenvalues are positive (so A is positive definite) or both are negative (so A is negative definite). In the latter case, $A_1 = (a_{11})$ must be negative, so we have $\Leftarrow$ in this case.

We prove $\Leftarrow$ for $n \ge 3$, using induction. The inductive hypothesis implies that if $\det A_\ell > 0$ for each $\ell \le n$, then A_{n-1} is positive definite. The next lemma then

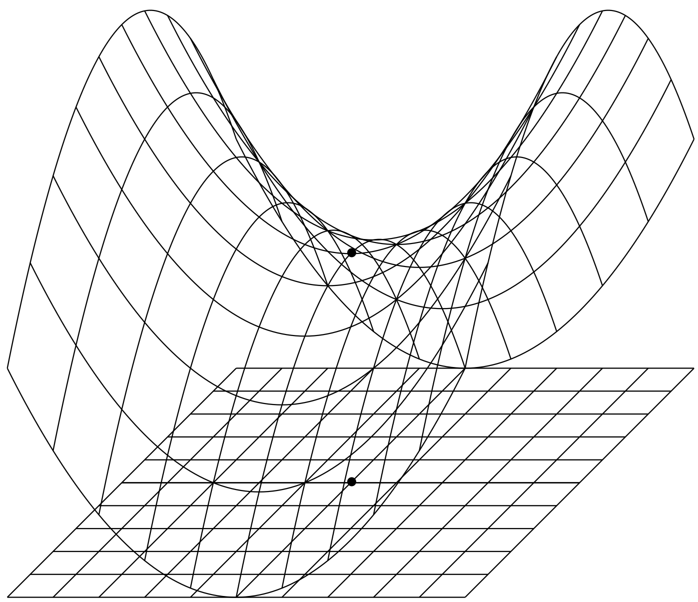

Figure 4.2.4. Critical point that is a saddle

guarantees that $A = A_n$ has at least $n-1$ positive eigenvalues. The hypothesis that $\det A > 0$ does not allow that the remaining eigenvalue be ≤ 0, so all the eigenvalues of A must be positive. Thus Proposition 4.2.6 is proven, once we have the following.

Lemma 4.2.7. *In the setting of Proposition 4.2.6, if A_{n-1} is positive definite, then $A = A_n$ has at least $n-1$ positive eigenvalues.*

Proof. Since A is symmetric, $\mathbb{R}^n$ has an orthonormal basis $v_1, \dots, v_n$ of eigenvectors of A; $Av_j = \lambda_j v_j$. See Appendix C.2. If the conclusion of the lemma is false, at least two of the eigenvalues, say λ_1, λ_2, are ≤ 0. Let $W = \operatorname{Span}(v_1, v_2)$, so

$$w \in W \Longrightarrow w \cdot Aw \le 0.$$

Since W has dimension 2, $\mathbb{R}^{n-1} \subset \mathbb{R}^n$ satisfies $\mathbb{R}^{n-1} \cap W \neq 0$, so there exists a nonzero $w \in \mathbb{R}^{n-1} \cap W$, and then

$$w \cdot A_{n-1} w = w \cdot Aw \le 0,$$

contradicting the hypothesis that A_{n-1} is positive definite. □

Remark. Given (4.2.40), we see by taking $A \mapsto -A$ that if A is a real, symmetric $n \times n$ matrix,

$$A \ \text{ negative definite } \Longleftrightarrow (-1)^\ell \det A_\ell > 0, \quad \forall\, \ell \in \{1, \dots, n\}. \tag{4.2.41}$$

Example. Consider the function $f : \mathbb{R}^2 \to \mathbb{R}$ defined by

$$f(x, y) = (\cos x)(\cos y). \tag{4.2.42}$$

We have

$$\nabla f(x, y) = -(\sin x \, \cos y, \cos x \, \sin y), \tag{4.2.43}$$

which vanishes at the following points:

$$(x, y) = (j\pi, k\pi), \quad (x, y) = ((j + \tfrac{1}{2})\pi, (k + \tfrac{1}{2})\pi), \quad j, k \in \mathbb{Z}. \tag{4.2.44}$$

We have

$$D^2 f(x, y) = \begin{pmatrix} -\cos x \, \cos y & \sin x \, \sin y \\ \sin x \, \sin y & -\cos x \, \cos y \end{pmatrix}. \tag{4.2.45}$$

Hence

$$\begin{aligned} D^2 f(j\pi, k\pi) &= -(\cos j\pi)(\cos k\pi) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \\ &= (-1)^{j+k+1} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \end{aligned} \tag{4.2.46}$$

and

$$\begin{aligned} D^2 f((j + \tfrac{1}{2})\pi, (k + \tfrac{1}{2})\pi) &= \sin(j + \tfrac{1}{2})\pi \, \sin(k + \tfrac{1}{2})\pi \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \\ &= (-1)^{j+k} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}. \end{aligned} \tag{4.2.47}$$

The matrix $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ has 1 as a double eigenvalue, and the eigenvalues of $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ are ± 1. Hence, for $j, k \in \mathbb{Z}$,

$$\begin{aligned} (j\pi, k\pi) \text{ is a local maximum if } j + k \text{ is even,} \\ \text{local minimum if } j + k \text{ is odd,} \end{aligned} \tag{4.2.48}$$

and

$$((j + \tfrac{1}{2})\pi, (k + \tfrac{1}{2})\pi) \ \text{ are all saddles.} \tag{4.2.49}$$

Further remainder formulas

We return to higher order power series formulas with remainder and complement Proposition 4.2.4. Let us go back to (4.2.13)–(4.2.14) and note that the integral in (4.2.14) is $1/(k+1)$ times a weighted average of $\varphi^{(k+1)}(s)$ over $s \in [0, t]$. Hence we can write

$$r_k(t) = \frac{1}{(k + 1)!} \varphi^{(k+1)}(\theta t), \quad \text{for some} \ \ \theta \in [0, 1],$$

if φ is of class C^{k+1}. This is the Lagrange form of the remainder. If φ is of class C^k, we can replace $k + 1$ by k in (4.2.13) and write

$$\varphi(t) = \varphi(0) + \varphi'(0)t + \cdots + \frac{1}{(k - 1)!} \varphi^{(k-1)}(0) t^{k-1} + \frac{1}{k!} \varphi^{(k)}(\theta t) t^k, \tag{4.2.50}$$

for some $\theta \in [0,1]$. Pluging (4.2.50) into (4.2.17) for $\varphi(t) = F(tx)$ gives

$$F(x) = \sum_{|J| \le k-1} \frac{1}{|J|!} F^{(J)}(0) x^J + \frac{1}{k!} \sum_{|J|=k} F^{(J)}(\theta x) x^J, \tag{4.2.51}$$

for some $\theta \in [0,1]$ (depending on x and on k, but not on J), when F is of class C^k on a neighborhood $B_r(0)$ of $0 \in \mathbb{R}^n$. Similarly, using the α-multi-index notation, we have, as an alternative to (4.2.28)–(4.2.29),

$$F(x) = \sum_{|\alpha| \le k-1} \frac{1}{\alpha!} F^{(\alpha)}(0) x^\alpha + \sum_{|\alpha|=k} \frac{1}{\alpha!} F^{(\alpha)}(\theta x) x^\alpha, \tag{4.2.52}$$

for some $\theta \in [0,1]$ (depending on x and on $|\alpha|$, but not on α), if $F \in C^k(B_r(0))$. Note also that

$$\begin{aligned} \frac{1}{2} \sum_{|J|=2} F^{(J)}(\theta x) x^J &= \frac{1}{2} \sum_{j,k=1}^{n} \frac{\partial^2 F}{\partial x_k \partial x_j}(\theta x) x_j x_k \\ &= \frac{1}{2}\, x \cdot D^2 F(\theta x) x, \end{aligned} \tag{4.2.53}$$

with $D^2F(y)$ as in (4.2.33), so if $F \in C^2(B_r(0))$, we have, as an alternative to (4.2.34),

$$F(x) = F(0) + DF(0)x + \frac{1}{2} x \cdot D^2 F(\theta x) x, \tag{4.2.54}$$

for some $\theta \in [0,1]$.

Multi-linear alternative to multi-index notation

We next complement the multi-index notations for higher derivatives of a function F by a multi-linear notation, defined as follows. If $k \in \mathbb{N}$, $F \in C^k(U)$, and $y \in U \subset \mathbb{R}^n$, set

$$D^k F(y)(u_1, \dots, u_k) = \partial_{t_1} \cdots \partial_{t_k} F(y + t_1 u_1 + \cdots + t_k u_k)\Big|_{t_1 = \cdots = t_k = 0}, \tag{4.2.55}$$

for $u_1, \dots, u_k \in \mathbb{R}^n$. For $k = 1$, this formula is equivalent to the definition of DF given at the beginning of this section. For $k = 2$, we have

$$D^2 F(y)(u, v) = u \cdot D^2 F(y) v, \tag{4.2.56}$$

with $D^2F(y)$ on the right as in (4.2.33). Generally, (4.2.55) defines $D^k F(y)$ as a symmetric, k-linear form in $u_1, \dots, u_k \in \mathbb{R}^n$.

We can relate (4.2.55) to the J-multi-index notation as follows. We start with

$$\partial_{t_1} F(y + t_1 u_1 + \cdots + t_k u_k) = \sum_{|J|=1} F^{(J)}(y + \Sigma t_j u_j) u_1^J, \tag{4.2.57}$$

and inductively obtain

$$\partial_{t_1} \cdots \partial_{t_k} F(y + \Sigma t_j u_j) = \sum_{|J_1| = \cdots = |J_k| = 1} F^{(J_1 + \cdots + J_k)}(y + \Sigma t_j u_j) u_1^{J_1} \cdots u_k^{J_k}, \tag{4.2.58}$$

hence

$$D^k F(y)(u_1, \dots, u_k) = \sum_{|J_1|=\cdots=|J_k|=1} F^{(J_1+\cdots+J_k)}(y) u_1^{J_1} \cdots u_k^{J_k}. \tag{4.2.59}$$

In particular, if $u_1 = \cdots = u_k = u$,

$$D^k F(y)(u, \dots, u) = \sum_{|J|=k} F^{(J)}(y) u^J. \tag{4.2.60}$$

Hence (4.2.51) yields the multi-linear Taylor formula with remainder

$$\begin{aligned} F(x) = F(0) + DF(0)x + \cdots + \frac{1}{(k-1)!} D^{k-1}F(0)(x, \dots, x) \\ + \frac{1}{k!} D^k F(\theta x)(x, \dots, x), \end{aligned} \tag{4.2.61}$$

for some $\theta \in [0,1]$, if $F \in C^k(B_r(0))$. In fact, rather than appealing to (4.2.51), we can note that

$$\begin{aligned} \varphi(t) = F(tx) \Longrightarrow \varphi^{(k)}(t) &= \partial_{t_1} \cdots \partial_{t_k} \varphi(t + t_1 + \cdots + t_k)\Big|_{t_1=\cdots=t_k=0} \\ &= D^k F(tx)(x, \dots, x), \end{aligned}$$

and obtain (4.2.61) directly from (4.2.50). We can also use the notation

$$D^j F(y) x^{\otimes j} = D^j F(y)(x, \dots, x), \tag{4.2.62}$$

with j copies of x within the last set of parentheses, and rewrite (4.2.61) as

$$\begin{aligned} F(x) = F(0) + DF(0)x + \cdots + \frac{1}{(k-1)!} D^{k-1}F(0)x^{\otimes(k-1)} \\ + \frac{1}{k!} D^k F(\theta x) x^{\otimes k}. \end{aligned} \tag{4.2.63}$$

Note how (4.2.61) and (4.2.63) generalize (4.2.54).

Convergent power series and their derivatives

Here we consider functions given by convergent power series, of the form

$$F(x) = \sum_{\alpha \ge 0} b_\alpha x^\alpha. \tag{4.2.64}$$

We allow $b_\alpha \in \mathbb{C}$, and take $x = (x_1, \dots, x_n) \in \mathbb{R}^n$, with x^α given by (4.2.12). Here is our first result.

Proposition 4.2.8. *Assume there exist $y \in \mathbb{R}^n$ and $C_0 < \infty$ such that*

$$|y_k| = a_k > 0, \ \forall\, k, \quad |b_\alpha y^\alpha| \le C_0, \ \forall\, \alpha. \tag{4.2.65}$$

Then, for each $\delta \in (0,1)$, the series (4.2.64) converges absolutely and uniformly on each set

$$R_\delta = \{x \in \mathbb{R}^n : |x_k| \le (1-\delta) a_k, \ \forall\, k\}. \tag{4.2.66}$$

The sum $F(x)$ is continuous on $\widetilde{R} = \{x \in \mathbb{R}^n : |x_k| < a_k, \ \forall\, k\}$.

Proof. We have

$$x \in R_\delta \Longrightarrow |b_\alpha x^\alpha| \le C_0(1-\delta)^{|\alpha|}, \quad \forall\, \alpha, \tag{4.2.67}$$

hence

$$\begin{aligned}\sum_{\alpha\ge 0} |b_\alpha x^\alpha| &\le C_0 \sum_{\alpha\ge 0}(1-\delta)^{|\alpha|} \\ &= C_0 \sum_{\alpha_1\ge 0}(1-\delta)^{\alpha_1}\cdots \sum_{\alpha_n\ge 0}(1-\delta)^{\alpha_n} \\ &= C_0\delta^{-n} < \infty.\end{aligned} \tag{4.2.68}$$

Thus the power series (4.2.64) is absolutely convergent whenever $x \in R_\delta$. We also have, for each $N \in \mathbb{N}$,

$$F(x) = \sum_{|\alpha|\le N} b_\alpha x^\alpha + R_N(x), \tag{4.2.69}$$

and, for $x \in R_\delta$,

$$\begin{aligned}|R_N(x)| &\le \sum_{|\alpha|>N} |b_\alpha x^\alpha| \\ &\le C_0 \sum_{|\alpha|>N}(1-\delta)^{|\alpha|} \\ &= \varepsilon_N \to 0 \ \text{ as } \ N \to \infty.\end{aligned} \tag{4.2.70}$$

This shows that $R_N(x) \to 0$ uniformly for $x \in R_\delta$, and completes the proof of Proposition 4.2.8. □

We next discuss differentiability of power series.

Proposition 4.2.9. *In the setting of Proposition 4.2.8, F is differentiable on $\widetilde{R}$ and, for each $j \in \{1,\dots,n\}$,*

$$\frac{\partial F}{\partial x_j}(x) = \sum_{\alpha\ge\varepsilon_j} \alpha_j b_\alpha x^{\alpha-\varepsilon_j}, \quad \forall\, x \in \widetilde{R}. \tag{4.2.71}$$

Here, we set $\varepsilon_j = (0,\dots,1,\dots,0)$, with the 1 in the jth slot. It is convenient to begin the proof of Proposition 4.2.9 with the following.

Lemma 4.2.10. *In the setting of Proposition 4.2.8, for each $j \in \{1,\dots,n\}$,*

$$G_j(x) = \sum_{\alpha\ge\varepsilon_j} \alpha_j b_\alpha x^{\alpha-\varepsilon_j} \tag{4.2.72}$$

is absoletely convergent for $x \in \widetilde{R}$, uniformly on R_δ for each $\delta \in (0,1)$, therefore defining G_j as a continuous function on $\widetilde{R}$.

Proof. Take $a = (a_1,\dots,a_n)$, with a_j as in (4.2.65). Given $x \in R_\delta$, we have

$$\begin{aligned}\sum_{\alpha\ge\varepsilon_j} \alpha_j |b_\alpha x^{\alpha-\varepsilon_j}| &\le \sum_{\alpha\ge\varepsilon_j} \alpha_j(1-\delta)^{|\alpha|-1}|b_\alpha a^{\alpha-\varepsilon_j}| \\ &\le \frac{C_0}{a_j(1-\delta)} \sum_{\alpha\ge 0} \alpha_j(1-\delta)^{|\alpha|},\end{aligned} \tag{4.2.73}$$

and this is

$$\le M_\delta < \infty, \quad \forall\, \delta \in (0,1). \tag{4.2.74}$$

This gives the asserted convergence on R_δ and hence defines the function G_j, continuous on $\widetilde{R}$. ☐

To prove Proposition 4.2.9, we need to show that

$$\frac{\partial F}{\partial x_j} = G_j \quad \text{on} \quad \widetilde{R}, \tag{4.2.75}$$

for each j. Let us use the notation

$$\widehat{x}_j = (x_1, \dots, x_{j-1}, 0, x_{j+1}, \dots, x_n) = x - x_j e_j, \tag{4.2.76}$$

where e_j is the jth standard basis vector of $\mathbb{R}^n$. Now, given $x \in R_\delta,\ \delta \in (0,1)$, the uniform convergence of (4.2.72) on R_δ implies

$$\begin{aligned}\int_0^{x_j} G_j(\widehat{x}_j + te_j)\, dt &= \sum_{\alpha \ge \varepsilon_j} \alpha_j b_\alpha \int_0^{x_j} (\widehat{x}_j + te_j)^{\alpha - \varepsilon_j}\, dt \\ &= \sum_{\alpha \ge \varepsilon_j} \alpha_j b_\alpha \alpha_j^{-1} x^\alpha \\ &= \sum_{\alpha \ge \varepsilon_j} b_\alpha x^\alpha \\ &= F(x) - F(\widehat{x}_j).\end{aligned} \tag{4.2.77}$$

Applying $\partial/\partial x_j$ to the left side of (4.2.77) and using the fundamental theorem of calculus then yields (4.2.75) as desired. This gives the identity (4.2.71). Since each G_j is continuous on $\widetilde{R}$, this implies F is differentiable on $\widetilde{R}$.

We can iterate Proposition 4.2.9, obtaining $\partial_k \partial_j F(x) = \partial_k G_j(x)$ as a convergent power series on $\widetilde{R}$, etc. In particular, we have the following.

Corollary 4.2.11. *In the setting of Proposition 4.2.8, we have $F \in C^\infty(\widetilde{R})$.*

Exercises

1. Considering the power series

$$f(x) = f(y) + f'(y)(x-y) + \cdots + \frac{f^{(j)}(y)}{j!}(x-y)^j + R_j(x,y),$$

show that

$$\frac{\partial R_j}{\partial y} = -\frac{1}{j!} f^{(j+1)}(y)(x-y)^j, \quad R_j(x,x) = 0.$$

Use this to derive (4.2.13)–(4.2.14).

We define "big oh" and "little oh" notation:

$$f(x) = O(x) \quad (\text{as } x \to 0) \Leftrightarrow \Big|\frac{f(x)}{x}\Big| \le C \quad \text{as} \quad x \to 0,$$
$$f(x) = o(x) \quad (\text{as } x \to 0) \Leftrightarrow \frac{f(x)}{x} \to 0 \quad \text{as} \quad x \to 0.$$

2. Let $\mathcal{O} \subset \mathbb{R}^n$ be open and $y \in \mathcal{O}$. Show that

$$f \in C^{k+1}(\mathcal{O}) \Rightarrow f(x) = \sum_{|\alpha| \le k} \frac{1}{\alpha!} f^{(\alpha)}(y)(x-y)^\alpha + O(|x-y|^{k+1}),$$
$$f \in C^{k}(\mathcal{O}) \Rightarrow f(x) = \sum_{|\alpha| \le k} \frac{1}{\alpha!} f^{(\alpha)}(y)(x-y)^\alpha + o(|x-y|^{k}).$$

3. Assume $G : U \to \mathcal{O},\ F : \mathcal{O} \to \Omega$. Show that

$$F, G \in C^1 \Longrightarrow F \circ G \in C^1. \tag{4.2.78}$$

More generally, show that, for $k \in \mathbb{N}$,

$$F, G \in C^k \Longrightarrow F \circ G \in C^k. \tag{4.2.79}$$

Hint. Write $H = F \circ G$, with $h_\ell(x) = f_\ell(g_1(x), \dots, g_n(x))$, and use (4.1.35) to get

$$\partial_j h_\ell(x) = \sum_{k=1}^n \partial_k f_\ell(g_1, \dots, g_n) \partial_j g_k. \tag{4.2.80}$$

Show that this yields (4.2.78). To proceed, deduce from (4.2.80) that

$$\begin{aligned} \partial_{j_1}\partial_{j_2} h_\ell(x) = &\sum_{k_1,k_2=1}^n \partial_{k_1}\partial_{k_2} f_\ell(g_1, \dots, g_n)(\partial_{j_1} g_{k_1})(\partial_{j_2} g_{k_2}) \\ &+ \sum_{k=1}^n \partial_k f_\ell(g_1, \dots, g_n) \partial_{j_1}\partial_{j_2} g_k. \end{aligned} \tag{4.2.81}$$

Use this to get (4.2.79) for $k = 2$. Proceeding inductively, show that there exist constants $C(\mu, J^\#, k^\#) = C(\mu, J_1, \dots, J_\mu, k_1, \dots, k_\mu)$ such that if $F, G \in C^k$ and $|J| \le k$,

$$h_\ell^{(J)}(x) = \sum C(\mu, J^\#, k^\#) g_{k_1}^{(J_1)} \cdots g_{k_\mu}^{(J_\mu)} f_\ell^{(k_1, \dots, k_\mu)}(g_1, \dots, g_n), \tag{4.2.82}$$

where the sum is over

$$\mu \le |J|,\ J_1 + \cdots + J_\mu \sim J,\ |J_\nu| \ge 1,$$

and $J_1 + \cdots + J_\mu \sim J$ means J is a rearrangement of $J_1 + \cdots + J_\mu$. Show that (4.2.79) follows from this.

4. Show that the map $\Phi : Gl(n, \mathbb{R}) \to Gl(n, \mathbb{R})$ given by $\Phi(X) = X^{-1}$ is C^k for each k, i.e., $\Phi \in C^\infty$.
Hint. Start with the material of Exercise 3. Write $D\Phi(X)Y = -X^{-1}YX^{-1}$ as

$$\partial_{\ell m}\Phi(X) = \frac{\partial}{\partial x_{\ell m}} \Phi(X) = D\Phi(X)E_{\ell m} = -\Phi(X)E_{\ell m}\Phi(X),$$

where $X = (x_{\ell m})$ and $E_{\ell m}$ has just one nonzero entry, at position (ℓ, m). Iterate this to get

$$\partial_{\ell_2 m_2}\partial_{\ell_1 m_1}\Phi(X) = -(\partial_{\ell_2 m_2}\Phi(X))E_{\ell_1 m_1}\Phi(X) - \Phi(X)E_{\ell_1 m_1}(\partial_{\ell_2 m_2}\Phi(X)),$$

and continue.

5. Define $g(x, y)$ on $\mathbb{R}^2$ by

$$g(x,y) = \begin{cases} \dfrac{xy^3}{x^2+y^2}, & (x,y) \neq (0,0), \\ 0, & (x,y) = (0,0). \end{cases}$$

Show that g is smooth on $\mathbb{R}^2 \setminus (0,0)$ and class C^1 on $\mathbb{R}^2$. Show that $\partial_x\partial_y g$ and $\partial_y\partial_x g$ exist at each point of $\mathbb{R}^2$, and are continuous on $\mathbb{R}^2 \setminus (0,0)$, but not on $\mathbb{R}^2$. Show that

$$\frac{\partial}{\partial y}\frac{\partial g}{\partial x}(0,0) = 1, \quad \frac{\partial}{\partial x}\frac{\partial g}{\partial y}(0,0) = 0.$$

6. Use the fact that $\det X$ is a polynomial in the matrix entries of X to show directly that $\det : M(n, \mathbb{R}) \to \mathbb{R}$ is continuous, and of class C^k for all k. Use the continuity of det plus the characterization

$$G\ell(n,\mathbb{R}) = \{X \in M(n,\mathbb{R}) : \det X \neq 0\}$$

to show that $G\ell(n, \mathbb{R})$ is open in $M(n, \mathbb{R})$.

7. Let $\Omega \subset \mathbb{R}^n$ be open, $f, g \in C^k(\Omega)$, real valued, $0 \in \Omega$. Write

$$f(x) = \sum_{|\beta|\le k} f_\beta x^\beta + o(x^k), \quad g(x) = \sum_{|\gamma|\le k} g_\gamma x^\gamma + o(x^k),$$

with

$$f_\beta = \frac{f^{(\beta)}(0)}{\beta!}, \quad g_\gamma = \frac{g^{(\gamma)}(0)}{\gamma!}.$$

Show that $h(x) = f(x)g(x)$ satisfies

$$h(x) = \sum_{|\beta|,|\gamma|\le k} f_\beta g_\gamma x^{\beta+\gamma} + o(x^k),$$

and deduce that, for $|\alpha| \le k$,

$$\frac{h^{(\alpha)}(0)}{\alpha!} = \sum_{\beta+\gamma=\alpha} f_\beta g_\gamma = \sum_{\beta+\gamma=\alpha} \frac{1}{\beta!\gamma!} f^{(\beta)}(0)g^{(\gamma)}(0).$$

From this, deduce that

$$\partial^\alpha(fg)(0) = \sum_{\beta+\gamma=\alpha} \frac{\alpha!}{\beta!\gamma!} f^{(\beta)}(0)g^{(\gamma)}(0).$$

Pass from this to the identity

$$\partial^\alpha(fg)(x) = \sum_{\beta+\gamma=\alpha} \frac{\alpha!}{\beta!\gamma!} f^{(\beta)}(x)g^{(\gamma)}(x), \tag{4.2.83}$$

for $x \in \Omega$. This identity is called the *Leibniz identity*.

8. Let $\mathcal{O} \subset \mathbb{R}^n$ be open. Take $K \subset \mathcal{O}$ compact, and $r > 0$ small enough that $x \in K,\ v \in \mathbb{R}^n,\ |v| \le r \Rightarrow x + v \in \mathcal{O}$. For $f \in C(\mathcal{O}),\ x \in K$, define

$$\tau_v f(x) = f(x+v).$$

Show that if $v, w \in \mathbb{R}^n,\ |v|, |w| \le r/2$, then, for $f \in C(\mathcal{O}),\ x \in K$,

$$\tau_w \tau_v f(x) = \tau_v \tau_w f(x). \tag{4.2.84}$$

Show that

$$\Delta_{j,h} f(x) = \frac{1}{h}\big[\tau_{he_j} - I\big] f(x).$$

Deduce from (4.2.84) that, if $|h| \le r/2,\ x \in K$,

$$\Delta_{j,h} \Delta_{k,h} f(x) = \Delta_{k,h} \Delta_{j,h} f(x).$$

Cf. (4.2.7). Also show that, if $f \in C^1(\mathcal{O})$,

$$\partial_k \tau_v f(x) = \tau_v \partial_k f(x),$$

and deduce that (cf. (4.2.5))

$$\partial_k \Delta_{j,h} f(x) = \Delta_{j,h} \partial_k f(x).$$

9. Consider the following function $f : \mathbb{R}^2 \to \mathbb{R}$:

$$f(x,y) = (\sin x)(\sin y).$$

Find all its critical points, and determine which of these are local maxima, local minima, and saddle points.

10. Define $f : \mathbb{R}^2 \to \mathbb{R}$ by

$$f(x) = e^{x_1} \cos x_2.$$

Compute $f^{(\alpha)}(x)$ for $|\alpha| \le 3$. Then write down

$$P(x) = \sum_{|\alpha| \le 3} \frac{1}{\alpha!} f^{(\alpha)}(0) x^\alpha.$$

11. Attack the computation of $P(x)$ in Exercise 10 using Exercise 7, starting with

$$e^{x_1} = 1 + x_1 + \frac{x_1^2}{2} + \frac{x_1^3}{3!} + \cdots,$$

and a similar expansion of $\cos x_2$.

12. Write down the power series about $(0,0)$ of

$$F(x,y) = \int_0^1 \frac{e^{xt}}{1+yt}\, dt.$$

Hint. Start by multiplying the power series of e^{xt} and $(1+yt)^{-1}$.

13. Show that, for $x = (x_1, \dots, x_n)$, with $|x_j| < 1$ for all j,

$$\sum_{\alpha \ge 0} x^\alpha = \frac{1}{1-x_1} \cdots \frac{1}{1-x_n}.$$

Hint. Write the left side as

$$\sum_{\alpha_1 \ge 0} x_1^{\alpha_1} \cdots \sum_{\alpha_n \ge 0} x_n^{\alpha_n}.$$

14. In this exercise, we take

$$\eta = (t, t, \dots, t) \in \mathbb{R}^n, \quad |t| < 1,$$

and consider

$$F(\eta) = \sum_{\alpha \ge 0} \eta^\alpha.$$

(a) Show that, for $|t| < 1$,

$$F(\eta) = \sum_{\alpha_1 \ge 0} t^{\alpha_1} \sum_{\alpha_2 \ge 0} t^{\alpha_2} \cdots \sum_{\alpha_n \ge 0} t^{\alpha_n} = (1-t)^{-n}.$$

(b) Show that

$$F(\eta) = \sum_{\alpha \ge 0} t^{|\alpha|} = \sum_{k=0}^{\infty} d_k(n) t^k,$$

where

$$\begin{aligned} d_k(n) &= \#\{\alpha = (\alpha_1, \dots, \alpha_n) : |\alpha| = k\} \\ &= \dim \mathcal{P}_k(\mathbb{R}^n), \end{aligned}$$

with

$$\mathcal{P}_k(\mathbb{R}^n) = \text{space of polynomials in } x \in \mathbb{R}^n, \text{ homogeneous of degree } k.$$

(c) Comparing results of (a) and (b), show that

$$\begin{aligned} d_k(n) &= \text{ coefficient of } t^k \text{ in } f_n(t) = (1-t)^{-n} \\ &= \frac{1}{k!} f_n^{(k)}(0) \\ &= \frac{n(n+1)\cdots(n+k-1)}{k!} \\ &= \binom{n+k-1}{k}. \end{aligned}$$

(d) If $\mathcal{P}^k(\mathbb{R}^n)$ = space of polynomials in $x \in \mathbb{R}^n$ of degree $\le k$, show that

$$\begin{aligned} \dim \mathcal{P}^k(\mathbb{R}^n) &= \dim \mathcal{P}_k(\mathbb{R}^{n+1}) \\ &= \binom{n+k}{k} = \binom{n+k}{n} \\ &= \frac{(k+n)(k+n-1)\cdots(k+1)}{n!}. \end{aligned}$$

4.3. Inverse function and implicit function theorem

The Inverse Function Theorem gives a condition under which a function can be locally inverted. This theorem and its corollary the Implicit Function Theorem are fundamental results in multivariable calculus. First we state the Inverse Function Theorem. Here, we assume $k \geq 1$.

Theorem 4.3.1. *Let F be a C^k map from an open neighborhood Ω of $p_0 \in \mathbb{R}^n$ to $\mathbb{R}^n$, with $q_0 = F(p_0)$. Suppose the derivative $DF(p_0)$ is invertible. Then there is a neighborhood U of p_0 and a neighborhood V of q_0 such that $F : U \to V$ is one-to-one and onto, and $F^{-1} : V \to U$ is a C^k map. (One says $F : U \to V$ is a diffeomorphism.)*

First we show that F is one-to-one on a neighborhood of p_0, under these hypotheses. In fact, we establish the following result, of interest in its own right.

Proposition 4.3.2. *Assume $\Omega \subset \mathbb{R}^n$ is open and convex, and let $f : \Omega \to \mathbb{R}^n$ be C^1. Assume that the symmetric part of $Df(u)$ is positive-definite, for each $u \in \Omega$. Then f is one-to-one on Ω.*

Proof. Take distinct points $u_1,\ u_2 \in \Omega$, and set $u_2 - u_1 = w$. Consider $\varphi : [0,1] \to \mathbb{R}$, given by

$$\varphi(t) = w \cdot f(u_1 + tw).$$

Then $\varphi'(t) = w \cdot Df(u_1+tw)w > 0$ for $t \in [0,1]$, so $\varphi(0) \neq \varphi(1)$. But $\varphi(0) = w \cdot f(u_1)$ and $\varphi(1) = w \cdot f(u_2)$, so $f(u_1) \neq f(u_2)$. ☐

To continue the proof of Theorem 4.3.1, let us set

$$f(u) = A\big(F(p_0 + u) - q_0\big), \quad A = DF(p_0)^{-1}. \tag{4.3.1}$$

Then $f(0) = 0$ and $Df(0) = I$, the identity matrix. We will show that f maps a neighborhood of 0 one-to-one and onto some neighborhood of 0. We can write

$$f(u) = u + R(u), \quad R(0) = 0, \quad DR(0) = 0, \tag{4.3.2}$$

and R is C^1. Pick $b > 0$ such that

$$\|u\| \leq 2b \Longrightarrow \|DR(u)\| \leq \frac{1}{2}. \tag{4.3.3}$$

Then $Df = I + DR$ has positive definite symmetric part on

$$B_{2b}(0) = \{u \in \mathbb{R}^n : \|u\| < 2b\},$$

so by Proposition 4.3.2,

$$f : B_{2b}(0) \longrightarrow \mathbb{R}^n \ \text{ is one-to-one.}$$

We will show that the range $f(B_{2b}(0))$ contains $B_b(0)$, that is to say, we can solve

$$f(u) = v, \tag{4.3.4}$$

given $v \in B_b(0)$, for some (unique) $u \in B_{2b}(0)$. This is equivalent to $u + R(u) = v$.

To get the solution, we set

$$T_v(u) = v - R(u). \tag{4.3.5}$$

Then solving (4.3.4) is equivalent to solving

$$T_v(u) = u. \tag{4.3.6}$$

We look for a fixed point

$$u = K(v) = f^{-1}(v). \tag{4.3.7}$$

Also, we want to show that $DK(0) = I$, i.e., that

$$K(v) = v + r(v), \quad r(v) = o(\|v\|). \tag{4.3.8}$$

(The "little oh" notation was introduced in (4.1.1)–(4.1.2), and studied in Exercise 8 of §4.1.) If we succeed in doing this, it follows that, for y close to q_0, $G(y) = F^{-1}(y)$ is defined. Also, taking

$$x = p_0 + u, \quad y = F(x), \quad v = f(u) = A(y - q_0),$$

as in (4.3.1), we have, via (4.3.8),

$$\begin{aligned} G(y) &= p_0 + u = p_0 + K(v) \\ &= p_0 + K(A(y - q_0)) \\ &= p_0 + A(y - q_0) + o(\|y - q_0\|). \end{aligned}$$

Hence G is differentiable at q_0 and

$$DG(q_0) = A = DF(p_0)^{-1}. \tag{4.3.9}$$

A parallel argument, with p_0 replaced by a nearby x and $y = F(x)$, gives

$$DG(y) = DF(G(y))^{-1}. \tag{4.3.10}$$

Thus our task is to solve (4.3.6). To do this, we use the following general result, known as the Contraction Mapping Theorem.

Theorem 4.3.3. *Let X be a complete metric space, and let $T : X \to X$ satisfy*

$$dist(Tx, Ty) \le r \; dist(x, y), \tag{4.3.11}$$

for some $r < 1$. (We say T is a contraction.) Then T has a unique fixed point x. For any $y_0 \in X$, $T^k y_0 \to x$ as $k \to \infty$.

Proof. Pick $y_0 \in X$ and let $y_k = T^k y_0$. Then $\operatorname{dist}(y_k, y_{k+1}) \le r^k \operatorname{dist}(y_0, y_1)$, so

$$\begin{aligned} \operatorname{dist}(y_k, y_{k+m}) &\le \operatorname{dist}(y_k, y_{k+1}) + \cdots + \operatorname{dist}(y_{k+m-1}, y_{k+m}) \\ &\le \big(r^k + \cdots + r^{k+m-1}\big) \operatorname{dist}(y_0, y_1) \\ &\le r^k \big(1 - r\big)^{-1} \operatorname{dist}(y_0, y_1). \end{aligned} \tag{4.3.12}$$

It follows that (y_k) is a Cauchy sequence, so it converges; $y_k \to x$. Since $Ty_k = y_{k+1}$ and T is continuous, it follows that $Tx = x$, i.e., x is a fixed point. Uniqueness of the fixed point is clear from the estimate $\operatorname{dist}(Tx, Tx') \le r \operatorname{dist}(x, x')$, which implies $\operatorname{dist}(x, x') = 0$ if x and x' are fixed points. This proves Theorem 4.3.3. □

Returning to the task of solving (4.3.6), having b as in (4.3.3), we claim that

$$\begin{aligned} \|v\| \le b &\Longrightarrow T_v : \overline{B_{2\|v\|}(0)} \to X_v \\ &\Longrightarrow T_v : X_v \to X_v, \end{aligned} \tag{4.3.13}$$

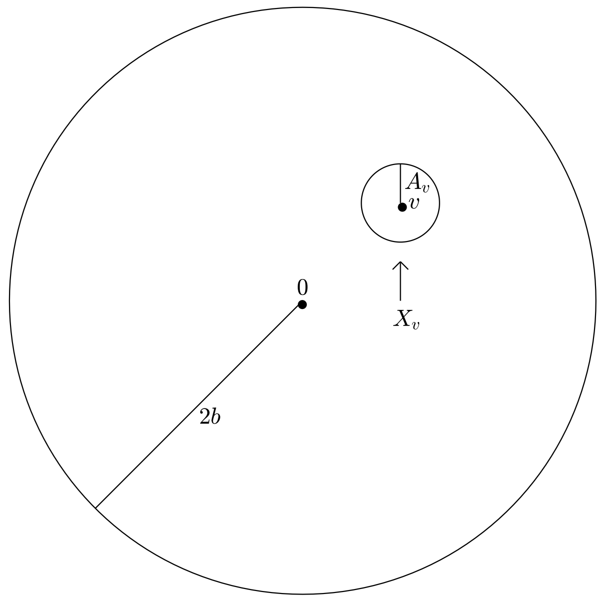

Figure 4.3.1. $T_v : X_v \to X_v$

where

$$X_v = \{u \in \overline{B_{2b}(0)} : \|u - v\| \le A_v\}, \quad A_v = \sup_{\|w\| \le 2\|v\|} \|R(w)\|. \tag{4.3.14}$$

See Figure 4.3.1. Note from (4.3.2)–(4.3.3) that

$$\|w\| \le 2b \Longrightarrow \|R(w)\| \le \frac{1}{2}\|w\|, \quad \text{and} \quad \|R(w)\| = o(\|w\|). \tag{4.3.15}$$

Hence

$$\|v\| \le b \Longrightarrow A_v \le \|v\|, \quad \text{and} \quad A_v = o(\|v\|). \tag{4.3.16}$$

In particular, when $\|v\| \le b$,

$$\begin{aligned} \|w - v\| \le A_v &\Longrightarrow \|w - v\| \le \|v\| \\ &\Longrightarrow \|w\| \le 2\|v\| \le 2b, \end{aligned} \tag{4.3.17}$$

which in turn implies $w \in X_v$. In addition,

$$\begin{aligned} \|u\| \le 2\|v\| &\Longrightarrow \|R(u)\| \le A_v \\ &\Longrightarrow \|T_v(u) - v\| \le A_v, \end{aligned} \tag{4.3.18}$$

giving the first implication in (4.3.13). Furthermore, via (4.3.17),

$$u \in X_v \Longrightarrow \|u\| \le 2\|v\|,$$

so

$$X_v \subset \overline{B_{2\|v\|}(0)}, \tag{4.3.19}$$

and we have the second implication in (4.3.13).

As for the contraction property, given $u_j \in X_b,\ \|v\| \le b$,

$$\begin{aligned}\|T_v(u_1) - T_v(u_2)\| &= \|R(u_2) - R(u_1)\| \\ &\le \frac{1}{2}\|u_1 - u_2\|,\end{aligned} \tag{4.3.20}$$

the last inequality by (4.3.3), so the map (4.3.13) is a contraction. Hence, by Theorem 4.3.3, there is a unique fixed point, $u = K(v) \in X_v$. Also, since $u \in X_v$,

$$\|K(v) - v\| \le A_v = o(\|v\|). \tag{4.3.21}$$

Thus we have (4.3.8). This establishes the existence of the inverse function $G = F^{-1} : V \to U$, and we have the formula (4.3.10) for the derivative DG. Since G is differentiable on V, it is certainly continuous, so (4.3.10) implies DG is continuous, given $F \in C^1(U)$.

To finish the proof of the Inverse Function Theorem and show that G is C^k if F is C^k, for $k \ge 2$, one uses an inductive argument. See Exercise 6 at the end of this section for an approach to this last argument.

Thus if DF is invertible on the domain of F, F is a local diffeomorphism. Stronger hypotheses are needed to guarantee that F is a global diffeomorphism onto its range. Proposition 4.3.2 provides one tool for doing this. Here is a slight strengthening.

Corollary 4.3.4. *Assume $\Omega \subset \mathbb{R}^n$ is open and convex, and that $F : \Omega \to \mathbb{R}^n$ is C^1. Assume there exist $n \times n$ matrices A and B such that the symmetric part of $A\,DF(u)\,B$ is positive definite for each $u \in \Omega$. Then F maps Ω diffeomorphically onto its image, an open set in $\mathbb{R}^n$.*

Proof. Exercise. □

We make a comment about solving the equation $F(x) = y$, under the hypotheses of Theorem 4.3.1, when y is close to q_0. The fact that finding the fixed point for T_v in (4.3.13) is accomplished by taking the limit of $T_v^k(v)$ implies that, when y is sufficiently close to q_0, the sequence (x_k), defined by

$$x_0 = p_0, \quad x_{k+1} = x_k + DF(p_0)^{-1}\big(y - F(x_k)\big), \tag{4.3.22}$$

converges to the solution x. An analysis of the rate at which $x_k \to x$, and $F(x_k) \to y$, can be made by applying F to (4.3.22), yielding

$$\begin{aligned}F(x_{k+1}) &= F(x_k + DF(p_0)^{-1}(y - F(x_k)) \\ &= F(x_k) + DF(x_k)DF(p_0)^{-1}(y - F(x_k)) \\ &\quad + R(x_k, DF(p_0)^{-1}(y - F(x_k))),\end{aligned}$$

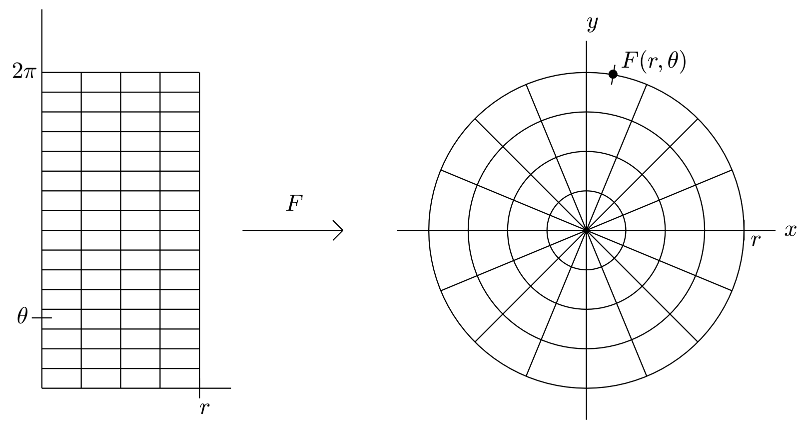

Figure 4.3.2. Polar coordinates on $\mathbb{R}^2$

and hence

$$
\begin{aligned}
y - F(x_{k+1}) &= \big(I - DF(x_k)DF(p_0)^{-1}\big)(y - F(x_k)) \\
&\quad + \widetilde{R}(x_k, y - F(x_k)),
\end{aligned} \tag{4.3.23}
$$

with $\|\widetilde{R}(x_k, y - F(x_k))\| = o(\|y - F(x_k)\|)$.

It turns out that replacing $DF(p_0)^{-1}$ by $DF(x_k)^{-1}$ in (4.3.22) yields a faster approximation. This method, known as Newton's method, is described in the exercises.

We consider some examples of maps to which Theorem 4.3.1 applies. First, we look at polar coordinates on $\mathbb{R}^2$ (previewed in (3.1.48)):

$$
F : (0, \infty) \times \mathbb{R} \longrightarrow \mathbb{R}^2, \quad F(r, \theta) = \begin{pmatrix} r\cos\theta \\ r\sin\theta \end{pmatrix} = \begin{pmatrix} x(r, \theta) \\ y(r, \theta) \end{pmatrix}. \tag{4.3.24}
$$

See Figure 4.3.2. We have

$$
DF(r, \theta) = \begin{pmatrix} \partial_r x & \partial_\theta x \\ \partial_r y & \partial_\theta y \end{pmatrix} = \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix}, \tag{4.3.25}
$$

so

$$
\det DF(r, \theta) = r\cos^2\theta + r\sin^2\theta = r. \tag{4.3.26}
$$

Hence $DF(r, \theta)$ is invertible for all $(r, \theta) \in (0, \infty) \times \mathbb{R}$. Theorem 4.3.1 implies that each $(r_0, \theta_0) \in (0, \infty) \times \mathbb{R}$ has a neighborhood U and $(x_0, y_0) = (r_0\cos\theta_0, r_0\sin\theta_0)$

has a neighborhood V such that F is a smooth diffeomorphism of U onto V. In this simple situation, it can be verified directly that

$$F : (0, \infty) \times (-\pi, \pi) \longrightarrow \mathbb{R}^2 \setminus \{(x, 0) : x \le 0\} \tag{4.3.27}$$

is a smooth diffeomorphism.

Note that $DF(1, 0) = I$ in (4.3.25). Let us check the domain of applicability of Proposition 4.3.2. The symmetric part of $DF(r, \theta)$ in (4.3.25) is

$$S(r, \theta) = \begin{pmatrix} \cos\theta & \frac{1}{2}(1-r)\sin\theta \\ \frac{1}{2}(1-r)\sin\theta & r\cos\theta \end{pmatrix}. \tag{4.3.28}$$

By Proposition 4.2.6, this is positive definite if and only if

$$\cos\theta > 0, \tag{4.3.29}$$

and

$$\det S(r, \theta) = r\cos^2\theta - \frac{1}{4}(1-r)^2 \sin^2\theta > 0. \tag{4.3.30}$$

Now (4.3.29) holds for $\theta \in (-\pi/2, \pi/2)$, but not on all of $(-\pi, \pi)$. Furthermore, (4.3.30) holds for (r, θ) in a neighborhood of $(r_0, \theta_0) = (1, 0)$, but it does not hold on all of $(0, \infty) \times (-\pi/2, \pi/2)$. We see that Proposition 4.3.2 does not capture the full force of the diffeomorphism property of (4.3.27).

We move on to another example. As in §4.1, we can extend Theorem 4.3.1, replacing $\mathbb{R}^n$ by a finite dimensional real vector space, isometric to a Euclidean space, such as $M(n, \mathbb{R}) \approx \mathbb{R}^{n^2}$. Consider the matrix exponential

$$\mathrm{Exp} : M(n, \mathbb{R}) \longrightarrow M(n, \mathbb{R}), \quad \mathrm{Exp}(X) = e^X = \sum_{k=0}^{\infty} \frac{1}{k!} X^k. \tag{4.3.31}$$

Smoothness of Exp follows from Corollary 4.2.11. See §C.4 for more. Since

$$\mathrm{Exp}(Y) = I + Y + \frac{1}{2} Y^2 + \cdots, \tag{4.3.32}$$

we have

$$D\,\mathrm{Exp}(0)Y = Y, \quad \forall\, Y \in M(n, \mathbb{R}), \tag{4.3.33}$$

so $D\,\mathrm{Exp}(0)$ is invertible. Then Theorem 4.3.1 implies that there exist a neighborhod U of $0 \in M(n, \mathbb{R})$ and a neighborhood V of $I \in M(n, \mathbb{R})$ such that $\mathrm{Exp} : U \to V$ is a smooth diffeomorphism.

We move from the inverse function theorem to the implicit function theorem. To motivate the result, we consider the following example. Take $a > 0$ and consider the equation

$$x^2 + y^2 = a^2, \quad F(x, y) = x^2 + y^2. \tag{4.3.34}$$

Note that

$$DF(x, y) = (2x \;\; 2y), \quad D_x F(x, y) = 2x, \quad D_y F(x, y) = 2y. \tag{4.3.35}$$

The equation (4.3.34) defines y "implicitly" as a smooth function of x if $|x| < a$. Explicitly,

$$|x| < a \Longrightarrow y = \sqrt{a^2 - x^2}, \tag{4.3.36}$$

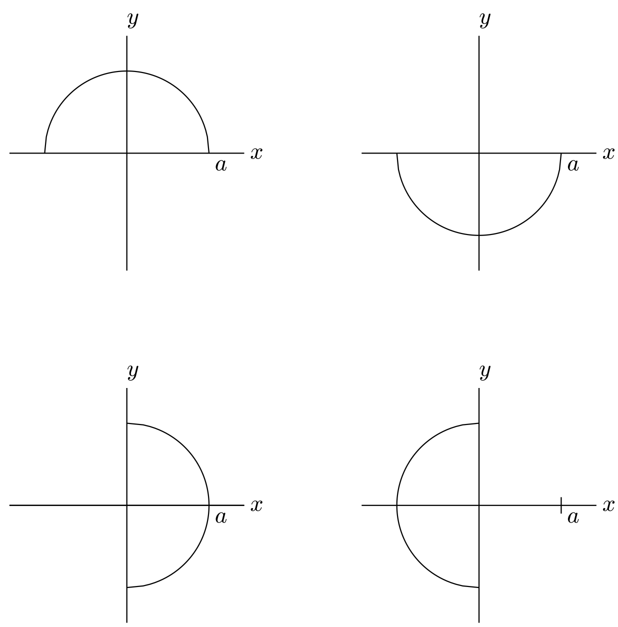

Figure 4.3.3. Functions defined implicitly by $x^2 + y^2 = a^2$

or alternatively $y = -\sqrt{a^2 - x^2}$. Similarly, (4.3.34) defines x implicitly as a smooth function of y if $|y| < a$; explicitly

$$|y| < a \Longrightarrow x = \sqrt{a^2 - y^2}, \tag{4.3.37}$$

or alternatively $x = -\sqrt{a^2 - y^2}$. See Figure 4.3.3 for an illustration. Now, given $x_0 \in \mathbb{R},\ a > 0$, there exists $y_0 \in \mathbb{R}$ such that $F(x_0, y_0) = a^2$ if and only if $|x_0| \le a$. Furthermore,

$$\text{given} \quad F(x_0, y_0) = a^2, \quad D_y F(x_0, y_0) \neq 0 \Leftrightarrow |x_0| < a. \tag{4.3.38}$$

Similarly, given $y_0 \in \mathbb{R}$, there exists x_0 such that $F(x_0, y_0) = a^2$ if and only if $|y_0| \le a$, and

$$\text{given} \quad F(x_0, y_0) = a^2, \quad D_x F(x_0, y_0) \neq 0 \Leftrightarrow |x_0| < a. \tag{4.3.39}$$

Note also that, whenever $(x, y) \in \mathbb{R}^2$ and $F(x, y) = a^2 > 0$,

$$DF(x, y) \neq 0, \tag{4.3.40}$$

so either $D_x F(x, y) \neq 0$ or $D_y F(x, y) \neq 0$, and, as seen above whenever $(x_0, y_0) \in \mathbb{R}^2$ and $F(x_0, y_0) = a^2 > 0$, we can solve $F(x, y) = a^2$ for either y as a smooth function of x for x near x_0 or for x as a smooth function of y for y near y_0.

We move from these observations to the next result, the Implicit Function Theorem.

Theorem 4.3.5. *Suppose U is a neighborhood of $x_0 \in \mathbb{R}^m$, V a neighborhood of $y_0 \in \mathbb{R}^\ell$, and we have a C^k map*

$$F : U \times V \longrightarrow \mathbb{R}^\ell, \quad F(x_0, y_0) = u_0. \tag{4.3.41}$$

Assume $D_y F(x_0, y_0)$ is invertible. Then the equation $F(x, y) = u_0$ defines $y = g(x, u_0)$ for x near x_0 (satisfying $g(x_0, u_0) = y_0$) with g a C^k map.

Proof. Consider $H : U \times V \to \mathbb{R}^m \times \mathbb{R}^\ell$ defined by

$$H(x, y) = \big(x, F(x, y)\big). \tag{4.3.42}$$

(Actually, regard (x, y) and $(x, F(x, y))$ as column vectors.) We have

$$DH = \begin{pmatrix} I & 0 \\ D_x F & D_y F \end{pmatrix}. \tag{4.3.43}$$

Thus $DH(x_0, y_0)$ is invertible, so $G = H^{-1}$ exists, on a neighborhood of (x_0, u_0), and is C^k, by the Inverse Function Theorem. Let us set

$$G(x, u) = (\xi(x, u), g(x, u)). \tag{4.3.44}$$

Then

$$\begin{aligned} H \circ G(x, u) &= H(\xi(x, u), g(x, u)) \\ &= (\xi(x, u), F(\xi(x, u), g(x, u)). \end{aligned} \tag{4.3.45}$$

Since $H \circ G(x, u) = (x, u)$, we have $\xi(x, u) = x$, so

$$G(x, u) = (x, g(x, u)) \tag{4.3.46}$$

and hence

$$H \circ G(x, u) = (x, F(x, g(x, u)), \tag{4.3.47}$$

hence

$$F(x, g(x, u)) = u. \tag{4.3.48}$$

Note that $G(x_0, u_0) = (x_0, y_0)$, so $g(x_0, u_0) = y_0$, and g is the desired map. ☐

Here is an example where Theorem 4.3.5 applies. Set

$$F : \mathbb{R}^4 \longrightarrow \mathbb{R}^2, \quad F(u, v, x, y) = \begin{pmatrix} x(u^2 + v^2) \\ xu + yv \end{pmatrix}. \tag{4.3.49}$$

We have

$$F(2, 0, 1, 1) = \begin{pmatrix} 4 \\ 2 \end{pmatrix}. \tag{4.3.50}$$

Note that

$$D_{u,v} F(u, v, x, y) = \begin{pmatrix} 2xu & 2xv \\ x & y \end{pmatrix}, \tag{4.3.51}$$

hence

$$D_{u,v} F(2, 0, 1, 1) = \begin{pmatrix} 4 & 0 \\ 1 & 1 \end{pmatrix} \tag{4.3.52}$$

is invertible, so Theorem 4.3.5 (with (u, v) in place of y and (x, y) in place of x) implies that the equation

$$F(u, v, x, y) = \begin{pmatrix} 4 \\ 2 \end{pmatrix} \tag{4.3.53}$$

defines smooth functions

$$u = u(x, y), \quad v = v(x, y), \tag{4.3.54}$$

for (x, y) near $(x_0, y_0) = (1, 1)$, satisfying (4.3.53), with $(u(1, 1), v(1, 1)) = (2, 0)$.

Let us next focus on the case $\ell = 1$ of Theorem 4.3.5, so

$$z = (x, y) \in \mathbb{R}^n, \quad x \in \mathbb{R}^{n-1}, \quad y \in \mathbb{R}, \quad F(z) \in \mathbb{R}. \tag{4.3.55}$$

Then $D_y F = \partial_y F$. If $F(x_0, y_0) = u_0$, Theorem 4.3.5 says that if

$$\partial_y F(x_0, y_0) \neq 0, \tag{4.3.56}$$

then one can solve

$$F(x, y) = u_0 \ \text{ for } \ y = g(x, u_0), \tag{4.3.57}$$

for x near x_0 (satisfying $g(x_0, u_0) = y_0$), with g a C^k function. This phenomenon was illustrated in (4.3.34)–(4.3.38). To generalize the observations involving (4.3.39)–(4.3.40), we note the following. Set $(x, y) = z = (z_1, \dots, z_n)$, $z_0 = (x_0, y_0)$. The condition (4.3.56) is that $\partial_{z_n} F(z_0) \neq 0$. Now a simple permutation of variables allows us to pick $j \in \{1, \dots, n\}$ and modify our assumption to

$$\partial_{z_j} F(z_0) \neq 0, \quad F(z_0) = u_0, \tag{4.3.58}$$

and deduce that one can solve

$$F(z) = u_0, \quad \text{for } \ z_j = g(z_1, \dots, z_{j-1}, z_{j+1}, \dots, z_n). \tag{4.3.59}$$

Let us record this result, changing notation and replacing z by x.

Proposition 4.3.6. *Let Ω be a neighborhood of $x_0 \in \mathbb{R}^n$. Asume we have a C^k function*

$$F : \Omega \longrightarrow \mathbb{R}, \quad F(x_0) = u_0, \tag{4.3.60}$$

and assume

$$DF(x_0) \neq 0, \quad \textit{i.e.,} \ \ (\partial_1 F(x_0), \dots, \partial_n F(x_0)) \neq 0. \tag{4.3.61}$$

Then there exists $j \in \{1, \dots, n\}$ such that one can solve $F(x) = u_0$ for

$$x_j = g(x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_n), \tag{4.3.62}$$

with $(x_{10}, \dots, x_{j0}, \dots, x_{n0}) = x_0$, for a C^k function g.

Remark. For $F : \Omega \to \mathbb{R}$, it is common to denote $DF(x)$ by $\nabla F(x)$,

$$\nabla F(x) = (\partial_1 F(x), \dots, \partial_n F(x)). \tag{4.3.63}$$

Here is an example to which Proposition 4.3.6 applies. Using the notation $(x, y) = (x_1, x_2)$, set

$$F : \mathbb{R}^2 \longrightarrow \mathbb{R}, \quad F(x, y) = x^2 + y^2 - x. \tag{4.3.64}$$

Then

$$\nabla F(x,y) = (2x-1, 2y), \tag{4.3.65}$$

which vanishes if and only if $x = 1/2,\ y = 0$. Hence Proposition 4.3.6 applies if and only if $(x_0, y_0) \neq (1/2, 0)$.

Let us give an example involving a real valued function on $M(n,\mathbb{R})$, namely

$$\det : M(n,\mathbb{R}) \longrightarrow \mathbb{R}. \tag{4.3.66}$$

As indicated in Exercise 15 of §4.1, if $\det X \neq 0$,

$$D\det(X)Y = (\det X)\,\mathrm{Tr}(X^{-1}Y), \tag{4.3.67}$$

so

$$\det X \neq 0 \Longrightarrow D\det(X) \neq 0. \tag{4.3.68}$$

We deduce that, if

$$X_0 \in M(n,\mathbb{R}), \quad \det X_0 = a \neq 0, \tag{4.3.69}$$

then, writing

$$X = (x_{jk})_{1\le j,k\le n}, \tag{4.3.70}$$

we can say that there exist $\mu, \nu \in \{1, \dots, n\}$ such that the equation

$$\det X = a \tag{4.3.71}$$

has a smooth solution of the form

$$x_{\mu\nu} = g\Big(x_{\alpha\beta} : (\alpha,\beta) \neq (\mu,\nu)\Big), \tag{4.3.72}$$

such that, if the argument of g consists of the matrix entries of X_0 other than the μ,ν entry, then the left side of (4.3.72) is the μ,ν entry of X_0.

Let us redo the determinant calculation, in case $n = 2$, taking

$$X = \begin{pmatrix} x & y \\ z & w \end{pmatrix}, \quad \det X = \Phi(x,y,z,w) = xw - yz. \tag{4.3.73}$$

We have

$$\nabla\Phi(x,y,z,w) = (w, -z, -y, x), \tag{4.3.74}$$

which is nonvanishing whenever $X \neq 0$. This is more precise than (4.3.68), which indeed we can improve in general. If we use Exercise 14 of §4.1 instead of Exercise 15, we get

$$D\det(X)Y = \mathrm{Tr}\big(\mathrm{Cof}(X)^t Y\big), \tag{4.3.75}$$

so in general we can strengthen (4.3.68) to

$$X \in M(n,\mathbb{R}),\ \mathrm{Cof}(X) \neq 0 \Longrightarrow D\det(X) \neq 0. \tag{4.3.76}$$

We return to the setting of Theorem 4.3.5, with ℓ not necessarily equal to 1. In notation parallel to that of (4.3.58), we assume F is a C^k map,

$$F : \Omega \longrightarrow \mathbb{R}^\ell, \quad F(z_0) = u_0, \tag{4.3.77}$$

where Ω is a neighborhood of z_0 in $\mathbb{R}^n$. We assume

$$DF(z_0) : \mathbb{R}^n \longrightarrow \mathbb{R}^\ell \ \text{ is surjective.} \tag{4.3.78}$$

Then, upon reordering the variables $z = (z_1, \dots, z_n)$, we can write $z = (x, y),\ x = (x_1, \dots, x_{n-\ell}),\ y = (y_1, \dots, y_\ell)$, such that $D_y F(z_0)$ is invertible, and Theorem 4.3.5 applies. Thus (for this reordering of variables), we have a C^k solution to

$$F(x, y) = u_0, \quad y = g(x, u_0), \tag{4.3.79}$$

satisfying $y_0 = g(x_0, u_0),\ z_0 = (x_0, y_0)$.

To give one example to which this result applies, we take another look at $F : \mathbb{R}^4 \to \mathbb{R}^2$ in (4.3.49). We have

$$DF(u, v, x, y) = \begin{pmatrix} 2xu & 2xv & u^2 + v^2 & 0 \\ x & y & u & v \end{pmatrix}. \tag{4.3.80}$$

The reader is invited to determine for which $(u, v, x, y) \in \mathbb{R}^4$ the matrix on the right side of (4.3.80) has rank 2. See Exercise 14 below.

Here is another example, involving a map defined on $M(n, \mathbb{R})$. Set

$$F : M(n, \mathbb{R}) \longrightarrow \mathbb{R}^2, \quad F(X) = \begin{pmatrix} \det X \\ \operatorname{Tr} X \end{pmatrix}. \tag{4.3.81}$$

Parallel to (4.3.67), if $\det X \neq 0,\ Y \in M(n, \mathbb{R})$,

$$DF(X)Y = \begin{pmatrix} (\det X) \operatorname{Tr}(X^{-1}Y) \\ \operatorname{Tr} Y \end{pmatrix}. \tag{4.3.82}$$

Hence, given $\det X \neq 0$, $DF(X) : M(n, \mathbb{R}) \to \mathbb{R}^2$ is surjective if and only if

$$L : M(n, \mathbb{R}) \to \mathbb{R}^2, \quad LY = \begin{pmatrix} \operatorname{Tr}(X^{-1}Y) \\ \operatorname{Tr} Y \end{pmatrix} \tag{4.3.83}$$

is surjective. This is seen to be the case if and only if X is not a scalar multiple of the identity $I \in M(n, \mathbb{R})$. See Exercise 15 below.

Exercises

1. Suppose $F : U \to \mathbb{R}^n$ is a C^2 map, $p \in U$, open in $\mathbb{R}^n$, and $DF(p)$ is invertible. With $q = F(p)$, define a map N on a neighborhood of p by

$$N(x) = x + DF(x)^{-1}\bigl(q - F(x)\bigr). \tag{4.3.84}$$

Show that there exists $\varepsilon > 0$ and $C < \infty$ such that, for $0 \le r < \varepsilon$,

$$\|x - p\| \le r \Longrightarrow \|N(x) - p\| \le C\, r^2.$$

Conclude that, if $\|x_1 - p\| \le r$ with $r < \min(\varepsilon, 1/2C)$, then $x_{j+1} = N(x_j)$ defines a sequence converging very rapidly to p. This is the basis of *Newton's method*, for solving $F(p) = q$ for p.
Hint. Apply F to both sides of (2.73).

2. Applying Newton's method to $f(x) = 1/x$, show that you get a fast approximation to division using only addition and multiplication.
Hint. Carry out the calculation of $N(x)$ in this case and notice a "miracle."

3. Identify $\mathbb{R}^2$ with $\mathbb{C}$ via $z = x + iy$, as in Exercise 9 of §4.1. Let $U \subset \mathbb{R}^2$ be open, $F : U \to \mathbb{R}^2$ be C^1. Assume $p \in U$, $DF(p)$ invertible. If $F^{-1} : V \to U$ is given as in Theorem 4.3.1, show that F^{-1} is holomorphic provided F is.

4. Let $\mathcal{O} \subset \mathbb{R}$ be open. We say a function $f \in C^\infty(\mathcal{O})$ is real analytic provided that, for each $x_0 \in \mathcal{O}$, we have a convergent power series expansion

$$f(x) = \sum_{\alpha \ge 0} \frac{1}{\alpha!} f^{(\alpha)}(x_0)(x - x_0)^\alpha, \tag{4.3.85}$$

valid in a neighborhood of x_0. Show that we can let x be complex in (4.3.85), and obtain an extension of f to a neighborhood of $\mathcal{O}$ in $\mathbb{C}$. Show that the extended function is holomorphic, i.e., satisfies the Cauchy-Riemann equations.
Hint. Use Proposition 4.2.9.
Remark. It can be shown that, conversely, any holomorphic function has a power series expansion. See [**17**]. For the next exercise, assume this as known.

5. Let $\mathcal{O} \subset \mathbb{R}$ be open, $p \in \mathcal{O}$, $f : \mathcal{O} \to \mathbb{R}$ be real analytic, with $Df(p)$ invertible. Take $f^{-1} : V \to U$ as in Theorem 4.3.1. Show f^{-1} is real analytic.
Hint. Consider a holomorphic extension $F : \Omega \to \mathbb{C}$ of f and apply Exercise 3.

6. Use (4.3.10) to show that if a C^1 diffeomorphism has a C^1 inverse G, and if actually F is C^k, then also G is C^k.
Hint. Use induction on k. Write (4.3.10) as

$$\mathcal{G}(x) = \Phi \circ \mathcal{F} \circ G(x),$$

with $\Phi(X) = X^{-1}$, as in Exercises 3 and 10 of §4.1, $\mathcal{G}(x) = DG(x), \mathcal{F}(x) = DF(x)$. Apply Exercise 9 of §4.1 to show that, in general

$$G, \mathcal{F}, \Phi \in C^\ell \Longrightarrow \mathcal{G} \in C^\ell.$$

Deduce that if one is given $F \in C^k$ and one knows that $G \in C^{k-1}$, then this result applies to give $\mathcal{G} = DG \in C^{k-1}$, hence $G \in C^k$.

7. Show that there is a neighborhood $\mathcal{O}$ of $(1, 0) \in \mathbb{R}^2$ and there are functions $u, v, w \in C^1(\mathcal{O})$ ($u = u(x, y)$, etc.) satisfying the equations

$$\begin{aligned} u^3 + v^3 - xw^3 &= 0, \\ u^2 + yw^2 + v &= 1, \\ xu + yvw &= 1, \end{aligned}$$

for $(x, y) \in \mathcal{O}$, and satisfying

$$u(1, 0) = 1, \quad v(1, 0) = 0, \quad w(1, 0) = 1.$$

Hint. Define $F : \mathbb{R}^5 \to \mathbb{R}^3$ by

$$F(u, v, w, x, y) = \begin{pmatrix} u^3 + v^3 - xw^3 \\ u^2 + yw^2 + v \\ xu + yvw \end{pmatrix},$$

Then $F(1,0,1,1,0) = (0,1,1)^t$. Evaluate the 3×3 matrix $D_{u,v,w}F(1,0,1,1,0)$. Compare (4.3.49)–(4.3.54).

8. Consider $F : M(n,\mathbb{R}) \to M(n,\mathbb{R})$, given by $F(X) = X^2$. Show that F is a diffeomorphism of a neighborhood of the identity matrix I onto a neighborhood of I. Show that F is *not* a diffeomorphism of a neighborhood of

$$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

onto a neighborhood of I (in case $n = 2$).

9. Prove Corollary 4.3.4.

10. Let $f : \mathbb{R}^2 \to \mathbb{R}^3$ be a C^1 map. Assume $f(0) = (0,0,0)$ and

$$\frac{\partial f}{\partial x}(0) \times \frac{\partial f}{\partial y}(0) = (0,0,1).$$

Show that there exist neighborhoods $\mathcal{O}$ and Ω of $0 \in \mathbb{R}^2$ and a C^1 map $u : \Omega \to \mathbb{R}$ such that the image of $\mathcal{O}$ under f in $\mathbb{R}^3$ is the graph of u over Ω.
Hint. Let $\Pi : \mathbb{R}^3 \to \mathbb{R}^2$ be $\Pi(x,y,z) = (x,y)$, and consider

$$\varphi(x,y) = \Pi(f(x,y)), \quad \varphi : \mathbb{R}^2 \to \mathbb{R}^2.$$

Show that $D\varphi(0) : \mathbb{R}^2 \to \mathbb{R}^2$ is invertible, and apply the inverse function theorem. Then let u be the z-component of $f \circ \varphi^{-1}$.

11. Generalize Exercise 10 to the setting where $f : \mathbb{R}^m \to \mathbb{R}^n$ ($m < n$) is C^1 and

$$Df(0) : \mathbb{R}^m \longrightarrow \mathbb{R}^n \ \text{ is injective.}$$

REMARK. For related results, see the opening paragraphs of §6.1.

12. Let $\Omega \subset \mathbb{R}^n$ be open and contain p_0. Assume $F : \overline{\Omega} \to \mathbb{R}^n$ is continuous and $F(p_0) = q_0$. Assume F is C^1 on Ω and $DF(x)$ is invertible for all $x \in \Omega$. Finally, assume there exists $R > 0$ such that

$$x \in \partial\Omega \Longrightarrow \|F(x) - q_0\| \geq R. \tag{4.3.86}$$

See Figure 4.3.4. Show that

$$F(\Omega) \supset B_{R/2}(q_0). \tag{4.3.87}$$

Hint. Given $y_0 \in B_{R/2}(q_0)$, use compactness to show that there exists $x_0 \in \overline{\Omega}$ such that

$$\|F(x_0) - y_0\| = \inf_{x \in \overline{\Omega}} \|F(x) - y_0\|.$$

Use the hypothesis (4.3.86) to show that $x_0 \in \Omega$. If $F(x_0) \neq y_0$, use

$$F(x_0 + tz) = F(x_0) + tDF(x_0)z + o(\|tz\|),$$

to produce $z \in \mathbb{R}^n$ (say $DF(x_0)z = y_0 - F(x_0)$) such that $F(x_0 + tz)$ is closer to y_0 than $F(x_0)$ is, for small $t > 0$. Contradiction.

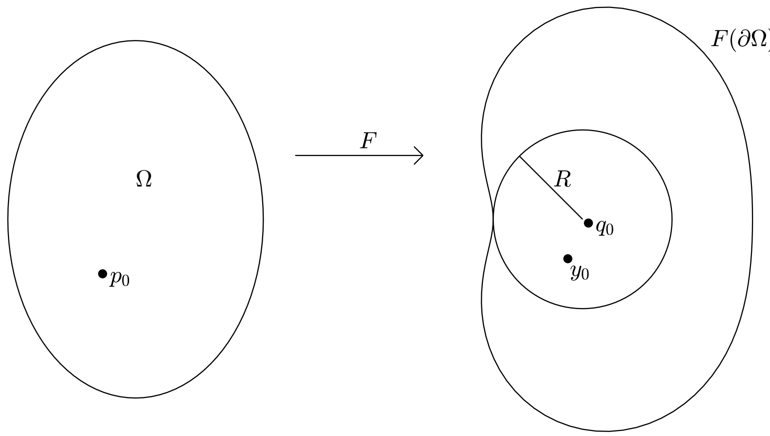

Figure 4.3.4. $F(\Omega)$ contains $B_R(q_0)$

13. Do Exercise 12 with the conclusion (4.3.87) strengthened to

$$F(\Omega) \supset B_R(q_0). \tag{4.3.88}$$

Hint. It suffices to show that $F(\Omega) \supset B_S(q_0)$ for each $S < R$. Given such S, produce a diffeomorphism $\varphi : \mathbb{R}^n \to \mathbb{R}^n$ such that Exercise 12 applies to $\varphi \circ F$, and yields the desired conclusion.

14. In the setting of Exercise 12, take

$$\Omega = B_r(p_0),$$

and assume there exists $a > 0$ such that

$$v \cdot DF(x)v \geq a\|v\|^2, \quad \forall\, x \in \Omega,\ v \in \mathbb{R}^n.$$

As before, F is assumed continuous on $\overline{\Omega}$, C^1 on Ω, and $F(p_0) = q_0$. Adapt the proof of Proposition 4.3.2 to show that, for $x, y \in \Omega$,

$$(x - y) \cdot [F(x) - F(y)] \geq a\|x - y\|^2,$$

hence

$$\|F(x) - F(y)\| \geq a\|x - y\|.$$

Deduce that

$$x \in \partial\Omega \Longrightarrow \|F(x) - q_0\| \geq ar,$$

and conclude that F maps Ω one-one and onto a set

$$F(\Omega) \supset B_{ar}(q_0).$$

15. Show that the 2×4 matrix D in (4.3.80) has rank 2 whenever $v \neq 0$. In case $v = 0$, the matrix becomes

$$D = \begin{pmatrix} 2xu & 0 & u^2 & 0 \\ x & y & u & 0 \end{pmatrix}.$$

Determine when this has rank 2.

16. Let $u, v \in \mathbb{R}^n$, and define

$$L : \mathbb{R}^m \to \mathbb{R}^2, \quad Ly = \begin{pmatrix} u \cdot y \\ v \cdot y \end{pmatrix}.$$

Show that L is surjective if and only if u and v are linearly independent. Relate this to the analysis of (4.3.83).
Hint. Use $M(n, \mathbb{R}) \approx \mathbb{R}^{n^2}$, with inner product $\langle S, T \rangle = \operatorname{Tr} S^t T$. Write L in (4.3.83) as

$$LY = \begin{pmatrix} \langle U, Y \rangle \\ \langle V, Y \rangle \end{pmatrix}, \quad U = (X^{-1})^t, \ V = I.$$

Chapter 5

Multivariable integral calculus

This central chapter develops integral calculus on domains in $\mathbb{R}^n$, taking up the multidimensional Riemann integral. The basic definition is quite parallel to the one-dimensional case, but a number of fundamental results, while parallel in statement to the one-dimensional case, require more elaborate demonstrations in higher dimensions. This chapter is one of the most demanding in this text, and it is in a sense the heart of the course.

We start in §5.1 with the integral of a function defined on a *cell* in $\mathbb{R}^n$, i.e., a product of n intervals. This is done via partitions of a cell R, and a passage to the limit as the partitions become finer. When the limit of upper and lower sums of a bounded function $f : R \to \mathbb{R}$ exist and coincide, we say f is Riemann integrable on R, and take the limit to be its integral. Continuous functions on R are seen to be integrable. If $\Omega \subset \mathbb{R}^n$ is a more general bounded set, and $f : \Omega \to \mathbb{R}$, we take a cell R such that $\Omega \subset R$ and extend f by 0 on $R \setminus \Omega$. Such a construction makes it crucial to examine the integrability of discontinuous functions on R. We show that if $f : R \to \mathbb{R}$ is bounded and the set of points of discontinuity of f is negligible in some sense, then f is Riemann integrable. One simple result along these lines is given in §5.1, and a sharper result along these lines is given later in the chapter.

One central result of §5.1 is the reduction of multiple integrals to iterated integrals. This reduction is essential for computations, and we illustrate it with a variety of examples. Another central result is the change of variable formula for multidimensional integrals. Important special cases include transforming 2D integrals to polar coordinates, and 3D integrals to spherical polar coordinates. We illustrate the use of these two results with computations of volumes of balls in $\mathbb{R}^n$, particularly for $n = 2$ and 3, but also for higher n. We will find in the next chapter that the change of variable formula for the integral is a very important ingredient for developing the integral on surfaces.

In §5.2, we apply methods developed in §5.1 to study mean values of functions defined on a contented domain $\mathcal{O} \subset \mathbb{R}^n$, with emphasis on results on the center of mass of $\mathcal{O}$.

Extending the scope of §5.1, we treat unbounded integrable functions in §5.3. A key result established here is a *monotone convergence theorem*.

In §5.4 we introduce the concept of outer measure and sharpen the integrability condition of §5.1, showing that a sufficient condition that a bounded function be Riemann integrable is that its set of points of discontinuity have outer measure zero.

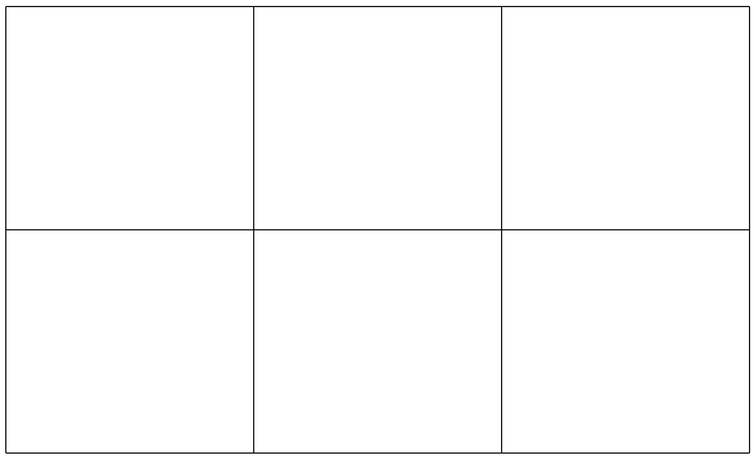

Figure 5.1.1. Partition of a cell, $\mathcal{P}_1$

5.1. The Riemann integral in n variables

We define the Riemann integral of a bounded function $f : R \to \mathbb{R}$, where $R \subset \mathbb{R}^n$ is a cell, i.e., a product of intervals $R = I_1 \times \cdots \times I_n$, where $I_\nu = [a_\nu, b_\nu]$ are intervals in $\mathbb{R}$. Recall that a partition of an interval $I = [a, b]$ is a finite collection of subintervals $\{J_k : 0 \le k \le N\}$, disjoint except for their endpoints, whose union is I. We can take $J_k = [x_k, x_{k+1}]$, where

$$a = x_0 < x_1 < \cdots < x_N < x_{N+1} = b. \tag{5.1.1}$$

Now, if one has a partition of each I_ν into $J_{\nu 1} \cup \cdots \cup J_{\nu, N(\nu)}$, then a partition $\mathcal{P}$ of R consists of the cells

$$R_\alpha = J_{1\alpha_1} \times J_{2\alpha_2} \times \cdots \times J_{n\alpha_n}, \tag{5.1.2}$$

where $0 \le \alpha_\nu \le N(\nu)$. See Figure 5.1.1. For such a partition, define

$$\operatorname{maxsize}(\mathcal{P}) = \max_\alpha \operatorname{diam} R_\alpha, \tag{5.1.3}$$

where $(\operatorname{diam} R_\alpha)^2 = \ell(J_{1\alpha_1})^2 + \cdots + \ell(J_{n\alpha_n})^2$. Here, $\ell(J)$ denotes the length of an interval J. Each cell has n-dimensional volume

$$V(R_\alpha) = \ell(J_{1\alpha_1}) \cdots \ell(J_{n\alpha_n}). \tag{5.1.4}$$

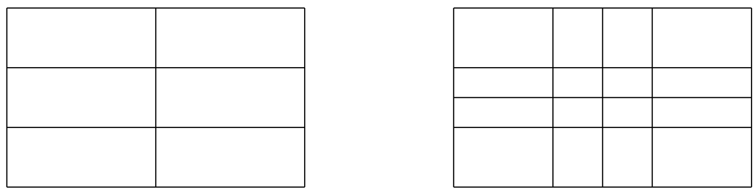

Figure 5.1.2. Second partition, $\mathcal{P}_2$, and common refinement, $\mathcal{Q} \succ \mathcal{P}_j$

Sometimes we use $V_n(R_\alpha)$ for emphasis on the dimension. We also use $A(R)$ for $V_2(R)$, and, of course, $\ell(R)$ for $V_1(R)$.

We set

$$\overline{I}_{\mathcal{P}}(f) = \sum_\alpha \sup_{R_\alpha} f(x)\, V(R_\alpha), \tag{5.1.5}$$
$$\underline{I}_{\mathcal{P}}(f) = \sum_\alpha \inf_{R_\alpha} f(x)\, V(R_\alpha).$$

Note that $\underline{I}_{\mathcal{P}}(f) \le \overline{I}_{\mathcal{P}}(f)$. These quantities should approximate the Riemann integral of f, if the partition $\mathcal{P}$ is sufficiently "fine."

To be more precise, if $\mathcal{P}$ and $\mathcal{Q}$ are two partitions of R, we say $\mathcal{Q}$ refines $\mathcal{P}$, and write $\mathcal{Q} \succ \mathcal{P}$, if each partition of each interval factor I_ν of R involved in the definition of $\mathcal{P}$ is further refined in order to produce the partitions of the factors I_ν, used to define $\mathcal{Q}$, via (5.1.2). It is an exercise to show that any two partitions of R have a common refinement. See Figure 5.1.2. Note also that

$$\mathcal{Q} \succ \mathcal{P} \Longrightarrow \overline{I}_{\mathcal{Q}}(f) \le \overline{I}_{\mathcal{P}}(f), \text{ and } \underline{I}_{\mathcal{Q}}(f) \ge \underline{I}_{\mathcal{P}}(f). \tag{5.1.6}$$

Consequently, if $\mathcal{P}_j$ are any two partitions of R and $\mathcal{Q}$ is a common refinement, we have

$$\underline{I}_{\mathcal{P}_1}(f) \le \underline{I}_{\mathcal{Q}}(f) \le \overline{I}_{\mathcal{Q}}(f) \le \overline{I}_{\mathcal{P}_2}(f). \tag{5.1.7}$$

Now, whenever $f : R \to \mathbb{R}$ is bounded, the following quantities are well defined:

$$\overline{I}(f) = \inf_{\mathcal{P}\in\Pi(R)} \overline{I}_{\mathcal{P}}(f), \quad \underline{I}(f) = \sup_{\mathcal{P}\in\Pi(R)} \underline{I}_{\mathcal{P}}(f), \tag{5.1.8}$$

where $\Pi(R)$ is the set of all partitions of R, as defined above. Clearly, by (5.1.7), $\underline{I}(f) \le \overline{I}(f)$. We then say that f is Riemann integrable (on R) provided $\overline{I}(f) = \underline{I}(f)$, and in such a case, we set

$$\int_R f(x)\, dV(x) = \overline{I}(f) = \underline{I}(f). \tag{5.1.9}$$

We will denote the set of Riemann integrable functions on R by $\mathcal{R}(R)$. If $\dim R = 2$, we will often use $dA(x)$ instead of $dV(x)$ in (5.1.9). For general n, we might also use simply dx.

We derive some basic properties of the Riemann integral. First, the proof of the Darboux theorem in §1.2 readily extends, to give:

Proposition 5.1.1. *Let $\mathcal{P}_\nu$ be any sequence of partitions of R such that*

$$\text{maxsize}\,(\mathcal{P}_\nu) = \delta_\nu \to 0. \tag{5.1.10}$$

Then, if $f : R \to \mathbb{R}$ is bounded,

$$\overline{I}_{\mathcal{P}_\nu}(f) \to \overline{I}(f), \quad \text{and} \quad \underline{I}_{\mathcal{P}_\nu}(f) \to \underline{I}(f). \tag{5.1.11}$$

Consequently, if $\xi_{\nu\alpha}$ is any choice of one point in each cell $R_{\nu\alpha}$ in the partition $\mathcal{P}_\nu$, then, whenever $f \in \mathcal{R}(R)$,

$$\int_R f(x)\, dV(x) = \lim_{\nu\to\infty} \sum_\alpha f(\xi_{\nu\alpha})\, V(R_{\nu\alpha}). \tag{5.1.12}$$

This is the multidimensional Darboux theorem. The sums that arise in (5.1.12) are Riemann sums.

Also, we can extend the proof of additivity of the integral in §1.1, to obtain:

Proposition 5.1.2. *If $f_j \in \mathcal{R}(R)$ and $c_j \in \mathbb{R}$, then $c_1 f_1 + c_2 f_2 \in \mathcal{R}(R)$, and*

$$\int_R (c_1 f_1 + c_2 f_2)\, dV = c_1 \int_R f_1\, dV + c_2 \int_R f_2\, dV. \tag{5.1.13}$$

Next, we establish an integrability result analogous to Proposition 1.2.2.

Proposition 5.1.3. *If f is continuous on R, then $f \in \mathcal{R}(R)$.*

Proof. As in the proof of Proposition 1.2.2, we have that,

$$\text{maxsize}\,(\mathcal{P}) \le \delta \Longrightarrow \overline{I}_{\mathcal{P}}(f) - \underline{I}_{\mathcal{P}}(f) \le \omega(\delta) \cdot V(R), \tag{5.1.14}$$

where $\omega(\delta)$ is a modulus of continuity for f on R. This proves the proposition. □

Content, volume, and integrability

When the number of variables exceeds one, it becomes crucial to identify some nice classes of *discontinuous* functions on R that are Riemann integrable. A useful tool for this is the following notion of size of a set $S \subset R$, called *content*. Extending the notion from §1.2, we define "upper content" cont^+ and "lower content" cont^- by

$$\text{cont}^+(S) = \overline{I}(\chi_S), \quad \text{cont}^-(S) = \underline{I}(\chi_S), \tag{5.1.15}$$

where χ_S is the characteristic function of S. We say S has content, or "is contented," if these quantities are equal, which happens if and only if $\chi_S \in \mathcal{R}(R)$, in which case the common value of $\text{cont}^+(S)$ and $\text{cont}^-(S)$ is

$$V(S) = \int_R \chi_S(x)\, dV(s), \tag{5.1.16}$$

which we call the volume of S. For $S \subset \mathbb{R}^n$, we might denote this by $V_n(S)$, to emphasize the dimension. When $n = 2$, we might denote this quantity by $A(S)$, and call is the *area* of S. We mention that, if $S = I_1 \times \cdots \times I_n$ is a cell, it is readily verified that the definitions in (5.1.5), (5.1.8), and (5.1.15) yield

$$\text{cont}^+(S) = \text{cont}^-(S) = \ell(I_1) \cdots \ell(I_n),$$

so the definition of $V(S)$ given by (5.1.16) is consistent with that given in (5.1.4).

An equivalent characterization of upper content is

$$\text{cont}^+(S) = \inf \Big\{ \sum_{k=1}^N V(R_k) : S \subset R_1 \cup \cdots \cup R_N \Big\}, \tag{5.1.17}$$

where R_k are cells contained in R. In a literal translation of (5.1.15) the R_α in (5.1.17) should be part of a partition $\mathcal{P}$ of R, as defined above, but if $\{R_1, \dots, R_N\}$ are any cells in R, they can be chopped up into smaller cells, some perhaps thrown away, to yield a finite cover of S by cells in a partition of R, so one gets the same result.

It is an exercise to see that, for any set $S \subset R$,

$$\text{cont}^+(S) = \text{cont}^+(\overline{S}), \tag{5.1.18}$$

where $\overline{S}$ is the closure of S.

We note that, generally, for a bounded function f on R,

$$\underline{I}(f) + \overline{I}(1 - f) = V(R). \tag{5.1.19}$$

This follows directly from (5.1.5). In particular, given $S \subset R$,

$$\text{cont}^-(S) + \text{cont}^+(R \setminus S) = V(R). \tag{5.1.20}$$

Using this together with (5.1.18), with S and $R \setminus S$ switched, we have

$$\text{cont}^-(S) = \text{cont}^-(\overset{\circ}{S}), \tag{5.1.21}$$

where $\overset{\circ}{S}$ denotes the interior of S. The difference $\overline{S} \setminus \overset{\circ}{S}$ is called the boundary of S, and denoted bS.

Note that if $S \subset R$, and $\mathcal{P}$ is a partition of R, we can classify each cell in $\mathcal{P}$ as either contained in $\overset{\circ}{S}$, intersecting bS, or disjoint from $\overline{S}$. It follows that

$$\overline{I}_{\mathcal{P}}(\chi_{\overline{S}}) = \underline{I}_{\mathcal{P}}(\chi_{\overset{\circ}{S}}) + \overline{I}_{\mathcal{P}}(\chi_{bS}). \tag{5.1.22}$$

Taking partitions $\mathcal{P} = \mathcal{P}_\nu$ with maxsize $\to 0$ and applying the Darboux theorem, we obtain in the limit that

$$\text{cont}^+(\overline{S}) = \text{cont}^-(\overset{\circ}{S}) + \text{cont}^+(bS). \tag{5.1.23}$$

Taking into account (5.1.18) and (5.1.21), we have:

Proposition 5.1.4. *If $S \subset R$, then S is contented if and only if cont$^+(bS) = 0$.*

If a set $\Sigma \subset R$ has the property that $\text{cont}^+(\Sigma) = 0$, we say that Σ has content zero, or is a *nil* set. Clearly Σ is nil if and only if $\overline{\Sigma}$ is nil. It follows easily from Proposition 5.1.2 that, if Σ_j are nil, $1 \le j \le K$, then $\bigcup_{j=1}^K \Sigma_j$ is nil.

If $S_1, S_2 \subset R$ and $S = S_1 \cup S_2$, then $\overline{S} = \overline{S}_1 \cup \overline{S}_2$ and $\overset{\circ}{S} \supset \overset{\circ}{S}_1 \cup \overset{\circ}{S}_2$. Hence $bS \subset b(S_1) \cup b(S_2)$. It follows then from Proposition 5.1.4 that, if S_1 and S_2 are contented, so is $S_1 \cup S_2$. Clearly, if S_j are contented, so are $S_j^c = R \setminus S_j$. It follows that, if S_1 and S_2 are contented, so is $S_1 \cap S_2 = \big(S_1^c \cup S_2^c\big)^c$. A family $\mathcal{F}$ of subsets of R is called an *algebra* of subsets of R provided the following conditions hold:

$$\begin{aligned} &R \in \mathcal{F}, \\ &S_j \in \mathcal{F} \Rightarrow S_1 \cup S_2 \in \mathcal{F}, \text{ and} \\ &S \in \mathcal{F} \Rightarrow R \setminus S \in \mathcal{F}. \end{aligned}$$

Algebras of sets are automatically closed under finite intersections also. We see that:

Proposition 5.1.5. *The family of contented subsets of R is an algebra of sets.*

The following result specifies a useful class of Riemann integrable functions. For a sharper result, see Proposition 5.4.1.

Proposition 5.1.6. *If $f : R \to \mathbb{R}$ is bounded and the set S of points of discontinuity of f is a nil set, then $f \in \mathcal{R}(R)$.*

Proof. Suppose $|f| \le M$ on R, and take $\varepsilon > 0$. Take a partition $\mathcal{P}$ of R, and write $\mathcal{P} = \mathcal{P}' \cup \mathcal{P}''$, where cells in $\mathcal{P}'$ do not meet $\overline{S}$, and cells in $\mathcal{P}''$ do intersect $\overline{S}$. Since $\text{cont}^+(\overline{S}) = 0$, we can pick $\mathcal{P}$ so that the cells in $\mathcal{P}''$ have total volume $\le \varepsilon$. Now f is continuous on each cell in $\mathcal{P}'$. Further refining the partition if necessary, we can assume that f varies by $\le \varepsilon$ on each cell in $\mathcal{P}'$. Thus

$$\overline{I}_{\mathcal{P}}(f) - \underline{I}_{\mathcal{P}}(f) \le \big[V(R) + 2M\big]\varepsilon. \tag{5.1.24}$$

This proves the proposition. □

To give an example, suppose $K \subset R$ is a closed set such that bK is nil. Let $f : K \to \mathbb{R}$ be continuous. Define $\widetilde{f} : R \to \mathbb{R}$ by

$$\begin{aligned} \widetilde{f}(x) = f(x) &\quad \text{for} \quad x \in K, \\ 0 &\quad \text{for} \quad x \in R \setminus K. \end{aligned} \tag{5.1.25}$$

Then the set of points of discontinuity of $\widetilde{f}$ is contained in bK. Hence $\widetilde{f} \in \mathcal{R}(R)$. We set

$$\int_K f \, dV = \int_R \widetilde{f} \, dV. \tag{5.1.26}$$

In connection with this, we note the following fact, whose proof is an exercise. Suppose R and $\widetilde{R}$ are cells, with $R \subset \widetilde{R}$. Suppose that $g \in \mathcal{R}(R)$ and that $\tilde{g}$ is defined on $\widetilde{R}$, to be equal to g on R and to be 0 on $\widetilde{R} \setminus R$. Then

$$\tilde{g} \in \mathcal{R}(\widetilde{R}), \quad \text{and} \quad \int_R g\, dV = \int_{\widetilde{R}} \tilde{g}\, dV. \tag{5.1.27}$$

This can be shown by an argument involving refining any given pair of partitions of R and $\widetilde{R}$, respectively, to a pair of partitions $\mathcal{P}_R$ and $\mathcal{P}_{\widetilde{R}}$ with the property that each cell in $\mathcal{P}_R$ is a cell in $\mathcal{P}_{\widetilde{R}}$.

The following describes an important class of sets $S \subset \mathbb{R}^n$ that have content zero.

Proposition 5.1.7. *Let $\Sigma \subset \mathbb{R}^{n-1}$ be a closed bounded set and let $g : \Sigma \to \mathbb{R}$ be continuous. Then the graph of g,*

$$\mathfrak{G} = \big\{ \big(x, g(x)\big) : x \in \Sigma \big\}$$

is a nil subset of $\mathbb{R}^n$.

Proof. Put Σ in a cell $R_0 \subset \mathbb{R}^{n-1}$. Suppose $|g| \le M$ on Σ. Take $N \in \mathbb{Z}^+$ and set $\varepsilon = M/N$. Pick a partition $\mathcal{P}_0$ of R_0, sufficiently fine that g varies by at most ε on each set $\Sigma \cap R_\alpha$, for any cell $R_\alpha \in \mathcal{P}_0$. Partition the interval $I = [-M, M]$ into $2N$ equal intervals J_ν, of length ε. Then $\{R_\alpha \times J_\nu\} = \{Q_{\alpha\nu}\}$ forms a partition of $R_0 \times I$. Now, over each cell $R_\alpha \in \mathcal{P}_0$, there lie at most 2 cells $Q_{\alpha\nu}$ which intersect $\mathfrak{G}$, so $\text{cont}^+(\mathfrak{G}) \le 2\varepsilon \cdot V(R_0)$. Letting $N \to \infty$, we have the proposition. ☐

Similarly, for any $j \in \{1, \dots, n\}$, the graph of x_j as a continuous function of the complementary variables is a nil set in $\mathbb{R}^n$. So are finite unions of such graphs. Such sets arise as boundaries of many ordinary-looking regions in $\mathbb{R}^n$.

Here is a further class of nil sets.

Proposition 5.1.8. *Let $\mathcal{O} \subset \mathbb{R}^n$ be open and let $S \subset \mathcal{O}$ be a compact nil subset. Assume $f : \mathcal{O} \to \mathbb{R}^n$ is a Lipschitz map. Then $f(S)$ is a nil subset of $\mathbb{R}^n$.*

Proof. The Lipschitz hypothesis on f is that there exists $L < \infty$ such that, for $p, q \in \mathcal{O}$,

$$|f(p) - f(q)| \le L|p - q|.$$

If we cover S with k cells (in a partition), of total volume $\le \alpha$, each cubical with edgesize δ, then $f(S)$ is covered by k sets of diameter $\le L\sqrt{n}\delta$, hence it can be covered by k cubical cells of edgesize $L\sqrt{n}\delta$, having total volume $\le (L\sqrt{n})^n\alpha$. From this we have the (not very sharp) general bound

$$\text{cont}^+\big(f(S)\big) \le (L\sqrt{n})^n \ \text{cont}^+(S), \tag{5.1.28}$$

which proves the proposition. ☐

Iterated integrals

In evaluating n-dimensional integrals, it is usually convenient to reduce them to *iterated integrals*. Such results go under the label of Fubini Theorems. Here is one simple example of such a result.

Proposition 5.1.9. *Let $X \subset \mathbb{R}^k$, $Y \subset \mathbb{R}^\ell$, $R = X \times Y \subset \mathbb{R}^n$ be cells, $k + \ell = n$. Let $f : X \times Y \to \mathbb{R}$ be continuous. Form*

$$\varphi(x) = \int\limits_Y f(x, y)\, dV_\ell(y). \tag{5.1.29}$$

Then $\varphi \in C(X)$, and

$$\int\limits_R f\, dV_n = \int\limits_X \varphi\, dV_k. \tag{5.1.30}$$

Proof. We know f is uniformly continuous on R. Let $\omega(\delta)$ be a modulus of continuity. If $x_j \in X$, we have

$$\begin{aligned} |\varphi(x_1) - \varphi(x_2)| &\le \int\limits_Y |f(x_1, y) - f(x_2, y)|\, dV_\ell(y) \\ &\le \omega(|x_1 - x_2|) V_\ell(Y), \end{aligned} \tag{5.1.31}$$

so φ is continuous on X.

To proceed, take $\varepsilon > 0$ and pick $\delta > 0$ such that $\omega(\delta) < \varepsilon$. Take partitions $\mathcal{X} = \{X_\alpha\}$ of X, $\mathcal{Y} = \{Y_\beta\}$ of Y, and $\mathcal{P} = \{R_{\alpha\beta} = X_\alpha \times Y_\beta\}$ of $R = X \times Y$ into cells, such that

$$\text{maxsize}(\mathcal{P}) \le \delta. \tag{5.1.32}$$

The same upper bound holds for maxsize$(\mathcal{X})$ and maxsize$(\mathcal{Y})$.

Let $\xi_\alpha \in X_\alpha$, $\eta_\beta \in Y_\beta$, and $\zeta_{\alpha\beta} = (\xi_\alpha, \eta_\beta) \in R_{\alpha\beta}$ denote the centers of these cells. We have

$$\Bigl| \int\limits_R f\, dV_n - \sum_{\alpha,\beta} f(\xi_\alpha, \eta_\beta)\, V_n(R_{\alpha\beta}) \Bigr| \le V_n(R)\varepsilon. \tag{5.1.33}$$

Also, for each α,

$$\Bigl| \varphi(\xi_\alpha) - \sum_\beta f(\xi_\alpha, \eta_\beta)\, V_\ell(Y_\beta) \Bigr| \le V_\ell(Y)\varepsilon. \tag{5.1.34}$$

Furthermore, by (5.1.31),

$$\Bigl| \int\limits_X \varphi\, dV_k - \sum_\alpha \varphi(\xi_\alpha)\, V_k(X_\alpha) \Bigr| \le V_k(X) V_\ell(Y)\varepsilon. \tag{5.1.35}$$

From (5.1.34)–(5.1.37), we have

$$\Bigl| \int\limits_X \varphi\, dV_k - \sum_{\alpha,\beta} f(\xi_\alpha, \eta_\beta)\, V(X_\alpha) V(Y_\beta) \Bigr| \le 2V_n(R)\varepsilon. \tag{5.1.36}$$

Comparison with (5.1.33) gives

$$\Big|\int_R f\,dV_n - \int_X \varphi\,dV_k\Big| \le 3V(R)\varepsilon. \tag{5.1.37}$$

Taking $\varepsilon \to 0$ gives the asserted identity (5.1.30). □

For applications, it is crucial to obtain results parallel to (5.1.30) in cases where f is not continuous on $X \times Y$. Here is a useful result of this nature.

Theorem 5.1.10. *Let $\Sigma \subset \mathbb{R}^{n-1}$ be a closed, bounded contented set and let $g_j : \Sigma \to \mathbb{R}$ be continuous, with $g_0(x) < g_1(x)$ on Σ. Take*

$$\Omega = \big\{(x,y) \in \mathbb{R}^n : x \in \Sigma,\ g_0(x) \le y \le g_1(x)\big\}. \tag{5.1.38}$$

Then Ω is a contented set in $\mathbb{R}^n$. If $f : \Omega \to \mathbb{R}$ is continuous, then

$$\varphi(x) = \int_{g_0(x)}^{g_1(x)} f(x,y)\,dy \tag{5.1.39}$$

is continuous on Σ, and

$$\int_\Omega f\,dV_n = \int_\Sigma \varphi\,dV_{n-1}, \tag{5.1.40}$$

i.e.,

$$\int_\Omega f\,dV_n = \int_\Sigma \left(\int_{g_0(x)}^{g_1(x)} f(x,y)\,dy\right) dV_{n-1}(x). \tag{5.1.41}$$

Proof. The continuity of φ in (5.1.39) is straightforward.

Put Σ in a cell $R \subset \mathbb{R}^{n-1}$. If $A \le g_0 < g_1 \le B$, set $I = [A,B]$. Then Ω is contained in the cell $Q = R \times I \subset \mathbb{R}^n$. We will work with a partition $\mathcal{P} = \{R_\alpha\}$ of R, with properties to be specified shortly.

Let us note that

$$\text{cont}^+(b\Omega) = 0 \quad \text{and} \quad \text{cont}^+(b\Sigma) = 0. \tag{5.1.42}$$

Since the sets of discontinuities of f (extended by 0 on $Q \setminus \Omega$) and of φ (extended by 0 on $R \setminus \Sigma$) lie in these boundaries, it follows that f and φ are Riemann integrable, so both sides of (5.1.40) are well defined. We will prove (5.1.40) by chopping Q into pieces. We will apply Proposition 5.1.9 to the bulk of these pieces, and show that the contribution of the remaining pieces is vanishingly small.

Let $\omega(\delta)$ be a modulus of continuity for g_0 and g_1. Pick $\varepsilon > 0$. Then pick $\delta_1 > 0$ and the partition $\mathcal{P}$ so that

$$\omega(\delta_1) \le \varepsilon, \quad \text{and} \quad \text{maxsize}(\mathcal{P}) \le \delta_1. \tag{5.1.43}$$

If necessary, shrink δ_1 to be sufficiently small that all the cells in $\mathcal{P}$ that intersect $b\Sigma$ have

$$\text{total } (n-1)\text{-dimensional volume} \le \varepsilon. \tag{5.1.44}$$

Denote by $\mathcal{P}'$ the collection of cells in $\mathcal{P}$ that lie in the interior $\overset{\circ}{\Sigma}$ of Σ. We have

$$\Big|\int_{\Omega} f\,dV_n - \sum_{R_\alpha\in\mathcal{P}'}\int_{R_\alpha\times I} f\,dV_n\Big| \le M(B-A)\varepsilon, \tag{5.1.45}$$

where

$$M = \max_{\Omega} |f|. \tag{5.1.46}$$

We also have

$$\Big|\int_{\Sigma} \varphi\,dV_{n-1} - \sum_{R_\alpha\in\mathcal{P}'}\int_{R_\alpha} \varphi\,dV_{n-1}\Big| \le M(B-A)\varepsilon, \tag{5.1.47}$$

since $|\varphi| \le M(B-A)$.

Now, for each $R_\alpha \in \mathcal{P}'$, take

$$A_\alpha = \max_{R_\alpha} g_0, \quad B_\alpha = \min_{R_\alpha} g_1, \quad I_\alpha = [A_\alpha, B_\alpha]. \tag{5.1.48}$$

Assume ε is so small that each $A_\alpha < B_\alpha$. Then, for each $R_\alpha \in \mathcal{P}'$,

$$\Big|\int_{R_\alpha\times I} f\,dV_n - \int_{R_\alpha\times I_\alpha} f\,dV_n\Big| \le 2M\varepsilon V_{n-1}(R_\alpha), \tag{5.1.49}$$

and

$$x\in R_\alpha \Longrightarrow \Big|\varphi(x) - \int_{A_\alpha}^{B_\alpha} f(x,y)\,dy\Big| \le 2M\varepsilon. \tag{5.1.50}$$

Now Proposition 5.1.9 yields

$$\int_{R_\alpha\times I_\alpha} f\,dV_n = \int_{R_\alpha}\Big(\int_{A_\alpha}^{B_\alpha} f(x,y)\,dy\Big)\,dV_{n-1}(x), \tag{5.1.51}$$

so, by (5.1.50),

$$\Big|\int_{R_\alpha\times I_\alpha} f\,dV_n - \int_{R_\alpha} \varphi\,dV_{n-1}\Big| \le 2M\varepsilon V_{n-1}(R_\alpha), \tag{5.1.52}$$

and hence, taking into account (5.1.49),

$$\Big|\sum_{R_\alpha\in\mathcal{P}'}\Big(\int_{R_\alpha\times I} f\,dV_n - \int_{R_\alpha} \varphi\,dV_{n-1}\Big)\Big| \le 4M\varepsilon V(R). \tag{5.1.53}$$

Therefore, also bringing in (5.1.44), we have

$$\Big|\int_{\Omega} f\,dV_n - \int_{\Sigma} \varphi\,dV_{n-1}\Big| \le K\varepsilon. \tag{5.1.54}$$

Taking $\varepsilon \to 0$ yields the asserted identity (5.1.39). □

Remark. A little more work allows one to replace the hypothesis $g_0(x) < g_1(x)$ in Theorem 5.1.10 by $g_0(x) \le g_1(x)$. We leave this as a task for the reader.

To present some applications of Theorem 5.1.10, we take the unit disk in $\mathbb{R}^2$,

$$D = \{(x, y) \in \mathbb{R}^2 : x \in [-1, 1],\ |y| \le \sqrt{1 - x^2}\}. \tag{5.1.55}$$

In this case we see from Theorem 5.1.10 that, if $f \in C(D)$,

$$\int_D f\, dA = \int_{-1}^{1} \Bigl(\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} f(x, y)\, dy \Bigr)\, dx. \tag{5.1.56}$$

In particular, the area of D is given by

$$A(D) = \int_{-1}^{1} 2\sqrt{1 - x^2}\, dx. \tag{5.1.57}$$

The change of variable $x = \sin t$ gives

$$A(D) = 2 \int_{-\pi/2}^{\pi/2} \cos^2 t\, dt. \tag{5.1.58}$$

Using the identity $\cos 2t = 2\cos^2 t - 1$, we obtain

$$A(D) = \int_{-\pi/2}^{\pi/2} (1 + \cos 2t)\, dt = \pi, \tag{5.1.59}$$

as the formula for the area of the unit disk $D \subset \mathbb{R}^2$. See Exercise 23 for another approach.

Extending the last computation, we highlight the general application of Theorem 5.1.10 to the computation of areas and volumes.

Corollary 5.1.11. *Take $\Sigma \subset \mathbb{R}^{n-1}$, $g_j : \Sigma \to \mathbb{R}$, and $\Omega \subset \mathbb{R}^n$ as in Theorem 5.1.10. Then*

$$V(\Omega) = \int_\Sigma [g_1(x) - g_0(x)]\, dx. \tag{5.1.60}$$

Specializing to $n = 2$ (as we did in (5.1.55)–(5.1.59)), we have for the area under the curve $y = g(x)$, for continuous $g : [a, b] \to (0, \infty)$, the formula

$$A = \int_a^b g(x)\, dx, \tag{5.1.61}$$

which is familiar from first-year calculus (though, in such a course, one might have seen a less precise definition of area).

Let us move on to higher dimensional volume, such as the volume of the n-dimensional ball:

$$B^n = \{x \in \mathbb{R}^n : |x| \le 1\}. \tag{5.1.62}$$

We can apply (5.1.60) to write

$$V(B^n) = 2 \int_{B^{n-1}} \sqrt{1 - |x|^2}\, dx. \tag{5.1.63}$$

For example,

$$V(B^3) = 2 \int_D \sqrt{1 - |x|^2}\, dx, \tag{5.1.64}$$

where $D = B^2$ is the unit disk. In turn, an application of Theorem 5.1.10 gives

$$\int_D \sqrt{1-|x|^2}\,dx = \int_{-1}^{1}\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \sqrt{1-x^2-y^2}\,dy\,dx. \tag{5.1.65}$$

Taking $a^2 = 1 - x^2$, we write the inner integral as

$$\begin{aligned}\int_{-a}^{a} \sqrt{a^2-y^2}\,dy &= a^2 \int_{-1}^{1} \sqrt{1-s^2}\,ds \\ &= \frac{\pi}{2}a^2,\end{aligned} \tag{5.1.66}$$

using $y = as$ and the computation of (5.1.57). Hence

$$\int_D \sqrt{1-|x|^2}\,dx = \frac{\pi}{2}\int_{-1}^{1}(1-x^2)\,dx = \frac{2}{3}\pi, \tag{5.1.67}$$

and we get

$$V(B^3) = \frac{4}{3}\pi. \tag{5.1.68}$$

Another attack on the integral (5.1.64), using polar coordinates, will be discussed below.

Another approach to computing $V(B^n)$ will arise from the following generalization of Theorem 5.1.10.

Proposition 5.1.12. *Let $n = k+\ell$, and let $\Sigma \subset \mathbb{R}^k$ be a closed, bounded, contented set. Let $g_j : \Sigma \to [0,\infty)$ be continuous, and satisfy $g_0(x) \le g_1(x)$. Take*

$$\Omega = \{(x,y) \in \mathbb{R}^n : x \in \Sigma,\ y \in \mathbb{R}^\ell,\ g_0(x) \le |y| \le g_1(x)\}. \tag{5.1.69}$$

Then Ω is a contented set in $\mathbb{R}^n$. If $f : \Omega \to \mathbb{R}$ is continuous, then

$$\varphi(x) = \int_{g_0(x)\le|y|\le g_1(x)} f(x,y)\,dy \tag{5.1.70}$$

is continuous on Σ, and

$$\int_\Omega f\,dV_n = \int_\Sigma \varphi\,dV_k. \tag{5.1.71}$$

The reader can extend the proof of Theorem 5.1.10 to cover this result.

Before applying Proposition 5.1.11 to $V(B^n)$, we look at a class of 3D domains to which it applies, namely solids of rotation. Take a continuous function $g : [a,b] \to (0,\infty)$, and consider

$$\Omega = \{(x,y,z) : a \le x \le b,\ \sqrt{y^2+z^2} \le g(x)\}. \tag{5.1.72}$$

See Figure 5.1.3. This has the form (5.1.69), with $\Sigma = [a,b]$, $g_0 \equiv 0$, $g_1(x) = g(x)$. If $f : \Omega \to \mathbb{R}$ is continuous, then (5.1.70) leads to

$$\varphi(x) = \int_{|y|\le g(x)} f(x,y)\,dy. \tag{5.1.73}$$

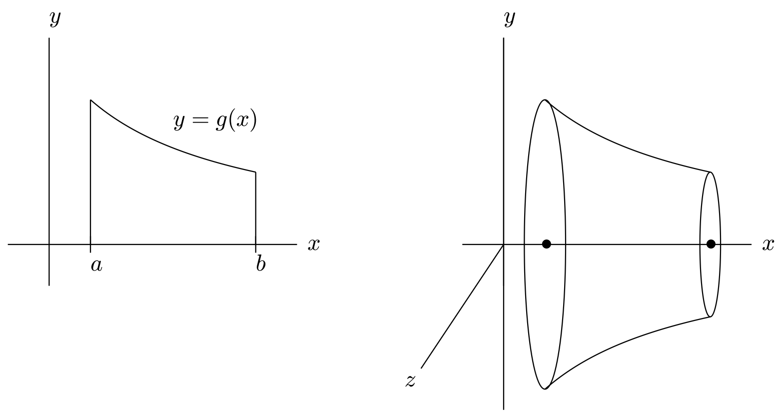

Figure 5.1.3. Area under a curve and solid of revolution

In particular, if $f = f(x)$, then $\varphi(x) = f(x)A(D_{g(x)})$, with

$$D_\rho = \{y \in \mathbb{R}^2 : |y| \le \rho\}, \quad A(D_\rho) = \pi\rho^2,$$

by (5.1.59), so, for Ω as in (5.1.72),

$$\int\limits_\Omega f(x)\,dx\,dy\,dz = \pi \int_a^b f(x)g(x)^2\,dx, \tag{5.1.74}$$

and taking $f \equiv 1$ gives

$$V(\Omega) = \pi \int_a^b g(x)^2\,dx. \tag{5.1.75}$$

The ball B^3 is the solid of revolution one gets with $g(x) = \sqrt{1-x^2}$, $[a,b] = [-1,1]$, so (5.1.75) yields an alternative derivation of (5.1.68):

$$V(B^3) = \pi \int_{-1}^1 (1-x^2)\,dx = \frac{4}{3}\pi. \tag{5.1.76}$$

Taking up the case B^n, we apply Proposition 5.1.12, with $\Sigma = [-1,1]$, $g_0 \equiv 0$, $g_1(x) = \sqrt{1-x^2}$, to obtain, for $f \in C(B^n)$,

$$\int_{B^n} f\,dV_n = \int_{-1}^{1} \Big(\int_{|y|\le\sqrt{1-x^2}} f(x,y)\,dy \Big)\,dx. \tag{5.1.77}$$

In particular,

$$V(B^n) = \int_{-1}^{1} V(B^{n-1}_{\sqrt{1-x^2}})\,dx, \tag{5.1.78}$$

where

$$B^{n-1}_r = \{y \in \mathbb{R}^{n-1} : |y| \le r\}. \tag{5.1.79}$$

Scaling gives

$$V(B^{n-1}_r) = V(B^{n-1}) r^{n-1}, \tag{5.1.80}$$

so we have the inductive result

$$V(B^n) = \beta_n V(B^{n-1}), \quad \beta_n = \int_{-1}^{1} (1-x^2)^{(n-1)/2}\,dx. \tag{5.1.81}$$

Applying this to $n = 3$, and using $V(B^2) = A(D) = \pi$, leads back to (5.1.76). To go one step further, we have

$$V(B^4) = \beta_4 V(B^3), \tag{5.1.82}$$

with

$$\begin{aligned} \beta_4 &= \int_{-1}^{1} (1-x^2)^{3/2}\,dx \\ &= 2\int_{0}^{\pi/2} \cos^4 t\,dt. \end{aligned} \tag{5.1.83}$$

One can attack this trigonometric integral by taking

$$2\cos^2 t = 1 + \cos 2t,$$

and squaring it. See Exercise 24 for an alternative approach.

In §6.1 we will give another approach to the calculation of $V(B^n)$, tied in with calculating the area of the sphere S^{n-1}. This will produce a unified formula, for all n, which involves the Gamma function.

Theorem 5.1.10 and Proposition 5.1.12 are designed to apply to the reduction of multiple integrals to iterated integrals on some fairly basic domains that one encounters. One can imagine other types of domains that are not covered by these two results. Rather than seek further results that apply to continuous integrands on more elaborate domains Ω, we will establish a rather general result in Theorem 5.1.15, after some further useful characterizations of the Riemann integral in Proposition 5.1.13 and Corollary 5.1.14.

Other characterizations of $\overline{I}(f)$ and $\underline{I}(f)$

At this point, it is useful to bring in some additional characterizations of $\overline{I}(f)$ and $\underline{I}(f)$. To do this, we introduce two classes of discontinuous functions on a cell R, which we denote $\mathfrak{C}(R)$ and $\mathrm{PK}(R)$. These are defined as follows.

Given a cell R and $f : R \to \mathbb{R}$, bounded, we say

$$f \in \mathfrak{C}(R) \Longleftrightarrow \text{ the set of discontinuities of } f \text{ is nil.} \tag{5.1.84}$$

Proposition 5.1.6 implies

$$\mathfrak{C}(R) \subset \mathcal{R}(R). \tag{5.1.85}$$

From the closure of the class of nil sets under finite unions it is clear that $\mathfrak{C}(R)$ is closed under sums and products, i.e., that $\mathfrak{C}(R)$ is an algebra of functions on R. We will denote by $\mathfrak{C}_c(\mathbb{R}^n)$ the set of bounded functions $f : \mathbb{R}^n \to \mathbb{R}$ such that f has compact support and its set of discontinuities is nil. Any $f \in \mathfrak{C}_c(\mathbb{R}^n)$ is supported in some cell R, and $f\big|_R \in \mathfrak{C}(R)$.

Next, given a cell $R \subset \mathbb{R}^n$ and $f : R \to \mathbb{R}$ bounded, we say

$$\begin{aligned} f \in \mathrm{PK}(R) \Longleftrightarrow\ & \exists \text{ a partition } \mathcal{P} \text{ of } R \text{ such that } f \text{ is constant} \\ & \text{on the interior of each cell } \ R_\alpha \in \mathcal{P}. \end{aligned} \tag{5.1.86}$$

The following will prove to be very useful in a number of applications.

Proposition 5.1.13. *Given a cell $R \subset \mathbb{R}^n$ and $f : R \to \mathbb{R}$ bounded,*

$$\begin{aligned} \overline{I}(f) &= \inf\Big\{ \int_R g\, dV : g \in \mathrm{PK}(R),\ g \geq f \Big\} \\ &= \inf\Big\{ \int_R g\, dV : g \in \mathfrak{C}(R),\ g \geq f \Big\} \\ &= \inf\Big\{ \int_R g\, dV : g \in C(R),\ g \geq f \Big\}. \end{aligned} \tag{5.1.87}$$

Similarly,

$$\begin{aligned} \underline{I}(f) &= \sup\Big\{ \int_R g\, dV : g \in \mathrm{PK}(R),\ g \leq f \Big\} \\ &= \sup\Big\{ \int_R g\, dV : g \in \mathfrak{C}(R),\ g \leq f \Big\} \\ &= \sup\Big\{ \int_R g\, dV : g \in C(R),\ g \leq f \Big\}. \end{aligned} \tag{5.1.88}$$

Proof. Denote the three quantities on the right side of (5.1.87) by $\overline{I}_1(f), \overline{I}_2(f)$, and $\overline{I}_3(f)$, respectively. The definition of $\overline{I}_1(f)$ is sufficiently close to that of $\overline{I}(f)$ in (5.1.8) that the identity

$$\overline{I}(f) = \overline{I}_1(f)$$

is apparent. Now $\overline{I}_2(f)$ is an inf over a larger class of functions g than that defining $\overline{I}_1(f)$, so

$$\overline{I}_2(f) \le \overline{I}_1(f).$$

On the other hand, $\overline{I}(g) \ge \overline{I}(f)$ for all g involved in defining $\overline{I}_2(f)$, so

$$\overline{I}_2(f) \ge \overline{I}(f),$$

hence $\overline{I}_2(f) = \overline{I}(f)$.

Next, $\overline{I}_3(f)$ is an inf over a smaller class of functions g than that defining $\overline{I}_2(f)$, so

$$\overline{I}_3(f) \ge \overline{I}(f).$$

On the other hand, given $\varepsilon > 0$ and $\psi \in \mathrm{PK}(R)$, one can readily find $g \in C(R)$ such that $g \ge \psi$ and $\int_R (g - \psi)\, dV < \varepsilon$. This implies

$$\overline{I}_3(f) \le \overline{I}(f) + \varepsilon,$$

for all $\varepsilon > 0$, and finishes the proof of (5.1.87). The proof of (5.1.88) is similar. □

Corollary 5.1.14. *Given a cell $R \subset \mathbb{R}^n$, let $f : R \to \mathbb{R}$ be bounded. Then $f \in \mathcal{R}(R)$ if and only if the following holds. For each $\varepsilon > 0$, there exist $g_0, g_1 \in C(R)$ satisfying*

$$g_0 \le f \le g_1, \quad \text{and} \quad \int\limits_R (g_1 - g_0)\, dV < \varepsilon. \tag{5.1.89}$$

More general Fubini-type theorem

We will make use of Proposition 5.1.13 to prove the following result, which is substantially more general than Theorem 5.1.10 and Proposition 5.1.12. As in Proposition 5.1.9, let $X \subset \mathbb{R}^k$, $Y \subset \mathbb{R}^\ell$, and $R = X \times Y \subset \mathbb{R}^n$ ($n = k + \ell$) be cells.

Theorem 5.1.15. *Let $f \in \mathcal{R}(X \times Y)$, and assume that, for each $x \in X$,*

$$g_x(y) = f(x, y) \quad \textit{defines} \quad g_x \in \mathcal{R}(Y). \tag{5.1.90}$$

Set

$$\varphi(x) = \int\limits_Y f(x, y)\, dV_\ell(y). \tag{5.1.91}$$

Then $\varphi \in \mathcal{R}(X)$, and

$$\int\limits_R f\, dV_n = \int\limits_X \varphi\, dV_k. \tag{5.1.92}$$

Proof. Pick $\varepsilon > 0$ and, using Corollary 5.1.14, take $u_j \in C(R)$ such that

$$u_0 \le f \le u_1, \quad \int\limits_R (u_1 - u_0)\, dV_n < \varepsilon. \tag{5.1.93}$$

Set

$$\psi_j(x) = \int\limits_Y u_j(x, y)\, dV_\ell(y), \quad x \in X. \tag{5.1.94}$$

Then $\psi_j \in C(X)$ and

$$\psi_0(x) \le \varphi(x) \le \psi_1(x), \quad \forall\, x \in X. \tag{5.1.95}$$

By Proposition 5.1.9, we have

$$\int\limits_R u_j\, dV_n = \int\limits_X \psi_j\, dV_k. \tag{5.1.96}$$

Hence, by (5.1.93),

$$\int\limits_X (\psi_1 - \psi_0)\, dV_k < \varepsilon. \tag{5.1.97}$$

Hence Corollary 5.1.14 implies that $\varphi \in \mathcal{R}(X)$, and we see that the two sides of (5.1.92) differ by a quantity that is $\le \varepsilon$, for all $\varepsilon > 0$. Thus identity must hold. □

Remark. The hypothesis (5.1.90) holds provided that the set of points in Y at which g_x is discontinuous has upper content zero.

Remark. See Exercise 22 below for an extension of Theorem 5.1.15.

Change of variables

We next take up the change of variables formula for multiple integrals, extending the one-variable formula discussed in Exercise 4 of §1.2. We begin with a result on linear changes of variables. The set of invertible real $n \times n$ matrices is denoted $Gl(n, \mathbb{R})$. In (5.1.98) and subsequent formulas, $\int f\, dV$ denotes $\int_R f\, dV$ for some cell R on which f is supported. The integral is independent of the choice of such a cell; cf. (5.1.27).

Proposition 5.1.16. *Let f be a continuous function with compact support in $\mathbb{R}^n$. If $A \in Gl(n, \mathbb{R})$, then*

$$\int f(x)\, dV = |\det A| \int f(Ax)\, dV. \tag{5.1.98}$$

Proof. Let $\mathcal{G}$ be the set of elements $A \in Gl(n, \mathbb{R})$ for which (5.1.98) is true. Clearly $I \in \mathcal{G}$. Using $\det A^{-1} = (\det A)^{-1}$, and $\det AB = (\det A)(\det B)$, we can conclude that $\mathcal{G}$ is a subgroup of $Gl(n, \mathbb{R})$, i.e., $\mathcal{G}$ is a subset of $Gl(n, \mathbb{R})$ possessing the two properties (5.1.102)–(5.1.103) listed below.

In more detail, for $A \in Gl(n, \mathbb{R})$, f as above, let

$$I_A(f) = \int f_A\, dV = I(f_A), \quad f_A(x) = f(Ax). \tag{5.1.99}$$

Then

$$A \in \mathcal{G} \Longleftrightarrow I_A(f) = |\det A|^{-1} I(f), \tag{5.1.100}$$

for all such f. We see that

$$I_{AB}(f) = I(f_{AB}) = I_B(f_A), \tag{5.1.101}$$

so

$$\begin{aligned} A, B \in \mathcal{G} \Longrightarrow I_{AB}(f) &= |\det B|^{-1} I(f_A) \\ &= |\det B|^{-1} |\det A|^{-1} I(f) = |\det AB|^{-1} I(f) \\ \Longrightarrow AB \in \mathcal{G}. \end{aligned} \tag{5.1.102}$$

Applying a similar argument to $I_{AA^{-1}}(f) = I(f)$, also yields the implication

$$A \in \mathcal{G} \Rightarrow A^{-1} \in \mathcal{G}. \tag{5.1.103}$$

To prove the proposition, it will therefore suffice to show that $\mathcal{G}$ contains all elements of the following 3 forms, since (as shown in the exercises on row reduction at the end of this section) the method of applying elementary row operations to reduce a matrix shows that any element of $Gl(n, \mathbb{R})$ is a product of a finite number of these elements. Here, $\{e_j : 1 \le j \le n\}$ denotes the standard basis of $\mathbb{R}^n$, and σ a permutation of $\{1, \dots, n\}$.

$$\begin{aligned} &A_1 e_j = e_{\sigma(j)}, \\ &A_2 e_j = c_j e_j, \quad c_j \neq 0 \\ &A_3 e_2 = e_2 + c e_1, \quad A_3 e_j = e_j \text{ for } j \neq 2. \end{aligned} \tag{5.1.104}$$

The proofs of (5.1.98) in the first two cases are elementary consequences of the definition of the Riemann integral, and can be left as exercises.

We show that (5.1.98) holds for transformations of the form A_3 by using Proposition 5.1.9 to reduce it to the case $n = 1$. Given $f \in C(\mathbb{R})$, compactly supported, and $b \in \mathbb{R}$, we clearly have

$$\int f(x)\,dx = \int f(x+b)\,dx. \tag{5.1.105}$$

Now, for the case $A = A_3$, with $x = (x_1, x')$, we have

$$\begin{aligned} \int f(x_1 + cx_2, x')\,dV_n(x) &= \int \Bigl(\int f(x_1 + cx_2, x')\,dx_1 \Bigr)\,dV_{n-1}(x') \\ &= \int \Bigl(\int f(x_1, x')\,dx_1 \Bigr)\,dV_{n-1}(x'), \end{aligned} \tag{5.1.106}$$

the second identity by (5.1.105). Thus we get (5.1.98) in case $A = A_3$, so the proposition is proved. □

It is desirable to extend Proposition 5.1.16 to more general Riemann-integrable functions. Say $f \in \mathcal{R}_c(\mathbb{R}^n)$ if f has compact support, say in some cell R, and $f \in \mathcal{R}(R)$. Also say $f \in C_c(\mathbb{R}^n)$ if f is continuous on $\mathbb{R}^n$, with compact support.

Proposition 5.1.17. *Given $A \in Gl(n, \mathbb{R})$, the identity* (5.1.98) *holds for all $f \in \mathcal{R}_c(\mathbb{R}^n)$.*

Proof. We have from Proposition 5.1.13 that, for each $\nu \in \mathbb{N}$, there exist $g_\nu, h_\nu \in C_c(\mathbb{R}^n)$ such that $h_\nu \le f \le g_\nu$ and, with $B = \int f\,dV$,

$$B - \frac{1}{\nu} \le \int h_\nu\,dV \le B \le \int g_\nu\,dV \le B + \frac{1}{\nu}.$$

Now Proposition 5.1.16 applies to g_ν and h_ν, so

$$B - \frac{1}{\nu} \le |\det A| \int h_\nu(Ax)\,dV \le B \le |\det A| \int g_\nu(Ax)\,dV \le B + \frac{1}{\nu}. \tag{5.1.107}$$

Furthermore, with $f_A(x) = f(Ax)$, we have $h_\nu(Ax) \le f_A(x) \le g_\nu(Ax)$, so (4.44) gives

$$B - \frac{1}{\nu} \le |\det A|\,\underline{I}(f_A) \le |\det A|\,\overline{I}(f_A) \le B + \frac{1}{\nu}, \tag{5.1.108}$$

for all ν, and leting $\nu \to \infty$ we obtain (5.1.98). □

Corollary 5.1.18. *If $\Sigma \subset \mathbb{R}^n$ is a compact, contented set and $A \in Gl(n, \mathbb{R})$, then $A(\Sigma) = \{Ax : x \in \Sigma\}$ is contented, and*

$$V\bigl(A(\Sigma)\bigr) = |\det A|\,V(\Sigma). \tag{5.1.109}$$

We now extend Proposition 5.1.16 to nonlinear changes of variables.

Proposition 5.1.19. *Let $\mathcal{O}$ and Ω be open in $\mathbb{R}^n$, $G : \mathcal{O} \to \Omega$ a C^1 diffeomorphism, and f a continuous function with compact support in Ω. Then*

$$\int_\Omega f(y)\,dV(y) = \int_{\mathcal{O}} f\bigl(G(x)\bigr)\,|\det DG(x)|\,dV(x). \tag{5.1.110}$$

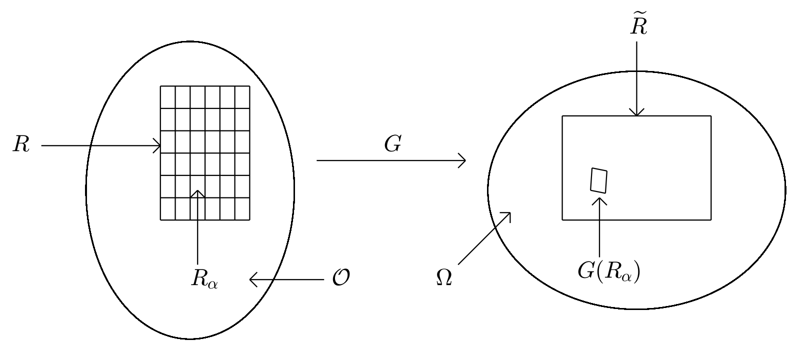

Figure 5.1.4. Image of a cell

Proof. It suffices to prove the result under the additional assumption that $f \geq 0$, which we make from here on. Also, using a partition of unity (see §6.6), we can write f as a finite sum of continuous functions with small supports, so it suffices to treat the case where f is supported in a cell $\widetilde{R} \subset \Omega$ and $f \circ G$ is supported in a cell $R \subset \mathcal{O}$. See Figure 5.1.4. Let $\mathcal{P} = \{R_\alpha\}$ be a partition of R. Note that for each $R_\alpha \in \mathcal{P}$, $bG(R_\alpha) = G(bR_\alpha)$, so $G(R_\alpha)$ is contented, in view of Propositions 5.1.4 and 5.1.8.

Let ξ_α be the center of R_α, and let $\widetilde{R}_\alpha = R_\alpha - \xi_\alpha$, a cell with center at the origin. Then

$$G(\xi_\alpha) + DG(\xi_\alpha)\big(\widetilde{R}_\alpha\big) = \eta_\alpha + H_\alpha \tag{5.1.111}$$

is an n-dimensional parallelepiped, each point of which is very close to a point in $G(R_\alpha)$, if R_α is small enough. To be precise, for $y \in \widetilde{R}_\alpha$,

$$G(\xi_\alpha + y) = \eta_\alpha + DG(\xi_\alpha)y + \Phi(\xi_\alpha, y)y,$$
$$\Phi(\xi_\alpha, y) = \int_0^1 \big[DG(\xi_\alpha + ty) - DG(\xi_\alpha)\big]\, dt.$$

See Figure 5.1.5.

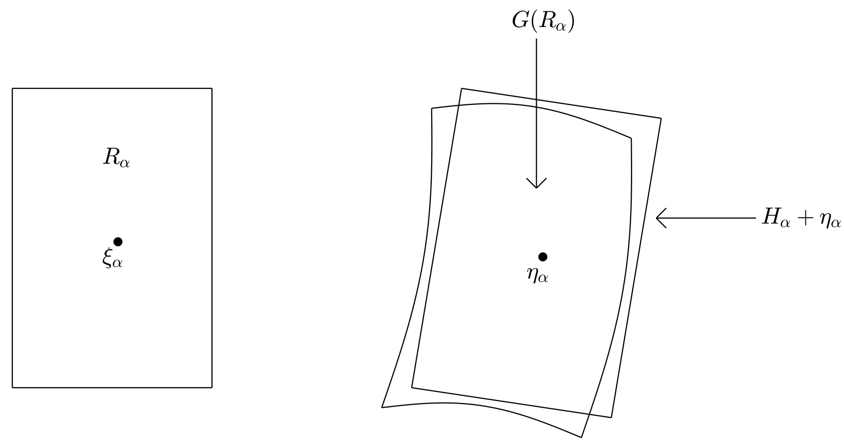

Figure 5.1.5. Cell image closeup

Consequently, given $\varepsilon > 0$, if $\delta > 0$ is small enough and $\operatorname{maxsize}(\mathcal{P}) \le \delta$, then we have

$$\eta_\alpha + (1+\varepsilon)H_\alpha \supset G(R_\alpha), \tag{5.1.112}$$

for all $R_\alpha \in \mathcal{P}$. Now, by (5.1.109),

$$V(H_\alpha) = |\det DG(\xi_\alpha)|\, V(R_\alpha). \tag{5.1.113}$$

Hence

$$V\big(G(R_\alpha)\big) \le (1+\varepsilon)^n |\det DG(\xi_\alpha)|\, V(R_\alpha). \tag{5.1.114}$$

Now we have

$$\begin{aligned} \int f\, dV &= \sum_\alpha \int\limits_{G(R_\alpha)} f\, dV \\ &\le \sum_\alpha \sup_{R_\alpha} f\circ G(x)\, V\big(G(R_\alpha)\big) \\ &\le (1+\varepsilon)^n \sum_\alpha \sup_{R_\alpha} f\circ G(x)\, |\det DG(\xi_\alpha)|\, V(R_\alpha). \end{aligned} \tag{5.1.115}$$

To see that the first line of (5.1.115) holds, note that $f\chi_{G(R_\alpha)}$ is Riemann integrable, by Proposition 5.1.6; note also that $\sum_\alpha f\chi_{G(R_\alpha)} = f$ except on a set of content zero. Then the additivity result in Proposition 5.1.2 applies. The first inequality in (5.1.115) is elementary; the second inequality uses (5.1.114) and $f \geq 0$. If we set

$$h(x) = f \circ G(x)\,|\det DG(x)|, \tag{5.1.116}$$

then we have

$$\sup_{R_\alpha} f \circ G(x)\,|\det DG(\xi_\alpha)| \leq \sup_{R_\alpha} h(x) + M\omega(\delta), \tag{5.1.117}$$

provided $|f| \leq M$ and $\omega(\delta)$ is a modulus of continuity for DG. Taking arbitrarily fine partitions, we get, in the limit $\delta \to 0$,

$$\int_\Omega f\,dV \leq \int_\mathcal{O} h\,dV. \tag{5.1.118}$$

If we apply this result, with G replaced by G^{-1}, $\mathcal{O}$ and Ω switched, and f replaced by h, given by (5.1.116), we have

$$\int_\mathcal{O} h\,dV \leq \int_\Omega h \circ G^{-1}(y)\,|\det DG^{-1}(y)|\,dV(y) = \int_\Omega f\,dV. \tag{5.1.119}$$

The inequalities (5.1.118) and (5.1.119) together yield the identity (5.1.110). ☐

We now extend Proposition 5.1.19 to more general Riemann integrable functions. Recall that $f \in \mathcal{R}_c(\mathbb{R}^n)$ if f has compact support, say in some cell R, and $f \in \mathcal{R}(R)$. If $\Omega \subset \mathbb{R}^n$ is open and $f \in \mathcal{R}_c(\mathbb{R}^n)$ has support in Ω, we say $f \in \mathcal{R}_c(\Omega)$. We also say $f \in \mathfrak{C}_c(\Omega)$ if $f \in \mathfrak{C}_c(\mathbb{R}^n)$ has support in Ω, and we say $f \in C_c(\Omega)$ if f is continuous with compact support in Ω.

Theorem 5.1.20. *Let $\mathcal{O}$ and Ω be open in $\mathbb{R}^n$, $G : \mathcal{O} \to \Omega$ a C^1 diffeomorphism. If $f \in \mathcal{R}_c(\Omega)$, then $f \circ G \in \mathcal{R}_c(\mathcal{O})$, and* (5.1.110) *holds.*

Proof. The proof is similar to that of Proposition 5.1.17. Given $\nu \in \mathbb{N}$, we have from Proposition 5.1.13 that there exist $g_\nu, h_\nu \in C_c(\Omega)$ such that $h_\nu \leq f \leq g_\nu$ and, with $B = \int_\Omega f\,dV$,

$$B - \frac{1}{\nu} \leq \int h_\nu\,dV \leq B \leq \int g_\nu\,dV \leq B + \frac{1}{\nu}.$$

Then Proposition 5.1.19 applies to h_ν and g_ν, so

$$\begin{aligned} B - \frac{1}{\nu} &\leq \int_\mathcal{O} h_\nu(G(x))|\det DG(x)|\,dV(x) \\ &\leq B \leq \int_\mathcal{O} g_\nu(G(x))|\det DG(x)|\,dV(x) \leq B + \frac{1}{\nu}. \end{aligned}$$

Now, with $f_G(x) = f(G(x))$, we have $h_\nu(G(x)) \leq f_G(x) \leq g_\nu(G(x))$, so

$$B - \frac{1}{\nu} \leq \underline{I}(f_G|\det DG|) \leq \overline{I}(f_G|\det DG|) \leq B + \frac{1}{\nu}, \tag{5.1.120}$$

for all ν, and letting $\nu \to \infty$, we obtain (5.1.110). ☐

Polar coordinates

The most frequently invoked case of the change of variable formula, in the case $n = 2$, involves the following change from Cartesian to polar coordinates:

$$x = r\cos\theta, \quad y = r\sin\theta. \tag{5.1.121}$$

Thus, take $G(r,\theta) = (r\cos\theta, r\sin\theta)$. We have

$$DG(r,\theta) = \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix}, \quad \det DG(r,\theta) = r. \tag{5.1.122}$$

For example, if $\rho \in (0,\infty)$ and

$$D_\rho = \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 \le \rho^2\}, \tag{5.1.123}$$

then, for $f \in C(D_\rho)$,

$$\int\limits_{D_\rho} f(x,y)\ dA = \int_0^\rho \int_0^{2\pi} f(r\cos\theta, r\sin\theta) r\, d\theta\, dr. \tag{5.1.124}$$

To get this, we first apply Proposition 5.1.19, with $\mathcal{O} = [\varepsilon,\rho] \times [0, 2\pi - \varepsilon]$, then apply Theorem 5.1.10, then let $\varepsilon \searrow 0$.

In case $f \equiv 1$, we have the formula for the area of the disk D_ρ,

$$A(D_\rho) = \int_0^\rho \int_0^{2\pi} r\, dr = \pi\rho^2, \tag{5.1.125}$$

obtaining, by different means, the area formula derived (for $\rho = 1$) in (5.1.59).

More generally, if f is a radial function, i.e.,

$$f(x,y) = g(r), \quad r = \sqrt{x^2+y^2}, \tag{5.1.126}$$

then (5.1.124) gives

$$\begin{aligned} \int\limits_{D_\rho} f\, dA &= \int_0^\rho \int_0^{2\pi} g(r) r\, d\theta\, dr \\ &= 2\pi \int_0^\rho g(r) r\, dr. \end{aligned} \tag{5.1.127}$$

We can apply this as follows to the computation of the volume of the unit ball $B^3 \subset \mathbb{R}^3$, which is given by

$$|z| \le \sqrt{1 - x^2 - y^2}, \quad (x,y) \in D. \tag{5.1.128}$$

By Theorem 5.1.10,

$$V(B^3) = 2 \int\limits_D \sqrt{1 - x^2 - y^2}\, dx\, dy. \tag{5.1.129}$$

Applying (5.1.127) yields

$$V(B^3) = \frac{4}{3}\pi, \tag{5.1.130}$$

recovering (5.1.76), by a third method. The task of going from (5.1.129) to (5.1.130) constitutes Exercise 6 below.

One can also apply (5.1.124) to non-radial functions. For example, given $j, k \in \mathbb{Z}^+$,

$$\begin{aligned} \int_D x^j y^k \, dA &= \int_0^1 \int_0^{2\pi} \cos^j \theta \sin^k \theta \, r^{j+k+1} \, d\theta \, dr \\ &= \frac{1}{j+k+2} \int_0^{2\pi} \cos^j \theta \sin^k \theta \, d\theta. \end{aligned} \tag{5.1.131}$$

One can apply polar coordinates to other regions, such as regions defined in polar coordinates by

$$\alpha(\theta) \le r \le \beta(\theta), \quad \theta_0 \le \theta \le \theta_1. \tag{5.1.132}$$

If this defines the region Ω, then

$$\int_\Omega f(x,y) \, dA = \int_{\theta_0}^{\theta_1} \int_{\alpha(\theta)}^{\beta(\theta)} f(r\cos\theta, r\sin\theta) r \, dr \, d\theta. \tag{5.1.133}$$

For example, the disk $\mathcal{D}_{a/2}$ of radius $a/2$, centered at $(a, 0)$, is defined by

$$0 \le r \le a\cos\theta, \quad -\frac{\pi}{2} \le \theta \le \frac{\pi}{2}, \tag{5.1.134}$$

(cf. (3.1.52)), and in this case

$$\int_{\mathcal{D}_{a/2}} f(x,y) \, dA = \int_{-\pi/2}^{\pi/2} \int_0^{a\cos\theta} f(r\cos\theta, r\sin\theta) r \, dr \, d\theta. \tag{5.1.135}$$

Note that taking $f \equiv 1$ yields

$$\begin{aligned} A(\mathcal{D}_{a/2}) &= \frac{a^2}{2} \int_{-\pi/2}^{\pi/2} \cos^2 \theta \, d\theta \\ &= \pi \Big(\frac{a}{2}\Big)^2, \end{aligned} \tag{5.1.136}$$

the last identity obtained as in (5.1.59).

Spherical polar coordinates on $\mathbb{R}^3$

On $\mathbb{R}^3$, we have *spherical polar coordinates*, given by

$$\begin{gathered} G(\rho, \theta, \psi) = (\rho \sin\theta \cos\psi, \rho\sin\theta\sin\psi, \rho\cos\theta), \\ \rho > 0, \quad 0 \le \theta \le \pi, \quad 0 \le \psi \le 2\pi. \end{gathered} \tag{5.1.137}$$

See Figure 5.1.6. We have

$$DG(\rho,\theta,\psi) = \begin{pmatrix} \sin\theta\cos\psi & \rho\cos\theta\cos\psi & -\rho\sin\theta\sin\psi \\ \sin\theta\sin\psi & \rho\cos\theta\sin\psi & \rho\sin\theta\cos\psi \\ \cos\theta & -\rho\sin\theta & 0 \end{pmatrix}. \tag{5.1.138}$$

A calculation, e.g., expanding by minors down the third column, gives

$$\det DG(\rho,\theta,\psi) = \rho^2 \sin\theta. \tag{5.1.139}$$

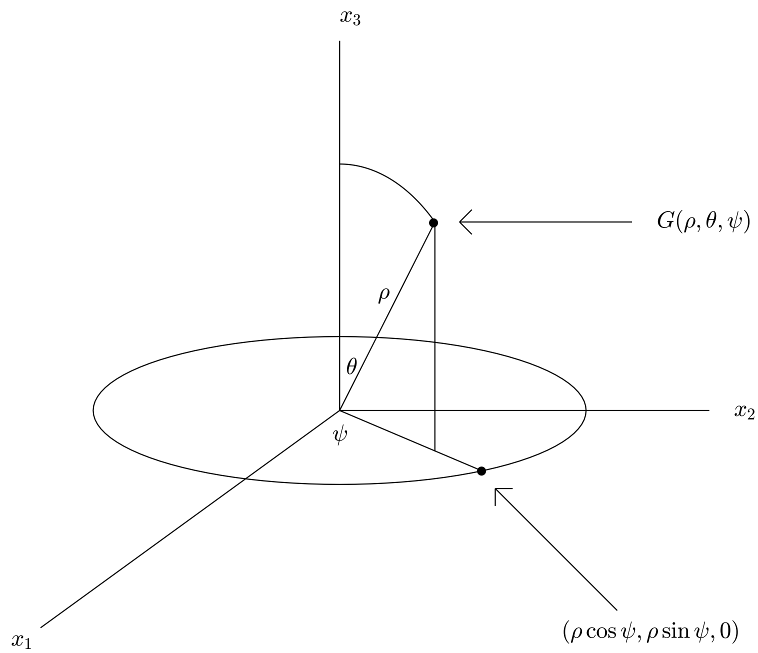

Figure 5.1.6. Spherical polar coordinates on $\mathbb{R}^3$

Hence, if $\mathcal{O} \subset \{(\rho,\theta,\psi) : \rho > 0, \theta \in [0,\pi], \psi \in [0,2\pi]\}$, and $\Omega = G(\mathcal{O})$, then

$$\int\limits_{\Omega} f(x)\,dV(x) = \int\limits_{\mathcal{O}} f(G(\rho,\theta,\psi))\rho^2 \sin\theta\,d\rho\,d\theta\,d\psi. \tag{5.1.140}$$

In particular, if $B^3 = \{x \in \mathbb{R}^3 : |x| \le 1\}$ is the unit ball in $\mathbb{R}^3$,

$$\int\limits_{B^3} f(x)\,dV(x) = \int_0^{2\pi}\int_0^{\pi}\int_0^1 f(G(\rho,\theta,\psi))\rho^2 \sin\theta\,d\rho\,d\theta\,d\psi. \tag{5.1.141}$$

Taking $f \equiv 1$ yields the volume formula

$$V(B^3) = \frac{4}{3}\pi, \tag{5.1.142}$$

recovering (5.1.130), by a fourth method. The tasks of checking (5.1.139) and going from (5.1.141) to (5.1.142) constitute Exercise 5 below.

Back to integrability

We have seen how Proposition 5.1.13 has been useful. The following result, to some degree a variant of Proposition 5.1.13, is also useful.

Lemma 5.1.21. *Let $F : R \to \mathbb{R}$ be bounded, $B \in \mathbb{R}$. Suppose that, for each $\nu \in \mathbb{Z}^+$, there exist $\Psi_\nu, \Phi_\nu \in \mathcal{R}(R)$ such that*

$$\Psi_\nu \le F \le \Phi_\nu \tag{5.1.143}$$

and

$$B - \delta_\nu \le \int\limits_R \Psi_\nu(x)\, dV(x) \le \int\limits_R \Phi_\nu(x)\, dV(x) \le B + \delta_\nu, \quad \delta_\nu \to 0. \tag{5.1.144}$$

Then $F \in \mathcal{R}(R)$ and

$$\int_R F(x)\, dV(x) = B. \tag{5.1.145}$$

Furthermore, if there exist $\Psi_\nu, \Phi_\nu \in \mathcal{R}(R)$ such that (5.1.143) *holds and*

$$\int\limits_R (\Phi_\nu(x) - \Psi_\nu(x))\, dV \le \delta_\nu \to 0, \tag{5.1.146}$$

then there exists B such that (5.1.144) *holds. Hence $F \in \mathcal{R}(R)$ and* (5.1.145) *holds.*

We next use Lemma 5.1.21 to establish the following useful result on products of Riemann integrable functions.

Proposition 5.1.22. *Given $f_1, f_2 \in \mathcal{R}(R)$, we have $f_1 f_2 \in \mathcal{R}(R)$.*

Proof. It suffices to prove this when $f_j \ge 0$. Take partitions $\mathcal{P}_\nu$ and functions $\psi_{j\nu}, \varphi_{j\nu} \ge 0$, constant in the interior of each cell in $\mathcal{P}_\nu$, such that

$$0 \le \psi_{j\nu} \le f_j \le \varphi_{j\nu} \le M,$$

and

$$\int \psi_{j\nu}\, dV, \quad \int \varphi_{j\nu}\, dV \longrightarrow \int f_j\, dV.$$

We apply Lemma 5.1.21 with

$$F = f_1 f_2, \quad \Psi_\nu = \psi_{1\nu}\psi_{2\nu}, \quad \Phi_\nu = \varphi_{1\nu}\varphi_{2\nu}.$$

Note that

$$\begin{aligned} \Phi_\nu - \Psi_\nu &= \varphi_{1\nu}(\varphi_{2\nu} - \psi_{2\nu}) + \psi_{2\nu}(\varphi_{1\nu} - \psi_{1\nu}) \\ &\le M(\varphi_{2\nu} - \psi_{2\nu}) + M(\varphi_{1\nu} - \psi_{1\nu}). \end{aligned}$$

Hence (5.1.146) holds, giving $F \in \mathcal{R}(R)$. □

As a consequence of Proposition 5.1.22, we can make the following construction. Assume R is a cell and $S \subset R$ is a contented set. If $f \in \mathcal{R}(R)$, we have $\chi_S f \in \mathcal{R}(R)$, by Proposition 5.1.22. We define

$$\int\limits_S f(x)\, dV(x) = \int\limits_R \chi_S(x) f(x)\, dV(x). \tag{5.1.147}$$

Note how his extends the scope of (5.1.26).

Integrals over $\mathbb{R}^n$

It is often useful to integrate a function whose support is not bounded. Generally, given a bounded function $f : \mathbb{R}^n \to \mathbb{R}$, we say

$$f \in \mathcal{R}(\mathbb{R}^n)$$

provided $f\big|_R \in \mathcal{R}(R)$ for each cell $R \subset \mathbb{R}^n$, and

$$\int\limits_R |f| \; dV \le C,$$

for some $C < \infty$, independent of R. If $f \in \mathcal{R}(\mathbb{R}^n)$, we set

$$\int\limits_{\mathbb{R}^n} f \; dV = \lim_{s\to\infty} \int\limits_{R_s} f \; dV, \quad R_s = \{x \in \mathbb{R}^n : |x_j| \le s, \forall\, j\}. \tag{5.1.148}$$

The existence of the limit in (5.1.148) can be established as follows. If $M < N$, then

$$\int\limits_{R_N} f \, dV - \int\limits_{R_M} f \, dV = \int\limits_{R_N \setminus R_M} f \, dV,$$

which is dominated in absolute value by $\int_{R_N \setminus R_M} |f| \, dV$. If $f \in \mathcal{R}(\mathbb{R}^n)$, then $a_N = \int_{R_N} |f| \, dV$ is a bounded monotone sequence, which hence converges, so

$$\int\limits_{R_N \setminus R_M} |f| \, dV = \int\limits_{R_N} |f| \, dV - \int\limits_{R_M} |f| \, dV \longrightarrow 0, \quad \text{as} \;\; M, N \to \infty.$$

The following simple but useful result is an exercise.

Proposition 5.1.23. *If K_ν is any sequence of compact contented subsets of $\mathbb{R}^n$ such that each R_s, for $s < \infty$, is contained in all K_ν for ν sufficiently large, i.e., $\nu \ge N(s)$, then, whenever $f \in \mathcal{R}(\mathbb{R}^n)$,*

$$\int\limits_{\mathbb{R}^n} f \; dV = \lim_{\nu\to\infty} \int\limits_{K_\nu} f \; dV. \tag{5.1.149}$$

Change of variables formulas and Fubini's Theorem extend to this case. For example, the limiting case of (5.1.124) as $\rho \to \infty$ is

(5.1.150)

$$f \in C(\mathbb{R}^2) \cap \mathcal{R}(\mathbb{R}^2) \Longrightarrow \int\limits_{\mathbb{R}^2} f(x,y) \; dA = \int_0^\infty \int_0^{2\pi} f(r\cos\theta, r\sin\theta) r \; d\theta \; dr.$$

To see this, use Proposition 5.1.22 with $K_\nu = D_\nu$, defined as in (5.1.123), to write

$$\int\limits_{\mathbb{R}^2} f(x,y) \, dA = \lim_{\nu\to\infty} \int\limits_{D_\nu} f(x,y) \, dA, \tag{5.1.151}$$

and apply (5.1.124) to write the integral on the right as

$$\int_0^\nu \int_0^{2\pi} f(r\cos\theta, r\sin\theta) r \, d\theta \, dr.$$

You get the right side of (5.1.150) in the limit $\nu \to \infty$.

The following is a good example. Take $f(x,y) = e^{-x^2-y^2}$. We have

$$\int_{\mathbb{R}^2} e^{-x^2-y^2}\, dA = \int_0^\infty \int_0^{2\pi} e^{-r^2} r\, d\theta\, dr = 2\pi \int_0^\infty e^{-r^2} r\, dr. \tag{5.1.152}$$

Now, methods of §1.2 allow the substitution $s = r^2$, so

$$\int_0^\infty e^{-r^2} r\, dr = \frac{1}{2} \int_0^\infty e^{-s}\, ds = \frac{1}{2}.$$

Hence

$$\int_{\mathbb{R}^2} e^{-x^2-y^2}\, dA = \pi. \tag{5.1.153}$$

On the other hand, Theorem 5.1.10 extends to give

$$\begin{aligned} \int_{\mathbb{R}^2} e^{-x^2-y^2}\, dA &= \int_{-\infty}^\infty \int_{-\infty}^\infty e^{-x^2-y^2}\, dy\, dx \\ &= \left(\int_{-\infty}^\infty e^{-x^2}\, dx \right) \left(\int_{-\infty}^\infty e^{-y^2}\, dy \right). \end{aligned} \tag{5.1.154}$$

Note that the two factors in the last product are equal. We deduce that

$$\int_{-\infty}^\infty e^{-x^2}\, dx = \sqrt{\pi}. \tag{5.1.155}$$

We can generalize (5.1.154), to obtain (via (5.1.155))

$$\int_{\mathbb{R}^n} e^{-|x|^2}\, dV_n = \left(\int_{-\infty}^\infty e^{-x^2}\, dx \right)^n = \pi^{n/2}. \tag{5.1.156}$$

The integrals (5.1.152)–(5.1.156) are called Gaussian integrals, and their evaluation has many uses. We shall see some in §6.1.

We record the following additivity result for the integral over $\mathbb{R}^n$, whose proof is also an exercise.

Proposition 5.1.24. *If $f, g \in \mathcal{R}(\mathbb{R}^n)$, then $f + g \in \mathcal{R}(\mathbb{R}^n)$, and*

$$\int_{\mathbb{R}^n} (f+g)\, dV = \int_{\mathbb{R}^n} f\, dV + \int_{\mathbb{R}^n} g\, dV. \tag{5.1.157}$$

Exercises

1. Show that any two partitions of a cell R have a common refinement.
Hint. Consider the argument given for the one-dimensional case in §1.2.

2. Write down a proof of the identity (5.1.18), i.e., $\text{cont}^+(S) = \text{cont}^+(\overline{S})$.

3. Write down the details of the argument giving (5.1.27), on the independence of the integral from the choice of cell containing K.

4. Write down a direct proof that the transformation formula (5.1.98) holds for those linear transformations of the form A_1 and A_2 in (5.1.104).
Hint. A_1 and A_2 each take a cell to another cell. Relate their volumes.

5. Consider spherical polar coordinates on $\mathbb{R}^3$, given by

$$x = \rho \ \sin\theta \ \cos\psi, \quad y = \rho \ \sin\theta \ \sin\psi, \quad z = \rho \ \cos\theta,$$

i.e., take

$$G(\rho, \theta, \psi) = \big(\rho \ \sin\theta \ \cos\psi, \rho \ \sin\theta \ \sin\psi, \rho \ \cos\theta\big).$$

See Figure 5.1.6. Show that

$$\det \, DG(\rho, \theta, \psi) = \rho^2 \sin\theta,$$

as asserted in (5.1.139). Verify the computation of $V(B^3)$ stated in (5.1.142).

6. If B is the unit ball in $\mathbb{R}^3$, show that Theorem 5.1.10 implies

$$V(B) = 2 \int\limits_D \sqrt{1 - |x|^2} \, dA(x),$$

where $D = \{x \in \mathbb{R}^2 : |x| \le 1\}$ is the unit disk. Use polar coordinates, as in (5.1.121)–(5.1.124), to compute this integral. Compare the result with that of Exercise 5.

7. Apply Corollary 5.1.18 and the answer to Exercises 5 and 6 to compute the volume of the ellipsoidal region in $\mathbb{R}^3$ defined by

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} \le 1,$$

given $a, b, c \in (0, \infty)$.

8. Prove Lemma 5.1.21.

9. If R is a cell and $S \subset R$ is a contented set, and $f \in \mathcal{R}(R)$, we have, via

Proposition 5.1.22,

$$\int_S f(x)\,dV(x) = \int_R \chi_S(x)f(x)\,dV(x).$$

Show that, if $S_j \subset R$ are contented and they are disjoint (or more generally $\mathrm{cont}^+(S_1 \cap S_2) = 0$), then, for $f \in \mathcal{R}(R)$,

$$\int_{S_1 \cup S_2} f(x)\,dV(x) = \int_{S_1} f(x)\,dV(x) + \int_{S_2} f(x)\,dV(x).$$

10. Establish the convergence result (5.1.149), for all $f \in \mathcal{R}(\mathbb{R}^n)$.

In Exercises 11–13, let $D_R = \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 \le R^2\}$, and compute the following integrals.

11. $\displaystyle\iint_{D_R} (1 + x^2 + y^2)^{-1}\,dx\,dy.$

12. $\displaystyle\iint_{D_R} \sin(x^2 + y^2)\,dx\,dy.$

13. $\displaystyle\iint_{D_R} e^{-(x^2+y^2)}\,dx\,dy.$

14. Theorem 5.1.10, relating multiple integrals and iterated integrals, played the following role in the proof of the change of variable formula (5.1.110). Namely, it was used to establish the identity (5.1.113) for the volume of the parallelepiped H_α, via an appeal to Corollary 5.1.18, hence to Proposition 5.1.16, whose proof relied on Theorem 5.1.10.

Try to establish Corollary 5.1.18 directly, without using Theorem 5.1.10, in the case when Σ is either a cell or the image of a cell under an element of $Gl(n,\mathbb{R})$.

In preparation for the next three exercises, review the proof of Proposition 1.2.15.

15. Assume $f \in \mathcal{R}(R)$, $|f| \le M$, and let $\varphi : [-M, M] \to \mathbb{R}$ be Lipschitz and monotone. Show directly from the definition that $\varphi \circ f \in \mathcal{R}(R)$.

16. If $\varphi : [-M, M] \to \mathbb{R}$ is continuous and piecewise linear, show that you can write $\varphi = \varphi_1 - \varphi_2$ with φ_j Lipschitz and monotone. Deduce that $f \in \mathcal{R}(R) \Rightarrow \varphi \circ f \in \mathcal{R}(R)$ when φ is piecewise linear.

17. Assume $u_\nu \in \mathcal{R}(R)$ and that $u_\nu \to u$ uniformly on R. Show that $u \in \mathcal{R}(R)$. Deduce that if $f \in \mathcal{R}(R)$, $|f| \le M$, and $\psi : [-M, M] \to \mathbb{R}$ is continuous, then $\psi \circ f \in \mathcal{R}(R)$.

18. Let $R \subset \mathbb{R}^n$ be a cell and let $f, g : R \to \mathbb{R}$ be bounded. Show that

$$\overline{I}(f + g) \le \overline{I}(f) + \overline{I}(g), \quad \underline{I}(f + g) \ge \underline{I}(f) + \underline{I}(g).$$

Hint. Look at the proof of Proposition 1.2.1.

19. Let $R \subset \mathbb{R}^n$ be a cell and let $f : R \to \mathbb{R}$ be bounded. Assume that for each $\varepsilon > 0$, there exist bounded $f_\varepsilon, g_\varepsilon$ such that

$$f = f_\varepsilon + g_\varepsilon, \quad f_\varepsilon \in \mathcal{R}(R), \quad \overline{I}(|g_\varepsilon|) \le \varepsilon.$$

Show that $f \in \mathcal{R}(R)$ and

$$\int\limits_R f_\varepsilon \, dV \longrightarrow \int\limits_R f \, dV.$$

Hint. Use Exercise 18.

20. Use the result of Exercise 19 to produce another proof of Proposition 5.1.6.

21. Behind (5.1.108) is the assertion that if R is a cell, g is supported on $K \subset R$, and $|g| \le M$, then $\overline{I}(|g|) \le M \operatorname{cont}^+(K)$. Prove this. More generally, if $g, h : R \to \mathbb{R}$ are bounded and $|g| \le M$, show that $\overline{I}(|gh|) \le M\overline{I}(|h|)$.

22. Establish the following Fubini-type theorem, and compare it with Theorem 5.1.15.

Proposition 5.1.25. *Let $A \subset \mathbb{R}^m$ and $B \subset \mathbb{R}^n$ be cells, and take $f \in \mathcal{R}(A \times B)$. For $x \in A$, define $f_x : B \to \mathbb{R}$ by $f_x(y) = f(x, y)$. Define $L_f, U_f : A \to \mathbb{R}$ by*

$$L_f(x) = \underline{I}(f_x), \quad U_f(x) = \overline{I}(f_x).$$

Then L_f and U_f belong to $\mathcal{R}(A)$, and

$$\int\limits_{A\times B} f \, dV = \int\limits_A L_f(x) \, dx = \int\limits_A U_f(x) \, dx.$$

Hint. Given $\varepsilon > 0$, use Proposition 5.1.13 to take $\varphi, \psi \in \mathrm{PK}(A \times B)$ such that

$$\varphi \le f \le \psi, \quad \int \psi \, dV - \int \varphi \, dV < \varepsilon.$$

With definitions of φ_x and ψ_x analogous to that of f_x, show that

$$\begin{aligned} \int\limits_{A\times B} \varphi \, dV &= \int\limits_A \varphi_x \, dx \le \underline{I}(L_f) \\ &\le \overline{I}(U_f) \le \int\limits_A \psi_x \, dx = \int\limits_{A\times B} \psi \, dV. \end{aligned}$$

Deduce that

$$\underline{I}(L_f) = \overline{I}(U_f),$$

and proceed.

23. In Chapter 3 we defined π as half the length of the unit circle, which in turn

led to

$$\pi = \int_{-1}^{1} \frac{dx}{\sqrt{1-x^2}}.$$

See Exercise 27 in §3.2. In (5.1.57), we saw that the area of the unit disk is given by

$$A(D) = 2B, \quad B = \int_{-1}^{1} \sqrt{1-x^2}\, dx.$$

Take the following route to evaluating B. Use integration by parts to write

$$\begin{aligned} B &= -\int_{-1}^{1} x \frac{d}{dx}\sqrt{1-x^2}\, dx \\ &= \int_{-1}^{1} \frac{x^2}{\sqrt{1-x^2}}\, dx. \end{aligned}$$

In turn, write the last integral as

$$\begin{aligned} &\int_{-1}^{1} \Bigl[1-(1-x^2)\Bigr] \frac{dx}{\sqrt{1-x^2}} \\ &= \int_{-1}^{1} \frac{dx}{\sqrt{1-x^2}} - \int_{-1}^{1} \sqrt{1-x^2}\, dx = \pi - B, \end{aligned}$$

to conclude that

$$2B = \pi,$$

recovering the formula (5.1.59).

24. In (5.1.82)–(5.1.83), we saw that

$$V(B^4) = \beta_4 V(B^3), \quad \beta_4 = \int_{-1}^{1} (1-x^2)^{3/2}\, dx.$$

Take the following route to evaluating β_4. Use integration by parts to write

$$\begin{aligned} \beta_4 &= -\int_{-1}^{1} x \frac{d}{dx}(1-x^2)^{3/2}\, dx \\ &= 3\int_{-1}^{1} x^2(1-x^2)^{3/2}\, dx. \end{aligned}$$

As in Exercise 23, set $x^2 = 1-(1-x^2)$ to obtain

$$\beta_4 = 3\int_{-1}^{1} \sqrt{1-x^2}\, dx - 3\beta_4,$$

and proceed to identify β_4, hence $V(B^4)$.

25. Generalize the recursion in Exercise 24 to treat β_n for more general n. Compute $V(B^5)$ and $V(B^6)$.

Exercises on row reduction and matrix products

We consider the following three types of row operations on an $n \times n$ matrix $A = (a_{jk})$. If σ is a permutation of $\{1, \dots, n\}$, let

$$\rho_\sigma(A) = (a_{\sigma(j)k}).$$

If $c = (c_1, \dots, c_j)$, and all c_j are nonzero, set

$$\mu_c(A) = (c_j^{-1} a_{jk}).$$

Finally, if $c \in \mathbb{R}$ and $\mu \neq \nu$, define

$$\varepsilon_{\mu\nu c}(A) = (b_{jk}), \quad b_{\nu k} = a_{\nu k} - c a_{\mu k},\ b_{jk} = a_{jk} \ \text{ for } \ j \neq \nu.$$

We relate these operations to left multiplication by matrices P_σ, M_c, and $E_{\mu\nu c}$, defined by the following actions on the standard basis $\{e_1, \dots, e_n\}$ of $\mathbb{R}^n$:

$$P_\sigma e_j = e_{\sigma(j)}, \quad M_c e_j = c_j e_j,$$

and

$$E_{\mu\nu c} e_\mu = e_\mu + c e_\nu, \quad E_{\mu\nu c} e_j = e_j \ \text{ for } \ j \neq \mu.$$

1. Show that

$$A = P_\sigma \rho_\sigma(A), \quad A = M_c \mu_c(A), \quad A = E_{\mu\nu c} \varepsilon_{\mu\nu c}(A).$$

2. Show that $P_\sigma^{-1} = P_{\sigma^{-1}}$.

3. Show that, if $\mu \neq \nu$, then $E_{\mu\nu c} = P_\sigma^{-1} E_{21c} P_\sigma$, for some permutation σ.

4. If $B = \rho_\sigma(A)$ and $C = \mu_c(B)$, show that $A = P_\sigma M_c C$. Generalize this to other cases where a matrix C is obtained from a matrix A via a sequence of row operations.

5. If A is an invertible, real $n \times n$ matrix (i.e., $A \in Gl(n, \mathbb{R})$), then the rows of A form a basis of $\mathbb{R}^n$. Use this to show that A can be transformed to the identity matrix via a sequence of row operations. Deduce that any $A \in Gl(n, \mathbb{R})$ can be written as a finite product of matrices of the form P_σ, M_c and $E_{\mu\nu c}$, hence as a finite product of matrices of the form listed in (5.1.104).

5.2. Mean values of functions and centers of mass

Let $S \subset \mathbb{R}^n$ be a contented set, with positive volume. If $f \in \mathcal{R}(S)$, we define

$$\operatorname{Avg}_S(f) = \frac{1}{V(S)} \int\limits_S f\,dV, \tag{5.2.1}$$

and call this the *average* (or mean value) of f over S. If S is understood, we use the notation $\operatorname{Avg}(f)$. Other common notations include

$$\overline{f} = \langle f \rangle = \operatorname{Avg}(f). \tag{5.2.2}$$

If we wish to record the dependence on S, we use, e.g., $\langle f \rangle_S$.

As a variant of (5.2.1), we can place a “mass distribution” $\mu \ge 0$ on S, satisfying $\mu \in \mathcal{R}(S)$, $\int_S \mu\,dV > 0$,

$$\operatorname{Avg}_{S,\mu}(f) = \frac{1}{M_\mu(S)} \int\limits_S f\mu\,dV, \quad M_\mu(S) = \int\limits_S \mu\,dV. \tag{5.2.3}$$

this can be recovered from objects of the form (5.2.1), as

$$\operatorname{Avg}_{S,\mu}(f) = \frac{\operatorname{Avg}_S(f\mu)}{\operatorname{Avg}_S(\mu)}, \tag{5.2.4}$$

and we will not dwell upon this generalization.

In our study of (5.2.1), we will particularly be interested in what arises by taking $f(x) = x$, which defines the *center of mass* of S,

$$\operatorname{CM}(S) \in \mathbb{R}^n. \tag{5.2.5}$$

In this case, f takes values in $\mathbb{R}^n$. We find it convenient to characterize $\operatorname{CM}(S)$ by the following formula:

$$\begin{aligned} v \cdot \operatorname{CM}(S) &= \operatorname{Avg}_S(v \cdot x) \\ &= \frac{1}{V(S)} \int\limits_S v \cdot x\,dV, \quad \forall\, v \in \mathbb{R}^n. \end{aligned} \tag{5.2.6}$$

To take an example, let us take $a, b, c, h \in \mathbb{R},\ a < b,\ h > 0$, and consider the triangle T_l, with vertices at $(a, 0), (b, 0)$, and (c, h), which is a triangle with base $b - a$ and height h; see Figure 5.2.1. We see that

$$\begin{aligned} \int\limits_{T_l} y\,dA &= \int_0^h (b-a)\Bigl(1 - \frac{y}{h}\Bigr) y\,dy \\ &= (b-a) \int_0^h \Bigl[y - \frac{y^2}{h}\Bigr]\,dy \\ &= \frac{(b-a)}{6} h^2, \end{aligned} \tag{5.2.7}$$

while $A(T_l) = (b - a)h/2$, so

$$\operatorname{Avg}_{T_l}(y) = \frac{h}{3}. \tag{5.2.8}$$

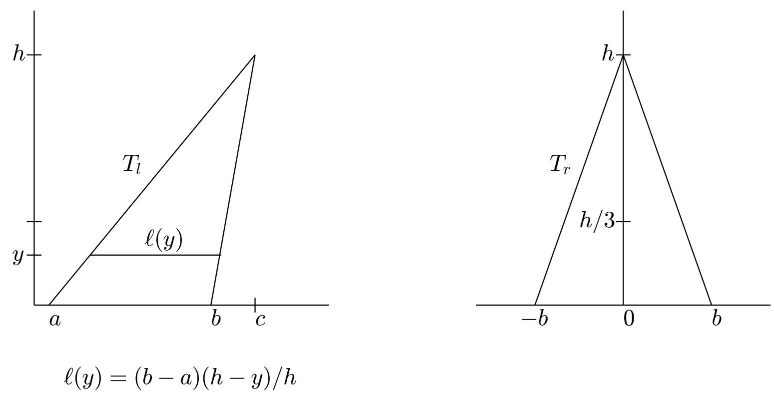

Figure 5.2.1. Triangles and their centers of mass

The calculation (5.2.8) specifies the y-component of $\mathrm{CM}(T_l)$. We want to completely specify this vector. Consider the special case T_r, pictured on the right side of Figure 5.2.1, with $a = -b,\ c = 0$. This is an isosceles triangle, having reflection symmetry across the y-axis, and it should be intuitively clear that the center of mass of T_r lies on this axis. This places

$$\mathrm{CM}(T_r) = \Bigl(0, \frac{h}{3}\Bigr). \tag{5.2.9}$$

Our next task is to come up with some mathematical results to back up this intuitive reasoning.

We can do this based on how $\mathrm{CM}(S)$ behaves when certain transformations are applied to S. First, there is translation:

$$\tau_v : \mathbb{R}^n \longrightarrow \mathbb{R}^n, \quad \tau_v(x) = x + v, \tag{5.2.10}$$

with $v \in \mathbb{R}^n$. It is easily verified that

$$\mathrm{CM}(\tau_v(S)) = \mathrm{CM}(S) + v. \tag{5.2.11}$$

Next we consider linear maps,

$$A : \mathbb{R}^n \longrightarrow \mathbb{R}^n, \quad A \in M(n, \mathbb{R}). \tag{5.2.12}$$

We will establish the following.

Proposition 5.2.1. *If $S \subset \mathbb{R}^n$ is a bounded, contented set, of positive volume, then*

$$A \in G\ell(n,\mathbb{R}) \Longrightarrow \mathrm{CM}(A(S)) = A\ \mathrm{CM}(S). \tag{5.2.13}$$

Proof. The change of variable formula gives

$$\int\limits_{A(S)} f(x)\,dV = |\det A| \int\limits_{S} f(Ax)\,dV, \tag{5.2.14}$$

for $f \in \mathcal{R}(A(S))$. Applying this to $f(x) = v \cdot x,\ v \in \mathbb{R}^n$, we have

$$\begin{aligned} \int\limits_{A(S)} v \cdot x\,dV &= |\det A| \int\limits_{S} v \cdot Ax\,dV \\ &= |\det A| \int\limits_{S} (A^t v) \cdot x\,dV. \end{aligned} \tag{5.2.15}$$

Divide both sides by $V(A(S))$. Since $V(A(S)) = |\det A| V(S)$, we obtain

$$v \cdot \mathrm{CM}(A(S)) = A^t v \cdot \mathrm{CM}(S), \quad \forall\, v \in \mathbb{R}^n, \tag{5.2.16}$$

which yields (5.2.13). □

The following corollary records how symmetries of a set S help locate its center of mass.

Corollary 5.2.2. *Let $S \subset \mathbb{R}^n$ be a bounded contented set of positive volume, and assume $A \in G\ell(n,\mathbb{R})$. Then*

$$A(S) = S \Longrightarrow A\ \mathrm{CM}(S) = \mathrm{CM}(S). \tag{5.2.17}$$

Proposition 5.2.1 and the associated corollary justify the reasoning behind the computation (5.2.9). It also leads to further results. Note that (5.2.9) is equivalent to

$$\mathrm{CM}(T_r) = \frac{1}{3}(a_1 + a_2 + a_3), \tag{5.2.18}$$

where $\{a_j : 1 \le j \le 3\}$ are the three vertices of T_r. This has the following generalization.

Proposition 5.2.3. *Consider*

$$T \subset \mathbb{R}^2, \ \text{triangle, with vertices } v_1, v_2, v_3, \tag{5.2.19}$$

and area $A(T) > 0$. Then

$$\mathrm{CM}(T) = \frac{1}{3}(v_1 + v_2 + v_3). \tag{5.2.20}$$

Proof. Applying (5.2.11), we can arrange that $v_1 = a_1$. Then $\{a_2 - a_1, a_3 - a_1\}$ and $\{v_2 - v_1, v_3 - v_1\}$ both form bases of $\mathbb{R}^2$, so there exists $A \in G\ell(2,\mathbb{R})$ taking the one basis to the other. In this way we get a transformation

$$X_{w,A} : \mathbb{R}^2 \longrightarrow \mathbb{R}^2, \quad X_{w,A}(x) = Ax + w, \tag{5.2.21}$$

such that

$$X_{w,A} v_j = a_j, \quad 1 \le j \le 3. \tag{5.2.22}$$

It follows that

$$X_{w,A}(T) = T_r, \tag{5.2.23}$$

and hence, by (5.2.11) and Proposition 5.2.1,

$$X_{w,A}\ \mathrm{CM}(T) = \mathrm{CM}(T_r). \tag{5.2.24}$$

On the other hand, (5.2.22) also implies

$$X_{w,A}\frac{v_1+v_2+v_3}{3} = \frac{a_1+a_2+a_3}{3}. \tag{5.2.25}$$

Thus (5.2.20) follows from (5.2.18). □

We move on to other classes of domains, and consider

$$B^n_+ = \{x \in B^n : x_n \ge 0\}, \tag{5.2.26}$$

where we recall that $B^n = \{x \in \mathbb{R}^n : |x| \le 1\}$ is the unit ball. We see that B^n_+ is invariant under rotation about the x_n-axis (and, for $n = 2$, invariant under reflection about the x_2-axis). Hence

$$\mathrm{CM}(B^n_+) \in \{(0, y) \in \mathbb{R}^n : 0 \le y \le 1\}. \tag{5.2.27}$$

Therefore, to compute the center of mass, it suffices to compute $\mathrm{Avg}_{B^n_+}(x_n)$. We have

$$\begin{aligned} \int\limits_{B^n_+} x_n\,dV &= \int\limits_{B^{n-1}} \Big(\int_0^{\sqrt{1-|x|^2}} x_n\,dx_n\Big)\,dx \\ &= \frac{1}{2}\int\limits_{B^{n-1}} (1-|x|^2)\,dx. \end{aligned} \tag{5.2.28}$$

We specialize to the cases $n = 2$ and $n = 3$, keeping in mind that, by previous calculations,

$$A(B^2_+) = \frac{\pi}{2}, \quad V(B^3_+) = \frac{2\pi}{3}. \tag{5.2.29}$$

For $n = 2$, the right side of (5.2.28) is

$$\frac{1}{2}\int_{-1}^{1} (1-x^2)\,dx = \frac{2}{3}, \tag{5.2.30}$$

so

$$\mathrm{Avg}_{B^2_+}(x_2) = \frac{4}{3\pi}, \quad \mathrm{CM}(B^2_+) = \Big(0, \frac{4}{3\pi}\Big). \tag{5.2.31}$$

For $n = 3$, the right side of (5.2.28) is

$$\begin{aligned} \frac{1}{2}\int\limits_{D} (1-|x|^2)\,dx &= \frac{1}{2}\int_0^{2\pi}\int_0^1 (1-r^2)r\,dr\,d\theta \\ &= \pi\int_0^1 (r-r^3)\,dr \\ &= \frac{\pi}{4}, \end{aligned} \tag{5.2.32}$$

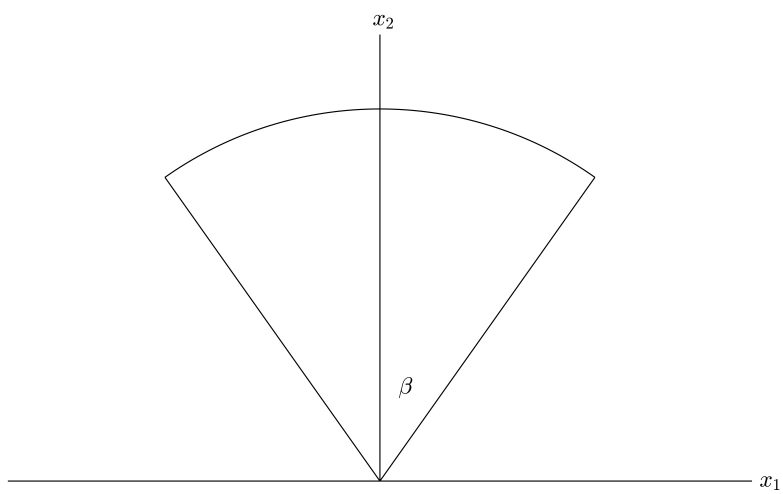

Figure 5.2.2. Pie slice

so

$$\operatorname{Avg}_{B^3_+}(x_3) = \frac{\pi}{4} \cdot \frac{3}{2\pi} = \frac{3}{8}, \quad \operatorname{CM}(B^3_+) = \Big(0, 0, \frac{3}{8}\Big). \tag{5.2.33}$$

See the exercises below for another approach to computing $\int_{B^n_+} x_n \, dV$. See also §6.1 for a method of computing

$$\int\limits_{B^{n-1}} \varphi(|x|) \, dx, \tag{5.2.34}$$

which is applicable to the right side of (5.2.28), for general n.

Generalizing the regions B^n_+, we take $\beta \in (0, \pi/2]$ and consider the sets

$$K^n_\beta = \{x = (x', x_n) \in B^n : x_n \ge (\cos\beta)|x'|\}. \tag{5.2.35}$$

See Figure 5.2.2 for an illustration in the case $n = 2$. The set K^2_β looks like a pie slice. The reader can imagine that K^3_β looks like a sno cone. Note that $K^n_{\pi/2} = B^n_+$. Again K^n_β has enough symmetry about the x_n-axis that we have

$$\operatorname{CM}(K^n_\beta) \in \{(0, y) : 0 < y < 1\}, \tag{5.2.36}$$

so we are left with the task of computing $\operatorname{Avg}_{K^n_\beta}(x_n)$. We carry this out for $n = 2, 3$.

In case $n = 2$, we use polar coordinates to write

$$A(K_\beta^2) = \int_{\pi/2-\beta}^{\pi/2+\beta} \int_0^1 r\, dr\, d\theta = \beta, \tag{5.2.37}$$

and

$$\begin{aligned} \int\limits_{K_\beta^2} x_2\, dA &= \int_{\pi/2-\beta}^{\pi/2+\beta} \int_0^1 r^2 \sin\theta\, dr\, d\theta \\ &= \frac{1}{3}(-\cos\theta)\Big|_{\pi/2-\beta}^{\pi/2+\beta} \\ &= \frac{2}{3}\sin\beta. \end{aligned} \tag{5.2.38}$$

Hence

$$\mathrm{Avg}_{K_\beta^2}(x_2) = \frac{2}{3}\,\frac{\sin\beta}{\beta}, \quad \mathrm{CM}(K_\beta^2) = \Big(0, \frac{2}{3}\,\frac{\sin\beta}{\beta}\Big). \tag{5.2.39}$$

In case $n = 3$, we use spherical polar coordinates to write

$$\int\limits_{K_\beta^3} f\, dV = \int_0^{2\pi} \int_0^\beta \int_0^1 f\, \rho^2 \sin\theta\, d\rho\, d\theta\, d\psi, \tag{5.2.40}$$

hence

$$\begin{aligned} V(K_\beta^3) &= 2\pi \int_0^\beta \int_0^1 \rho^2 \sin\theta\, d\rho\, d\theta \\ &= \frac{2\pi}{3} \int_0^\beta \sin\theta\, d\theta \\ &= \frac{2\pi}{3}(1 - \cos\beta). \end{aligned} \tag{5.2.41}$$

In addition,

$$\begin{aligned} \int\limits_{K_\beta^3} x_3\, dV &= 2\pi \int_0^\beta \int_0^1 \rho^3 \cos\theta \sin\theta\, d\rho\, d\theta \\ &= \frac{\pi}{2} \int_0^\beta \cos\theta \sin\theta\, d\theta \\ &= \frac{\pi}{4} \int_0^\beta \sin 2\theta\, d\theta \\ &= \frac{\pi}{8}(1 - \cos 2\beta). \end{aligned} \tag{5.2.42}$$

Hence

$$\mathrm{Avg}_{K_\beta^3}(x_3) = \frac{3}{2\pi} \cdot \frac{\pi}{8}\, \frac{1 - \cos 2\beta}{1 - \cos\beta} = \frac{3}{16}\, \frac{1 - \cos 2\beta}{1 - \cos\beta}. \tag{5.2.43}$$

Using the identity

$$1 - \cos 2\beta = 2\sin^2\beta, \tag{5.2.44}$$

we can rewrite the last quotient, and conclude that

$$\mathrm{CM}(K^3_\beta) = \Big(0, 0, \frac{3}{16}\Big(\frac{\sin\beta}{\sin\beta/2}\Big)^2\Big). \tag{5.2.45}$$

We have focused on bounded regions, but some unbounded regions have well defined centers of mass. To take a family of examples, we pick $a > 0$ and consider

$$\Omega = \{(x, y) \in \mathbb{R}^2 : x \geq 1, 0 \leq y \leq x^{-a}\}. \tag{5.2.46}$$

If $a > 1$, Ω has finite area,

$$A(\Omega) = \int_1^\infty x^{-a}\, dx = \frac{1}{a-1}. \tag{5.2.47}$$

In such a case,

$$\begin{aligned} \int_\Omega y\, dA &= \int_1^\infty \int_0^{x^{-a}} y\, dy\, dx \\ &= \frac{1}{2}\int_1^\infty x^{-2a}\, dx \\ &= \frac{1}{2(2a-1)}. \end{aligned} \tag{5.2.48}$$

Now $x|_\Omega$ is not bounded, so we need an extension of the results developed in §5.1. Such an extension is given in §5.3, one can peek ahead, or just see what is going on in the next calculation. For $x|_\Omega$ to have a finite integral, we will need to tighten our hypothesis on a to

$$a > 2. \tag{5.2.49}$$

Then

$$\begin{aligned} \int_\Omega x\, dA &= \lim_{R\to\infty} \int_1^R \int_0^{x^{-a}} x\, dy\, dx \\ &= \int_1^\infty x^{1-a}\, dx \\ &= \frac{1}{a-2}. \end{aligned} \tag{5.2.50}$$

We deduce that, as long as a satisfies (5.2.49), the center of mass of the domain Ω described in (5.2.46) is given by

$$\mathrm{CM}(\Omega) = \Big(\frac{a-1}{a-2}, \frac{a-1}{2(2a-1)}\Big). \tag{5.2.51}$$

We now bring in a calculation that ties in the center of mass of a region $\mathcal{O} \subset \mathbb{R}^2$ with the volume of the solid of revolution it generates. Here let

$$\mathcal{O} \subset \{(x, y) \in \mathbb{R}^2 : x > 0\}$$

be a smoothly bounded region, as illustrated in Figure 5.2.3, and let $\Omega \subset \mathbb{R}^3$ be the solid produced by rotating $\mathcal{O}$ about the y-axis in $\mathbb{R}^2$, so

$$\Omega = \{(x\cos\theta, y, x\sin\theta) : (x, y) \in \mathcal{O},\ \theta \in [0, 2\pi]\}. \tag{5.2.52}$$

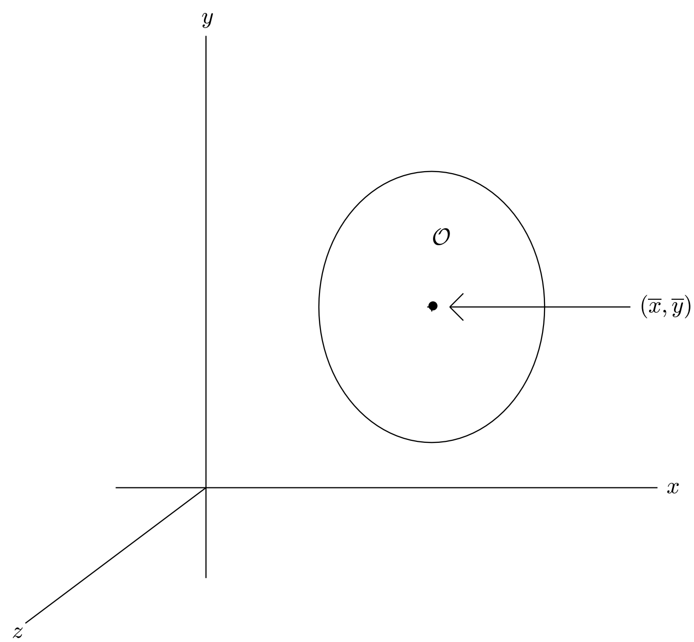

Figure 5.2.3. Setting for Pappus's theorem

That is to say, Ω is the image of

$$F : \mathcal{O} \times [0, 2\pi] \longrightarrow \mathbb{R}^3, \tag{5.2.53}$$

where

$$F(x, y, \theta) = (x \cos \theta, y, x \sin \theta). \tag{5.2.54}$$

A calculation, which we leave to the reader, gives

$$\det Df(x, y, \theta) = x. \tag{5.2.55}$$

Hence

$$\begin{aligned} V(\Omega) &= \int_0^{2\pi} \int_{\mathcal{O}} \det DF(x, y, \theta)\, dx\, dy\, d\theta \\ &= 2\pi \int_{\mathcal{O}} x\, dx\, dy. \end{aligned} \tag{5.2.56}$$

An alternative way to write this is

$$V(\Omega) = 2\pi \ \mathrm{Avg}_{\mathcal{O}}(x)\, A(\mathcal{O}). \tag{5.2.57}$$

This gives the following result, known as Pappus's theorem.

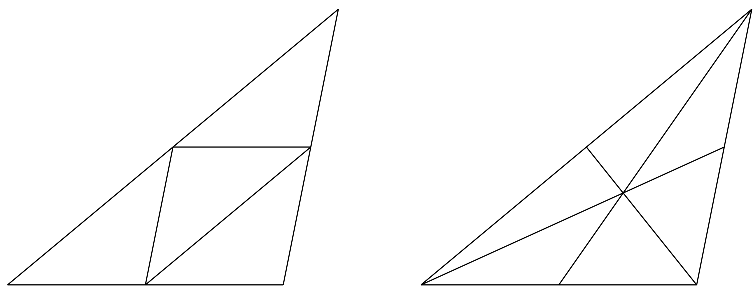

Figure 5.2.4. Triangles and their center of mass

Proposition 5.2.4. *Let $\mathcal{O} \subset \mathbb{R}^2$ be as described above, generating the solid of revolution about the y-axis $\Omega \subset \mathbb{R}^3$. Say*

$$\mathrm{CM}(\mathcal{O}) = (\overline{x}, \overline{y}). \tag{5.2.58}$$

Then

$$V(\Omega) = 2\pi\overline{x}\, A(\mathcal{O}). \tag{5.2.59}$$

Exercises

1. Let $T \subset \mathbb{R}^2$ be a triangle, and form $T' \subset T$, the triangle whose vertices consist of the midpoints of the edges of T. Show that

$$\mathrm{CM}(T') = \mathrm{CM}(T).$$

Hint. Use (5.2.20). (See the left half of Figure 5.2.4.)

2. If $T \subset \mathbb{R}^2$ is a triangle, and v is a vertex, show that $\mathrm{CM}(T)$ lies on the line segment from v to the midpoint of the edge opposite v. (See the right half of Figure 5.2.4.)

Hint. Reduce this to the case where T is isosceles, using reasoning parallel to the proof of Proposition 5.2.3.

3. Pursue the following approach to computing $\mathrm{Avg}_{B^n_+}(x_n)$. Start with

$$\int_{B^n_+} x_n\,dV = \int_0^1 V\Big(B^{n-1}_{\sqrt{1-x_n^2}}\Big) x_n\,dx_n$$
$$= V(B^{n-1}) \int_0^1 x_n (1-x_n^2)^{(n-1)/2}\,dx_n.$$

Use the substitution $s = x_n^2$ and show that

$$\mathrm{Avg}_{B^n_+}(x_n) = \frac{2}{n+1}\cdot\frac{1}{\beta_n}, \quad \beta_n = \frac{V(B^n)}{V(B^{n-1})}.$$

In cases $n = 2, 3$, compare this with the results (5.2.31) and (5.2.33).

4. Extending the scope of Exercise 3, for $a \in [-1, 1]$ set

$$B^n_{a+} = \{x \in B^n : x_n \ge a\},$$

and produce a formula for $\mathrm{CM}(B^n_{a+})$.

5. For $a > 0$, set

$$P^n_a = \{x = (x', x_n) \in \mathbb{R}^n : |x'|^2 \le x_n \le a^2\}.$$

Compute

$$\mathrm{CM}(P^n_a), \quad \text{for} \quad n = 2, 3.$$

6. For $\beta \in (0, \pi/2)$, consider the cone

$$C^n_\beta = \{x = (x', x_n) : (\cos\beta)|x'| \le x_n \le 1\}.$$

Compute

$$\mathrm{CM}(C^n_\beta), \quad \text{for} \quad n = 2, 3.$$

How does the result depend on β?

5.3. Unbounded integrable functions

There are lots of unbounded functions we would like to be able to integrate. For example, consider $f(x) = x^{-1/2}$ on $(0, 1]$ (defined any way you like at $x = 0$). Since, for $\varepsilon \in (0, 1)$,

$$\int_{\varepsilon}^{1} x^{-1/2}\, dx = 2 - 2\sqrt{\varepsilon}, \tag{5.3.1}$$

this has a limit as $\varepsilon \searrow 0$, and it is natural to set

$$\int_{0}^{1} x^{-1/2}\, dx = 2. \tag{5.3.2}$$

Sometimes (5.3.2) is called an "improper integral," but we do not consider that to be a proper designation. We aim for a treatment of the integral for a natural class of unbounded functions. To this end, we define a class $\mathcal{R}^{\#}(I)$ of not necessarily bounded "integrable" functions on I. The set I will stand for either $\mathbb{R}^n$ or a cell in $\mathbb{R}^n$.

To start, assume $f \geq 0$ on I, and for $A \in (0, \infty)$, set

$$\begin{aligned} f_A(x) = \ & f(x) \quad \text{if} \ \ f(x) \leq A, \\ & A, \quad \text{if} \ \ f(x) > A. \end{aligned} \tag{5.3.3}$$

(We hereby abandon the use of f_A as in the proof of Proposition 5.1.16.) We say $f \in \mathcal{R}^{\#}(I)$ provided

$$\begin{gathered} f_A \in \mathcal{R}(I), \quad \forall\, A < \infty, \ \ \text{and} \\ \exists \ \text{uniform bound} \ \int\limits_I f_A\, dV \leq M. \end{gathered} \tag{5.3.4}$$

If $f \geq 0$ satisfies (5.3.4), then $\int_I f_A\, dV$ increases monotonically to a finite limit as $A \nearrow +\infty$, and we call the limit $\int_I f\, dV$:

$$\int\limits_I f_A\, dV \nearrow \int\limits_I f\, dV, \quad \text{for} \ \ f \in \mathcal{R}^{\#}(I), \ f \geq 0. \tag{5.3.5}$$

If I is understood, we might just write $\int f\, dV$.

Remark. If $f \in \mathcal{R}(I)$ is ≥ 0, then $f_A \in \mathcal{R}(I)$ for all $A < \infty$. See the easy part of Exercise 15.

It is valuable to have the following.

Proposition 5.3.1. *If $f, g : I \to \mathbb{R}^+$ are in $\mathcal{R}^{\#}(I)$, then $f + g \in \mathcal{R}^{\#}(I)$, and*

$$\int\limits_I (f + g)\, dV = \int\limits_I f\, dV + \int\limits_I g\, dV. \tag{5.3.6}$$

Proof. To start, note that $(f + g)_A \leq f_A + g_A$. In fact,

$$(f + g)_A = (f_A + g_A)_A. \tag{5.3.7}$$

Hence $(f+g)_A \in \mathcal{R}(I)$ and $\int (f+g)_A \, dV \le \int f_A \, dV + \int g_A \, dV \le \int f \, dV + \int g \, dV$, so we have $f + g \in \mathcal{R}^\#(I)$ and

$$\int (f+g)\, dV \le \int f \, dV + \int g \, dV. \tag{5.3.8}$$

On the other hand, if $B > 2A$, then $(f+g)_B \ge f_A + g_A$, so

$$\int (f+g)\, dV \ge \int f_A \, dV + \int g_A \, dV, \tag{5.3.9}$$

for all $A < \infty$, and hence

$$\int (f+g)\, dV \ge \int f \, dV + \int g \, dV. \tag{5.3.10}$$

Together, (5.3.8) and (5.3.10) yield (5.3.6). □

Next, we take $f : I \to \mathbb{R}$ and set

$$f = f^+ - f^-, \quad f^+(x) = \begin{cases} f(x) & \text{if } f(x) \ge 0, \\ 0 & \text{if } f(x) < 0. \end{cases} \tag{5.3.11}$$

Then we say

$$f \in \mathcal{R}^\#(I) \Longleftrightarrow f^+, f^- \in \mathcal{R}^\#(I), \tag{5.3.12}$$

and set

$$\int_I f \, dV = \int_I f^+ \, dV - \int_I f^- \, dV, \tag{5.3.13}$$

where the two terms on the right are defined as in (5.3.5). To extend the additivity, we begin as follows

Lemma 5.3.2. *Assume that $g \in \mathcal{R}^\#(I)$ and that $g_j \ge 0$, $g_j \in \mathcal{R}^\#(I)$, and*

$$g = g_0 - g_1. \tag{5.3.14}$$

Then

$$\int g \, dV = \int g_0 \, dV - \int g_1 \, dV. \tag{5.3.15}$$

Proof. Take $g = g^+ - g^-$ as in (5.3.11). Then (5.3.14) implies

$$g^+ + g_1 = g_0 + g^-, \tag{5.3.16}$$

which by Proposition 5.1.24 yields

$$\int g^+ \, dV + \int g_1 \, dV = \int g_0 \, dV + \int g^- \, dV. \tag{5.3.17}$$

This implies

$$\int g^+ \, dV - \int g^- \, dV = \int g_0 \, dV - \int g_1 \, dV, \tag{5.3.18}$$

which yields (5.3.15). □

We now extend additivity.

Proposition 5.3.3. *Assume $f_1, f_2 \in \mathcal{R}^\#(I)$. Then $f_1 + f_2 \in \mathcal{R}^\#(I)$ and*

$$\int_I (f_1 + f_2)\, dV = \int_I f_1\, dV + \int_I f_2\, dV. \tag{5.3.19}$$

Proof. If $g = f_1 + f_2 = (f_1^+ - f_1^-) + (f_2^+ - f_2^-)$, then

$$g = g_0 - g_1, \quad g_0 = f_1^+ + f_2^+, \quad g_1 = f_1^- + f_2^-. \tag{5.3.20}$$

We have $g_j \in \mathcal{R}^\#(I)$, and then

$$\begin{aligned}\int (f_1 + f_2)\, dV &= \int g_0\, dV - \int g_1\, dV \\ &= \int (f_1^+ + f_2^+)\, dV - \int (f_1^- + f_2^-)\, dV \\ &= \int f_1^+\, dV + \int f_2^+\, dV - \int f_1^-\, dV - \int f_2^-\, dV,\end{aligned} \tag{5.3.21}$$

the first equality by Lemma 5.3.2, the second tautologically, and the third by Proposition 5.3.1. Since

$$\int f_j\, dV = \int f_j^+\, dV - \int f_j^-\, dV, \tag{5.3.22}$$

this gives (5.3.19). □

If $f : I \to \mathbb{C}$, we set $f = f_1 + i f_2,\ f_j : I \to \mathbb{R}$, and say $f \in \mathcal{R}^\#(I)$ if and only if f_1 and f_2 belong to $\mathcal{R}^\#(I)$. Then we set

$$\int f\, dV = \int f_1\, dV + i \int f_2\, dV. \tag{5.3.23}$$

Similar comments apply to $f : I \to \mathbb{R}^n$.

We next establish a useful result on products.

Proposition 5.3.4. *Assume $f \in \mathcal{R}^\#(\mathbb{R}^n)$, $g \in \mathcal{R}(\mathbb{R}^n)$, and $f, g \ge 0$. Then $fg \in \mathcal{R}^\#(\mathbb{R}^n)$ and*

$$\int f_A g\, dV \nearrow \int fg\, dV \quad \text{as} \quad A \nearrow +\infty. \tag{5.3.24}$$

Proof. Given the additivity properties just established, it would be equivalent to prove this with g replaced by $g + 1$, so we will assume from here that $g \ge 1$. Then

$$(fg)_A = (f_A g)_A. \tag{5.3.25}$$

By Proposition 5.1.22, $f_A g|_R \in \mathcal{R}(R)$ for each cell R. Hence (e.g., by the easy part of Exercise 15), $(f_A g)_A|_R \in \mathcal{R}(R)$ for each cell R. Thus

$$(fg)_A\big|_R \in \mathcal{R}(R). \tag{5.3.26}$$

Now there exists $K < \infty$ such that $1 \le g \le K$, so

$$f_A g \le K f_A, \quad \text{hence} \quad (fg)_A \le K f_A. \tag{5.3.27}$$

The hypothesis $f \in \mathcal{R}^\#(\mathbb{R}^n)$ implies there exists $M < \infty$ such that

$$\int_R f_A\, dV \le M, \tag{5.3.28}$$

for all $A < \infty$ and each cell R. Hence, by (5.3.27),

$$\sup_A \int_R (fg)_A \, dV \le MK, \tag{5.3.29}$$

independent of R. This implies $fg \in \mathcal{R}^\#(\mathbb{R}^n)$. By definition,

$$\int (fg)_A \, dV \nearrow \int fg \, dV, \quad \text{as} \ \ A \nearrow +\infty. \tag{5.3.30}$$

Meanwhile, clearly $f_A g \nearrow$ as $A \nearrow$, so the estimate (5.3.27) implies

$$\int f_A g \, dV \nearrow L, \quad \text{as} \ \ A \nearrow +\infty, \tag{5.3.31}$$

for some $L \in \mathbb{R}^+$. It remains to identify the limits in (5.3.30) and (5.3.31). Now (5.3.25) implies

$$(fg)_A \le f_A g, \quad \text{hence} \quad \int fg \, dV \le L. \tag{5.3.32}$$

Finally, since $f_A g \le fg$ and $f_A g \le KA$, we have

$$f_A g \le (fg)_B \ \ \text{for} \ \ B \ge KA. \tag{5.3.33}$$

This implies

$$L \le \sup_B \int (fg)_B \, dV = \int fg \, dV, \tag{5.3.34}$$

and hence we have (5.3.24). □

We now extend the change of variable formula in Theorem 5.1.20 to unbounded functions. It is convenient to introduce the following notation. Given an open set $\Omega \subset \mathbb{R}^n$, we say $f \in \mathcal{R}^\#_c(\Omega)$ provided $f \in \mathcal{R}^\#(\mathbb{R}^n)$ and f is supported on a compact subset of Ω.

Proposition 5.3.5. *Let $\mathcal{O}$ and Ω be open in $\mathbb{R}^n$, $G : \mathcal{O} \to \Omega$ a C^1 diffeomorphism. If $f \in \mathcal{R}^\#_c(\Omega)$, then $f \circ G \in \mathcal{R}^\#_c(\mathcal{O})$ and*

$$\int_\Omega f(y) \, dV(y) = \int_{\mathcal{O}} f(G(x)) |\det DG(x)| \, dV(x). \tag{5.3.35}$$

Proof. It suffices to establish this in case $f \ge 0$, which we assume from here. Then

$$\int_\Omega f_A \, dV \nearrow \int_\Omega f \, dV. \tag{5.3.36}$$

We set $\varphi = f \circ G$ and note that $f_A \circ G = \varphi_A$. Hence, by Theorem 5.1.20, for each $A \in (0, \infty)$,

$$\int_\Omega f_A(y) \, dV(y) = \int_{\mathcal{O}} \varphi_A(x) |\det DG(x)| \, dV(x). \tag{5.3.37}$$

If f is supported on a compact set $K \subset \Omega$, then φ_A is supported on $G^{-1}(K) \subset \mathcal{O}$, also compact, hence on which $|\det DG|$ has a positive lower bound. Hence an upper bound on the right side of (5.3.37) implies an upper bound on $\int \varphi_A \, dV$, independent of A, so $\varphi \in \mathcal{R}^{\#}(\mathbb{R}^n)$. Then Proposition 5.3.4 implies $\varphi |\det DG| \in \mathcal{R}^{\#}(\mathbb{R}^n)$ and

$$\int \varphi_A(x) |\det DG(x)| \, dV(x) \nearrow \int \varphi(x) |\det DG(x)| \, dV(x). \tag{5.3.38}$$

Together (5.3.36)–(5.3.38) yield (5.3.35). □

One also has versions of Proposition 5.3.5 where f need not have compact support. See Exercise 13 below for an example.

Our next result on a class of elements of $\mathcal{R}^{\#}(I)$ ties in closely with the example in (5.3.1). As before, I is either $\mathbb{R}^n$ or a cell in $\mathbb{R}^n$.

Proposition 5.3.6. *Let $f : I \to [0, \infty)$ and assume $f_A \in \mathcal{R}(I)$ for each $A < \infty$. Assume there are nested contented subsets of I:*

$$U_1 \supset U_2 \supset U_3 \supset \cdots, \quad V(U_\nu) \to 0. \tag{5.3.39}$$

Assume $f(1 - \chi_{U_\nu}) \in \mathcal{R}(I)$ for each ν and that there exists $C < \infty$ such that

$$\int_{I \setminus U_\nu} f \, dV = J_\nu \le C, \quad \forall \, \nu. \tag{5.3.40}$$

Then $f \in \mathcal{R}^{\#}(I)$ and

$$J_\nu \nearrow \int_I f \, dV. \tag{5.3.41}$$

Proof. The hypothesis (5.3.40) implies $J_\nu \nearrow J$ for some $J \in [0, \infty)$. Also, since $0 \le f_A \le f$, we have

$$\int_{I \setminus U_\nu} f_A \, dV \le J, \quad \forall \, v, A. \tag{5.3.42}$$

Furthermore,

$$\int_{U_\nu} f_A \, dV \le A V(U_\nu), \tag{5.3.43}$$

so

$$\int_I f_A \, dV \le J + A V(U_\nu), \quad \forall \, \nu, A, \tag{5.3.44}$$

hence

$$\int_I f_A \, dV \le J, \quad \forall A. \tag{5.3.45}$$

It follows that $f \in \mathcal{R}^{\#}(I)$ and

$$\int_I f \, dV \le J. \tag{5.3.46}$$

On the other hand,

$$\int\limits_I f\,dV \ge \int\limits_{I\setminus U_\nu} f\,dV = J_\nu, \tag{5.3.47}$$

for each ν, so we have (5.3.41). □

Monotone convergence theorem

We aim to establish a circle of results known as monotone convergence theorems. Here is the first result (which will be strengthened in Proposition 5.3.9).

Proposition 5.3.7. *Let $R \subset \mathbb{R}^n$ be a cell. Assume $f_k \in \mathcal{R}(R)$. Then*

$$f_k(x) \searrow 0 \quad \forall\, x \in R \Longrightarrow \int\limits_R f_k\,dV \searrow 0. \tag{5.3.48}$$

Proof. It suffices to assume $V(R) = 1$. Say $0 \le f_1 \le K$ on R, so also $0 \le f_k \le K$. We have

$$\int\limits_R f_k\,dV \searrow \alpha, \tag{5.3.49}$$

for some $\alpha \ge 0$, and we want to show that $\alpha = 0$. Suppose $\alpha > 0$. Pick a partition $\mathcal{P}_k$ of R such that $\underline{I}_{\mathcal{P}_k}(f_k) \ge \alpha/2$. Thus $f_k \ge \varphi_k \ge 0$ for some $\varphi_k \in \mathrm{PK}(R)$, constant on the interior of each cell in $\mathcal{P}_k$, with integral $\ge \alpha/2$. The contribution to $\int_R \varphi_k\,dV$ from the cells on which $\varphi_k \le \alpha/4$ is $\le \alpha/4$, so the contribution from the cells on which $\varphi_k \ge \alpha/4$ must be $\ge \alpha/4$. Since $\varphi_k \le K$ on R, it follows that the latter class of cells must have total volume $\ge \alpha/4K$. Consequently, for each k, there exists $S_k \subset R$, a finite union of cells in $\mathcal{P}_k$, such that

$$V(S_k) \ge \frac{\alpha}{4K}, \quad \text{and} \quad f_k \ge \frac{\alpha}{4} \quad \text{on} \quad S_k. \tag{5.3.50}$$

Then $f_\ell \ge \alpha/4$ on S_k for all $\ell \le k$. Hence, with

$$\mathcal{O}_\ell = \bigcup_{k \ge \ell} S_k, \tag{5.3.51}$$

we have

$$\mathrm{cont}^-(\mathcal{O}_\ell) \ge \frac{\alpha}{4K}, \quad f_\ell \ge \frac{\alpha}{4} \text{ on } \mathcal{O}_\ell. \tag{5.3.52}$$

The hypothesis $f_\ell \searrow 0$ on R implies

$$\mathcal{O}_\ell \searrow \emptyset \quad \text{as } \ell \nearrow \infty. \tag{5.3.53}$$

Without loss of generality, we can take S_k open in (5.3.50), hence each $\mathcal{O}_\ell$ is open. The conclusion of Proposition 5.3.7 is hence a consequence of the following, which implies that (5.3.52) and (5.3.53) are contradictory. □

Proposition 5.3.8. *If $\mathcal{O}_\ell \subset R$ are open sets, for $\ell \in \mathbb{N}$, then*

$$\mathcal{O}_\ell \searrow \emptyset \Longrightarrow \mathrm{cont}^-(\mathcal{O}_\ell) \searrow 0. \tag{5.3.54}$$

Proof. Assume $\mathcal{O}_\ell \searrow \emptyset$. If the conclusion of (5.3.54) fails, then

$$\operatorname{cont}^-(\mathcal{O}_\ell) \searrow b \tag{5.3.55}$$

for some $b > 0$. Passing to a subsequence if necessary, we can assume

$$\operatorname{cont}^-(\mathcal{O}_\ell) \le b + \delta_\ell, \quad \delta_\ell < 2^{-\ell} \cdot 10^{-9} \cdot b. \tag{5.3.56}$$

Then we can pick $K_\ell \subset \mathcal{O}_\ell$, a compact union of finitely many cells in a partition of R, such that

$$V(K_\ell) \ge b - \delta_\ell. \tag{5.3.57}$$

We claim that $\cap_\ell K_\ell \neq \emptyset$, which will provide a contradiction.

Place $K_1 \cup K_2$ in a finite union $\mathcal{C}_1$ of cells, contained in $\mathcal{O}_1$. We then have

$$\begin{aligned} V(K_1 \cap K_2) &\ge V(K_1) - V(\mathcal{C}_1 \setminus K_2) \\ &\ge b - (2\delta_1 + \delta_2), \end{aligned} \tag{5.3.58}$$

since $V(\mathcal{C}_1 \setminus K_2) = V(\mathcal{C}_1) - V(K_2) \le \operatorname{cont}^-(\mathcal{O}_1) - V(K_2) \le \delta_1 + \delta_2$. Next, place $(K_1 \cap K_2) \cup K_3$ in a finite union $\mathcal{C}_2$ of cells, contained in $\mathcal{O}_2$. Then

$$\begin{aligned} V(K_1 \cap K_2 \cap K_3) &\ge V(K_1 \cap K_2) - V(\mathcal{C}_2 \setminus K_3) \\ &\ge b - (2\delta_1 + \delta_2) - (2\delta_2 + \delta_3), \end{aligned} \tag{5.3.59}$$

since $V(\mathcal{C}_2 \setminus K_3) = V(\mathcal{C}_2) - V(K_3) \le \operatorname{cont}^-(\mathcal{O}_2) - V(K_3) \le \delta_2 + \delta_3$. Proceeding in this fashion, we get

$$V\Bigl(\bigcap_{\ell=1}^{k} K_\ell\Bigr) \ge b - \sum_{\ell=1}^{k} (2\delta_\ell + \delta_{\ell+1}) > 0, \quad \forall\, k. \tag{5.3.60}$$

Thus, $\widetilde{K}_k = \cap_{\ell=1}^k K_\ell$ is a decreasing sequence of nonempty compact sets. Hence

$$\bigcap_{\ell \ge 1} \mathcal{O}_\ell \supset \bigcap_{\ell \ge 1} K_\ell \neq \emptyset, \tag{5.3.61}$$

contradicting the hypothesis of (5.3.54). ☐

Having Proposition 5.3.7, we proceed to the following significant improvement.

Proposition 5.3.9. *Assume $f_k \in \mathcal{R}^\#(R)$. Then*

$$f_k(x) \searrow 0 \ \ \forall\, x \in R \Longrightarrow \int_R f_k \, dV \searrow 0. \tag{5.3.62}$$

Proof. Again we have (5.3.49) for some $\alpha \ge 0$ and again we want to show that $\alpha = 0$. For each $A \in (0, \infty)$ and each $k \in \mathbb{N}$, form $(f_k)_A$, as in (5.3.3). Thus $(f_k)_A \in \mathcal{R}(R)$, and the hypothesis of (5.3.62) implies $(f_k)_A \searrow 0$ as $k \nearrow \infty$. Thus, by Proposition 5.3.7,

$$\int_R (f_k)_A \, dV \searrow 0 \ \text{ as } \ k \to \infty, \ \text{ for each } \ A < \infty. \tag{5.3.63}$$

We note that

$$f_{k+1}(x) - (f_{k+1})_A(x) \le f_k(x) - (f_k)_A(x) \tag{5.3.64}$$

for each $x \in R$, $k \in \mathbb{N}$. In fact, if $f_k(x) \le A$ (so $f_{k+1}(x) \le A$), both sides of (5.3.64) are 0, if $f_{k+1}(x) \ge A$ (so $f_k(x) \ge A$), we get $f_{k+1}(x) - A \le f_k(x) - A$, and if $f_{k+1}(x) < A < f_k(x)$, we get $0 \le f_k(x) - A$. It follows that, for each $A < \infty$,

$$\int_R [f_k - (f_k)_A]\, dV \searrow \alpha, \quad \text{as} \quad k \to \infty. \tag{5.3.65}$$

However, for each $\delta > 0$, there exists $A = A(\delta) < \infty$ such that $\int_R [f_1 - (f_1)_A]\, dV \le \delta$. This forces $\alpha = 0$, and proves Proposition 5.3.9. ☐

Applying Proposition 5.3.9 to $f_k = g - g_k$, we have the following.

Corollary 5.3.10. *Assume $g, g_k \in \mathcal{R}^\#(R)$. Then*

$$g_k(x) \nearrow g(x) \quad \forall\, x \in R \Longrightarrow \int_R g_k\, dV \nearrow \int_R g\, dV. \tag{5.3.66}$$

Finally, we remove the support constraint.

Proposition 5.3.11. *Assume $g, g_k \in \mathcal{R}^\#(\mathbb{R}^n)$. Then*

$$g_k(x) \nearrow g(x) \quad \forall\, x \in \mathbb{R}^n \Longrightarrow \int_{\mathbb{R}^n} g_k\, dV \nearrow \int_{\mathbb{R}^n} g\, dV. \tag{5.3.67}$$

Proof. Clearly

$$\int_{\mathbb{R}^n} g_k\, dV \nearrow c, \quad \text{and} \quad c \le \int_{\mathbb{R}^n} g\, dV. \tag{5.3.68}$$

Now, given $\varepsilon > 0$, there is a cell $R \subset \mathbb{R}^n$ such that

$$\int_{\mathbb{R}^n \setminus R} (|g| + |g_1|)\, dV < \varepsilon, \tag{5.3.69}$$

and Corollary 5.3.10 gives

$$\int_R g_k\, dV \nearrow \int_R g\, dV. \tag{5.3.70}$$

We deduce that $c \ge \int_{\mathbb{R}^n} g\, dV - \varepsilon$ for all $\varepsilon > 0$, so (4.137) holds. ☐

In the Lebesgue theory of integration, there is a stronger result. Namely, if g_k are integrable on $\mathbb{R}^n$ and $g_k(x) \nearrow g(x)$ for each x, and if there is a uniform upper bound $\int_{\mathbb{R}^n} g_k\, dx \le B < \infty$, then g is integrable on $\mathbb{R}^n$ and the conclusion of (5.3.67) holds. Such a result can be found in [**16**].

Exercises

Given $B \subset \mathbb{R}^n$ and $f : B \to \mathbb{R}$, we say $f \in \mathcal{R}^\#(B)$ provided $g \in \mathcal{R}^\#(\mathbb{R}^n)$, where

$$\begin{aligned} g(x) = f(x) &\quad \text{for} \;\; x \in B, \\ 0 &\quad \text{for} \;\; x \notin B. \end{aligned}$$

1. Take $B = \{x \in \mathbb{R}^n : |x| \le 1/2\}$, and let $f : B \to \mathbb{R}^+$. Assume f is continuous on $B \setminus 0$. Show that

$$f \in \mathcal{R}^\#(B) \Leftrightarrow \int\limits_{|x|>\varepsilon} f\,dV \;\;\text{is bounded as}\;\; \varepsilon \searrow 0.$$

2. With $B \subset \mathbb{R}^n$ as in Exercise 1, define $q_b : B \to \mathbb{R}$ by

$$q_b(x) = \frac{1}{|x|^n \big|\log|x|\,\big|^b},$$

for $x \neq 0$. Say $q_b(0) = 0$. Show that $q_b \in \mathcal{R}^\#(B) \Leftrightarrow b > 1$.

3. Show that

$$f(x) = |x|^{-a} e^{-|x|^2} \in \mathcal{R}^\#(\mathbb{R}^n) \Longleftrightarrow a < n.$$

4. Compute

$$\int\limits_B \frac{\log|x|}{|x|^{n-1}}\,dV.$$

Hint. See Exercise 6 of §1.4.

5. Compute

$$\int\limits_{\mathbb{R}^n} |x|^{-a} e^{-|x|^2}\,dx,$$

for $a = n-1,\; n-2$.

6. Peek ahead to §6.1 and express the integral in Exercise 5 above in terms of the Gamma function, for general $a < n$.

7. Take

$$T = \{(x,y) \in \mathbb{R}^2 : 0 < y \le x \le 1\},$$

and define

$$f_a : T \to \mathbb{R}, \quad f_a(x,y) = x^{-a}.$$

Determine for which $a \in \mathbb{R}^+$ we have $f_a \in \mathcal{R}^\#(T)$, and compute

$$\int\limits_T x^{-a}\,dx\,dy,$$

for such a.

8. Let $D = \{x \in \mathbb{R}^2 : |x| < 1\}$. Compute

$$\int\limits_D (1 - |x|^2)^{-a} \, dA,$$

for $a < 1$. Also compute

$$\int\limits_D x_j (1 - |x|^2)^{-a} \, dA,$$

$$\int\limits_D x_1 x_2 (1 - |x|^2)^{-a} \, dA,$$

$$\int\limits_D x_j^2 (1 - |x|^2)^{-a} \, dA.$$

Hint. Use symmetries. Show that the last integral is independent of j, and sum over j.

5.4. Outer measure and Riemann integrability

Given a bounded set $S \subset \mathbb{R}^n$, its upper content is defined in (5.1.15) and an equivalent characterization given in (5.1.17). A related quantity is the *outer measure* of S, defined by

$$m^*(S) = \inf\Bigl\{\sum_{k\geq 1} V(R_k) : R_k \subset \mathbb{R}^n \text{ cells}, S \subset \bigcup_{k\geq 1} R_k\Bigr\}. \tag{5.4.1}$$

The difference between (5.1.17) and (5.4.1) is that in (5.1.17) we require the cover of S by cells to be finite and in (5.4.1) we allow any *countable* cover of S by cells. Clearly (5.4.1) is an inf over a larger collection of objects than (5.1.17), so

$$m^*(S) \leq \text{cont}^+(S). \tag{5.4.2}$$

We get the same result in (5.4.1) if we require

$$S \subset \bigcup_{k\geq 1} \overset{\circ}{R}_k \tag{5.4.3}$$

(just expand each R_k by a factor of $(1+2^{-k}\varepsilon)$). Since any open cover of a compact set has a finite subcover (see Proposition 2.1.8), it follows that

$$S \text{ compact } \Longrightarrow m^*(S) = \text{cont}^+(S). \tag{5.4.4}$$

On the other hand, it is readily verified from (5.4.1) that

$$S \text{ countable } \Longrightarrow m^*(S) = 0. \tag{5.4.5}$$

For example, if $R = \{x \in \mathbb{R}^n : 0 \leq x_j \leq 1, \ \forall\, j\}$, then

$$m^*(R \cap \mathbb{Q}^n) = 0, \quad \text{but} \quad \text{cont}^+(R \cap \mathbb{Q}^n) = 1, \tag{5.4.6}$$

the latter result by (5.1.18).

We now establish the following integrability criterion, which sharpens Proposition 5.1.6.

Proposition 5.4.1. *Let $f : R \to \mathbb{R}$ be bounded, and let $S \subset R$ be the set of points of discontinuity of f. Then*

$$m^*(S) = 0 \Longrightarrow f \in \mathcal{R}(R). \tag{5.4.7}$$

Proof. Assume $|f| \leq M$ and pick $\varepsilon > 0$. Take a countable collection $\{R_k\}$ of cells that are open (in R) such that $S \subset \cup_{k\geq 1} R_k$ and $\sum_{k\geq 1} V(R_k) < \varepsilon$. Now f is continuous at each $p \in R \setminus S$, so there exists a cell $R_p^\#$, open (in R), containing p, such that $\sup_{R_p^\#} f - \inf_{R_p^\#} f < \varepsilon$. Then $\{R_k : k \in \mathbb{N}\} \cup \{R_p^\# : p \in R \setminus S\}$ is an open cover of R. Since R is compact, there is a finite subcover, which we denote $\{R_1, \dots, R_N, R_1^\#, \dots, R_M^\#\}$. We have

$$\sum_{k=1}^{N} V(R_k) < \varepsilon, \quad \text{and} \quad \sup_{R_j^\#} f - \inf_{R_j^\#} f < \varepsilon, \quad \forall\, j \in \{1, \dots, M\}. \tag{5.4.8}$$

Recall that $R = I_1 \times \cdots \times I_n$ is a product of n closed, bounded intervals. Also each cell R_k and $R_j^\#$ is a product of intervals. For each $\nu \in \{1, \dots, n\}$, take the collection of all endpoints in the νth factor of each of these cells, and use these to form a partition of I_ν. Taking products yields a partition $\mathcal{P}$ of R. We can write

$$\begin{aligned} \mathcal{P} &= \{L_k : 1 \le k \le \mu\} \\ &= \Big(\bigcup_{k\in\mathcal{A}} L_k\Big) \cup \Big(\bigcup_{k\in\mathcal{B}} L_k\Big), \end{aligned} \tag{5.4.9}$$

where we say $k \in \mathcal{A}$ provided L_k is contained in a cell of the form $R_j^\#$ for some $j \in \{1, \dots, M\}$, as in (5.4.8). Consequently, if $k \in \mathcal{B}$, then $L_k \subset R_\ell$ for some $\ell \in \{1, \dots, N\}$, so

$$\bigcup_{k\in\mathcal{B}} L_k \subset \bigcup_{\ell=1}^{N} R_\ell. \tag{5.4.10}$$

We therefore have

$$\sum_{k\in\mathcal{B}} V(L_k) < \varepsilon, \quad \text{and} \quad \sup_{L_j} f - \inf_{L_j} f < \varepsilon, \quad \forall\, j \in \mathcal{A}. \tag{5.4.11}$$

It follows that

$$\begin{aligned} 0 \le \overline{I}_{\mathcal{P}}(f) - \underline{I}_{\mathcal{P}}(f) &< \sum_{k\in\mathcal{B}} 2MV(L_k) + \sum_{j\in\mathcal{A}} \varepsilon V(L_j) \\ &< 2\varepsilon M + \varepsilon V(R). \end{aligned} \tag{5.4.12}$$

Since ε can be taken arbitrarily small, this establishes that $f \in \mathcal{R}(R)$. $\square$

Remark. The condition (5.4.7) is sharp. That is, given $f : R \to \mathbb{R}$ bounded, $f \in \mathcal{R}(R) \Leftrightarrow m^*(S) = 0$. Proofs of this can be found in standard measure theory texts, such as [**16**].

Chapter 6

Calculus on surfaces

Having developed differential and integral calculus on open sets in n-dimensional Euclidean space, we now pursue notions of calculus on a higher level, for surfaces in $\mathbb{R}^n$, and more generally for a class of objects known as "manifolds."

In §6.1 we define the notion of a smooth m-dimensional surface in $\mathbb{R}^n$ and study properties of these objects. We associate to such a surface a "metric tensor," and make use of this to define the integral of functions on a surface. This includes the study of surface area. Examples include the computation of areas of higher dimensional spheres. We also explore integration on the group of rotations on $\mathbb{R}^n$, leading to the notion of "averaging over rotations." In this section, we see that the inverse function and implicit function theorems from §4.3 are of crucial importance for differential calculus on surfaces, and the change of variable formula from §5.1 is crucial for integral calculus on surfaces.

In §6.2 we discuss constrained maxima and minima, that is, extremal points for a smooth function $f : S \to \mathbb{R}$, where $S \subset \mathbb{R}^n$ is a smooth surface. We bring in the method of Lagrange multipliers to find these relative extrema.

In §6.3, we establish some important integral identities due to Gauss, Green, and Stokes. This class of identities can be thought of as the natural expression of the fundamental theorem of calculus in several variables. Here they are derived for domains $\Omega \subset \mathbb{R}^n$ (specializing to $n = 2$ for Green's formula and to $n = 3$ for Stokes' formula). They will be studied on a much more general level in the following chapter.

In §6.4, we introduce a class of objects more general than surfaces, called manifolds. Manifolds can also be endowed with metric tensors. These are called Riemannian manifolds, and one can again define the integral of functions.

We also have a section on polar decomposition of matrices, used to prove that the set $Gl_+(n, \mathbb{R})$ of $n \times n$ real matrices with positive determinant is connected, and a section on partitions of unity, useful to localize analysis on an n-dimensional surface (or manifold) to analysis on an open subset of $\mathbb{R}^n$.

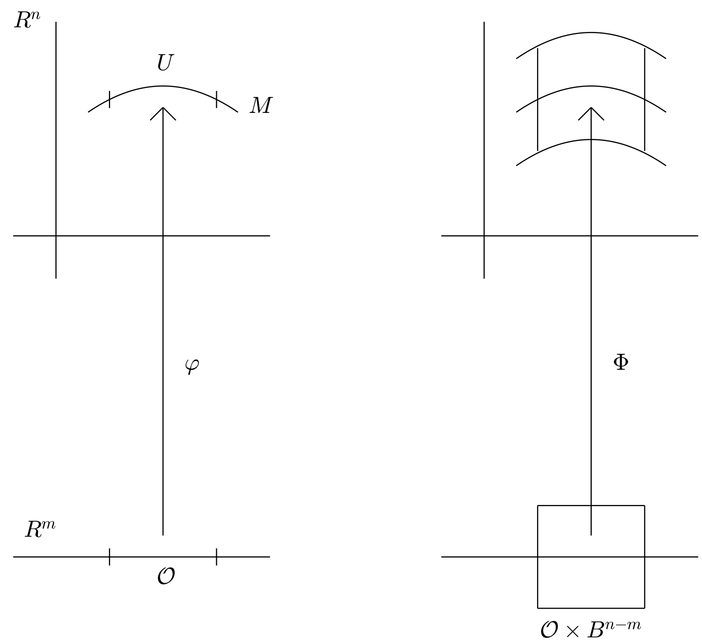

Figure 6.1.1. A coordinate chart φ and its thickening Φ

6.1. Surfaces and surface integrals

A smooth m-dimensional surface $M \subset \mathbb{R}^n$ is characterized by the following property. Given $p \in M$, there is a neighborhood U of p in M and a smooth map $\varphi : \mathcal{O} \to U$, from an open set $\mathcal{O} \subset \mathbb{R}^m$ bijectively to U, with injective derivative at each point, and continuous inverse $\varphi^{-1} : U \to \mathcal{O}$. Such a map φ is called a *coordinate chart* on M. We call $U \subset M$ a coordinate patch. If all such maps φ are smooth of class C^k, we say M is a surface of class C^k.

There is an abstraction of the notion of a surface, namely the notion of a *manifold*, which we will discuss in §6.4. Examples include projective spaces and other spaces obtained as quotients of surfaces.

If $\varphi : \mathcal{O} \to U$ is a C^k coordinate chart, such as described above, or more generally $\varphi : \mathcal{O} \to \mathbb{R}^n$ is a C^k map with injective derivative, and $\varphi(x_0) = p$, we set

$$T_pM = \operatorname{Range} D\varphi(x_0), \tag{6.1.1}$$

a linear subspace of $\mathbb{R}^n$ of dimension m, and we denote by N_pM its orthogonal complement. It is useful to consider the following map. Pick a linear isomorphism $A : \mathbb{R}^{n-m} \to N_pM$, let $B^{n-m} \subset \mathbb{R}^{n-m}$ be the unit ball, and define

$$\Phi : \mathcal{O} \times B^{n-m} \longrightarrow \mathbb{R}^n, \quad \Phi(x, z) = \varphi(x) + Az. \tag{6.1.2}$$

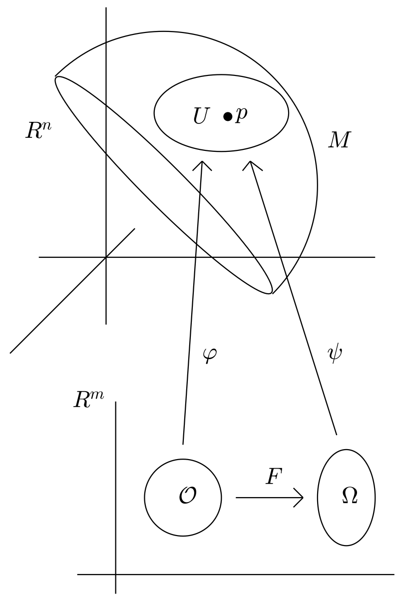

Figure 6.1.2. Coordinate charts

Thus Φ is a C^k map defined on an open subset of $\mathbb{R}^n$. We call Φ a *thickening* of φ. See Figure 6.1.1 for an illustration, with $m = 1,\ n = 2$. Note that

$$D\Phi(x_0, 0)\begin{pmatrix} v \\ w \end{pmatrix} = D\varphi(x_0)v + Aw, \tag{6.1.3}$$

so $D\Phi(x_0, 0) : \mathbb{R}^n \to \mathbb{R}^n$ is surjective, hence bijective, so the Inverse Function Theorem applies; Φ maps some neighborhood of $(x_0, 0)$ diffeomorphically onto a neighborhood of $p \in \mathbb{R}^n$.

Suppose there is another C^k coordinate chart, $\psi : \Omega \to U$. Since φ and ψ are by hypothesis one-to-one and onto, it follows that

$$F = \psi^{-1} \circ \varphi : \mathcal{O} \to \Omega \tag{6.1.4}$$

is a well defined map, which is one-to-one and onto. See Figure 6.1.2.

Example. Take the unit disk $D = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 < 1\}$, and define

$$\varphi : D \longrightarrow S^2, \quad \varphi(x, y) = (x, y, \sqrt{1 - x^2 - y^2}). \tag{6.1.5}$$

If we take $p = (0, 0, 1)$, then (6.1.2) becomes

$$\Phi(x, y, z) = (x, y, z + \sqrt{1 - x^2 - y^2}), \quad z \in (-1, 1). \tag{6.1.6}$$

Note that φ maps D one-to-one and onto the upper half of S^2, i.e., $\{(x,y,z) \in S^2 : z > 0\}$. Similarly one has a map

$$\psi : D \longrightarrow S^2, \quad \psi(u,v) = (\sqrt{1-u^2-v^2}, u, v), \tag{6.1.7}$$

which takes D one-to-one and onto the hemisphere $\{(x,y,z) \in S^2 : x > 0\}$. If we take

$$\mathcal{O} = \{(x,y) \in D : x > 0\}, \quad \Omega = \{(u,v) \in D : v > 0\}, \tag{6.1.8}$$

then

$$\varphi : \mathcal{O} \longrightarrow U, \quad \psi : \Omega \longrightarrow U, \tag{6.1.9}$$

where U is the intersection of these two hemishperes, i.e., $U = \{(x,y,z) \in S^2 : x > 0, z > 0\}$. In this case, we have $F : \mathcal{O} \to \Omega$, defined by $F(x,y) = (u,v)$. We see that $u = y,\ v = \sqrt{1-x^2-y^2}$, so

$$F(x,y) = (y, \sqrt{1-x^2-y^2}). \tag{6.1.10}$$

Note: if we want $p \in U$, we might take $p = (1,0,1)/\sqrt{2}$, and adjust the thickening map Φ accordingly.

Returning to generalities, we see from (6.1.4) that F and F^{-1} are continuous. In fact, we can say more.

Lemma 6.1.1. *Under the hypotheses above, F is a C^k diffeomorphism.*

Proof. It suffices to show that F and F^{-1} are C^k on a neighborhood of x_0 and y_0, respectively, where $\varphi(x_0) = \psi(y_0) = p$. Let us define a map Ψ in a fashion similar to (6.1.2). To be precise, we set $\widetilde{T}_pM = \operatorname{Range} D\psi(y_0)$, and let $\widetilde{N}_pM$ be its orthogonal complement. (Shortly we will show that $\widetilde{T}_pM = T_pM$, but we are not quite ready for that.) Then pick a linear isomorphism $B : \mathbb{R}^{n-m} \to \widetilde{N}_pM$ and consider

$$\Psi : \Omega \times \mathbb{R}^{n-m} \longrightarrow \mathbb{R}^n, \quad \Psi(y,z) = \psi(y) + Bz. \tag{6.1.11}$$

Again, Ψ is a C^k diffeomorphism from a neighborhood of $(y_0,0)$ onto a neighborhood of p. To be precise, there exist neighborhoods $\widetilde{\mathcal{O}}$ of $(x_0,0)$ in $\mathcal{O} \times \mathbb{R}^{n-m}$, $\widetilde{\Omega}$ of $(y_0,0)$ in $\Omega \times \mathbb{R}^{n-m}$, and $\widetilde{U}$ of p in $\mathbb{R}^n$ such that

$$\Phi : \widetilde{\mathcal{O}} \longrightarrow \widetilde{U}, \quad \text{and} \quad \Psi : \widetilde{\Omega} \longrightarrow \widetilde{U} \tag{6.1.12}$$

are C^k diffeomorphisms.

It follows that $\Psi^{-1} \circ \Phi : \widetilde{\mathcal{O}} \to \widetilde{\Omega}$ is a C^k diffeomorphism. Now note that, for $(x,0) \in \widetilde{\mathcal{O}}$ and $(y,0) \in \widetilde{\Omega}$,

$$\Psi^{-1} \circ \Phi(x,0) = \big(F(x),0\big), \quad \Phi^{-1} \circ \Psi(y,0) = \big(F^{-1}(y),0\big). \tag{6.1.13}$$

In fact, to verify the first identity in (6.1.13), we check that

$$\begin{aligned} \Psi(F(x),0) &= \psi(F(x)) + B0 \\ &= \psi(\psi^{-1} \circ \varphi(x)) \\ &= \varphi(x) \\ &= \Phi(x,0). \end{aligned} \tag{6.1.14}$$

The identities in (6.1.13) imply that F and F^{-1} have the desired regularity. $\square$

Thus, when there are two such coordinate charts, $\varphi : \mathcal{O} \to U,\ \psi : \Omega \to U$, we have a C^k diffeomorphism $F : \mathcal{O} \to \Omega$ such that

$$\varphi = \psi \circ F. \tag{6.1.15}$$

By the chain rule,

$$D\varphi(x) = D\psi(y)\, DF(x), \quad y = F(x). \tag{6.1.16}$$

In particular this implies that Range $D\varphi(x_0) =$ Range $D\psi(y_0)$, so T_pM in (6.1.1) is independent of the choice of coordinate chart. It is called the *tangent space* to M at p.

Remark. An application of the inverse function theorem related to the proof of Lemma 6.1.1 can be used to show that if $\mathcal{O} \subset \mathbb{R}^m$ is open, $m < n$, and $\varphi : \mathcal{O} \to \mathbb{R}^n$ is a C^k map such that $D\varphi(p) : \mathbb{R}^m \to \mathbb{R}^n$ is injective, $(p \in \mathcal{O})$, then there is a neighborhood $\widetilde{\mathcal{O}}$ of p in $\mathcal{O}$ such that the image of $\widetilde{\mathcal{O}}$ under φ is a C^k surface in $\mathbb{R}^n$. Compare Exercise 11 in §4.3.

Metric tensors

We next define an object called the *metric tensor* on M. Given a coordinate chart $\varphi : \mathcal{O} \to U$, there is associated an $m \times m$ matrix $G(x) = \big(g_{jk}(x)\big)$ of functions on $\mathcal{O}$, defined in terms of the inner product of vectors tangent to M :

$$g_{jk}(x) = D\varphi(x)e_j \cdot D\varphi(x)e_k = \frac{\partial \varphi}{\partial x_j} \cdot \frac{\partial \varphi}{\partial x_k} = \sum_{\ell=1}^{n} \frac{\partial \varphi_\ell}{\partial x_j} \frac{\partial \varphi_\ell}{\partial x_k}, \tag{6.1.17}$$

where $\{e_j : 1 \le j \le m\}$ is the standard orthonormal basis of $\mathbb{R}^m$. Equivalently,

$$G(x) = D\varphi(x)^t\, D\varphi(x). \tag{6.1.18}$$

We call (g_{jk}) the metric tensor of M, on U, with respect to the coordinate chart $\varphi : \mathcal{O} \to U$. Note that this matrix is positive-definite. From a coordinate-independent point of view, the metric tensor on M specifies inner products of vectors tangent to M, using the inner product of $\mathbb{R}^n$.

If we take another coordinate chart $\psi : \Omega \to U$, we want to compare (g_{jk}) with $H = (h_{jk})$, given by

$$h_{jk}(y) = D\psi(y)e_j \cdot D\psi(y)e_k, \quad \text{i.e.,} \quad H(y) = D\psi(y)^t\ D\psi(y). \tag{6.1.19}$$

As seen above we have a diffeomorphism $F : \mathcal{O} \to \Omega$ such that (6.1.15)–(6.1.16) hold. Consequently,

$$G(x) = DF(x)^t\, H(y)\, DF(x), \quad \text{for} \ \ y = F(x), \tag{6.1.20}$$

or equivalently,

$$g_{jk}(x) = \sum_{i,\ell} \frac{\partial F_i}{\partial x_j} \frac{\partial F_\ell}{\partial x_k}\, h_{i\ell}(y). \tag{6.1.21}$$

Surface integrals

We now define the notion of surface integral on M. If $f : M \to \mathbb{R}$ is a continuous function supported on U, we set

$$\int_M f\,dS = \int_{\mathcal{O}} f \circ \varphi(x)\,\sqrt{g(x)}\;dx, \tag{6.1.22}$$

where

$$g(x) = \det G(x). \tag{6.1.23}$$

We need to know that this is independent of the choice of coordinate chart $\varphi : \mathcal{O} \to U$. Thus, if we use $\psi : \Omega \to U$ instead, we want to show that (6.1.22) is equal to $\int_\Omega f \circ \psi(y)\,\sqrt{h(y)}\;dy$, where $h(y) = \det H(y)$. Indeed, since $f \circ \psi \circ F = f \circ \varphi$, we can apply the change of variable formula, Theorem 5.1.20, to get

$$\int_\Omega f \circ \psi(y)\,\sqrt{h(y)}\;dy = \int_{\mathcal{O}} f \circ \varphi(x)\,\sqrt{h(F(x))}\;|\det\,DF(x)|\;dx. \tag{6.1.24}$$

Now, (6.1.20) implies that

$$\sqrt{g(x)} = |\det\,DF(x)|\;\sqrt{h(y)}, \tag{6.1.25}$$

so the right side of (6.1.24) is seen to be equal to (6.1.22), and our surface integral is well defined, at least for f supported in a coordinate patch. More generally, if $f : M \to \mathbb{R}$ has compact support, write it as a finite sum of terms, each supported on a coordinate patch, and use (6.1.22) on each patch. Using (5.1.13), one readily verifies that

$$\int_M (f_1 + f_2)\,dS = \int_M f_1\,dS + \int_M f_2\,dS, \tag{6.1.26}$$

if $f_j : M \to \mathbb{R}$ are continuous functions with compact support.

Let us pause to consider the special cases $m = 1$ and $m = 2$. For $m = 1$, we are considering a curve in $\mathbb{R}^n$, say $\varphi : [a,b] \to \mathbb{R}^n$. Then $G(x)$ is a 1×1 matrix, namely $G(x) = |\varphi'(x)|^2$. If we denote the curve in $\mathbb{R}^n$ by γ, rather than M, the formula (6.1.22) becomes the *arc length* integral (compare (3.1.15))

$$\int_\gamma f\;ds = \int_a^b f \circ \varphi(x)\;|\varphi'(x)|\;dx. \tag{6.1.27}$$

In case $m = 2$, let us consider a surface $M \subset \mathbb{R}^3$, with a coordinate chart $\varphi : \mathcal{O} \to U \subset M$. For f supported in U, an alternative way to write the surface integral is

$$\int_M f\;dS = \int_{\mathcal{O}} f \circ \varphi(x)\;|\partial_1\varphi \times \partial_2\varphi|\;dx_1 dx_2, \tag{6.1.28}$$

where $u \times v$ is the cross product of vectors u and v in $\mathbb{R}^3$. To see this, we compare this integrand with the one in (6.1.22). In this case,

$$g = \det\begin{pmatrix} \partial_1\varphi\cdot\partial_1\varphi & \partial_1\varphi\cdot\partial_2\varphi \\ \partial_2\varphi\cdot\partial_1\varphi & \partial_2\varphi\cdot\partial_2\varphi \end{pmatrix} = |\partial_1\varphi|^2|\partial_2\varphi|^2 - (\partial_1\varphi\cdot\partial_2\varphi)^2. \tag{6.1.29}$$

Recall from (2.5.11) that $|u \times v| = |u|\ |v|\ |\sin\theta|$, where θ is the angle between u and v. Equivalently, since $u \cdot v = |u|\ |v|\ \cos\theta$,

$$|u \times v|^2 = |u|^2|v|^2\big(1 - \cos^2\theta\big) = |u|^2|v|^2 - (u \cdot v)^2. \tag{6.1.30}$$

Thus we see that $|\partial_1\varphi \times \partial_2\varphi| = \sqrt{g}$, in this case, and (6.1.28) is equivalent to (6.1.22). See Exercises 15–21 for applications of (6.1.28).

An important class of surfaces is the class of graphs of smooth functions. Let $u \in C^1(\Omega)$, for an open $\Omega \subset \mathbb{R}^{n-1}$, and let M be the graph of $z = u(x)$. The map $\varphi(x) = \big(x, u(x)\big)$ provides a natural coordinate system, in which the metric tensor formula (6.1.17) becomes

$$g_{jk}(x) = \delta_{jk} + \frac{\partial u}{\partial x_j}\frac{\partial u}{\partial x_k}. \tag{6.1.31}$$

If u is C^1, we see that g_{jk} is continuous. To calculate $g = \det(g_{jk})$, at a given point $p \in \Omega$, if $\nabla u(p) \neq 0$, rotate coordinates so that $\nabla u(p)$ is parallel to the x_1 axis. We obtain

$$\sqrt{g} = \big(1 + |\nabla u|^2\big)^{1/2}. \tag{6.1.32}$$

(See Exercise 31 for another take on this formula.) In particular, the $(n-1)$-dimensional volume of the surface M is given by

$$V_{n-1}(M) = \int\limits_M dS = \int\limits_\Omega \big(1 + |\nabla u(x)|^2\big)^{1/2}\, dx. \tag{6.1.33}$$

Particularly important examples of surfaces are the unit spheres S^{n-1} in $\mathbb{R}^n$,

$$S^{n-1} = \{x \in \mathbb{R}^n : |x| = 1\}. \tag{6.1.34}$$

Spherical polar coordinates on $\mathbb{R}^n$ are defined in terms of a smooth diffeomorphism

$$R : (0,\infty) \times S^{n-1} \longrightarrow \mathbb{R}^n \setminus 0, \quad R(r,\omega) = r\omega. \tag{6.1.35}$$

Let $(h_{\ell m})$ denote the metric tensor on S^{n-1} (induced from its inclusion in $\mathbb{R}^n$) with respect to some coordinate chart $\varphi : \mathcal{O} \to U \subset S^{n-1}$. Then we have a coordinate chart $\Phi : (0,\infty) \times \mathcal{O} \to \mathcal{U} \subset \mathbb{R}^n$ given by $\Phi(r,y) = r\varphi(y)$. Take $y_0 = r,\ y = (y_1, \dots, y_{n-1})$. In the coordinate system Φ the Euclidean metric tensor (e_{jk}) is given by

$$\begin{aligned} e_{00} &= \partial_0\Phi \cdot \partial_0\Phi = \varphi(y) \cdot \varphi(y) = 1, \\ e_{0j} &= \partial_0\Phi \cdot \partial_j\Phi = \varphi(y) \cdot \partial_j\varphi(y) = 0, \quad 1 \le j \le n-1, \\ e_{jk} &= r^2\partial_j\varphi \cdot \partial_k\varphi = r^2 h_{jk}, \quad 1 \le j,k \le n-1. \end{aligned}$$

The fact that $\varphi(y)\cdot\partial_j\varphi(y) = 0$ follows by applying $\partial/\partial y_j$ to the identity $\varphi(y)\cdot\varphi(y) \equiv 0$. To summarize,

$$\big(e_{jk}\big) = \begin{pmatrix} 1 & \\ & r^2 h_{\ell m} \end{pmatrix}. \tag{6.1.36}$$

Now (6.1.36) yields

$$\sqrt{e} = r^{n-1}\sqrt{h}. \tag{6.1.37}$$

We therefore have the following result for integrating a function in spherical polar coordinates.

$$\int\limits_{\mathbb{R}^n} f(x)\,dx = \int\limits_{S^{n-1}} \Bigl[\int_0^\infty f(r\omega)r^{n-1}\,dr\Bigr]\,dS(\omega). \tag{6.1.38}$$

See (5.1.137)–(5.1.141) for the case $n = 3$ (with special coordinates on S^2).

We next compute the $(n-1)$-dimensional area A_{n-1} of the unit sphere $S^{n-1} \subset \mathbb{R}^n$, using (6.1.38) together with the computation

$$\int\limits_{\mathbb{R}^n} e^{-|x|^2}\,dx = \pi^{n/2}, \tag{6.1.39}$$

from (5.1.156). First note that, whenever $f(x) = \varphi(|x|)$, (6.1.38) yields

$$\int\limits_{\mathbb{R}^n} \varphi(|x|)\,dx = A_{n-1}\int_0^\infty \varphi(r)r^{n-1}\,dr. \tag{6.1.40}$$

In particular, taking $\varphi(r) = e^{-r^2}$ and using (6.1.39), we have

$$\pi^{n/2} = A_{n-1}\int_0^\infty e^{-r^2}r^{n-1}\,dr = \tfrac{1}{2}A_{n-1}\int_0^\infty e^{-s}s^{n/2-1}\,ds, \tag{6.1.41}$$

where we used the substitution $s = r^2$ to get the last identity. We hence have

$$A_{n-1} = \frac{2\pi^{n/2}}{\Gamma(\frac{n}{2})}, \tag{6.1.42}$$

where $\Gamma(z)$ is Euler's Gamma function, defined for $z > 0$ by

$$\Gamma(z) = \int_0^\infty e^{-s}s^{z-1}\,ds. \tag{6.1.43}$$

We need to complement (6.1.42) with some results on $\Gamma(z)$ allowing a computation of $\Gamma(n/2)$ in terms of more familiar quantities. Of course, setting $z = 1$ in (6.1.43), we immediately get

$$\Gamma(1) = 1. \tag{6.1.44}$$

Also, setting $n = 1$ in (6.1.41), we have

$$\pi^{1/2} = 2\int_0^\infty e^{-r^2}\,dr = \int_0^\infty e^{-s}s^{-1/2}\,ds,$$

or

$$\Gamma\Bigl(\frac{1}{2}\Bigr) = \pi^{1/2}. \tag{6.1.45}$$

We can proceed inductively from (6.1.44)-(6.1.45) to a formula for $\Gamma(n/2)$ for any $n \in \mathbb{Z}^+$, using the following.

Lemma 6.1.2. *For all* $z > 0$,

$$\Gamma(z+1) = z\Gamma(z). \tag{6.1.46}$$

Proof. We can write

$$\Gamma(z+1) = -\int_0^\infty \Bigl(\frac{d}{ds}e^{-s}\Bigr)s^z\, ds = \int_0^\infty e^{-s}\,\frac{d}{ds}\bigl(s^z\bigr)\, ds,$$

the last identity by integration by parts. The last expression here is seen to equal the right side of (6.1.46). □

Consequently, for $k \in \mathbb{Z}^+$,

$$\Gamma(k) = (k-1)!, \quad \Gamma\Bigl(k+\frac{1}{2}\Bigr) = \Bigl(k-\frac{1}{2}\Bigr)\cdots\Bigl(\frac{1}{2}\Bigr)\pi^{1/2}. \tag{6.1.47}$$

Thus (6.1.42) can be rewritten

$$A_{2k-1} = \frac{2\pi^k}{(k-1)!}, \quad A_{2k} = \frac{2\pi^k}{\bigl(k-\frac{1}{2}\bigr)\cdots\bigl(\frac{1}{2}\bigr)}. \tag{6.1.48}$$

The rotation group, and averaging over rotations

We discuss another important example of a smooth surface, in the space $M(n,\mathbb{R}) \approx \mathbb{R}^{n^2}$ of real $n\times n$ matrices, namely

$$SO(n) = \{T \in M(n,\mathbb{R}) : T^tT = I,\ \det T > 0\}, \tag{6.1.49}$$

(hence $\det T = 1$). To obtain a coordinate system, we bring in the exponential map,

$$\operatorname{Exp} : M(n,\mathbb{R}) \longrightarrow M(n,\mathbb{R}), \tag{6.1.50}$$

defined by

$$\operatorname{Exp}(X) = e^X = \sum_{k=0}^\infty \frac{1}{k!}X^k. \tag{6.1.51}$$

As noted in (4.3.31)–(4.3.33), Exp is smooth and

$$D\operatorname{Exp}(0)Y = Y, \quad \forall\, Y \in M(n,\mathbb{R}), \tag{6.1.52}$$

Hence the Inverse Function Theorem implies that there is a ball Ω centered at 0 in $M(n,\mathbb{R})$ that is mapped diffeomorphically by Exp onto a neighborhood $\widetilde{\Omega}$ of I in $M(n,\mathbb{R})$. Now we have the identities

$$\operatorname{Exp} X^t = (\operatorname{Exp} X)^t, \quad \operatorname{Exp}(-X) = (\operatorname{Exp} X)^{-1}, \tag{6.1.53}$$

for all $X \in M(n,\mathbb{R})$ (see (C.4.7)), and these imply that

$$\operatorname{Exp} : \operatorname{Skew}(n) \longrightarrow SO(n), \tag{6.1.54}$$

where

$$\operatorname{Skew}(n) = \{X \in M(n,\mathbb{R}) : X^t = -X\}. \tag{6.1.55}$$

Since $\operatorname{Exp} : \Omega \to \widetilde{\Omega}$ is a diffeomorphism, we have, for $X \in \Omega,\ A = \operatorname{Exp} X \in \widetilde{\Omega}$,

$$A \in SO(n) \Longleftrightarrow X \in \operatorname{Skew}(n). \tag{6.1.56}$$

Thus there is a neighborhood $\mathcal{O}$ of 0 in $\operatorname{Skew}(n)$ that is mapped by Exp diffeomorphically onto a smooth surface $U \subset M(n,\mathbb{R})$, of dimension $m = n(n-1)/2$. Furthermore, U is a neighborhood of I in $SO(n)$.

For general $T \in SO(n)$, we can define maps

$$\varphi_T : \mathcal{O} \longrightarrow SO(n), \quad \varphi_T(A) = T\ \mathrm{Exp}(A), \tag{6.1.57}$$

and obtain coordinate charts in $SO(n)$, which is consequently a smooth surface of dimension $n(n-1)/2$ in $M(n, \mathbb{R})$. Note that $SO(n)$ is a closed bounded subset of $M(n, \mathbb{R})$; hence it is compact. We call $SO(n)$ the *rotation group* on $\mathbb{R}^n$

Note that, by (6.1.52)–(6.1.56), the tangent space to $SO(n)$ at the identity element I is

$$T_I SO(n) = \mathrm{Skew}(n). \tag{6.1.58}$$

Hence, for $U \in SO(n)$,

$$\begin{aligned} T_U SO(n) &= \{UA : A \in \mathrm{Skew}(n)\} \\ &= \{\widetilde{A}U : \widetilde{A} \in \mathrm{Skew}(n)\}. \end{aligned} \tag{6.1.59}$$

We use the inner product on $M(n, \mathbb{R})$ computed componentwise; equivalently,

$$\langle A, B \rangle = \mathrm{Tr}\,(B^t A) = \mathrm{Tr}\,(BA^t). \tag{6.1.60}$$

See §2.4. This produces a metric tensor on $SO(n)$. The surface integral over $SO(n)$ has the following important invariance property.

Proposition 6.1.3. *Given $f \in C\big(SO(n)\big)$, if we set*

$$\rho_T f(X) = f(XT), \quad \lambda_T f(X) = f(TX), \tag{6.1.61}$$

for $T, X \in SO(n)$, we have

$$\int\limits_{SO(n)} \rho_T f\ dS = \int\limits_{SO(n)} \lambda_T f\ dS = \int\limits_{SO(n)} f\ dS. \tag{6.1.62}$$

Proof. Given $T \in SO(n)$, the maps $R_T, L_T : M(n, \mathbb{R}) \to M(n, \mathbb{R})$ defined by $R_T(X) = XT,\ L_T(X) = TX$ are easily seen from (6.1.60) to be isometries. Thus they yield maps of $SO(n)$ to itself which preserve the metric tensor, proving (6.1.62). ☐

Since $SO(n)$ is compact, its total volume $V\big(SO(n)\big) = \int_{SO(n)} 1\ dS$ is finite. We define the integral with respect to "Haar measure"

$$\int\limits_{SO(n)} f(g)\, dg = \frac{1}{V\big(SO(n)\big)} \int\limits_{SO(n)} f\, dS. \tag{6.1.63}$$

This is used in many arguments involving "averaging over rotations."

Extended notion of coordinates

Basic calculus as developed in this text so far has involved maps from one Euclidean space to another, of the type $F : \mathbb{R}^n \to \mathbb{R}^m$. It is convenient and useful to extend our setting to $F : V \to W$, where V and W are general finite-dimensional real vector spaces. There is the following notion of the derivative.

Let V and W be as above, and let $\Omega \subset V$ be open. We say $F : \Omega \to W$ is differentiable at $x \in \Omega$ provided there exists a linear map $L : V \to W$ such that, for $y \in V$ small,

$$F(x+y) = F(x) + Ly + r(x,y), \tag{6.1.64}$$

with $r(x,y) \to 0$ faster than $y \to 0$, i.e.,

$$\frac{\|r(x,y)\|}{\|y\|} \longrightarrow 0 \quad \text{as} \quad y \to 0. \tag{6.1.65}$$

For this to be meaningful, we need *norms* on V and W. Often these norms come from inner products. See Appendix C.1 for a discussion of inner product spaces. If (6.1.64)–(6.1.65) hold, we set $DF(x) = L$, and call the linear map

$$DF(x) : V \longrightarrow W$$

the derivative of F at x. We say F is C^1 if $DF(x)$ is continuous in x. Notions of F in C^k are produced in analogy with the situation in §4.1. Of course, we can reduce all this to the setting of §4.1 by picking bases of V and W.

Often such V and W arise as linear subspaces of $\mathbb{R}^n$, such as T_pM in (6.1.1), or $V = N_pM$, mentioned right below that. As noted there, we can take a linear isomorphism of such V with $\mathbb{R}^k$ for some k, and keep working in the context of maps between such standard Euclidean spaces, as in (6.1.2). However, it can be convenient to avoid this distraction, and, for example, replace (6.1.2) by

$$\Phi : \mathcal{O} \times N_pM \longrightarrow \mathbb{R}^n, \quad \Phi(x,z) = \varphi(x) + z, \tag{6.1.66}$$

and (6.1.3) by

$$D\Phi(x_0,0)\begin{pmatrix} v \\ w \end{pmatrix} = D\varphi(x_0)v + w. \tag{6.1.67}$$

In order to carry out Lemma 6.1.1 in this setting, we want the following version of the Inverse Function Theorem.

Proposition 6.1.4. *Let V and W be real vector spaces, each of dimension n. Let F be a C^k map from an open neighborhood Ω of $p_0 \in V$ to W, with $q_0 = F(p_0)$, $k \geq 1$. Assume the derivative*

$$DF(p_0) : V \to W \quad \textit{is an isomorphism.}$$

Then there exist a neighborhood U of p_0 and a neighborhood $\widetilde{U}$ of q_0 such that $F : U \to \widetilde{U}$ is one-to-one and onto, and $F^{-1} : \widetilde{U} \to U$ is a C^k map.

While Proposition 6.1.4 is apparently an extension of Theorem 4.3.1, there is no extra work required to prove it. One can simply take linear isomorphisms $A : \mathbb{R}^n \to V$ and $B : \mathbb{R}^n \to W$ and apply Theorem 4.3.1 to the map $G(x) = B^{-1}F(Ax)$. Thus Proposition 6.1.4 is not a technical improvement of Theorem 4.3.1, but it is a useful conceptual extension.

With this in mind, we can define the notion of an m-dimensional surface $M \subset V$ (an n-dimensional vector space) as follows. Take a vector space W, of dimension m. Given $p \in M$, we require there to be a neighborhood U of p in M and a smooth map $\varphi : \mathcal{O} \to U$, from an open set $\mathcal{O} \subset W$ bijectively to U, with an injective derivative at each point. We call such a map a coordinate chart. If all such maps are smooth of class C^k, we say M is a surface of class C^k. As a further wrinkle, we could take different vector spaces W_p for different $p \in M$, as long as they all have dimension m. The reader is invited to formulate the appropriate modification of Lemma 6.1.1 in this setting.

Submersions

Let V and W be finite dimensional real vector spaces, $\Omega \subset V$ open, and $F : \Omega \to W$ a C^k map, $k \geq 1$. We say F is a *submersion* provided that, for each $x \in \Omega$, $DF(x) : V \to W$ is surjective. (This requires $\dim V \geq \dim W$.) We establish the following Submersion Mapping Theorem, which the reader might recognize as a variant of the Implicit Function Theorem. In the statement, $\ker T$ denotes the null space

$$\ker T = \{v \in V : Tv = 0\},$$

if $T : V \to W$ is a linear transformation.

Proposition 6.1.5. *With V, W, and $\Omega \subset V$ as above, assume $F : \Omega \to W$ is a C^k map, $k \geq 1$. Fix $p \in W$, and consider*

$$S = \{x \in V : F(x) = p\}. \tag{6.1.68}$$

Assume that, for each $x \in S$, $DF(x) : V \to W$ is surjective. Then S is a C^k surface in Ω. Furthermore, for each $x \in S$,

$$T_x S = \ker DF(x). \tag{6.1.69}$$

Proof. Given $q \in S$, set $K_q = \ker DF(q)$ and define

$$G_q : V \longrightarrow W \oplus K_q, \quad G_q(x) = (F(x), P_q(x - q)), \tag{6.1.70}$$

where P_q is a projection of V onto K_q. Note that

$$G_q(q) = (F(q), 0) = (p, 0). \tag{6.1.71}$$

Also

$$DG_q(x) = (DF(x), P_q), \quad x \in V. \tag{6.1.72}$$

We claim that

$$DG_q(q) = (DF(q), P_q) : V \to W \oplus K_q \text{ is an isomorphism.} \tag{6.1.73}$$

This is a special case of the following general observation. □

Lemma 6.1.6. *If $A : V \to W$ is a surjective linear map and P is a projection of V onto $\ker A$, then*

$$(A, P) : V \longrightarrow W \oplus \ker A \quad \textit{is an isomorphism.} \tag{6.1.74}$$

We postpone the proof of this lemma and proceed with the proof of Proposition 6.1.5. Having (6.1.73), we can apply the Inverse Function Theorem (Proposition 6.1.4) to obtain a neighborhood U of q in V and a neighborhood $\mathcal{O}$ of $(p, 0)$ in $W \oplus K_q$ such that $G_q : U \to \mathcal{O}$ is bijective, with C^k inverse

$$G_q^{-1} : \mathcal{O} \longrightarrow U, \quad G_q^{-1}(p, 0) = q. \tag{6.1.75}$$

By (6.1.70), given $x \in U$,

$$x \in S \Longleftrightarrow G_q(x) = (p, v), \text{ for some } v \in K_q. \tag{6.1.76}$$

Hence $S \cap U$ is the image under the C^k diffeomorphism G_q^{-1} of $\mathcal{O} \cap \{(p, v) : v \in K_q\}$. Hence S is smooth of class C^k and $\dim T_q S = \dim K_q$. It follows from the chain rule that $T_q S \subset K_q$, so the dimension count yields $T_q S = K_q$. This proves Proposition 6.1.5. Note that we have the following coordinate chart on a neighborhood of $q \in S$:

$$\psi_q(v) = G_q^{-1}(p, v), \quad \psi_q : \Omega_q \to S, \tag{6.1.77}$$

where Ω_q is a neighborhood of 0 in $T_q S = K_q = \ker DF(q)$.

It remains to prove Lemma 6.1.6. Indeed, given that $A : V \to W$ is surjective, the fundamental theorem of linear algebra implies $\dim V = \dim(W \oplus \ker A)$, and it is clear that (A, P) in (6.1.74) is injective, so the isomorphism property follows.

Remark. In case $V = \mathbb{R}^n$ and $W = \mathbb{R}$, $DF(x)$ is typically denoted $\nabla F(x)$, the hypothesis on $DF(x)$ becomes $\nabla F(x) \neq 0$, and (6.1.69) is equivalent to the assertion that $\dim S = n - 1$ and, for $x \in S$,

$$\nabla F(x) \perp T_x S. \tag{6.1.78}$$

Compare the discussion following Proposition 4.3.6.

Example. Take $F : \mathbb{R}^n \to \mathbb{R}$ to be $F(x) = |x|^2$, so the unit sphere S^{n-1} is given by

$$S^{n-1} = \{x \in \mathbb{R}^n : F(x) = 1\}. \tag{6.1.79}$$

We have

$$\nabla F(x) = 2x, \tag{6.1.80}$$

so ∇F is nowhere vanishing on S^{n-1}. Thus (6.1.69) implies that, for $x \in S^{n-1}$,

$$T_x S^{n-1} = \{v \in \mathbb{R}^n : x \cdot v = 0\}. \tag{6.1.81}$$

We bring in another surface, called the *tangent bundle* of S^{n-1},

$$TS^{n-1} = \{(x, v) \in \mathbb{R}^n \times \mathbb{R}^n : x \in S^{n-1}, v \in T_x S^{n-1}\} \tag{6.1.82}$$

so, by (6.1.81),

$$TS^{n-1} = \{(x, v) \in \mathbb{R}^{2n} : F(x, v) = (1, 0)^t\}, \tag{6.1.83}$$

where

$$F : \mathbb{R}^{2n} \to \mathbb{R}^2, \quad F(x, v) = \begin{pmatrix} |x|^2 \\ x \cdot v \end{pmatrix}. \tag{6.1.84}$$

We see that

$$DF(x, v) : \mathbb{R}^{2n} \longrightarrow \mathbb{R}^2 \tag{6.1.85}$$

is given by the $2n \times 2$ matrix

$$DF(x, v) = \begin{pmatrix} 2x^t & 0 \\ v^t & x^t \end{pmatrix}, \quad x^t = (x_1, \dots, x_n). \tag{6.1.86}$$

We claim that $D = DF(x,v)$ is surjective when $(x,v) \in TS^{n-1}$. To check surjectivity, we examine

$$\begin{aligned} DD^t &= \begin{pmatrix} 2x^t & 0 \\ v^t & x^t \end{pmatrix} \begin{pmatrix} 2x & x \\ 0 & x \end{pmatrix} \\ &= \begin{pmatrix} 4|x|^2 & 2x\cdot v \\ 2x\cdot v & |x|^2 \end{pmatrix} \in M(2,\mathbb{R}). \end{aligned} \tag{6.1.87}$$

Hence

$$(x,v) \in TS^{n-1} \Longrightarrow DD^t = \begin{pmatrix} 4 & 0 \\ 0 & 1 \end{pmatrix}, \tag{6.1.88}$$

and Proposition 6.1.5 applies. The set TS^{n-1} is a smooth, $(2n-2)$-dimensional surface.

We next look at a related surface, the *unit sphere bundle* of S^{n-1}, defined by

$$SS^{n-1} = \{(x,v) \in TS^{n-1} : |v| = 1\}, \tag{6.1.89}$$

that is,

$$SS^{n-1} = \{(x,v) \in \mathbb{R}^n \times \mathbb{R}^n : F(x,v) = (1,1,0)^t\}, \tag{6.1.90}$$

where

$$F : \mathbb{R}^{2n} \to \mathbb{R}^3, \quad F(x,v) = \begin{pmatrix} |x|^2 \\ |v|^2 \\ x\cdot v \end{pmatrix}. \tag{6.1.91}$$

We assume $n \ge 2$. We will show that Proposition 6.1.5 applies, to yield that SS^{n-1} is a smooth $(2n-3)$-dimensional surface. Indeed, we have

$$DF(x,v) : \mathbb{R}^{2n} \longrightarrow \mathbb{R}^3, \tag{6.1.92}$$

given by

$$DF(x,v) = \begin{pmatrix} 2x^t & 0 \\ 0 & 2v^t \\ v^t & x^t \end{pmatrix}, \tag{6.1.93}$$

again with $x^t = (x_1, \dots, x_n)$. We claim that $D = DF(x,v)$ is surjective when $(x,v) \in SS^{n-1}$. To see this, we compute

$$\begin{aligned} DD^t &= \begin{pmatrix} 2x^t & 0 \\ 0 & 2v^t \\ v^t & x^t \end{pmatrix} \begin{pmatrix} 2x & 0 & v \\ 0 & 2v & x \end{pmatrix} \\ &= \begin{pmatrix} 4|x|^2 & 0 & 2x\cdot v \\ 0 & 4|v|^2 & 2x\cdot v \\ 2x\cdot v & 2x\cdot v & |x|^2+|v|^2 \end{pmatrix}. \end{aligned} \tag{6.1.94}$$

Hence

$$(x,v) \in SS^{n-1} \Rightarrow DD^t = \begin{pmatrix} 4 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 2 \end{pmatrix}, \tag{6.1.95}$$

and we have surjectivity. Proposition 6.1.5 implies that SS^{n-1} is a smooth surface of dimension $2n-3$. Note that, for $(a,b)^t \in \mathbb{R}^{2n}$,

$$DF(x,v)\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 2x\cdot a \\ 2v\cdot b \\ v\cdot a + x\cdot b \end{pmatrix}. \tag{6.1.96}$$

Hence, for $(x,v) \in SS^{n-1}$,

$$T_{(x,v)}SS^{n-1} = \Big\{ \begin{pmatrix} a \\ b \end{pmatrix} \in \mathbb{R}^{2n} : x\cdot a = v\cdot b = v\cdot a + x\cdot b = 0 \Big\}, \tag{6.1.97}$$

a linear subspace of $\mathbb{R}^{2n}$ whose orthogonal complement is

$$N_{(x,v)}SS^{n-1} = \operatorname{Span}\Big\{ \begin{pmatrix} x \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ v \end{pmatrix}, \begin{pmatrix} v \\ x \end{pmatrix} \Big\}. \tag{6.1.98}$$

Note that, when $(x,v) \in SS^{n-1}$, the three vectors on the right side of (6.1.98) are mutually orthogonal.

We next illustrate Proposition 6.1.5 with another proof that

$$SO(n) \subset M(n,\mathbb{R}) \tag{6.1.99}$$

is a smooth surface, different from the argument involving (6.1.54)–(6.1.57). To get this, we take

$$V = M(n,\mathbb{R}), \quad W = \operatorname{Sym}(n) = \{A \in M(n,\mathbb{R}) : A = A^t\}, \tag{6.1.100}$$

and

$$F : V \longrightarrow W, \quad F(X) = X^tX. \tag{6.1.101}$$

Now, given $X, Y \in V$, Y small,

$$F(X+Y) = X^tX + X^tY + Y^tX + O(\|Y\|^2), \tag{6.1.102}$$

so

$$DF(X)Y = X^tY + Y^tX. \tag{6.1.103}$$

We claim that

$$X \in SO(n) \Longrightarrow DF(X) : M(n,\mathbb{R}) \to \operatorname{Sym}(n) \ \text{ is surjective.} \tag{6.1.104}$$

Indeed, given $A \in \operatorname{Sym}(n)$, i.e., $A \in M(n,\mathbb{R})$ and $A^t = A$, and $X \in SO(n)$, we have

$$Y = \frac{1}{2}XA \Longrightarrow DF(X)Y = A. \tag{6.1.105}$$

This establishes (6.1.104), so Proposition 6.1.5 applies. Again we conclude that $SO(n)$ is a smooth surface in $M(n,\mathbb{R})$. By (6.1.69), the tangent space at $X \in SO(n)$ is

$$T_XSO(n) = \ker DF(X) = \{Y \in M(n,\mathbb{R}) : X^tY + Y^tX = 0\}. \tag{6.1.106}$$

If we write $Y = XB$, we see that the defining condition is $B + B^t = 0$, so

$$T_XSO(n) = \{XB : B \in \operatorname{Skew}(n)\}. \tag{6.1.107}$$

Riemann integrable functions on a surface

Let $M \subset \mathbb{R}^n$ be an m-dimensional surface, smooth of class C^1. We define the class $\mathcal{R}_c(M)$ of compactly supported Riemann integrable functions as follows, guided by Proposition 5.1.13. If $f : M \to \mathbb{R}$ is bounded and has compact support, we set

$$\begin{aligned} \overline{I}(f) &= \inf\Big\{\int_M g\, dS : g \in C_c(M),\, g \ge f\Big\}, \\ \underline{I}(f) &= \sup\Big\{\int_M h\, dS : h \in C_c(M),\, h \le f\Big\}, \end{aligned} \tag{6.1.108}$$

where $C_c(M)$ denotes the set of continuous functions on M with compact support. Then

$$f \in \mathcal{R}_c(M) \Longleftrightarrow \overline{I}(f) = \underline{I}(f), \tag{6.1.109}$$

and if such is the case, we denote the common value by $\int_M f\, dS$. It follows readily from the definition and arguments produced in §5.1 that

$$\begin{aligned} f_1, f_2 \in \mathcal{R}_c(M) \Longrightarrow f_1 + f_2 \in \mathcal{R}_c(M) \ \text{ and} \\ \int_M (f_1 + f_2)\, dS = \int_M f_1\, dS + \int_M f_2\, dS. \end{aligned} \tag{6.1.110}$$

In fact, using (6.1.26) for functions that are continuous on M with compact support, one obtains from the definition (6.1.108) that, if $f_j : M \to \mathbb{R}$ are bounded and have compact support,

$$\overline{I}(f_1 + f_2) \le \overline{I}(f_1) + \overline{I}(f_2), \quad \underline{I}(f_1 + f_2) \ge \underline{I}(f_1) + \underline{I}(f_2),$$

which yields (6.1.110). Also one can modify the proof of Proposition 5.1.22 to show that

$$f \in \mathcal{R}_c(M),\ u \in C(M) \Longrightarrow uf \in \mathcal{R}_c(M). \tag{6.1.111}$$

Furthermore, if $\varphi : \mathcal{O} \to U \subset M$ is a coordinate chart and $f \in \mathcal{R}_c(U)$, then an application of Proposition 5.1.13 gives

$$f \circ \varphi \in \mathcal{R}_c(\mathcal{O}), \ \text{ and } \int_M f\, dS = \int_{\mathcal{O}} f(\varphi(x))\sqrt{g(x)}\, dx, \tag{6.1.112}$$

with $g(x)$ as in (6.1.22)–(6.1.23). Given any $f \in \mathcal{R}_c(M)$, we can take a continuous partition of unity $\{u_j\}$, write $f = \sum_j f_j = \sum_j u_j f$, and use (6.1.110)–(6.1.112) to express $\int_M f\, dS$ as a sum of integrals over coordinate charts.

If $\Sigma \subset M$ has compact closure, then

$$\operatorname{cont}^+ \Sigma = \overline{I}(\chi_\Sigma), \tag{6.1.113}$$

and Σ is contented if and only if $\chi_\Sigma \in \mathcal{R}_c(M)$. In such a case, (6.1.113) is the area of Σ. Given $f : M \to \mathbb{R}$, bounded and compactly supported, in parallel with

(5.1.84) we say

$$f \in \mathfrak{C}_c(M) \Leftrightarrow \text{the set } \Sigma \text{ of points of discontinuity of } f \text{ satisfies } \operatorname{cont}^+ \Sigma = 0. \tag{6.1.114}$$

We have

$$\mathfrak{C}_c(M) \subset \mathcal{R}_c(M), \tag{6.1.115}$$

and (again parallel to Proposition 5.1.13) if $f : M \to \mathbb{R}$ is bounded and compactly supported,

$$\begin{aligned} \overline{I}(f) &= \inf\Big\{ \int_M g\, dS : g \in \mathfrak{C}_c(M),\, g \ge f \Big\}, \\ \underline{I}(f) &= \sup\Big\{ \int_M h\, dS : h \in \mathfrak{C}_c(M),\, h \le f \Big\}. \end{aligned} \tag{6.1.116}$$

One can proceed from here to define the spaces

$$\mathcal{R}(M), \quad \mathcal{R}^{\#}(M), \tag{6.1.117}$$

and establish properties of functions in these spaces, in analogy with work in §5.1 on $\mathcal{R}(\mathbb{R}^n)$ and $\mathcal{R}^{\#}(\mathbb{R}^n)$. We leave such an investigation to the reader.

Exercises

1. The map $x : \mathbb{R} \times \mathbb{R} \to S^2$ given by

$$x(\theta, \psi) = (\sin\theta\cos\psi, \sin\theta\sin\psi, \cos\theta)$$

is a smooth map of $\mathbb{R} \times \mathbb{R}$ onto S^2. See Figure 6.1.3. Produce the metric tensor and area element on S^2 in these coordinates. Show that

$$\int_{S^2} f\, dS = \int_0^\pi \int_0^{2\pi} f(\sin\theta\sin\psi, \sin\theta\sin\psi, \cos\theta) \sin\theta\, d\psi\, d\theta.$$

Deduce that

$$A(S^2) = 2\pi \int_0^\pi \sin\theta\, d\theta = 4\pi.$$

Compare this with the formula (6.1.42) for A_{n-1}, with $n = 3$.

2. Apply (6.1.40) with $\varphi = \chi_{[0,1]}$ to compute the volume of the unit ball $B^n = \{x \in \mathbb{R}^n : |x| \le 1\}$. Compare the result with other approaches to the computation of $V(B^n)$ given in Chapter 5.

3. Taking the upper half of the sphere S^n to be the graph of $x_{n+1} = (1 - |x|^2)^{1/2}$, for $x \in B^n$, the unit ball in $\mathbb{R}^n$, deduce from (6.1.33) and (6.1.40) that

$$A_n = 2A_{n-1} \int_0^1 \frac{r^{n-1}}{\sqrt{1-r^2}}\, dr = 2A_{n-1} \int_0^{\pi/2} (\sin\,\theta)^{n-1}\, d\theta.$$

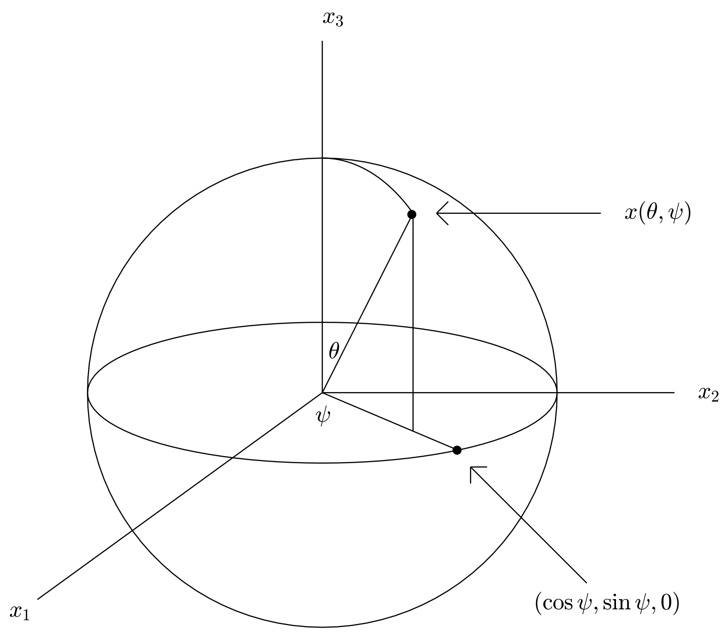

Figure 6.1.3. Spherical coordinates on S^2

Use this to get an alternative derivation of the formula (6.1.48) for A_n.
Hint. Rewrite this formula as

$$A_n = A_{n-1} b_{n-1}, \quad b_k = \int_0^\pi \sin^k \theta \, d\theta.$$

To analyze b_k, you can write, on the one hand,

$$b_{k+2} = b_k - \int_0^\pi \sin^k \theta \, \cos^2 \theta \, d\theta,$$

and on the other, using integration by parts,

$$b_{k+2} = \int_0^\pi \cos \theta \, \frac{d}{d\theta} \sin^{k+1} \theta \, d\theta.$$

Deduce that

$$b_{k+2} = \frac{k+1}{k+2} \, b_k.$$

4. Suppose M is a surface in $\mathbb{R}^n$ of dimension 2, and $\varphi : \mathcal{O} \to U \subset M$ is a coordinate chart, with $\mathcal{O} \subset \mathbb{R}^2$. Set $\varphi_{jk}(x) = \big(\varphi_j(x), \varphi_k(x)\big)$, so $\varphi_{jk} : \mathcal{O} \to \mathbb{R}^2$. Show that the formula (6.1.22) for the surface integral is equivalent to

$$\int\limits_M f \, dS = \int\limits_{\mathcal{O}} f \circ \varphi(x) \sqrt{\sum_{j<k} \Big(\det D\varphi_{jk}(x)\Big)^2} \, dx.$$

Hint. Show that the quantity under $\sqrt{}$ is equal to (6.1.29).

5. If M is an m-dimensional surface, $\varphi : \mathcal{O} \to M \subset M$ a coordinate chart, for $J = (j_1, \dots, j_m)$ set

$$\varphi_J(x) = \big(\varphi_{j_1}(x), \dots, \varphi_{j_m}(x)\big), \quad \varphi_J : \mathcal{O} \to \mathbb{R}^m.$$

Show that the formula (6.1.22) is equivalent to

$$\int_M f \, dS = \int_{\mathcal{O}} f \circ \varphi(x) \sqrt{\sum_{j_1 < \cdots < j_m} \Big(\det D\varphi_J(x)\Big)^2} \, dx.$$

Hint. Reduce to the following. For fixed $x_0 \in \mathcal{O}$, the quantity under $\sqrt{}$ is equal to $g(x)$ at $x = x_0$, in the case $D\varphi(x_0) = \big(D\varphi_1(x_0), \dots, D\varphi_m(x_0), 0, \dots, 0\big)$.

6. Let M be the graph in $\mathbb{R}^{n+1}$ of $x_{n+1} = u(x)$, $x \in \mathcal{O} \subset \mathbb{R}^n$. Show that, for $p = (x, u(x)) \in M$, T_pM (given as in ((6.1.1)) has a 1-dimensional orthogonal complement N_pM, spanned by $(-\nabla u(x), 1)$. We set $N = \big(1 + |\nabla u|^2\big)^{-1/2}(-\nabla u, 1)$, and call it the (upward-pointing) unit normal to M.

7. Let M be as in Exercise 6, and define N as done there. Show that, for a continuous function $f : M \to \mathbb{R}^{n+1}$,

$$\int_M f \cdot N \, dS = \int_{\mathcal{O}} f\big(x, u(x)\big) \cdot \big(-\nabla u(x), 1\big) \, dx.$$

The left side is often denoted $\int_M f \cdot d\mathbf{S}$.

8. Let M be a 2-dimensional surface in $\mathbb{R}^3$, covered by a single coordinate chart, $\varphi : \mathcal{O} \to M$. Suppose $f : M \to \mathbb{R}^3$ is continuous. Show that, if $\int_M f \cdot d\mathbf{S}$ is defined as in Exercise 7, then

$$\int_M f \cdot d\mathbf{S} = \int_{\mathcal{O}} f\big(\varphi(x)\big) \cdot (\partial_1 \varphi \times \partial_2 \varphi) \, dx.$$

9. Consider a symmetric $n \times n$ matrix $A = (a_{jk})$ of the form $a_{jk} = v_j v_k$. Show that the range of A is the one-dimensional space spanned by $v = (v_1, \dots, v_n)$ (if this is nonzero). Deduce that A has exactly one nonzero eigenvalue, namely $\lambda = |v|^2$. Use this to give another derivation of (6.1.32) from (6.1.31).
Hint. Show that $Ae_j = v_j v$, for each j.

10. Let $\Omega \subset \mathbb{R}^n$ be open and $u : \Omega \to \mathbb{R}$ be a C^k map. Fix $c \in \mathbb{R}$ and consider

$$S = \{x \in \Omega : u(x) = c\}.$$

Assume $S \neq \emptyset$ and that $\nabla u(x) \neq 0$ for all $x \in S$.

As seen after Proposition 6.1.5, S is a C^k surface of dimension $n - 1$, and, for each $p \in S$, T_pS has a 1-dimensional orthogonal complement N_pS spanned by

$\nabla u(p)$. Assume now that there is a C^k map $\varphi : \mathcal{O} \to \mathbb{R}$, with $\mathcal{O} \subset \mathbb{R}^{n-1}$ open, such that $u(x', \varphi(x')) = c$, and that $x' \mapsto (x', \varphi(x'))$ parametrizes S. Show that

$$\int_S f \, dS = \int_{\mathcal{O}} f \, \frac{|\nabla u|}{|\partial_n u|} \, dx',$$

where the functions in the integrand on the right are evaluated at $(x', \varphi(x'))$. *Hint*. Compare the formula in Exercise 6 for N with the fact that $\pm N = \nabla u / |\nabla u|$, and keep in mind the formula (6.1.33).

In the next exercises, we study $\operatorname{Exp} tJ = e^{tJ}$, where

$$J = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

See §C.4 for basic material on the matrix exponential.

11. Show that if $v \in \mathbb{R}^2$, then

$$\frac{d}{dt} \|e^{tJ} v\|^2 = 2 e^{tJ} v \cdot J e^{tJ} v = 0,$$

and deduce that $\|e^{tJ} v\| = \|v\|$ for all $v \in \mathbb{R}^2$, $t \in \mathbb{R}$.

12. Define $c(t)$ and $s(t)$ by

$$e^{tJ} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} c(t) \\ s(t) \end{pmatrix}.$$

Show that the identity $(d/dt) e^{tJ} = J e^{tJ}$ implies

$$c'(t) = -s(t), \quad s'(t) = c(t).$$

Deduce that $(c(t), s(t))$ is a unit speed curve, starting at $(c(0), s(0)) = (1, 0)$, with initial velocity $(c'(0), s'(0)) = (0, 1)$, and tracing out the unit circle $x^2 + y^2 = 1$ in $\mathbb{R}^2$. See Figure 6.1.4. Compare the derivation of (3.2.39).

13. Using Exercise 12 and (6.1.27), show that for $t > 0$, the curve $\gamma : [0, t] \to \mathbb{R}^2$ given by $\gamma(\tau) = (c(\tau), s(\tau))$ has length t. As discussed in §3.1, in trigonometry the line segments from $(0, 0)$ to $(1, 0)$ and from $(0, 0)$ to $(c(t), s(t))$ are said to meet at an angle, measured in radians, equal to the length of this curve, i.e., to t radians. Then the geometric definitions of the trigonometric functions $\cos t$ and $\sin t$ yield

$$\cos t = c(t), \quad \sin t = s(t). \tag{6.1.118}$$

Deduce that

$$e^{tJ} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} \cos t \\ \sin t \end{pmatrix}, \tag{6.1.119}$$

and from this, using $e^{tJ} J = J e^{tJ}$, that

$$e^{tJ} = \begin{pmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{pmatrix} = (\cos t) I + (\sin t) J. \tag{6.1.120}$$

Compare Euler's formula (3.2.39), and also (3.3.21).

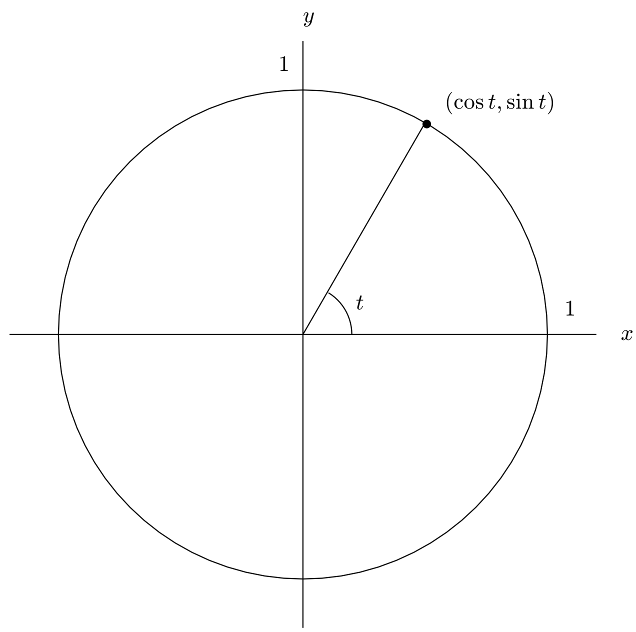

Figure 6.1.4. Unit circle

14. The following result in linear algebra is established in Proposition C.2.8 of Appendix C.2.

Proposition. If $A : \mathbb{R}^n \to \mathbb{R}^n$ is orthogonal, so $A^t A = I$, then $\mathbb{R}^n$ has an orthonormal basis in which the matrix representation of A consists of blocks

$$\begin{pmatrix} c_j & -s_j \\ s_j & c_j \end{pmatrix}, \quad c_j^2 + s_j^2 = 1,$$

plus perhaps an identity matrix block if 1 is an eigenvalue of A, and a block that is $-I$ if -1 is an eigenvalue of A.

Use this and (6.1.120) to prove that

$$\text{Exp} : \text{Skew}(n) \longrightarrow SO(n) \ \text{ is onto.} \tag{6.1.121}$$

,

In the next exercise, $\mathcal{T}$ denotes the "inner tube" obtained as follows. Take the circle in the (y, z)-plane, centered at $y = a, z = 0$, of radius b, with $0 < b < a$. Rotate this circle about the z-axis. Then $\mathcal{T}$ is the surface so swept out. See Figure 6.1.5.

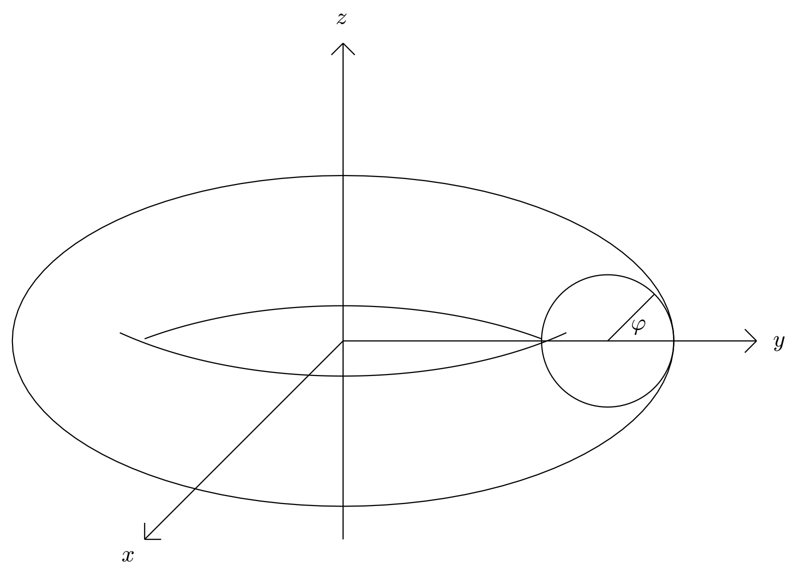

Figure 6.1.5. Inner tube

15. Define $\psi : \mathbb{R}^2 \to \mathbb{R}^3$ by $\psi(\theta, \varphi) = (x(\theta, \varphi), y(\theta, \varphi), z(\theta, \varphi))$, with

$$\begin{aligned} x(\theta, \varphi) &= (a + b\cos\varphi)\cos\theta, \\ y(\theta, \varphi) &= (a + b\cos\varphi)\sin\theta, \\ z(\theta, \varphi) &= b\sin\varphi. \end{aligned}$$

Show that ψ maps $[0, 2\pi] \times [0, 2\pi]$ onto $\mathcal{T}$.
Show that $|\partial_\theta \psi \times \partial_\varphi \psi| = b(a + b\cos\varphi)$.
Using (6.1.28), show that

$$\operatorname{Area} \mathcal{T} = 4\pi^2 ab.$$

16. In the setting of Exercise 15, compute the following integrals.

$$\int_{\mathcal{T}} x^2 \, dS, \quad \int_{\mathcal{T}} y^2 \, dS, \quad \int_{\mathcal{T}} z^2 \, dS.$$

In the next exercise, M is a surface of revolution, obtained by taking the graph of a function $y = f(x)$, $a \le x \le b$ (assuming $f > 0$) and rotating it about the x-axis, in $\mathbb{R}^3$.

17. Define $\psi : [a,b] \times \mathbb{R} \to \mathbb{R}^3$ by $\psi(s,t) = (s, f(s)\cos t, f(s)\sin t)$.
Show that ψ maps $[a,b] \times [0, 2\pi]$ onto M.
Show that $|\partial_s\psi \times \partial_t\psi| = f(s)\sqrt{1 + f'(s)^2}$.
Using (6.1.28), show that if $u : M \to \mathbb{R}$ is continuous,

$$\int\limits_M u\,dS = \int_0^{2\pi}\int_a^b u\big(s, f(s)\cos t, f(s)\sin t\big) f(s)\sqrt{1+f'(s)^2}\,ds\,dt.$$

In particular,

$$\text{Area } M = 2\pi \int_a^b f(s)\sqrt{1+f'(s)^2}\,ds.$$

Remark. As seen in §5.1, if

$$\Omega = \{(x,y,z) \in \mathbb{R}^3 : a \le x \le b, y^2 + z^2 \le f(x)^2\},$$

then

$$\operatorname{Vol}\Omega = \pi \int_a^b f(s)^2\,ds.$$

17A. In the setting of Exercise 17, take $f(s) = 1/s,\ a = 1,\ b > 1$. Write down the integrals for Area M and Vol Ω. Compute the limits of these quantities as $b \to \infty$.

18. Consider the ellipsoid of revolution $\mathcal{E}_a$, given for $a > 0$ by

$$\frac{x^2}{a^2} + y^2 + z^2 = 1.$$

Use the method of Exercise 17 to show that

$$\text{Area } \mathcal{E}_a = 4\pi \int_0^a \sqrt{1 - \beta s^2}\,ds, \quad \beta = \frac{1}{a^2} - \frac{1}{a^4}.$$

19. Given $a, b, c > 0$, consider the ellipsoid $\mathcal{E}(a,b,c)$, given by

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1.$$

Using (6.1.33), write down a formula for the area of $\mathcal{E}(a,b,c)$ as an integral over the region

$$E_{a,b} = \Big\{(x,y) \in \mathbb{R}^2 : \frac{x^2}{a^2} + \frac{y^2}{b^2} \le 1\Big\}.$$

20. Consider the parabolic curve

$$\gamma(t) = \Big(t, \frac{t^2}{2}\Big).$$

Show that the length of $\gamma([0,x])$ is

$$\ell(x) = \int_0^x \sqrt{1+t^2}\,dt.$$

Evaluate this integral using the substitution $t = \sinh u$.

21. Let M be the surface of revolution obtained by taking the graph of the function $y = e^x,\ a \le x \le b$, and rotating it about the x-axis in $\mathbb{R}^3$. Show that Exercise 17 yields

$$\operatorname{Area} M = 2\pi \int_a^b e^s \sqrt{1+e^{2s}}\, ds.$$

Taking $t = e^s$, show that this is equal to

$$2\pi \int_\alpha^\beta \sqrt{1+t^2}\, dt, \quad \alpha = e^a,\ \beta = e^b.$$

Relate this to Exercise 20.

In the next exercise, M is an n-dimensional "surface of revolution" given by a smooth map

$$\psi : [a,b] \times S^{n-1} \longrightarrow M \subset \mathbb{R} \times \mathbb{R}^n$$

of the form

$$\psi(s,\omega) = (s, f(s)\omega).$$

A coordinate chart $\varphi : \Omega \to S^{n-1}$, with Ω open in $\mathbb{R}^{n-1}$, gives rise to a coordinate chart

$$\tilde{\psi} : [a,b] \times \Omega \longrightarrow M, \quad \tilde{\psi}(s,x) = (s, f(s)\varphi(x)).$$

We set $x_0 = s,\ x = (x_1, \dots, x_{n-1})$.

22. Show that, in $\tilde{\psi}$-coordinates, the metric tensor of M takes the form (g_{jk}), for $0 \le j,k \le n-1$, with the following components:

$$\begin{aligned} g_{00} &= \frac{\partial \tilde{\psi}}{\partial x_0} \cdot \frac{\partial \tilde{\psi}}{\partial x_0} = 1 + f'(s)^2, \\ g_{0j} &= \frac{\partial \tilde{\psi}}{\partial x_0} \cdot \frac{\partial \tilde{\psi}}{\partial x_j} = 0, \quad \text{for } \ 1 \le j \le n-1, \\ g_{jk} &= \frac{\partial \tilde{\psi}}{\partial x_j} \cdot \frac{\partial \tilde{\psi}}{\partial x_k} = f(s)^2 h_{jk}, \quad \text{for } \ 1 \le j,k \le n-1, \end{aligned}$$

where $(h_{\ell m})$ is the metric tensor of S^{n-1} in the φ-coordinates. Otherwise said,

$$(g_{jk}) = \begin{pmatrix} 1 + f'(s)^2 & \\ & f(s)^2 h_{\ell m} \end{pmatrix}.$$

Compare (6.1.36).

23. In the setting of Exercise 22, deduce that if $u : M \to R$ is continuous,

$$\int\limits_M u\, dS = \int_a^b \int\limits_{S^{n-1}} u(s, f(s)\omega) f(s)^{n-1} \sqrt{1 + f'(s)^2}\, dS(\omega)\, ds.$$

In particular, with A_{n-1} as in (6.1.40)–(6.1.42),

$$\operatorname{Area} M = A_{n-1} \int_a^b f(s)^{n-1} \sqrt{1 + f'(s)^2}\, ds.$$

Note how this generalizes the conclusion of Exercise 17.

24. In the setting of Exercises 22–23, let $M = S^n$, with $f(s) = \sqrt{1-s^2}$. Show that

$$\int\limits_{S^n} u(x_0)\, dS(x) = A_{n-1} \int_{-1}^{1} u(s)(1-s^2)^{(n-2)/2}\, ds.$$

25. Let $\psi : SO(n) \to M(k, \mathbb{R})$ be continuous and satisfy the following properties:

$$\psi(gh) = \psi(g)\psi(h), \quad \psi(g^{-1}) = \psi(g)^{-1},$$

for all $g, h \in SO(n)$. We say ψ is a representation of $SO(n)$ on $\mathbb{R}^k$. Form

$$P = \int\limits_{SO(n)} \psi(g)\, dg, \quad P \in M(k, \mathbb{R}),$$

using the integral (6.1.63) (but here with a matrix valued integrand). Show that

$$P : \mathbb{R}^k \longrightarrow V, \ \text{ and } \ v \in V \Rightarrow Pv = v,$$

where

$$V = \{v \in \mathbb{R}^k : \psi(g)v = v,\ \forall\, g \in SO(n)\}.$$

Thus P is a projection of $\mathbb{R}^k$ onto V.
Hint. With $h \in SO(n)$ arbitrary, express $\psi(h)P$ as the integral $\int \psi(hg)\, dg$, and apply (6.1.62).

26. In the setting of Exercise 25, show that

$$\dim V = \int\limits_{SO(n)} \chi(g)\, dg, \quad \chi(g) = \mathrm{Tr}\ \psi(g).$$

27. Given $u \in C(\mathbb{R}^n)$, define $\mathcal{A}u \in C(\mathbb{R}^n)$ by

$$\mathcal{A}u(x) = \int\limits_{SO(n)} u(gx)\, dg.$$

Show that $\mathcal{A}u$ is a radial function, in fact

$$\mathcal{A}u(x) = \mathcal{S}u(|x|), \quad \text{with } \ \mathcal{S}u(r) = \frac{1}{A_{n-1}} \int\limits_{S^{n-1}} u(r\omega)\, dS(\omega).$$

28. In the setting of Exercise 27, show that if $u(x)$ is a polynomial in x, then $\mathcal{S}u(r)$ is a polynomial in r.
Hint. Show that $\mathcal{A}u(x)$ is a polynomial in x. Look at $\mathcal{A}u(re_1)$.

29. Let M be a C^1 surface, $K \subset M$ compact. Let $\varphi_j : \mathcal{O}_j \to U_j$ be coordinate charts on M and assume $K \subset \cup_{j=1}^{k} U_j$. Take $v_j \in C_c(U_j)$ such that $\sum_1^k v_j > 0$ on K.

Let $f : M \to \mathbb{R}$ be bounded and supported on K. Show that

$$f \in \mathcal{R}_c(M) \Longleftrightarrow (v_j f) \circ \varphi_j \in \mathcal{R}_c(\mathcal{O}_j), \text{ for each } j.$$

Here, use the definition (6.1.108)–(6.1.109) for $\mathcal{R}_c(M)$ and define $\mathcal{R}_c(\mathcal{O}_j)$ as in §5.1.

30. Let $M \subset \mathbb{R}^n$ be a compact, m-dimensional, C^1 surface. We define a contented partition of M to be a finite collection $\mathcal{P} = \{\Sigma_k\}$ of contented subsets of M such that

$$M = \bigcup_k \Sigma_k, \quad \text{cont}^+(\Sigma_j \cap \Sigma_k) = 0, \quad \forall\, j \neq k.$$

We say

$$\operatorname{maxsize} \mathcal{P} = \max_k \operatorname{diam}(\Sigma_k),$$

where $\operatorname{diam} \Sigma_k = \sup_{x,y \in \Sigma_k} \|x - y\|$. Establish the following variant of the Darboux theorem (Proposition 5.1.1).

Proposition. Let $\mathcal{P}_\nu = \{\Sigma_{k\nu} : 1 \le k \le N(\nu)\}$ be a sequence of contented partitions of M such that $\operatorname{maxsize} \mathcal{P}_\nu \to 0$. Pick points $\xi_{k\nu} \in \Sigma_{k\nu}$. Then, given $f \in \mathcal{R}(M)$, we have

$$\int_M f\, dS = \lim_{\nu \to \infty} \sum_{k=1}^{N(\nu)} f(\xi_{j\nu})\, V(\Sigma_{k\nu}),$$

where $V(\Sigma_{k\nu}) = \int_M \chi_{\Sigma_{k\nu}}\, dS$ is the content of $\Sigma_{k\nu}$.

Hint. First treat the case $f \in C(M)$. Use the material in (6.1.108)–(6.1.116) to extend this to $f \in \mathcal{R}(M)$.

31. We desire to compute $\det G$ when $G = (g_{jk})$ is an $m \times m$ matrix given by

$$g_{jk} = \delta_{jk} + v_j v_k.$$

Compare (6.1.31). In other words,

$$G = I + T, \quad T = (t_{jk}), \quad t_{jk} = v_j v_k.$$

(a) Let $v \in \mathbb{R}^m$ have components v_j. Show that, for $w \in \mathbb{R}^m$, $Tw = (v \cdot w)v$.
(b) Deduce that T has one nonzero eigenvalue, $|v|^2$.
(c) Deduce that one eigenvalue of G is $1 + |v|^2$, and the other $m - 1$ eigenvalues are 1.
(d) Deduce that $g = \det G = 1 + |v|^2$, so $\sqrt{g} = \sqrt{1 + |v|^2}$. Compare (6.1.32), with $v = \nabla u$.

Stereographic projection

32. With $(x', x_n) \in \mathbb{R}^n$, show that

$$\mathcal{S}(x', x_n) = \frac{1}{1 - x_n}\, x'$$

defines a diffeomorphism

$$\mathcal{S} : S^{n-1} \setminus \{e_n\} \longrightarrow \mathbb{R}^{n-1},$$

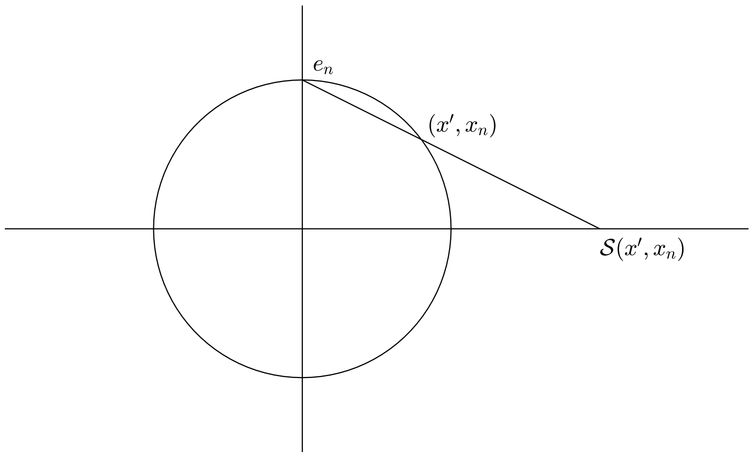

Figure 6.1.6. Stereographic projection

with inverse $F : \mathbb{R}^{n-1} \to S^{n-1} \setminus \{e_n\}$ given by

$$F(y) = \frac{1}{1+|y|^2}(2y, |y|^2 - 1).$$

The map $\mathcal{S}$ is called stereographic projection. See Figure 6.1.6.

33. In the setting of Exercise 32, show that

$$DF(y)^t DF(y) = \frac{4}{(1+|y|^2)^2}\, I,$$

where I is the $(n-1) \times (n-1)$ identity matrix.

34. Show that if $u \in C(S^{n-1})$, then

$$\int\limits_{S^{n-1}} u\, dS = \int\limits_{\mathbb{R}^{n-1}} u(F(x))\Big(\frac{2}{1+|y|^2}\Big)^{n-1}\, dy.$$

35. Deduce from Exercise 34 that the area of S^{n-1} is

$$\begin{aligned} A_{n-1} &= \int\limits_{\mathbb{R}^{n-1}} \Bigl(\frac{2}{1+|y|^2}\Bigr)^{n-1}\,dy \\ &= A_{n-2}\int_0^\infty \Bigl(\frac{2}{1+r^2}\Bigr)^{n-1} r^{n-2}\,dr. \end{aligned}$$

Compare computations in Exercise 3. Compute this last integral in the cases $n = 2$ and $n = 3$.

36. Define the alternative stereographic projection

$$\widetilde{\mathcal{S}} : S^{n-1} \setminus \{-e_n\} \longrightarrow \mathbb{R}^{n-1}, \quad \widetilde{\mathcal{S}}(x', x_n) = \mathcal{S}(x', -x_n),$$

and compute $\widetilde{F} = \widetilde{\mathcal{S}}^{-1}$ and

$$\mathcal{S} \circ \widetilde{\mathcal{S}}^{-1} : \mathbb{R}^{n-1} \setminus 0 \longrightarrow \mathbb{R}^{n-1} \setminus 0.$$

6.2. Constrained maxima and minima – Lagrange multipliers

Here we tackle the following sort of problem. Let $M \subset \mathbb{R}^n$ be a smooth, m-dimensional surface, and let $f : M \to \mathbb{R}$ be a smooth function (say of class C^1). We want to classify the points at which f assumes a maximum or minimum, or more generally a local maximum or minimum. Still more generally, we want to define and study the *critical points* of f. To get oriented, we mention the example

$$M = S^2 = \{x \in \mathbb{R}^3 : |x| = 1\}, \quad f(x) = x_1 x_2. \tag{6.2.1}$$

In our general study, we concentrate on the situation where

$$\begin{gathered} M \subset \mathcal{O}, \text{ open in } \mathbb{R}^n, \quad g : \mathcal{O} \to \mathbb{R}, \text{ smooth}, \\ f = g\big|_M. \end{gathered} \tag{6.2.2}$$

Let us take $p \in M$ and consider when we can say f has a local maximum or minimum (or other critical point) at p. Clearly, if we have a coordinate chart

$$\varphi : \Omega \longrightarrow M, \quad \varphi(x_0) = p, \tag{6.2.3}$$

then f has a local max (or min) at p if and only if

$$u = f \circ \varphi : \Omega \longrightarrow \mathbb{R} \tag{6.2.4}$$

has a local max (or min) at x_0. As we know, when this holds,

$$Du(x_0) = 0. \tag{6.2.5}$$

More generally, we say f has a critical point at p provided (6.2.5) holds. Results of §6.1 imply that if $\psi : \Omega_0 \to M$ is another coordinate chart, satisfying $\psi(y_0) = p$, then the condition (6.2.5) holds if and only if

$$Dv(y_0) = 0, \quad v = f \circ \psi. \tag{6.2.6}$$

Indeed, we can define

$$Df(p) : T_pM \longrightarrow \mathbb{R} \tag{6.2.7}$$

as the unique linear map satisfying

$$\begin{aligned} Du(x_0) &= Df(p)D\varphi(x_0), \quad \text{hence} \\ Dv(y_0) &= Df(p)D\psi(y_0), \end{aligned} \tag{6.2.8}$$

making use of the fact that $D\varphi(x_0)$ and $D\psi(y_0)$ are both isomorphisms of $\mathbb{R}^m$ onto T_pM, together with (6.1.15)–(6.1.16). We have that

$$\begin{gathered} f \text{ has a critical point at } p \in M \\ \Longleftrightarrow Df(p) : T_pM \to \mathbb{R} \text{ is } 0. \end{gathered} \tag{6.2.9}$$

In case (6.2.2) holds, the fact that

$$f \circ \varphi = g \circ \varphi \tag{6.2.10}$$

implies

$$Df(p) = Dg(p)\big|_{T_pM}, \tag{6.2.11}$$

and we have the following.

Proposition 6.2.1. *Assume $M \subset \mathcal{O} \subset \mathbb{R}^n$ and (6.2.2) holds. Take $p \in M$. Then p is a critical point of f if and only if*

$$Dg(p)v = 0, \quad \forall\, v \in T_pM. \tag{6.2.12}$$

We now specialize to the following situation. Take $\mathcal{O} \subset \mathbb{R}^n$, open. Suppose we have a smooth function

$$h : \mathcal{O} \longrightarrow \mathbb{R}, \quad \nabla h(x) \neq 0,\ \forall\, x \in \mathcal{O}. \tag{6.2.13}$$

We take $a \in \mathbb{R}$ in the image of h and set

$$M = \{x \in \mathcal{O} : h(x) = a\}. \tag{6.2.14}$$

By Proposition 6.1.5 (and the remark containing (6.1.78)), M is a smooth surface, of dimension $m = n - 1$, and, for $p \in M$,

$$T_pM = \{v \in \mathbb{R}^n : v \perp \nabla h(p)\}. \tag{6.2.15}$$

In this setting, the content of (6.2.12) is that

$$v \perp \nabla g(p),\ \forall\, v \in T_pM, \tag{6.2.16}$$

which, in concert with (6.2.15), is equivalent to the condition that

$$\nabla g(p) \parallel \nabla h(p), \tag{6.2.17}$$

that is,

$$\exists\, \lambda \in \mathbb{R} \text{ such that } \nabla g(p) = \lambda \nabla h(p). \tag{6.2.18}$$

We record the result.

Proposition 6.2.2. *Assume $M \subset \mathcal{O} \subset \mathbb{R}^n$ is an $(n-1)$-dimensional surface given by* (6.2.14)*, with h smooth and satisfying* (6.2.13)*, and take*

$$g : \mathcal{O} \longrightarrow \mathbb{R}, \quad f = g\big|_M. \tag{6.2.19}$$

Then a point $p \in M$ is a critical point of f if and only if (6.2.18) *holds.*

The real number λ connecting the two vectors $\nabla g(p)$ and $\nabla h(p)$ in (6.2.18) is called a *Lagrange multiplier*. The method of finding critical points of $f = g|_M$ by seeking such λ is called the method of Lagrange multipliers.

To illustrate this method, we return to the example presented in (6.2.1). That is, we take

$$\begin{gathered} \mathcal{O} = \mathbb{R}^3 \setminus 0, \quad h(x) = |x|^2, \quad M = \{x \in \mathcal{O} : h(x) = 1\}, \\ g(x) = x_1x_2, \quad f = g\big|_M. \end{gathered} \tag{6.2.20}$$

Then

$$\begin{aligned} \nabla h(x) &= 2(x_1, x_2, x_3), \\ \nabla g(x) &= (x_2, x_1, 0), \end{aligned} \tag{6.2.21}$$

and we seek points $x \in S^2$ where these two vectors are parallel. Since $n = 3$, we can find them by computing the cross product:

$$\begin{aligned} \frac{1}{2}\nabla h(x) \times \nabla g(x) &= \det \begin{pmatrix} i & j & k \\ x_1 & x_2 & x_3 \\ x_2 & x_1 & 0 \end{pmatrix} \\ &= (-x_1x_3, x_2x_3, x_1^2 - x_2^2). \end{aligned} \tag{6.2.22}$$

For this to vanish, we require

$$\begin{gathered} x_1 = \pm x_2, \quad \text{and} \\ \text{either } x_3 = 0 \text{ or } x_1 = x_2 = 0. \end{gathered} \tag{6.2.23}$$

Note that if $x \in S^2$ and $x_1 = x_2 = 0$, then $x_3 = \pm 1$. Thus we have 6 critical points of f on S^2:

$$\pm\frac{1}{2}(\sqrt{2}, \sqrt{2}, 0), \quad \pm\frac{1}{2}(\sqrt{2}, -\sqrt{2}, 0), \quad \pm(0, 0, 1). \tag{6.2.24}$$

We see that the values of f at these 3 pairs of points are, respectively,

$$\frac{1}{2}, \quad -\frac{1}{2}, \quad 0. \tag{6.2.25}$$

In particular, f has its maximum at the first pair of points in (6.2.24) and its minimum at the second pair of points.

We move to a more general class of examples, namely

$$\begin{gathered} M = S^{n-1} = \{x \in \mathbb{R}^n : h(x) = 1\}, \quad h(x) = |x|^2, \\ g(x) = x \cdot Ax, \quad A = A^t \in M(n, \mathbb{R}), \quad f = g\big|_M. \end{gathered} \tag{6.2.26}$$

In this setting,

$$\nabla h(x) = 2x, \quad \nabla g(x) = 2Ax, \tag{6.2.27}$$

and the Lagrange multiplier condition becomes

$$Ax = \lambda x, \quad \lambda \in \mathbb{R}, \ |x| = 1. \tag{6.2.28}$$

In other words, x should be an eigenvector of A. As shown in Proposition C.2.3, the condition $A = A^t$ implies $\mathbb{R}^n$ has an orthonormal basis of eigenvectors,

$$\{v_1, \dots, v_n\}, \quad v_j \cdot v_k = \delta_{jk}, \quad Av_j = \lambda_j v_j, \quad \lambda_1 \le \cdots \le \lambda_n. \tag{6.2.29}$$

Now, if the eigenvalues $\{\lambda_1, \dots, \lambda_n\}$ are all distinct, then f has $2n$ critical points,

$$\{\pm v_j : 1 \le j \le n\}, \quad f(v_j) = \lambda_j. \tag{6.2.30}$$

We have

$$\min f = \lambda_1, \quad \max f = \lambda_n. \tag{6.2.31}$$

On the other hand, if λ_j has multiplicity d_j, so

$$\mathcal{E}(A, \lambda_j) = \{v \in \mathbb{R}^n : Av = \lambda_j v\} \text{ has dimension } d_j, \tag{6.2.32}$$

then there is a $(d_j - 1)$-dimensional sphere of critical points,

$$\mathcal{E}(A, \lambda_j) \cap S^{n-1}, \quad \text{where} \quad f = \lambda_j. \tag{6.2.33}$$

We note that the example (6.2.1) is a special case of (6.2.26), with $n = 3$ and

$$2A = \begin{pmatrix} 0 & 1 & \\ 1 & 0 & \\ & & 0 \end{pmatrix}. \tag{6.2.34}$$

We now take up an example in which $g(x)$ is a cubic polynomial. Consider

$$\begin{gathered} M = S^2 = \{x \in \mathbb{R}^3 : h(x) = 1\}, \quad h(x) = |x|^2, \\ g(x) = x_1 x_2 x_3, \quad f = g\big|_M. \end{gathered} \tag{6.2.35}$$

Here

$$\nabla h(x) = 2x, \quad \nabla g(x) = (x_2x_3, x_1x_3, x_1x_2), \tag{6.2.36}$$

and the Lagrange multiplier condition becomes

$$\frac{x_1x_2x_3}{x_j} = 2\lambda x_j, \quad 1 \le j \le 3. \tag{6.2.37}$$

which implies

$$x_2x_2x_3 = 2\lambda x_1^2 = 2\lambda x_2^2 = 2\lambda x_3^2. \tag{6.2.38}$$

This leads to

$$\lambda = 0, \quad \text{or} \;\; x_1^2 = x_2^2 = x_3^2. \tag{6.2.39}$$

Now,

$$\lambda = 0 \Longrightarrow x_2x_3 = x_1x_3 = x_1x_2 = 0, \tag{6.2.40}$$

and, given $x \in S^2$, this implies

$$x = \pm i, \;\pm j, \;\text{or} \;\pm k \quad \text{(6 critical points)}, \tag{6.2.41}$$

where (i, j, k) is the standard basis of $\mathbb{R}^3$. On the other hand, if $x \in S^2$,

$$x_1^2 = x_2^2 = x_3^2 \Longrightarrow x_j^2 = \frac{1}{3}, \;\forall\, j, \tag{6.2.42}$$

and this implies

$$x = \Big(\pm\frac{\sqrt{3}}{3}, \pm\frac{\sqrt{3}}{3}, \pm\frac{\sqrt{3}}{3}\Big) \quad \text{(8 critical points)}. \tag{6.2.43}$$

Together, (6.2.41) and (6.2.43) give the 14 critical points of f. We have

$$\min f = -\frac{\sqrt{3}}{9}, \quad \max f = \frac{\sqrt{3}}{9}. \tag{6.2.44}$$

Averaging rotations

This class of examples involves a surface of higher codimension, namely

$$M = SO(n) \subset M(n, \mathbb{R}), \tag{6.2.45}$$

a surface of dimension $n(n-1)/2$ in the n^2-dimensional vector space $M(n, \mathbb{R})$. Suppose we are given

$$A_1, \dots, A_N \in SO(n). \tag{6.2.46}$$

We want to identify an element of $SO(n)$ that represents an "average" of these rotations A_j.

Part of our task is to produce a reasonable definition of "average" in this context. If we simply average in the vector space $M(n, \mathbb{R})$, we get

$$\frac{1}{N}\overline{A}, \quad \overline{A} = A_1 + \cdots + A_N. \tag{6.2.47}$$

However, typically this element of $M(n,\mathbb{R})$ does not belong to $SO(n)$. To formulate a notion of average that will work for averaging over sets that are not linear spaces, we start with the observation that $\overline{A}/N$ is obtained as the minimizer of

$$\psi(X) = \sum_{j=1}^{N} \|X - A_j\|^2, \tag{6.2.48}$$

if we minimize over all $X \in M(n,\mathbb{R})$. Here the norm is given by

$$\|T\|^2 = \langle T, T\rangle, \quad \langle S, T\rangle = \operatorname{Tr} S^t T. \tag{6.2.49}$$

See §2.4. Guided by this, we make the following

Definition. Given $S = \{A_1, \dots, A_N\} \subset SO(n)$, an element $X \in SO(n)$ that minimizes (6.2.48) over $SO(n)$ is said to be an R-average of S.

Certainly (6.2.48) has a minimum over $SO(n)$, though the minimizer might not be unique. If the minimizer is unique, we say it is *the* R-average.

We tackle the problem of characterizing R-averages of sets of elements of $SO(n)$. To analyze (6.2.48), write

$$\begin{aligned} \|X - A_j\|^2 &= \operatorname{Tr}(X^t - A_j^t)(X - A_j) \\ &= \operatorname{Tr}(X^t X - X^t A_j - A_j^t X + A_j^t A_j) \\ &= 2n - 2 \operatorname{Tr} A_j^t X, \end{aligned} \tag{6.2.50}$$

using $X^t X = A_j^t A_j = I$. Hence we have

$$\psi(X) = 2nN - 2 \operatorname{Tr} \overline{A}^t X, \quad \overline{A} = A_1 + \cdots + A_N. \tag{6.2.51}$$

Thus the problem of minimizing (6.2.48) over $SO(n)$ is equivalent to the following problem:

$$\text{Maximize } \operatorname{Tr} \overline{A}^t X \text{ over } X \in SO(n). \tag{6.2.52}$$

The function we want to maximize is

$$\varphi : SO(n) \longrightarrow \mathbb{R}, \quad \varphi(X) = \operatorname{Tr} \overline{A}^t X. \tag{6.2.53}$$

More generally, we look for the critical points of φ. By Proposition 6.2.1, $X \in SO(n)$ is a critical point of φ if and only if

$$D\varphi(X)V = 0, \quad \forall\, V \in T_X SO(n). \tag{6.2.54}$$

By (6.1.59), or (6.1.107),

$$T_X SO(n) = \{XB : B \in \operatorname{Skew}(n)\}. \tag{6.2.55}$$

Now we have

$$D\varphi(X)V = \operatorname{Tr} \overline{A}^t V. \tag{6.2.56}$$

Since $\operatorname{Skew}(n)$ and

$$\operatorname{Sym}(n) = \{A \in M(n,\mathbb{R}) : A^t = A\} \tag{6.2.57}$$

are orthogonal complements in $M(n,\mathbb{R})$ with respect to the inner product $\langle A, B\rangle = \operatorname{Tr} A^t B$, we see that

$$\begin{aligned} &X \in SO(n) \text{ is a critical point of } \varphi \\ \Longleftrightarrow\ &\operatorname{Tr} \overline{A}^t X B = 0, \quad \forall\, B \in \operatorname{Skew}(n) \\ \Longleftrightarrow\ &\overline{A}^t X \in \operatorname{Sym}(n). \end{aligned} \tag{6.2.58}$$

To proceed, we will discuss further the case

$$\det \overline{A} > 0. \tag{6.2.59}$$

In such a case, it is shown in §6.5 that there are unique matrices

$$\begin{aligned} &Q \in \mathcal{P}(n) = \{P \in \operatorname{Sym}(n) : x \cdot Px > 0,\ \forall\, x \in \mathbb{R}^n \setminus 0\}, \\ &U \in SO(n), \end{aligned} \tag{6.2.60}$$

such that

$$\overline{A} = UQ, \tag{6.2.61}$$

so

$$\overline{A}^t X = QU^t X. \tag{6.2.62}$$

Noting that $X, U \in SO(n) \Rightarrow U^t X \in SO(n)$, we bring in the following.

Lemma 6.2.3. *Given $Q \in \mathcal{P}(n),\ T \in SO(n)$,*

$$\operatorname{Tr} QT \le \operatorname{Tr} Q, \tag{6.2.63}$$

with equality if and only if $T = I$.

Proof. It follows from Proposition C.2.3 that $\mathbb{R}^n$ has an orthonormal basis $v_1, \dots, v_n$ consisting of eigenvectors of Q, $Qv_j = \lambda_j v_j,\ \lambda_j > 0$. Then

$$\operatorname{Tr} QT = \sum_j v_j \cdot QTv_j = \sum_j \lambda_j v_j \cdot Tv_j. \tag{6.2.64}$$

We have $v_j \cdot Tv_j \le 1$, with equality if and only if $Tv_j = v_j$, given $T \in SO(n)$. This yields (6.2.63). ☐

Thus we can solve our minimization problem, under the hypothesis (6.2.59).

Corollary 6.2.4. *Given A_j in* (6.2.46) *and $\overline{A}$ in* (6.2.47)*, if $\det \overline{A} > 0$, then there is a unique $X \in SO(n)$ minimizing ψ over $SO(n)$. It is given by*

$$X = U = \overline{A} Q^{-1}, \tag{6.2.65}$$

with Q and U specified in (6.2.61)*. Hence X in* (6.2.65) *is the R-average of $\{A_j : 1 \le j \le N\}$.*

Other cases that arise, in addition to (6.2.59), are

$$\det \overline{A} < 0, \quad \det \overline{A} = 0. \tag{6.2.66}$$

See §A.2 of [**20**] for a discussion of minimizers for ψ over $SO(n)$ (which might not be unique) in these cases.

Exercises

1. Find the point on the paraboloid $M \subset \mathbb{R}^3$ given by $x_3 = x_1^2 + x_2^2$ that is closest to the point $(1, 0, 0)$.

2. The last equivalence in (6.2.58) uses the fact that we have an orthogonal direct sum

$$M(n, \mathbb{R}) = \operatorname{Sym}(n) \oplus \operatorname{Skew}(n).$$

Prove this.
Hint. For orthogonality, take $S \in \operatorname{Sym}(n),\ A \in \operatorname{Skew}(n)$, and compare

$$\langle S, A\rangle = \operatorname{Tr} SA = \operatorname{Tr} AS$$

with

$$\operatorname{Tr} SA = \operatorname{Tr}(SA)^t = \operatorname{Tr} A^t S^t.$$

For the decomposition, write $T = (T + T^t)/2 + (T - T^t)/2$.

3. With $\psi(X)$ as in (6.2.48), show that, for arbitrary $X \in M(n, \mathbb{R})$,

$$\begin{aligned}\psi(X) &= N\|X\|^2 - 2\langle \overline{A}, X\rangle + \sum \|A_j\|^2 \\ &= N\Big\|X - \frac{1}{N}\overline{A}\Big\|^2 - \frac{1}{N}\|\overline{A}\|^2 + \sum \|A_j\|^2,\end{aligned}$$

and if you minimize over $X \in M(n, \mathbb{R})$, the minimum is achieved at

$$X = \frac{1}{N}\overline{A}.$$

4. Produce explicit subsets $\{A_1, \dots, A_N\} \subset SO(3)$ such that $\overline{A} = A_1 + \cdots + A_N$ satisfies

$$\begin{aligned}\det \overline{A} &> 0, \\ \det \overline{A} &< 0, \\ \det \overline{A} &= 0.\end{aligned}$$

See if you can come up with variants of Corollary 6.2.4 that cover the latter two cases.

Given a smooth surface $M \subset \mathbb{R}^n,\ p \in M$, denote the orthogonal complement of T_pM in $\mathbb{R}^n$ by

$$N_pM = (T_pM)^\perp.$$

5. In the setting of Proposition 6.2.1, show that $p \in M$ is a critical point of $f = g|_M$ if and only if

$$\nabla g(p) \in N_pM.$$

6. Show that, for $X \in SO(n)$, the orthogonal complement of $T_X SO(n)$ is
$$N_X SO(n) = \{XS : S \in \mathrm{Sym}(n)\}.$$

6.3. Formulas of Gauss, Green, and Stokes

In this section we establish the following integral identity, known as the Gauss divergence theorem, and obtain from it formulas of Green and Stokes. Here is the first result.

Theorem 6.3.1. *Let $\Omega \subset \mathbb{R}^n$ be a bounded open set, with a C^1 smooth boundary $\partial\Omega$. Denote by $N(x)$ the outward-pointing unit normal vector to $\partial\Omega$ at $x \in \partial\Omega$. If F is a smooth vector field on $\overline{\Omega}$, then*

$$\int_{\Omega} (\operatorname{div} F)\, dx = \int_{\partial\Omega} N \cdot F\, dS. \tag{6.3.1}$$

Here $\operatorname{div} F$ denotes the *divergence* of the vector field F, given by

$$\operatorname{div} F = \frac{\partial f_1}{\partial x_1} + \cdots + \frac{\partial f_n}{\partial x_n}, \tag{6.3.2}$$

if $F = (f_1, \dots, f_n)$.

To get started, for each $p \in \partial\Omega$, there is a neighborhood U of p in $\mathbb{R}^n$, a rotation of coordinate axes, and a C^1 function $u : \mathcal{O} \to \mathbb{R}$, defined on an open set $\mathcal{O} \subset \mathbb{R}^{n-1}$, such that

$$\Omega \cap U = \{x \in \mathbb{R}^n : x_n = u(x'), x' \in \mathcal{O}\} \cap U, \tag{6.3.3}$$

where $x = (x', x_n)$, $x' = (x_1, \dots, x_{n-1})$. We will obtain Theorem 6.3.1 from the following.

Proposition 6.3.2. *In the setting of Theorem 6.3.1, if $f \in C^1(\overline{\Omega})$ and e is an element of $\mathbb{R}^n$,*

$$\int_{\Omega} e \cdot \nabla f(x)\, dx = \int_{\partial\Omega} (e \cdot N) f\, dS. \tag{6.3.4}$$

In fact, taking $\{e_j\}$ to be the standard orthonormal basis of $\mathbb{R}^n$, replacing e by e_j and f by f_j, and summing, we have (6.3.1) as a consequence of (6.3.4).

To prove (6.3.4), after applying a partition of unity (see §6.6), we may as well suppose f is supported in such a set U as appears in (6.3.3). In such a case,

$$N = (1 + |\nabla u|^2)^{-1/2}(-\nabla u, 1). \tag{6.3.5}$$

Thus we have

$$\begin{aligned}
\int_{\Omega} \frac{\partial f}{\partial x_n}\, dx &= \int_{\mathcal{O}} \Big(\int_{x_n \le u(x')} \partial_n f(x', x_n)\, dx_n \Big) dx' \\
&= \int_{\mathcal{O}} f(x', u(x'))\, dx' \\
&= \int_{\partial\Omega} (e_n \cdot N) f\, dS.
\end{aligned} \tag{6.3.6}$$

The first identity in (6.3.6) follows from Theorem 5.1.10, the second identity from the Fundamental Theorem of Calculus, and the third identity from the identification

$$dS = (1 + |\nabla u|^2)^{1/2}\, dx', \tag{6.3.7}$$

established in (6.1.32). We use the standard basis $\{e_1, \dots, e_n\}$ of $\mathbb{R}^n$.

Such an argument works when e_n is replaced by any constant vector e with the property that we can represent $\partial\Omega \cap U$ as the graph of a function $y_n = \tilde{u}(y')$, with the y_n-axis parallel to e. In particular, it works for $e = e_n + ae_j$, for $1 \le j \le n-1$ and for $|a|$ sufficiently small. Thus we have

$$\int\limits_{\Omega} (e_n + ae_j) \cdot \nabla f(x)\, ds = \int\limits_{\partial\Omega} (e_n + ae_j) \cdot N\, f\, dS. \tag{6.3.8}$$

If we subtract (6.3.6) from this and divide the result by a, we obtain (6.3.4) for $e = e_j$, for all j, and hence (6.3.4) holds in general. The proof of Proposition 6.3.2, and hence of Theorem 6.3.1, is complete. □

We next specialize Theorem 6.3.1 to the case $n = 2$, and derive a classical Green theorem. If $\Omega \subset \mathbb{R}^2$ is a smoothly bounded open set, its boundary $\partial\Omega$ consists of a finite number of simple closed curves, of the form $\gamma : [a, b] \to \mathbb{R}^2$. We parametrize each such curve so that the unit tangent satisfies

$$T(x) = \frac{1}{|\gamma'(t)|}\gamma'(t) = JN(x), \quad x = \gamma(t), \tag{6.3.9}$$

where, as in §3.3,

$$J = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}. \tag{6.3.10}$$

See Figure 6.3.1.

Note. Here we take the opposite sign convention from what was used in §3.3. There we took $N = JT$.

Replacing F by JF in (6.3.1), we have

$$\begin{aligned} \int\limits_{\Omega} (\operatorname{div} JF)\, dx &= \int\limits_{\partial\Omega} N \cdot JF\, ds \\ &= -\int\limits_{\partial\Omega} F \cdot T\, ds. \end{aligned} \tag{6.3.11}$$

If γ is a boundary curve, we have from (6.3.9) and (6.1.27) (or (3.1.15)) that

$$\int\limits_{\gamma} F \cdot T\, ds = \int_a^b F(\gamma(t)) \cdot \gamma'(t)\, dt. \tag{6.3.12}$$

Now the integral on the right side of (6.3.12) can be cast as a *path integral* (or *line integral*), which we define as follows, in the n-dimensional setting. If $\gamma : [a, b] \to \mathbb{R}^n$ is a C^1 curve, we set

$$\int\limits_{\gamma} f_1\, dx_1 + \cdots + f_n\, dx_n = \int_a^b F(\gamma(t)) \cdot \gamma'(t)\, dt, \tag{6.3.13}$$

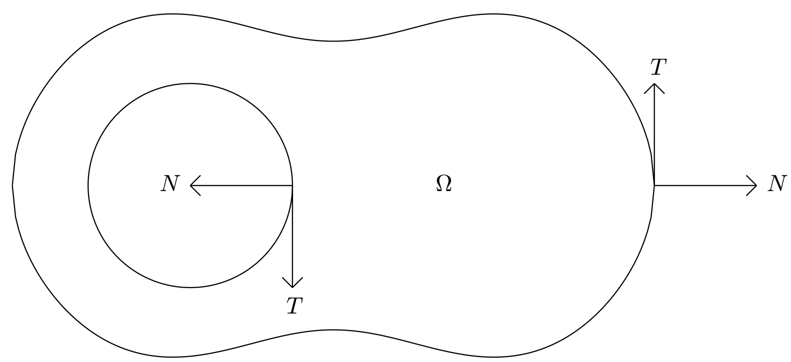

Figure 6.3.1. Smoothly bounded planar domain

for $F = (f_1, \dots, f_n)$.

With this notation, the identity (6.3.11) can be written as

$$\int\limits_{\Omega} \Big(\frac{\partial f_2}{\partial x_1} - \frac{\partial f_1}{\partial x_2} \Big)\, dx_1\, dx_2 = \int\limits_{\partial\Omega} f_1\, dx_1 + f_2\, dx_2. \tag{6.3.14}$$

Switching notation to $(x_1, x_2) = (x, y)$ and $(f_1, f_2) = (f, g)$, we have the following standard presentation of Green's theorem.

Proposition 6.3.3. *If $\Omega \subset \mathbb{R}^2$ is a smoothly bounded open set, and $f, g \in C^1(\overline{\Omega})$, then*

$$\int\limits_{\Omega} \Big(\frac{\partial g}{\partial x} - \frac{\partial f}{\partial y} \Big)\, dx\, dy = \int\limits_{\partial\Omega} f\, dx + g\, dy. \tag{6.3.15}$$

Remarks. See Appendix D for applications of Green's theorem to the study of complex differentiable functions.

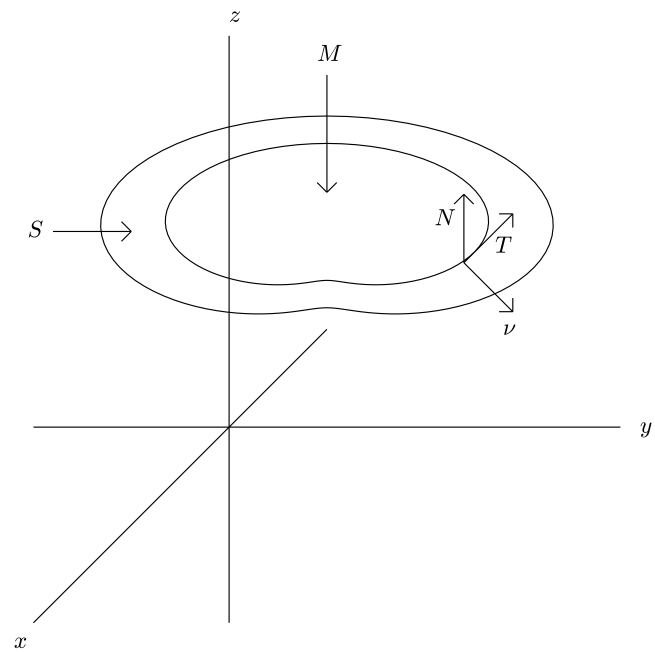

Figure 6.3.2. Set-up for the Stokes formula

We move to the Stokes formula. To formulate this, we bring in the notion of the curl of a vector field $F = (f, g, h)$ on an open set $\mathcal{O}$ in $\mathbb{R}^3$. We set

$$\begin{aligned} \operatorname{curl} F &= \det \begin{pmatrix} i & j & k \\ \partial_x & \partial_y & \partial_z \\ f & g & h \end{pmatrix} \\ &= (\partial_y h - \partial_z g, \partial_z f - \partial_x h, \partial_x g - \partial_y f). \end{aligned} \tag{6.3.16}$$

Here $\{i, j, k\}$ denotes the standard basis of $\mathbb{R}^3$. To give a special case, which strongly ties in with Green's formula, suppose $\mathcal{O}$ contains the planar domain

$$U = \{(x, y, 0) : (x, y) \in \Omega\}, \tag{6.3.17}$$

where $\Omega \subset \mathbb{R}^2$ is a smoothly bounded open set. Then

$$(\operatorname{curl} F) \cdot k = \frac{\partial g}{\partial x} - \frac{\partial f}{\partial y}, \tag{6.3.18}$$

and Green's formula (6.3.15) can be written

$$\int_U (\operatorname{curl} F) \cdot k \, dA = \int_{\partial U} (F \cdot T) \, ds. \tag{6.3.19}$$

Now let $S \subset \mathbb{R}^3$ be a smooth surface, and let $\overline{M} \subset S$ be a smoothly bounded subset. Assume there is a smooth unit normal field N on S. Parametrize the boundary curves that make up ∂M so that the unit tangent $T(x)$ at each $x \in \partial M$ satisfies

$$T(x) \times N(x) = \nu(x) \text{ is the outward pointing normal to } \partial M \text{ in } S. \tag{6.3.20}$$

See Figure 6.3.2. Here is the Stokes formula.

Proposition 6.3.4. *If F is a C^1 vector field on a neighbprhood $\mathcal{O}$ of $\overline{M}$ in $\mathbb{R}^3$, then*

$$\int_M (\operatorname{curl} F) \cdot N \, dS = \int_{\partial M} (F \cdot T) \, ds. \tag{6.3.21}$$

We begin by treating the following special case. Assume $\overline{M}$ is the graph in $\mathbb{R}^3$ of a smooth function

$$z = u(x, y), \quad (x, y) \in \overline{\mathcal{O}}, \tag{6.3.22}$$

where $\mathcal{O}$ is a smoothly bounded open set in $\mathbb{R}^2$. In such a case, we take

$$N = (1 + |\nabla u|^2)^{-1/2} \Big(-\frac{\partial u}{\partial x}, -\frac{\partial u}{\partial y}, 1 \Big), \tag{6.3.23}$$

at $(x, y, z) = (x, y, u(x, y)) \in M$. Also,

$$dS = (1 + |\nabla u|^2)^{1/2} \, dx \, dy, \tag{6.3.24}$$

so

$$\begin{aligned} \int_M (\operatorname{curl} F) \cdot N \, dS = \int_{\mathcal{O}} \Big[\Big(\frac{\partial F_3}{\partial y} - \frac{\partial F_2}{\partial z} \Big) \Big(-\frac{\partial u}{\partial x} \Big) + \Big(\frac{\partial F_1}{\partial z} - \frac{\partial F_3}{\partial x} \Big) \Big(-\frac{\partial u}{\partial y} \Big) \\ + \Big(\frac{\partial F_3}{\partial x} - \frac{\partial F_1}{\partial y} \Big) \Big] \, dx \, dy, \end{aligned} \tag{6.3.25}$$

where $\partial F_j / \partial x$, etc., are evaluated at $(x, y, z) = (x, y, u(x, y))$. On the other hand, if

$$\sigma(t) = (x(t), y(t)), \quad a \le t \le b, \tag{6.3.26}$$

parametrizes a boundary curve of $\partial \mathcal{O}$, then

$$\gamma(t) = (x(t), y(t), u(x(t), y(t))) \tag{6.3.27}$$

parametrizes the corresponding boundary curve of $\overline{M}$, and we have

$$\begin{aligned} \int_{\partial M} (F \cdot T) \, ds &= \int_a^b F(\sigma(t), u(\sigma(t))) \cdot \gamma'(t) \, dt \\ &= \int_{\partial \mathcal{O}} \Big(\widetilde{F}_1 + \widetilde{F}_3 \frac{\partial u}{\partial x} \Big) \, dx + \Big(\widetilde{F}_2 + \widetilde{F}_3 \frac{\partial u}{\partial y} \Big) \, dy, \end{aligned} \tag{6.3.28}$$

where

$$\widetilde{F}_j(x, y) = F_j(x, y, u(x, y)). \tag{6.3.29}$$

Now apply Green's theorem, with

$$f = \widetilde{F}_1 + \widetilde{F}_3 \frac{\partial u}{\partial x}, \quad g = \widetilde{F}_2 + \widetilde{F}_3 \frac{\partial u}{\partial y}. \tag{6.3.30}$$

One verifies that $\partial_x g - \partial_y f$ is equal to the integrand on the right side of (6.3.25) (see Exercise 7 below). This implies that the right sides of (6.3.25) and (6.3.28) are equal, and we have (6.3.21) for this class of surfaces.

A similar argument works when $\overline{M}$ is the graph in $\mathbb{R}^3$ of a smooth function

$$y = v(x, z), \quad \text{or} \quad x = w(y, z), \tag{6.3.31}$$

over smoothly bounded open sets in the (x, z)-plane or the (y, z)-plane, respectively. Now, in the general case of Proposition 6.3.4, one can use a partition of unity to write F as a finite sum of vector fields supported in portions of $\overline{M}$ for which (6.3.22) or one of the cases in (6.3.31) can be arranged. In this fashion we have Proposition 6.3.4 in general.

We return to the setting of Theorem 6.3.1 and obtain some further integral identities. First we apply (6.3.1) with F replaced by uX, where X is a vector field and u is real valued. We have the following "derivation" identity:

$$\operatorname{div} uX = u \operatorname{div} X + X \cdot \nabla u, \tag{6.3.32}$$

which follows easily from (6.3.2). Theorem 6.3.1 gives

$$\int_{\Omega} (\operatorname{div} X) u \, dV + \int_{\Omega} X \cdot \nabla u \, dV = \int_{\partial\Omega} (N \cdot X) u \, dS. \tag{6.3.33}$$

Replacing u by uv and using the derivation identity $\nabla(uv) = v\nabla u + u\nabla v$, we have

$$\int_{\Omega} \big[(X \cdot \nabla u)v + u(X \cdot \nabla v)\big] \, dV = -\int_{\Omega} (\operatorname{div} X) uv \, dV + \int_{\partial\Omega} (N \cdot X) uv \, dS. \tag{6.3.34}$$

It is useful to apply (6.3.33) to a gradient vector field X. Applying div to $X = \nabla v$ defines the *Laplace operator*:

$$\Delta v = \operatorname{div} \nabla v = \frac{\partial^2 v}{\partial x_1^2} + \cdots + \frac{\partial^2 v}{\partial x_n^2}. \tag{6.3.35}$$

Now setting $X = \nabla v$ in (6.3.33), we have $X \cdot N = N \cdot \nabla v$, which we call the normal derivative of v, and denote $\partial v / \partial \nu$. Hence

$$\int_{\Omega} u(\Delta v) \, dV = -\int_{\Omega} (\nabla u \cdot \nabla v) \, dV + \int_{\partial\Omega} u \frac{\partial v}{\partial \nu} \, dS. \tag{6.3.36}$$

If we interchange the roles of u and v and subtract, we have

$$\int_{\Omega} u(\Delta v) \, dV = \int_{\Omega} (\Delta u) v \, dV + \int_{\partial\Omega} \Big[u \frac{\partial v}{\partial \nu} - \frac{\partial u}{\partial \nu} v \Big] \, dS. \tag{6.3.37}$$

Formulas (6.3.36)–(6.3.37) are also called Green formulas. Applications are brought up in the exercises below.

Exercises

1. Let X and Y be smooth vector fields on an open set $\Omega \subset \mathbb{R}^3$. Show that

$$Y \cdot \operatorname{curl} X - X \cdot \operatorname{curl} Y = \operatorname{div}(X \times Y).$$

2. In the setting of Exercise 1, assume $\overline{\Omega}$ is compact and smoothly bounded, and that X and Y are C^1 on $\overline{\Omega}$. Show that

$$\int_\Omega X \cdot \operatorname{curl} Y \, dx = \int_\Omega Y \cdot \operatorname{curl} X \, dx,$$

provided either X is normal to $\partial\Omega$ or X is parallel to Y on $\partial\Omega$.

3. Show that, with $x = r\omega \in \mathbb{R}^n$, $\omega \in S^{n-1}$,

$$u(x) = f(|x|) \Longrightarrow \Delta u(r\omega) = f''(r) + \frac{n-1}{r} f'(r).$$

4. We say $f \in C^2(\Omega)$ is harmonic on $\Omega \subset \mathbb{R}^n$ if $\Delta f = 0$ on Ω. Show that

$$|x|^{-(n-2)} \ \text{ is harmonic on } \ \mathbb{R}^n \setminus 0.$$

In case $n = 2$, show that

$$\log |x| \ \text{ is harmonic on } \ \mathbb{R}^2 \setminus 0.$$

In Exercise 5, we take $n \geq 3$ and consider

$$\begin{aligned} Gf(x) &= \frac{1}{C_n} \int_{\mathbb{R}^n} \frac{f(y)}{|x-y|^{n-2}} \, dy \\ &= \frac{1}{C_n} \int_{\mathbb{R}^n} \frac{f(x-y)}{|y|^{n-2}} \, dy, \end{aligned}$$

with $C_n = -(n-2)A_{n-1}$.

5. Assume $f \in C_0^2(\mathbb{R}^n)$. Let $\Omega_\varepsilon = \mathbb{R}^n \setminus B_\varepsilon$, where $B_\varepsilon = \{x \in \mathbb{R}^n : |x| < \varepsilon\}$. Verify that

$$\begin{aligned} C_n \Delta Gf(0) &= \lim_{\varepsilon \to 0} \int_{\Omega_\varepsilon} \Delta f(x) \cdot |x|^{2-n} \, dx \\ &= \lim_{\varepsilon \to 0} \int_{\Omega_\varepsilon} [\Delta f(x) \cdot |x|^{2-n} - f(x) \Delta |x|^{2-n}] \, dx \\ &= -\lim_{\varepsilon \to 0} \int_{\partial\Omega_\varepsilon} \Big[\varepsilon^{2-n} \frac{\partial f}{\partial r} - (2-n)\varepsilon^{1-n} f \Big] \, dS \\ &= -(n-2)A_{n-1} f(0), \end{aligned}$$

using (6.3.37) for the third identity. Use this to show that

$$\Delta G f(x) = f(x).$$

6. Work out the analogue of Exercise 5 in case $n = 2$ and

$$Gf(x) = \frac{1}{2\pi} \int_{\mathbb{R}^2} f(y) \log |x - y| \, dy.$$

7. For f and g given by (6.3.30), compute $\partial g/\partial x$ and $\partial f/\partial y$, and verify that

$$\int_{\mathcal{O}} \Bigl(\frac{\partial g}{\partial x} - \frac{\partial f}{\partial y} \Bigr) \, dx \, dy$$

is equal to the right side of (6.3.25). Ingredients in the calculation include

$$\frac{\partial}{\partial x} F_j(x, y, u(x, y)) = \frac{\partial F_j}{\partial x}(x, y, u(x, y)) + \frac{\partial F_j}{\partial z}(x, y, u(x, y)) \frac{\partial u}{\partial x},$$

and the counterpart for the application of $\partial/\partial y$. Another ingredient involves the identity

$$\frac{\partial^2 u}{\partial x \, \partial y} = \frac{\partial^2 u}{\partial y \, \partial x}.$$

6.4. Projective spaces, quotient surfaces, and manifolds

Real projective space $\mathbb{P}^{n-1}$ is obtained from the sphere S^{n-1} by identifying each pair of antipodal points:

$$\mathbb{P}^{n-1} = S^{n-1}/\sim, \tag{6.4.1}$$

where

$$x \sim y \Longleftrightarrow x = \pm y, \tag{6.4.2}$$

for $x, y \in S^{n-1} \subset \mathbb{R}^n$. More generally, if $M \subset \mathbb{R}^n$ is an m-dimensional surface, smooth of class C^k, satisfying

$$0 \notin M, \quad x \in M \Rightarrow -x \in M, \tag{6.4.3}$$

we define

$$\mathbb{P}(M) = M/\sim, \tag{6.4.4}$$

using the equivalence relation (6.4.2). Note that M has the metric space structure $d(x,y) = \|x-y\|$, and then $\mathbb{P}(M)$ becomes a metric space with metric

$$d([x],[y]) = \min\{d(x',y') : x' \in [x],\ y' \in [y]\}, \tag{6.4.5}$$

or, in view of (6.4.2),

$$d([x],[y]) = \min\{d(x,y), d(x,-y)\}. \tag{6.4.6}$$

Here, $x \in M$ and $[x] \in \mathbb{P}(M)$ is its associated equivalence class. The map $x \mapsto [x]$ is a continuous map

$$\rho : M \longrightarrow \mathbb{P}(M). \tag{6.4.7}$$

It has the following readily established property.

Lemma 6.4.1. *Each $p \in \mathbb{P}(M)$ has an open neighborhood $U \subset \mathbb{P}(M)$ such that $\rho^{-1}(U) = U_0 \cup U_1$ is the disjoint union of two open subsets of M, and, for $j = 0, 1$, $\rho : U_j \to U$ is a homeomorphism, i.e., it is continuous, one-to-one, and onto, with continuous inverse.*

Given $p \in \mathbb{P}(M)$, $\{p_0, p_1\} = \rho^{-1}(p)$, let U_0 be a neighborhood of p_0 in M for which there is a C^k coordinate chart $\varphi_0 : \mathcal{O} \to U_0$ ($\mathcal{O} \subset \mathbb{R}^m$ open). Then $\varphi_1(x) = -\varphi_0(x)$ gives a coordinate chart $\varphi_1 : \mathcal{O} \to U_1$ onto a neighborhood U_1 of $p_1 \in M$. If U_0 is picked small enough, U_0 and U_1 are disjoint. The projection ρ maps U_0 and U_1 homeomorphically onto a neighborhood U of p in $\mathbb{P}(M)$, and we have "coordinate charts"

$$\rho \circ \varphi_j : \mathcal{O} \longrightarrow U. \tag{6.4.8}$$

In fact, $\rho \circ \varphi_1 = \rho \circ \varphi_0$. If $\psi_0 : \Omega \to U_0$ is another C^k coordinate chart, then, as in Lemma 6.1.1, we have a C^k diffeomorphism $F : \mathcal{O} \to \Omega$ such that $\psi_0 \circ F = \varphi_0$. Similarly $\psi_1 \circ F = \varphi_1$, with $\psi_1(x) = -\psi_0(x)$, and we have $\rho \circ \psi_j \circ F = \rho \circ \varphi_j$.

The structure just placed on the "quotient surface" $\mathbb{P}(M)$ makes it a *manifold*, an object we now define.

Given a metric space X, we say X has the structure of a C^k manifold of dimension m provided the following conditions hold. First, for each $p \in X$, we have an open neighborhood U_p of p in X, an open set $\mathcal{O}_p \subset \mathbb{R}^m$, and a homeomorphism

$$\varphi_p : \mathcal{O}_p \longrightarrow U_p. \tag{6.4.9}$$

Next, if also $q \in X$ and $U_{pq} = U_p \cap U_q \neq \emptyset$, then the homeomorphism from $\mathcal{O}_{pq} = \varphi_p^{-1}(U_{pq})$ to $\mathcal{O}_{qp} = \varphi_q^{-1}(U_{pq})$,

$$F_{pq} = \varphi_q^{-1} \circ \varphi_p\big|_{\mathcal{O}_{pq}}, \tag{6.4.10}$$

is a C^k diffeomorphism. As before, we call the maps $\varphi_p : \mathcal{O}_p \to U_p \subset X$ coordinate charts on X.

A metric tensor on a C^k manifold X is defined by positive-definite, symmetric $m \times m$ matrices $G_p \in C^{k-1}(\mathcal{O}_p)$, satisfying the compatibility condition

$$G_p(x) = DF_{pq}(x)^t G_q(y) DF_{pq}(x), \tag{6.4.11}$$

for

$$x \in \mathcal{O}_{pq} \subset \mathcal{O}_p, \quad y = F_{pq}(x) \in \mathcal{O}_{qp} \subset \mathcal{O}_q. \tag{6.4.12}$$

We then set

$$g_p = \det G_p \in C^{k-1}(\mathcal{O}_p), \tag{6.4.13}$$

satisfying

$$\sqrt{g_p(x)} = |\det DF_{pq}(x)| \sqrt{g_q(y)}, \tag{6.4.14}$$

for x and y as in (6.4.12). If $f : X \to \mathbb{R}$ is a continuous function supported in U_p, we set

$$\int_X f\, dS = \int_{\mathcal{O}_p} f(\varphi_p(x)) \sqrt{g_p(x)}\, dx. \tag{6.4.15}$$

As in (6.1.24)–(6.1.25), this leads to a well defined integral $\int_X f\, dS$ for $f \in C_c(X)$, obtained by writing f as a finite sum of continuous functions supported on various coordinate patches U_p. From here we can develop the class of functions $\mathcal{R}_c(X)$ and their integrals over X, in a fashion parallel to that done above when X is a surface in $\mathbb{R}^n$.

The quotient surfaces $\mathbb{P}(M)$ are examples of C^k manifolds as defined above. They get natural metric tensors with the property that ρ in (6.4.7) is a local isometry. In such a case,

$$\int_{\mathbb{P}(M)} f\, dS = \frac{1}{2} \int_M f \circ \rho\, dS. \tag{6.4.16}$$

Another important quotient manifold is the "flat torus"

$$\mathbb{T}^n = \mathbb{R}^n / \mathbb{Z}^n. \tag{6.4.17}$$

Here the equivalence relation on $\mathbb{R}^n$ is $x \sim y \Leftrightarrow x - y \in \mathbb{Z}^n$. Natural local coordinates on $\mathbb{T}^n$ are given by the projection $\rho : \mathbb{R}^n \to \mathbb{T}^n$, restricted to sufficiently small open sets in $\mathbb{R}^n$. The quotient $\mathbb{T}^n$ gets a natural metric tensor for which ρ is a local isometry.

Given two C^k manifolds X and Y, a continuous map $\psi : X \to Y$ is said to be smooth of class C^k provided that for each $p \in X$, there are neighborhoods U of p and $\widetilde{U}$ of $q = \psi(p)$, and coordinate charts $\varphi_1 : \mathcal{O} \to U$, $\varphi_2 : \widetilde{\mathcal{O}} \to \widetilde{U}$, such that $\varphi_2^{-1} \circ \psi \circ \varphi_1 : \mathcal{O} \to \widetilde{\mathcal{O}}$ is a C^k map. We say ψ is a C^k diffeomorphism if it is one-to-one and onto and $\psi^{-1} : Y \to X$ is a C^k map. If X is a C^k manifold and $M \subset \mathbb{R}^n$ a C^k surface, a C^k diffeomorphism $\psi : X \to M$ is called a C^k embedding of X into $\mathbb{R}^n$.

Here is an embedding of $\mathbb{T}^n$ into $\mathbb{R}^{2n}$:

$$\psi(x) = \sum_{j=1}^{n} (\cos 2\pi x_j) e_j + \sum_{j=1}^{n} (\sin 2\pi x_j) e_{n+j}. \tag{6.4.18}$$

A priori, $\psi : \mathbb{R}^n \to \mathbb{R}^{2n}$, but $\psi(x) = \psi(y)$ whenever $x - y \in \mathbb{Z}^n$, so this naturally induces a smooth map $\mathbb{T}^n \to \mathbb{R}^{2n}$, which can be seen to be an embedding.

If $M \subset \mathbb{R}^n$ is an m-dimensional surface satisfying (6.4.3), an embedding of $\mathbb{P}(M)$ into $M(n, \mathbb{R})$ can be constructed via the map

$$\psi : \mathbb{R}^n \longrightarrow M(n, \mathbb{R}), \quad \psi(x) = xx^t. \tag{6.4.19}$$

Note that

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \Longrightarrow xx^t = \begin{pmatrix} x_1^2 & \cdots & x_1 x_j & \cdots & x_1 x_n \\ \vdots & & \vdots & & \vdots \\ x_n x_1 & \cdots & x_n x_j & \cdots & x_n^2 \end{pmatrix}. \tag{6.4.20}$$

We need a couple of lemmas.

Lemma 6.4.2. *For ψ as in* (6.4.19), *$x, y \in \mathbb{R}^n$,*

$$\psi(x) = \psi(y) \Longleftrightarrow x = \pm y. \tag{6.4.21}$$

Proof. The map ψ is characterized by $\psi(x)e_j = x_j x$, where x is as in (6.4.20) and $\{e_j\}$ is the standard basis of $\mathbb{R}^n$. It follows that if $x \neq 0$, $\psi(x)$ has exactly one nonzero eigenvalue, namely $|x|^2$, and $\psi(x)x = |x|^2 x$. Thus $\psi(x) = \psi(y)$ implies that $|x|^2 = |y|^2$ and that x and y are parallel. Thus $x = ay$ and $a = \pm 1$. □

Lemma 6.4.3. *In the setting of Lemma 6.4.2, if $x \neq 0$,*

$$D\psi(x) : \mathbb{R}^n \longrightarrow M(n, \mathbb{R}) \quad \text{is injective.} \tag{6.4.22}$$

Proof. A calculation gives

$$D\psi(x)v = xv^t + vx^t. \tag{6.4.23}$$

Thus, if $v \in \ker D\psi(x)$,

$$xv^t = -vx^t. \tag{6.4.24}$$

Both sides are rank 1 elements of $M(n, \mathbb{R})$. The range of the left side is spanned by x and that of the right side is spanned by v, so $v = ax$ for some $a \in \mathbb{R}$. Then (6.4.24) becomes

$$axx^t = -axx^t, \tag{6.4.25}$$

which implies $a = 0$ if $x \neq 0$. □

REMARK. Here is a refinement of Lemma 6.4.3. Using the inner product on $M(n,\mathbb{R})$ given by (6.1.60), we can calculate

$$\langle D\psi(x)v, D\psi(x)v\rangle = 2\big(|x|^2|v|^2 + (x\cdot v)^2\big). \tag{6.4.26}$$

Lemmas 6.4.2 and 6.4.3 imply that if $M \subset \mathbb{R}^n$ is an m-dimensional surface satisfying (6.4.3), then $\psi|_M$ yields an embedding of $\mathbb{P}(M)$ into $M(n,\mathbb{R})$. Denote the image surface by $M^\#$. As we see from (6.4.26), this embedding is not typically an isometry. However, if $M = S^{n-1}$ and v is tangent to S^{n-1} at x, then $v\cdot x = 0$, and (6.4.26) implies that in this case the embedding of $\mathbb{P}^{n-1}$ into $M(n,\mathbb{R})$ is an isometry, up to a factor of 2.

It is the case that if X is any C^k manifold that is a countable union of compact sets, then X can be embedded into $\mathbb{R}^n$ for some n. In case X is compact, this is not very hard to prove, using local coordinate charts and smooth cutoffs, and the interested reader might take a crack at it. If X is provided with a metric tensor, this embedding might not preserve this metric tensor. If it does, one calls is an isometric embedding. It is the case that any such manifold has an isometric embedding into $\mathbb{R}^n$ for some n (if k is sufficiently large). This result is the famous Nash embedding theorem, and its proof is quite difficult. For X compact and C^∞, a proof is given in Chapter 14 of [**18**].

Exercises

1. In case $n = 3$, show that the map $\psi : M \to M(3,\mathbb{R})$ given by (6.4.19)–(6.4.20) is equivalent to

$$\psi^\# : M \to \mathbb{R}^6, \quad \psi^\#(x) = (x_jx_k : 1 \le j \le k \le 3).$$

Deduce that

$$\psi^b : S^2 \to \mathbb{R}^5, \quad \psi^b(x) = (x_1^2 - x_2^2, x_1^2 - x_3^2, x_1x_2, x_2x_3, x_3x_1)$$

has image that is diffeomorphic to $\mathbb{P}^2$.
Hint. $x_1^2 + x_2^2 + x_3^2 \equiv 1$ on S^2.

2. The map

$$x : \mathbb{R}\times\mathbb{R} \longrightarrow S^2,$$

given by

$$x(\theta,\psi) = (\sin\theta\cos\psi, \sin\theta\sin\psi, \cos\theta),$$

is a smooth map of $\mathbb{R}\times\mathbb{R}$ onto S^2, giving *spherical coordinates*. See Figure 6.1.3. Show that this gives rise to a diffeomorphism

$$x : (0,\pi)\times\mathbb{R}/2\pi\mathbb{Z} \longrightarrow S^2 \setminus \{\pm e_3\}.$$

3. Compose the map x of Exercise 2 with the map, given by Exericse 1, of S^2 onto a surface in $\mathbb{R}^5$ that is diffeomorphic to $\mathbb{P}^2$.

6.5. Polar decomposition of matrices

We define the spaces Sym(n) and $\mathcal{P}(n)$ by

$$\begin{aligned} \operatorname{Sym}(n) &= \{A \in M(n,\mathbb{R}) : A = A^t\}, \\ \mathcal{P}(n) &= \{A \in \operatorname{Sym}(n) : x \cdot Ax > 0,\ \forall\, x \in \mathbb{R}^n \setminus 0\}. \end{aligned} \tag{6.5.1}$$

It is easy to show that $\mathcal{P}(n)$ is an open, convex subset of the linear space Sym(n). We aim to prove the following result.

Proposition 6.5.1. *Given $A \in Gl_+(n,\mathbb{R})$, there exist unique $U \in SO(n)$ and $Q \in \mathcal{P}(n)$ such that*

$$A = UQ. \tag{6.5.2}$$

The representation (6.5.2) is called the polar decomposition of A. Note that

$$(UQ)^t UQ = QU^t UQ = Q^2, \tag{6.5.3}$$

so if the identity (6.5.2) were to hold, we would have

$$A^t A = Q^2. \tag{6.5.4}$$

Note also that

$$A \in Gl(n,\mathbb{R}) \Longrightarrow A^t A \in \mathcal{P}(n), \tag{6.5.5}$$

since $x \cdot A^t A x = (Ax) \cdot (Ax) = |Ax|^2$.

To prove Proposition 6.5.1, we bring in the following basic result of linear algebra. See Appendix C.2.

Proposition 6.5.2. *Given $B \in \operatorname{Sym}(n)$, there is an orthonormal basis of $\mathbb{R}^n$ consisting of eigenvectors of B, with eigenvalues $\lambda_j \in \mathbb{R}$. Equivalently, there exists $V \in SO(n)$ such that*

$$B = VDV^{-1}, \tag{6.5.6}$$

with

$$D = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}, \tag{6.5.7}$$

$\lambda_j \in \mathbb{R}$.

If $B \in \mathcal{P}(n)$, then each $\lambda_j > 0$. We can then set

$$Q = V \begin{pmatrix} \lambda_1^{1/2} & & \\ & \ddots & \\ & & \lambda_n^{1/2} \end{pmatrix} V^{-1}, \tag{6.5.8}$$

and obtain the following.

Corollary 6.5.3. *Given $B \in \mathcal{P}(n)$, there is a unique $Q \in \mathcal{P}(n)$ satisfying*

$$Q^2 = B. \tag{6.5.9}$$

We say $Q = B^{1/2}$.

To obtain the decomposition (6.5.2), we set

$$Q = (A^t A)^{1/2}, \quad U = AQ^{-1}. \tag{6.5.10}$$

Note that

$$U^t U = Q^{-1} A^t A Q^{-1} = Q^{-1} Q^2 Q^{-1} = I, \tag{6.5.11}$$

and $(\det U)(\det Q) = \det A > 0$, so $\det U > 0$, and hence $U \in SO(n)$, as desired. By (6.5.4) and Corollary 6.5.3, the factor $Q \in \mathcal{P}(n)$ in (6.5.2) is unique, and hence so is the factor U.

We can use Proposition 6.5.1 to prove the following.

Proposition 6.5.4. *The set $Gl_+(n, \mathbb{R})$ is connected. In fact, given $A \in Gl_+(n, \mathbb{R})$, there is a smooth path $\gamma : [0, 1] \to Gl_+(n, \mathbb{R})$ such that $\gamma(0) = I$ and $\gamma(1) = A$.*

Proof. To start, we have that

$$\text{Exp} : \text{Skew}(n) \longrightarrow SO(n) \ \text{ is onto.} \tag{6.5.12}$$

See Exercise 14 below for this (or Corollary C.2.9). Hence, with $A = UQ$ as in (6.5.2), we have a smooth path $\alpha(t) = \text{Exp}(tS)$, $\alpha : [0, 1] \to SO(n)$, such that $\alpha(0) = I$ and $\alpha(1) = U$. Since $\mathcal{P}(n)$ is a convex subset of $\text{Sym}(n)$, we can take $\beta(t) = (1-t)I + tQ$, obtaining a smooth path $\beta : [0, 1] \to \mathcal{P}(n)$, such that $\beta(0) = I$ and $\beta(1) = Q$. Then

$$\gamma(t) = \alpha(t)\beta(t) \tag{6.5.13}$$

does the trick. $\square$

Exercises

1. Establish the following counterpart to Proposition 6.5.1. Set

$$Gl_-(n, \mathbb{R}) = \{A \in M(n, \mathbb{R}) : \det A < 0\},$$

$$O^-(n) = \{U \in O(n) : \det U = -1\}.$$

Proposition. Given $A \in Gl_-(n, \mathbb{R})$, there exist unique $U \in O^-(n)$ and $Q \in \mathcal{P}(n)$ such that $A = UQ$.

Hint. As in the proof of Proposition 6.5.1, take $A^t A = Q^2$.

6.6. Partitions of unity

In the text we have made occasional use of partitions of unity, and we include some material on this topic here. We begin by defining and constructing a continuous partition of unity on a compact metric space, subordinate to a open cover $\{U_j : 1 \le j \le N\}$ of X. By definition, this is a family of continuous functions $\varphi_j : X \to \mathbb{R}$ such that

$$\varphi_j \ge 0, \quad \operatorname{supp} \varphi_j \subset U_j, \quad \sum_j \varphi_j = 1. \tag{6.6.1}$$

To construct such a partition of unity, we do the following. First, it can be shown that there is an open cover $\{V_j : 1 \le j \le N\}$ of X and open sets W_j such that

$$\overline{V}_j \subset W_j \subset \overline{W}_j \subset U_j. \tag{6.6.2}$$

Given this, let $\psi_j(x) = \operatorname{dist}(x, X \setminus W_j)$. Then ψ_j is continuous, $\operatorname{supp} \psi_j \subset \overline{W}_j \subset U_j$, and ψ_j is strictly positive on $\overline{V}_j$. Hence $\Psi = \sum_j \psi_j$ is continuous and strictly positive on X, and we see that

$$\varphi_j(x) = \Psi(x)^{-1} \psi_j(x) \tag{6.6.3}$$

yields such a partition of unity.

We indicate how to construct the sets V_j and W_j used above, starting with V_1 and W_1. Note that the set $K_1 = X \setminus (U_2 \cup \cdots \cup U_N)$ is a compact subset of U_1. Assume it is nonempty; otherwise just throw U_1 out and relabel the sets U_j. Now set

$$V_1 = \{x \in U_1 : \operatorname{dist}(x, K_1) < \tfrac{1}{3} \operatorname{dist}(x, X \setminus U_1)\},$$

and

$$W_1 = \{x \in U_1 : \operatorname{dist}(x, K_1) < \tfrac{2}{3} \operatorname{dist}(x, X \setminus U_1)\}.$$

To construct V_2 and W_2, proceed as above, but use the cover $\{U_2, \dots, U_N, V_1\}$. Continue until done.

Given a smooth compact surface M (perhaps with boundary), covered by coordinate patches U_j $(1 \le j \le N)$, one can construct a *smooth* partition of unity on M, subordinate to this cover. The main additional tool for this is the construction of a function $\psi \in C_0^\infty(\mathbb{R}^n)$ such that

$$\psi(x) = 1 \ \text{ for } \ |x| \le \frac{1}{2}, \quad \psi(x) = 0 \ \text{ for } \ |x| \ge 1. \tag{6.6.4}$$

One way to get this is to start with the function on $\mathbb{R}$ given by

$$\begin{aligned} f_0(x) = e^{-1/x} \quad & \text{for} \ \ x > 0, \\ 0 \qquad & \text{for} \ \ x \le 0. \end{aligned} \tag{6.6.5}$$

It is an exercise to show that

$$f_0 \in C^\infty(\mathbb{R}).$$

Now the function

$$f_1(x) = f_0(x) f_0(\tfrac{1}{2} - x)$$

belongs to $C^\infty(\mathbb{R})$ and is zero outside the interval $[0, 1/2]$. See Figure 6.6.1.

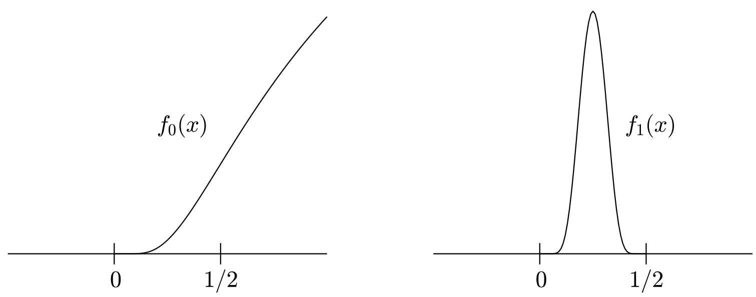

Figure 6.6.1. The bump function $f_1(x) = f_0(x)f_0(1/2 - x)$

Hence the function

$$f_2(x) = \int_{-\infty}^{x} f_1(s)\, ds$$

belongs to $C^\infty(\mathbb{R})$, is zero for $x \leq 0$, and equals some positive constant (say C_2) for $x \geq 1/2$. Then

$$\psi(x) = \frac{1}{C_2}\, f_2(1 - |x|)$$

is a function on $\mathbb{R}^n$ with the desired properties.

With this function in hand, to construct the smooth partition of unity mentioned above is an exercise we recommend to the reader.

Appendix A

Foundational material on the real numbers

One foundation for a course in analysis is a solid understanding of the real number system. This appendix provides a development of $\mathbb{R}$. It presupposes an understanding of basic algebraic results on the set $\mathbb{Q}$ of rational numbers, and derives the structure of $\mathbb{R}$ from there.

Section A.1 deals with infinite sequences, including convergent sequences and "Cauchy sequences." This prepares the way for §A.2. Here we construct the set $\mathbb{R}$ of real numbers, as "ideal limits" of rational numbers. We extend basic algebraic results from $\mathbb{Q}$ to $\mathbb{R}$. Furthermore, we establish the result that $\mathbb{R}$ is "complete," i.e., Cauchy sequences always have limits in $\mathbb{R}$.

Section A.3 establishes further metric properties of $\mathbb{R}$ and various subsets, with an emphasisis on the notion of *compactness*. The completeness property established in §A.2 plays a crucial role here.

Section A.4 introduces the set $\mathbb{C}$ of complex numbers and establishes basic algebraic and metric properties of $\mathbb{C}$. While some introductory treatments of analysis avoid complex numbers, we embrace them, and consider their use in basic analysis too precious to omit.

A.1. Infinite sequences

In this section, we discuss infinite sequences. For now, we deal with sequences of rational numbers, but we will not explicitly state this restriction below. In fact, once the set of real numbers is constructed in §A.2, the results of this section will be seen to hold also for sequences of real numbers.

Definition. A sequence (a_j) is said to converge to a limit a provided that, for any $n \in \mathbb{N}$, there exists $K(n)$ such that

$$j \geq K(n) \Longrightarrow |a_j - a| < \frac{1}{n}. \tag{A.1.1}$$

We write $a_j \to a$, or $a = \lim\ a_j$, or perhaps $a = \lim_{j\to\infty}\ a_j$.

Here, we define the absolute value $|x|$ of x by

$$\begin{aligned} |x| = &\ \ x \ \text{ if } \ x \geq 0, \\ &-x \ \text{ if } \ x < 0. \end{aligned} \tag{A.1.2}$$

The absolute value function has various simple properties, such as $|xy| = |x| \cdot |y|$, which follow readily from the definition. One basic property is the triangle inequality:

$$|x+y| \leq |x| + |y|. \tag{A.1.3}$$

In fact, if either x and y are both positive or they are both negative, one has equality in (A.1.3). If x and y have opposite signs, then $|x+y| \leq \max(|x|, |y|)$, which in turn is dominated by the right side of (A.1.3).

Proposition A.1.1. *If $a_j \to a$ and $b_j \to b$, then*

$$a_j + b_j \to a + b, \tag{A.1.4}$$

and

$$a_j b_j \to ab. \tag{A.1.5}$$

If furthermore, $b_j \neq 0$ for all j and $b \neq 0$, then

$$a_j / b_j \to a/b. \tag{A.1.6}$$

Proof. To see (A.1.4), we have, by (A.1.3),

$$|(a_j + b_j) - (a+b)| \leq |a_j - a| + |b_j - b|. \tag{A.1.7}$$

To get (A.1.5), we have

$$\begin{aligned} |a_j b_j - ab| &= |(a_j b_j - a b_j) + (a b_j - ab)| \\ &\leq |b_j| \cdot |a_j - a| + |a| \cdot |b - b_j|. \end{aligned} \tag{A.1.8}$$

The hypotheses imply $|b_j| \leq B$, for some B, and hence the criterion for convergence is readily verified. To get (A.1.6), we have

$$\left| \frac{a_j}{b_j} - \frac{a}{b} \right| \leq \frac{1}{|b| \cdot |b_j|} \big\{ |b| \cdot |a - a_j| + |a| \cdot |b - b_j| \big\}. \tag{A.1.9}$$

The hypotheses imply $1/|b_j| \leq M$ for some M, so we also verify the criterion for convergence in this case. □

We next define the concept of a Cauchy sequence.

Definition. A sequence (a_j) is a Cauchy sequence provided that, for any $n \in \mathbb{N}$, there exists $K(n)$ such that

$$j, k \geq K(n) \Longrightarrow |a_j - a_k| \leq \frac{1}{n}. \tag{A.1.10}$$

It is clear that any convergent sequence is Cauchy. On the other hand, we have:

Proposition A.1.2. *Each Cauchy sequence is bounded.*

Proof. Take $n = 1$ in the definition above. Thus, if (a_j) is Cauchy, there is a K such that $j, k \geq K \Rightarrow |a_j - a_k| \leq 1$. Hence, $j \geq K \Rightarrow |a_j| \leq |a_K| + 1$, so, for all j,

$$|a_j| \leq M, \quad M = \max\big(|a_1|, \dots, |a_{K-1}|, |a_K| + 1\big).$$

□

Now, the arguments proving Proposition A.1.1 also establish:

Proposition A.1.3. *If (a_j) and (b_j) are Cauchy sequences, so are $(a_j + b_j)$ and $(a_j b_j)$. Furthermore, if, for all j, $|b_j| \geq c$ for some $c > 0$, then (a_j/b_j) is Cauchy.*

The following proposition is a bit deeper than the first three.

Proposition A.1.4. *If (a_j) is bounded, i.e., $|a_j| \leq M$ for all j, then it has a Cauchy subsequence.*

Proof. We may as well assume $M \in \mathbb{N}$. Now, either $a_j \in [0, M]$ for infinitely many j or $a_j \in [-M, 0]$ for infinitely many j. Let I_1 be any one of these two intervals containing a_j for infinitely many j, and pick $j(1)$ such that $a_{j(1)} \in I_1$. Write I_1 as the union of two closed intervals, of equal length, sharing only the midpoint of I_1. Let I_2 be any one of them with the property that $a_j \in I_2$ for infinitely many j, and pick $j(2) > j(1)$ such that $a_{j(2)} \in I_2$. Continue, picking $I_\nu \subset I_{\nu-1} \subset \cdots \subset I_1$, of length $M/2^{\nu-1}$, containing a_j for infinitely many j, and picking $j(\nu) > j(\nu - 1) > \cdots > j(1)$ such that $a_{j(\nu)} \in I_\nu$. See Figure A.1.1 for an illustration of a possible scenario. Setting $b_\nu = a_{j(\nu)}$, we see that (b_ν) is a Cauchy subsequence of (a_j), since, for all $k \in \mathbb{N}$,

$$|b_{\nu+k} - b_\nu| \leq M/2^{\nu-1}.$$

□

Here is a significant variant of Proposition A.1.4.

Proposition A.1.5. *Each bounded monotone sequence (a_j) is Cauchy.*

Proof. To say (a_j) is monotone is to say that either (a_j) is increasing, i.e., $a_j \leq a_{j+1}$ for all j, or that (a_j) is decreasing, i.e., $a_j \geq a_{j+1}$ for all j. For the sake of argument, assume (a_j) is increasing.

By Proposition A.1.4, there is a subsequence $(b_\nu) = (a_{j(\nu)})$ that is Cauchy. Thus, given $n \in \mathbb{N}$, there exists $K(n)$ such that

$$\mu, \nu \geq K(n) \Longrightarrow |a_{j(\nu)} - a_{j(\mu)}| < \frac{1}{n}. \tag{A.1.11}$$

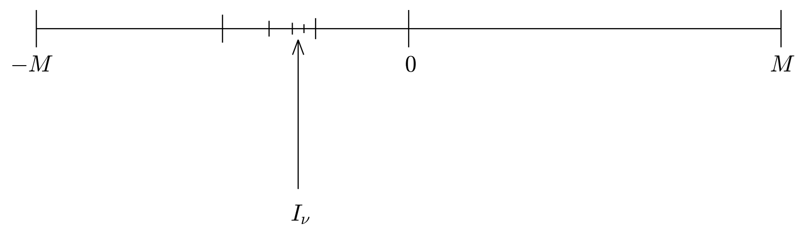

Figure A.1.1. Nested intervals containing a_j for infinitely many j

Now, if $\nu_0 \ge K(n)$ and $k \ge j \ge j(\nu_0)$, pick ν_1 such that $j(\nu_1) \ge k$. Then

$$a_{j(\nu_0)} \le a_j \le a_k \le a_{j(\nu_1)},$$

so

$$k \ge j \ge j(\nu_0) \Longrightarrow |a_j - a_k| < \frac{1}{n}. \tag{A.1.12}$$

□

Second proof. Again, we assume (a_j) is increasing. If (a_j) is not Cauchy, then there exists $n \in \mathbb{N}$ such that, for each j,

$$a_\ell > a_j + \frac{1}{n}, \quad \text{for some } \ \ell > j. \tag{A.1.13}$$

Hence there exist $j_\nu, k_\nu \nearrow \infty$ such that

$$j_\nu < k_\nu < j_{\nu+1} < k_{\nu+1} < \cdots, \tag{A.1.14}$$

and

$$a_{k_\nu} - a_{j_\nu} > \frac{1}{n}, \quad \forall\, \nu. \tag{A.1.15}$$

It follows that

$$a_{k_\nu} > a_1 + \frac{\nu}{n}, \quad \forall\, \nu \in \mathbb{N}, \tag{A.1.16}$$

which contradicts the hypothesis that (a_j) is bounded. □

We give a few simple but basic examples of convergent sequences.

Proposition A.1.6. *If* $|a| < 1$, *then* $a^j \to 0$.

Proof. Set $b = |a|$; it suffices to show that $b^j \to 0$. Consider $c = 1/b > 1$, hence $c = 1 + y,\ y > 0$. We claim that

$$c^j = (1+y)^j \ge 1 + jy,$$

for all $j \ge 1$. In fact, this clearly holds for $j = 1$, and if it holds for $j = k$, then

$$c^{k+1} \ge (1+y)(1+ky) > 1 + (k+1)y.$$

Hence, by induction, the estimate is established. Consequently,

$$b^j < \frac{1}{jy},$$

so the appropriate analogue of (A.1.1) holds, with $K(n) = Kn$, for any integer $K > 1/y$. □

Proposition A.1.6 enables us to establish the following result on geometric series.

Proposition A.1.7. *If* $|x| < 1$ *and*

$$a_j = 1 + x + \cdots + x^j, \tag{A.1.17}$$

then

$$a_j \to \frac{1}{1-x}. \tag{A.1.18}$$

Proof. Note that $xa_j = x + x^2 + \cdots + x^{j+1}$, so $(1-x)a_j = 1 - x^{j+1}$, i.e.,

$$a_j = \frac{1 - x^{j+1}}{1-x}.$$

The conclusion follows from Proposition A.1.6. □

Note in particular that

$$0 < x < 1 \Longrightarrow 1 + x + \cdots + x^j < \frac{1}{1-x}. \tag{A.1.19}$$

It is an important mathematical fact that not every Cauchy sequence of rational numbers has a rational number as limit. We give one example here. Consider the sequence

$$a_j = \sum_{\ell=0}^{j} \frac{1}{\ell!}. \tag{A.1.20}$$

Then (a_j) is increasing, and

$$a_{n+j} - a_n = \sum_{\ell=n+1}^{n+j} \frac{1}{\ell!} < \frac{1}{n!}\Big(\frac{1}{n+1} + \frac{1}{(n+1)^2} + \cdots + \frac{1}{(n+1)^j}\Big),$$

since $(n+1)(n+2)\cdots(n+j) > (n+1)^j$. Using (A.1.19), we have

$$a_{n+j} - a_n < \frac{1}{(n+1)!}\,\frac{1}{1-\frac{1}{n+1}} = \frac{1}{n!}\cdot\frac{1}{n}. \tag{A.1.21}$$

Hence (a_j) is Cauchy. Taking $n = 2$, we see that

$$j > 2 \Longrightarrow 2\tfrac{1}{2} < a_j < 2\tfrac{3}{4}. \tag{A.1.22}$$

Proposition A.1.8. *The sequence* (A.1.20) *cannot converge to a rational number.*

Proof. Assume $a_j \to m/n$ with $m, n \in \mathbb{N}$. By (A.1.22), we must have $n > 2$. Now, write

$$\frac{m}{n} = \sum_{\ell=0}^{n} \frac{1}{\ell!} + r, \quad r = \lim_{j\to\infty} (a_{n+j} - a_n). \tag{A.1.23}$$

Multiplying both sides of (A.1.23) by $n!$ gives

$$m(n-1)! = A + r \cdot n! \tag{A.1.24}$$

where

$$A = \sum_{\ell=0}^{n} \frac{n!}{\ell!} \in \mathbb{N}. \tag{A.1.25}$$

Thus the identity (A.1.23) forces $r \cdot n! \in \mathbb{N}$, while (A.1.21) implies

$$0 < r \cdot n! \le 1/n. \tag{A.1.26}$$

This contradiction proves the proposition. ☐

Exercises

1. Show that
$$\lim_{k\to\infty} \frac{k}{2^k} = 0,$$
and more generally for each $m \in \mathbb{N}$,
$$\lim_{k\to\infty} \frac{k^m}{2^k} = 0.$$
Hint. See Exercise 3.

2. Show that
$$\lim_{k\to\infty} \frac{2^k}{k!} = 0,$$
and more generally for each $b \in \mathbb{N}$,
$$\lim_{k\to\infty} \frac{b^k}{k!} = 0.$$

3. Suppose a sequence (a_j) has the property that there exist
$$r < 1, \quad K \in \mathbb{N}$$

such that

$$j \geq K \Longrightarrow \left| \frac{a_{j+1}}{a_j} \right| \leq r.$$

Show that $a_j \to 0$ as $j \to \infty$. How does this result apply to Exercises 1 and 2?

4. If (a_j) satisfies the hypotheses of Exercise 3, show that there exists $M < \infty$ such that

$$\sum_{j=1}^{k} |a_j| \leq M, \quad \forall\, k.$$

Remark. This yields the *ratio test* for infinite series.

5. Show that you get the same criterion for convergence if (A.1.1) is replaced by

$$j \geq K(n) \Longrightarrow |a_j - a| < \frac{5}{n}.$$

Generalize, and note the relevance for the proof of Proposition A.1.1. Apply the same observation to the criterion (A.1.10) for (a_j) to be Cauchy.

A.2. The real numbers

We think of a real number as a quantity that can be specified by a process of approximation arbitrarily closely by rational numbers. Thus, we define an element of $\mathbb{R}$ as an equivalence class of Cauchy sequences of rational numbers, where we define

$$(a_j) \sim (b_j) \Longleftrightarrow a_j - b_j \to 0. \tag{A.2.1}$$

Proposition A.2.1. *This is an equivalence relation.*

Proof. This is a straightforward consequence of Proposition A.1.1. In particular, to see that

$$(a_j) \sim (b_j),\ (b_j) \sim (c_j) \Longrightarrow (a_j) \sim (c_j), \tag{A.2.2}$$

just use (A.1.4) of Proposition A.1.1 to write

$$a_j - b_j \to 0,\ b_j - c_j \to 0 \Longrightarrow a_j - c_j \to 0.$$

□

We denote the equivalence class containing a Cauchy sequence (a_j) by $[(a_j)]$. We then define addition and multiplication on $\mathbb{R}$ to satisfy

$$\begin{aligned} [(a_j)] + [(b_j)] &= [(a_j + b_j)], \\ [(a_j)] \cdot [(b_j)] &= [(a_j b_j)]. \end{aligned} \tag{A.2.3}$$

Proposition A.1.3 states that the sequences $(a_j + b_j)$ and $(a_j b_j)$ are Cauchy if (a_j) and (b_j) are. To conclude that the operations in (A.2.3) are well defined, we need:

Proposition A.2.2. *If Cauchy sequences of rational numbers are given which satisfy $(a_j) \sim (a'_j)$ and $(b_j) \sim (b'_j)$, then*

$$(a_j + b_j) \sim (a'_j + b'_j), \tag{A.2.4}$$

and

$$(a_j b_j) \sim (a'_j b'_j). \tag{A.2.5}$$

The proof is a straightforward variant of the proof of parts (A.1.4)-(A.1.5) in Proposition A.1.1, with due account taken of Proposition A.1.2. For example, $a_j b_j - a'_j b'_j = a_j b_j - a_j b'_j + a_j b'_j - a'_j b'_j$, and there are uniform bounds $|a_j| \le A$, $|b'_j| \le B$, so

$$\begin{aligned} |a_j b_j - a'_j b'_j| &\le |a_j| \cdot |b_j - b'_j| + |a_j - a'_j| \cdot |b'_j| \\ &\le A|b_j - b'_j| + B|a_j - a'_j|. \end{aligned}$$

There is a natural injection

$$\mathbb{Q} \hookrightarrow \mathbb{R}, \quad a \mapsto [(a, a, a, \dots)], \tag{A.2.6}$$

whose image we identify with $\mathbb{Q}$. This map preserves addition and multiplication.

If $x = [(a_j)]$, we define

$$-x = [(-a_j)]. \tag{A.2.7}$$

For $x \neq 0$, we define x^{-1} as follows. First, to say $x \neq 0$ is to say there exists $n \in \mathbb{N}$ such that $|a_j| \geq 1/n$ for infinitely many j. Since (a_j) is Cauchy, this implies that there exists K such that $|a_j| \geq 1/2n$ for all $j \geq K$. Now, if we set $\alpha_j = a_{K+j}$, we have $(\alpha_j) \sim (a_j)$; we propose to set

$$x^{-1} = [(\alpha_j^{-1})]. \tag{A.2.8}$$

We claim that this is well defined. First, by Proposition A.1.3, (α_j^{-1}) is Cauchy. Furthermore, if for such x we also have $x = [(b_j)]$, and we pick K so large that also $|b_j| \geq 1/2n$ for all $j \geq K$, and set $\beta_j = b_{K+j}$, we claim that

$$(\alpha_j^{-1}) \sim (\beta_j^{-1}). \tag{A.2.9}$$

Indeed, we have

$$|\alpha_j^{-1} - \beta_j^{-1}| = \frac{|\beta_j - \alpha_j|}{|\alpha_j| \cdot |\beta_j|} \leq 4n^2 |\beta_j - \alpha_j|, \tag{A.2.10}$$

so (A.2.9) holds.

It is now a straightforward exercise to verify the basic algebraic properties of addition and multiplication in $\mathbb{R}$, given that these results hold in $\mathbb{Q}$. We state the result.

Proposition A.2.3. *Given $x, y, z \in \mathbb{R}$, the following algebraic properties hold.*

$$\begin{aligned} x + y &= y + x, \\ (x + y) + z &= x + (y + z), \\ x + 0 &= x, \\ x + (-x) &= 0, \\ x \cdot y &= y \cdot x, \\ (x \cdot y) \cdot z &= x \cdot (y \cdot z), \\ x \cdot 1 &= x, \\ x \cdot 0 &= 0, \\ x \cdot (-1) &= -x, \\ x \cdot (y + z) &= x \cdot y + x \cdot z. \end{aligned}$$

Furthermore,

$$x \neq 0 \Longrightarrow x \cdot x^{-1} = 1.$$

We define $x - y = x + (-y)$ and, if $y \neq 0$, $x/y = x \cdot y^{-1}$.

We now define an order relation on $\mathbb{R}$, assuming it is known on $\mathbb{Q}$. Take $x \in \mathbb{R}$, $x = [(a_j)]$. From the discussion above of x^{-1}, we see that, if $x \neq 0$, then one and only one of the following holds. Either, for some $n, K \in \mathbb{N}$,

$$j \geq K \Longrightarrow a_j \geq \frac{1}{2n}, \tag{A.2.11}$$

or, for some $n, K \in \mathbb{N}$,

$$j \geq K \Longrightarrow a_j \leq -\frac{1}{2n}. \tag{A.2.12}$$

If $(a_j) \sim (b_j)$ and (A.2.11) holds for a_j, it also holds for b_j (perhaps with different n and K), and ditto for (A.2.12). If (A.2.11) holds, we say $x \in \mathbb{R}^+$ (and we say $x > 0$), and if (A.2.12) holds we say $x \in \mathbb{R}^-$ (and we say $x < 0$). Clearly $x > 0$ if and only if $-x < 0$. It is also clear that the map $\mathbb{Q} \hookrightarrow \mathbb{R}$ in (A.2.6) preserves the order relation.

Thus we have the disjoint union

$$\mathbb{R} = \mathbb{R}^+ \cup \{0\} \cup \mathbb{R}^-, \quad \mathbb{R}^- = -\mathbb{R}^+. \tag{A.2.13}$$

Also,

$$x, y \in \mathbb{R}^+ \Longrightarrow x + y, xy \in \mathbb{R}^+. \tag{A.2.14}$$

We define

$$x < y \Longleftrightarrow y - x \in \mathbb{R}^+. \tag{A.2.15}$$

If $x = [(a_j)]$ and $y = [(b_j)]$, we see from (A.2.11)–(A.2.12) that

$$\begin{aligned} x < y \Longleftrightarrow & \text{ for some } n, K \in \mathbb{N}, \\ & j \geq K \Rightarrow b_j - a_j \geq \frac{1}{n} \quad \Big(\text{i.e., } a_j \leq b_j - \frac{1}{n}\Big). \end{aligned} \tag{A.2.16}$$

The relation (A.2.15) can also be written $y > x$. Similarly we define $x \leq y$ and $y \leq x$, in the obvious fashions.

The following results are straightforward.

Proposition A.2.4. *For elements of* $\mathbb{R}$, *we have*

$$x_1 < y_1, \ x_2 < y_2 \Longrightarrow x_1 + x_2 < y_1 + y_2, \tag{A.2.17}$$

$$x < y \Longleftrightarrow -y < -x, \tag{A.2.18}$$

$$0 < x < y, \ a > 0 \Longrightarrow 0 < ax < ay, \tag{A.2.19}$$

$$0 < x < y \Longrightarrow 0 < y^{-1} < x^{-1}. \tag{A.2.20}$$

Proof. The results (A.2.17) and (A.2.19) follow from (A.2.14); consider, for example, $a(y - x)$. The result (A.2.18) follows from (A.2.13). To prove (A.2.20), first we see that $x > 0$ implies $x^{-1} > 0$, as follows: if $-x^{-1} > 0$, the identity $x \cdot (-x^{-1}) = -1$ contradicts (A.2.14). As for the rest of (A.2.20), the hypotheses imply $xy > 0$, and multiplying both sides of $x < y$ by $a = (xy)^{-1}$ gives the result, by (A.2.19). □

As in (A.1.2), define $|x|$ by

$$\begin{aligned} |x| = & \ \ x \ \text{ if } \ x \geq 0, \\ & -x \ \text{ if } \ x < 0. \end{aligned} \tag{A.2.21}$$

Note that

$$x = [(a_j)] \Longrightarrow |x| = [(|a_j|)]. \tag{A.2.22}$$

It is straightforward (compare (A.1.3)) to verify

$$|xy| = |x| \cdot |y|, \quad |x + y| \leq |x| + |y|. \tag{A.2.23}$$

We now show that $\mathbb{R}$ has the Archimedean property.

Proposition A.2.5. *Given $x \in \mathbb{R}$, there exists $k \in \mathbb{Z}$ such that*

$$k - 1 < x \le k. \tag{A.2.24}$$

Proof. It suffices to prove (A.2.24) assuming $x \in \mathbb{R}^+$. Otherwise, work with $-x$. Say $x = [(a_j)]$ where (a_j) is a Cauchy sequence of rational numbers. By Proposition A.1.2, there exists $M \in \mathbb{Q}$ such that $|a_j| \le M$ for all j. We also have $M \le \ell$ for some $\ell \in \mathbb{N}$. Hence the set $S = \{\ell \in \mathbb{N} : \ell \ge x\}$ is nonempty. Then taking k to be the smallest element of S gives (A.2.24). □

Proposition A.2.6. *Given any real $\varepsilon > 0$, there exists $n \in \mathbb{N}$ such that $\varepsilon > 1/n$.*

Proof. Using Proposition A.2.5, pick $n > 1/\varepsilon$ and apply (A.2.20). Alternatively, use the reasoning given above (A.2.8). □

We are now ready to consider sequences of elements of $\mathbb{R}$.

Definition. A sequence (x_j) converges to x if and only if, for any $n \in \mathbb{N}$, there exists $K(n)$ such that

$$j \ge K(n) \Longrightarrow |x_j - x| < \frac{1}{n}. \tag{A.2.25}$$

In this case, we write $x_j \to x$, or $x = \lim\ x_j$.

The sequence (x_j) is Cauchy if and only if, for any $n \in \mathbb{N}$, there exists $K(n)$ such that

$$j, k \ge K(n) \Longrightarrow |x_j - x_k| < \frac{1}{n}. \tag{A.2.26}$$

We note that it is typical to phrase the definition above in terms of picking any real $\varepsilon > 0$ and demanding that, e.g., $|x_j - x| < \varepsilon$, for large j. The equivalence of the two definitions follows from Proposition A.2.6.

As in Proposition A.1.2, we have that every Cauchy sequence is bounded.

Next, the proof of Proposition A.1.1 extends to the present case, yielding:

Proposition A.2.7. *If $x_j \to x$ and $y_j \to y$, then*

$$x_j + y_j \to x + y, \tag{A.2.27}$$

and

$$x_j y_j \to xy. \tag{A.2.28}$$

If furthermore $y_j \ne 0$ for all j and $y \ne 0$, then

$$x_j / y_j \to x/y. \tag{A.2.29}$$

It is clear that, if each $x_j \in \mathbb{Q}$, then the notion that (x_j) is Cauchy given above coincides with that in §A.1. If also $x \in \mathbb{Q}$, the notion that $x_j \to x$ also coincides with that given in §A.1. Here is another natural but useful observation.

Proposition A.2.8. *If each $a_j \in \mathbb{Q}$, and $x \in \mathbb{R}$, then*

$$a_j \to x \Longleftrightarrow x = [(a_j)]. \tag{A.2.30}$$

Proof. First assume $x = [(a_j)]$. In particular, (a_j) is Cauchy. Now, given m, we have from (A.2.16) that

$$\begin{aligned} |x - a_k| < \frac{1}{m} &\Longleftrightarrow \exists\, K, n \text{ such that } j \geq K \Rightarrow |a_j - a_k| < \frac{1}{m} - \frac{1}{n} \\ &\Longleftarrow \exists\, K \text{ such that } j \geq K \Rightarrow |a_j - a_k| < \frac{1}{2m}. \end{aligned} \tag{A.2.31}$$

On the other hand, since (a_j) is Cauchy, for each $m \in \mathbb{N}$, there exists $K(m)$ such that

$$j, k \geq K(m) \Rightarrow |a_j - a_k| < \frac{1}{2m}. \tag{A.2.32}$$

Hence, by (A.2.31),

$$k \geq K(m) \Longrightarrow |x - a_k| < \frac{1}{m}. \tag{A.2.33}$$

This shows that $x = [(a_j)] \Rightarrow a_j \to x$.

For the converse, if $a_j \to x$, then (a_j) is Cauchy, so we have $[(a_j)] = y \in \mathbb{R}$. The previous argument implies $a_j \to y$. But

$$|x - y| \leq |x - a_j| + |a_j - y|, \quad \forall\, j, \tag{A.2.34}$$

so $x = y$. Thus $a_j \to x \Rightarrow x = [(a_j)]$. □

So far, statements made about $\mathbb{R}$ have emphasized similarities of its properties with corresponding properties of $\mathbb{Q}$. The crucial difference between these two sets of numbers is given by the following result, known as the completeness property.

Theorem A.2.9. *If (x_j) is a Cauchy sequence of real numbers, then there exists $x \in \mathbb{R}$ such that $x_j \to x$.*

Proof. Take $x_j = [(a_{j\ell} : \ell \in \mathbb{N})]$ with $a_{j\ell} \in \mathbb{Q}$. Using (A.2.30), take $a_{j,\ell(j)} = b_j \in \mathbb{Q}$ such that

$$|x_j - b_j| \leq 2^{-j}. \tag{A.2.35}$$

Then (b_j) is Cauchy, since $|b_j - b_k| \leq |x_j - x_k| + 2^{-j} + 2^{-k}$. Now, let

$$x = [(b_j)]. \tag{A.2.36}$$

It follows that

$$|x_j - x| \leq |x_j - b_j| + |x - b_j| \leq 2^{-j} + |x - b_j|, \tag{A.2.37}$$

which tends to 0, again by (A.2.30). Hence $x_j \to x$. □

If we combine Theorem A.2.9 with the argument behind Proposition A.1.4, we obtain the following important result, known as the Bolzano-Weierstrass Theorem.

Theorem A.2.10. *Each bounded sequence of real numbers has a convergent subsequence.*

Proof. If $|x_j| \leq M$, the proof of Proposition A.1.4 applies without change to show that (x_j) has a Cauchy subsequence. By Theorem A.2.9, that Cauchy subsequence converges. □

Similarly, adding Theorem A.2.9 to the argument behind Proposition A.1.5 yields:

Proposition A.2.11. *Each bounded monotone sequence* (x_j) *of real numbers converges.*

A related property of $\mathbb{R}$ can be described in terms of the notion of the "supremum" of a set.

Definition. If $S \subset \mathbb{R}$, one says that $x \in \mathbb{R}$ is an upper bound for S provided $x \geq s$ for all $s \in S$, and one says

$$x = \sup S \tag{A.2.38}$$

provided x is an upper bound for S and further $x \leq x'$ whenever x' is an upper bound for S. One also says x is the least upper bound of S, and writes $x = \operatorname{lub} S$.

For some sets, such as $S = \mathbb{Z}$, there is no $x \in \mathbb{R}$ satisfying (A.2.38). However, there is the following result, known as the supremum property.

Proposition A.2.12. *If* S *is a nonempty subset of* $\mathbb{R}$ *that has an upper bound, then there is a real* $x = \sup S$.

Proof. We use an argument similar to the one in the proof of Proposition A.1.4. Let x_0 be an upper bound for S, pick s_0 in S, and consider

$$I_0 = [s_0, x_0] = \{y \in \mathbb{R} : s_0 \leq y \leq x_0\}.$$

If $x_0 = s_0$, then already $x_0 = \sup S$. Otherwise, I_0 is an interval of nonzero length, $L = x_0 - s_0$. In that case, divide I_0 into two equal intervals, having in common only the midpoint; say $I_0 = I_0^\ell \cup I_0^r$, where I_0^r lies to the right of I_0^ℓ.

Let $I_1 = I_0^r$ if $S \cap I_0^r \neq \emptyset$, and otherwise let $I_1 = I_0^\ell$. Note that $S \cap I_1 \neq \emptyset$. Let x_1 be the right endpoint of I_1, and pick $s_1 \in S \cap I_1$. Note that x_1 is also an upper bound for S.

Continue, constructing

$$I_\nu \subset I_{\nu-1} \subset \cdots \subset I_0,$$

where I_ν has length $2^{-\nu} L$, such that the right endpoint x_ν of I_ν satisfies

$$x_\nu \geq s, \quad \forall\, s \in S, \tag{A.2.39}$$

and such that $S \cap I_\nu \neq \emptyset$, so there exist $s_\nu \in S$ such that

$$x_\nu - s_\nu \leq 2^{-\nu} L. \tag{A.2.40}$$

The sequence (x_ν) is bounded and monotone (decreasing) so, by Proposition A.2.11, it converges; $x_\nu \to x$. By (A.2.39), we have $x \geq s$ for all $s \in S$, and by (6.34) we have $x - s_\nu \leq 2^{-\nu} L$. Hence x satisfies (A.2.38). □

We turn to infinite series $\sum_{k=0}^\infty a_k$, with $a_k \in \mathbb{R}$. We say this series converges if and only if the sequence of partial sums

$$S_n = \sum_{k=0}^n a_k \tag{A.2.41}$$

converges:

$$\sum_{k=0}^{\infty} a_k = A \Longleftrightarrow S_n \to A \quad \text{as} \quad n \to \infty. \tag{A.2.42}$$

The following is a useful condition guaranteeing convergence.

Proposition A.2.13. *The infinite series* $\sum_{k=0}^{\infty} a_k$ *converges provided*

$$\sum_{k=0}^{\infty} |a_k| < \infty, \tag{A.2.43}$$

i.e., there exists $B < \infty$ *such that* $\sum_{k=0}^{n} |a_k| \le B$ *for all* n.

Proof. The triangle inequality (the second part of (A.2.23)) gives, for $\ell \ge 1$,

$$\begin{aligned} |S_{n+\ell} - S_n| &= \Big| \sum_{k=n+1}^{n+\ell} a_k \Big| \\ &\le \sum_{k=n+1}^{n+\ell} |a_k|, \end{aligned} \tag{A.2.44}$$

and we claim this tends to 0 as $n \to \infty$, uniformly in $\ell \ge 1$, provided (A.2.43) holds. In fact, if the right side of (A.2.44) fails to go to 0 as $n \to \infty$, there exists $\varepsilon > 0$ and infinitely many $n_\nu \to \infty$ and $\ell_\nu \in \mathbb{N}$ such that

$$\sum_{k=n_\nu+1}^{n_\nu+\ell_\nu} |a_k| \ge \varepsilon. \tag{A.2.45}$$

We can pass to a subsequence and assume $n_{\nu+1} > n_\nu + \ell_\nu$. Then

$$\sum_{k=n_1+1}^{n_\nu+\ell_\nu} |a_k| \ge \nu\varepsilon, \tag{A.2.46}$$

for all ν, contradicting the bound by B that follows from (A.2.43). Thus (A.2.43) $\Rightarrow$ (S_n) is Cauchy. Convergence follows, by Theorem A.2.9. $\square$

Alternative presentation. Set

$$T_n = \sum_{k=0}^{n} |a_k|.$$

The hypothesis (A.2.43) implies that (T_n) is a bounded monotone sequence. Then Proposition A.2.11 implies (T_n) is a Cauchy sequence. But (A.2.44) precisely says

$$|S_{n+\ell} - S_n| \le |T_{n+\ell} - T_n|.$$

$\square$

When (A.2.43) holds, we say the series $\sum_{k=0}^{\infty} a_k$ is *absolutely convergent.*

The following result on *alternating series* gives another sufficient condition for convergence.

Proposition A.2.14. *Assume* $a_k > 0,\ a_k \searrow 0$. *Then*

$$\sum_{k=0}^{\infty} (-1)^k a_k \tag{A.2.47}$$

is convergent.

Proof. Denote the partial sums by $S_n,\ n \ge 0$. We see that, for $m \in \mathbb{N}$,

$$S_{2m+1} \le S_{2m+3} \le S_{2m+2} \le S_{2m}. \tag{A.2.48}$$

Iterating this, we have, as $m \to \infty$,

$$S_{2m} \searrow \alpha, \quad S_{2m+1} \nearrow \beta, \quad \beta \le \alpha, \tag{A.2.49}$$

and

$$S_{2m} - S_{2m+1} = a_{2m+1}, \tag{A.2.50}$$

hence $\alpha = \beta$, and convergence is established. ☐

Here is an example:

$$\sum_{k=0}^{\infty} \frac{(-1)^k}{k+1} = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \cdots \quad \text{is convergent.}$$

This series is not absolutely convergent (cf. Exercise 6 below). Using Exercise 1 of §3.2 and an additional argument, one can show the sum is $\log 2$.

Exercises

1. Verify Proposition A.2.3.

2. If $S \subset \mathbb{R}$, we say that $x \in \mathbb{R}$ is a lower bound for S provided $x \le s$ for all $s \in S$, and we say

$$x = \inf S, \tag{A.2.51}$$

provided x is a lower bound for S and further $x \ge x'$ whenever x' is a lower bound for S. Mirroring Proposition A.2.12, show that if $S \subset \mathbb{R}$ is a nonempty set that has a lower bound, then there is a real $x = \inf S$.

3. Given a real number $\xi \in (0,1)$, show it has an infinite decimal expansion, i.e., there exist $b_k \in \{0, 1, \dots, 9\}$ such that

$$\xi = \sum_{k=1}^{\infty} b_k \cdot 10^{-k}. \tag{A.2.52}$$

Hint. Start by breaking $[0,1]$ into ten subintervals of equal length, and picking one to which ξ belongs.

4. Show that if $0 < x < 1$, then $x^n \to 0$ as $n \to \infty$ (as in Proposition A.1.6), and

$$\sum_{k=0}^{\infty} x^k = \frac{1}{1-x} < \infty. \tag{A.2.53}$$

Hint. We have

$$\sum_{k=0}^{n} x^k = \frac{1-x^{n+1}}{1-x}, \quad x \neq 1.$$

The series (A.2.53) is called a geometric series.

5. Assume $a_k > 0$ and $a_k \searrow 0$. Show that

$$\sum_{k=1}^{\infty} a_k < \infty \Longleftrightarrow \sum_{k=0}^{\infty} b_k < \infty, \tag{A.2.54}$$

where

$$b_k = 2^k a_{2^k}. \tag{A.2.55}$$

Hint. Use the following observations:

$$\frac{1}{2}b_2 + \frac{1}{2}b_3 + \cdots \leq (a_3 + a_4) + (a_5 + a_6 + a_7 + a_8) + \cdots, \quad \text{and}$$
$$(a_3 + a_4) + (a_5 + a_6 + a_7 + a_8) + \cdots \leq b_1 + b_2 + \cdots .$$

6. Deduce from Exercise 5 that the harmonic series $1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \cdots$ diverges, i.e.,

$$\sum_{k=1}^{\infty} \frac{1}{k} = \infty. \tag{A.2.56}$$

7. Deduce from Exercises 4–5 that

$$p > 1 \Longrightarrow \sum_{k=1}^{\infty} \frac{1}{k^p} < \infty. \tag{A.2.57}$$

To start, take $p \in \mathbb{N}$. See §1.1 to define k^p for $p \in \mathbb{Q}$, and §3.2 to define k^p for $p \in \mathbb{R}$.

8. Given $a, b \in \mathbb{R} \setminus 0,\ k \in \mathbb{Z}$, show that

$$a^{j+k} = a^j a^k, \quad a^{jk} = (a^j)^k, \quad (ab)^j = a^j b^j, \quad \forall\, j, k \in \mathbb{Z}.$$

9. Given $k \in \mathbb{N}$, show that, for $x_j \in \mathbb{R}$,

$$x_j \to x \Longrightarrow x_j^k \to x^k.$$

Hint. Use Proposition A.2.7.

10. Given $x_j, x, y \in \mathbb{R}$, show that

$$x_j \geq y\ \forall\, j, \quad x_j \to x \Longrightarrow x \geq y.$$

11. Given the alternating series $\sum(-1)^k a_k$ as in Proposition A.2.14 (with $a_k \searrow 0$), with sum S, show that, for each N,

$$\sum_{k=0}^{N}(-1)^k a_k = S + r_N, \quad |r_N| \le |a_{N+1}|.$$

12. Generalize Exercises 3–4 of §A.1 as follows. Suppose a sequence (a_j) in $\mathbb{R}$ has the property that there exist $r < 1$ and $K \in \mathbb{N}$ such that

$$j \ge K \Longrightarrow \left|\frac{a_{j+1}}{a_j}\right| \le r.$$

Show that there exists $M < \infty$ such that

$$\sum_{j=1}^{k} |a_j| \le M, \quad \forall k \in \mathbb{N}.$$

Conclude that $\sum_{k=1}^{\infty} a_k$ is convergent. This is the *ratio test* for convergence.

13. Show that, for each $x \in \mathbb{R}$,

$$\sum_{k=1}^{\infty} \frac{1}{k!} x^k$$

is convergent.

14. Let (b_j) be a Cauchy sequence of rational numbers, $y = [(b_j)],\ c \in \mathbb{Q}$. Show that

$$|b_j| \le c \ \ \forall\, j \Longrightarrow |y| \le c.$$

15. Produce an alternative presentation of the proof of the implication

$$x = [(a_j)] \Longrightarrow a_j \to x$$

in Proposition A.2.8 along the following lines.
Show that, for each k,

$$x - a_k = [(b_j)],$$

with

$$b_j = a_{j+k} - a_k.$$

Then, using Exercise 14, deduce that, if

$$|a_\ell - a_k| \le \frac{1}{m}, \quad \forall\, k, \ell \ge K(m),$$

then

$$k \ge K(m) \Longrightarrow |x - a_k| \le \frac{1}{m}.$$

A.3. Metric properties of $\mathbb{R}$

We discuss a number of notions and results related to convergence in $\mathbb{R}$. Recall that a sequence of points (p_j) in $\mathbb{R}$ converges to a limit $p \in \mathbb{R}$ (we write $p_j \to p$) if and only if for every $\varepsilon > 0$ there exists N such that

$$j \geq N \Longrightarrow |p_j - p| < \varepsilon. \tag{A.3.1}$$

A set $S \subset \mathbb{R}$ is said to be *closed* if and only if

$$p_j \in S, \ p_j \to p \Longrightarrow p \in S. \tag{A.3.2}$$

The complement $\mathbb{R} \setminus S$ of a closed set S is *open*. Alternatively, $\Omega \subset \mathbb{R}$ is open if and only if, given $q \in \Omega$, there exists $\varepsilon > 0$ such that $B_\varepsilon(q) \subset \Omega$, where

$$B_\varepsilon(q) = \{p \in \mathbb{R} : |p - q| < \varepsilon\}, \tag{A.3.3}$$

so q cannot be a limit of a sequence of points in $\mathbb{R} \setminus \Omega$.

In particular, the interval

$$[a, b] = \{x \in \mathbb{R} : a \leq x \leq b\} \tag{A.3.4}$$

is closed, and the interval

$$(a, b) = \{x \in \mathbb{R} : a < x < b\} \tag{A.3.5}$$

is open.

We define the *closure* $\overline{S}$ of a set $S \subset \mathbb{R}$ to consist of all points $p \in \mathbb{R}$ such that $B_\varepsilon(p) \cap S \neq \emptyset$ for all $\varepsilon > 0$. Equivalently, $p \in \overline{S}$ if and only if there exists an infinite sequence (p_j) of points in S such that $p_j \to p$. For example, the closure of the interval (a, b) is the interval $[a, b]$.

An important property of $\mathbb{R}$ is *completeness*, which we recall is defined as follows. A sequence (p_j) of points in $\mathbb{R}$ is called a Cauchy sequence if and only if

$$|p_j - p_k| \longrightarrow 0, \quad \text{as} \ \ j, k \to \infty. \tag{A.3.6}$$

It is easy to see that if $p_j \to p$ for some $p \in \mathbb{R}$, then (A.3.6) holds. The completeness property is the converse, given in Theorem A.2.9, which we recall here.

Theorem A.3.1. *If (p_j) is a Cauchy sequence in $\mathbb{R}$, then it has a limit.*

Completeness provides a path to the following key notion of *compactness*. A nonempty set $K \subset \mathbb{R}$ is said to be compact if and only if the following property holds.

$$\begin{array}{l}\text{Each infinite sequence } (p_j) \text{ in } K \text{ has a subsequence} \\ \text{that converges to a point in } K.\end{array} \tag{A.3.7}$$

It is clear that if K is compact, then it must be closed. It must also be bounded, i.e., there exists $R < \infty$ such that $K \subset B_R(0)$. Indeed, if K is not bounded, there exist $p_j \in K$ such that $|p_{j+1}| \geq |p_j| + 1$. In such a case, $|p_j - p_k| \geq 1$ whenever $j \neq k$, so (p_j) cannot have a convergent subsequence. The following converse statement is a key result.

Theorem A.3.2. *If a nonempty $K \subset \mathbb{R}$ is closed and bounded, then it is compact.*

Clearly every nonempty closed subset of a compact set is compact, so Theorem A.3.2 is a consequence of:

Proposition A.3.3. *Each closed bounded interval* $I = [a, b] \subset \mathbb{R}$ *is compact.*

Proof. This is a direct consequence of the Bolzano-Weierstrass theorem, Theorem A.2.10. □

Let $K \subset \mathbb{R}$ be compact. Since K is bounded from above and from below, we have well defined real numbers

$$b = \sup K, \quad a = \inf K, \tag{A.3.8}$$

the first by Proposition A.2.12, and the second by a similar argument (cf. Exercise 2 of §A.2). Since a and b are limits of elements of K, we have $a, b \in K$. We use the notation

$$b = \max K, \quad a = \min K. \tag{A.3.9}$$

We next discuss continuity. If $S \subset \mathbb{R}$, a function

$$f : S \longrightarrow \mathbb{R} \tag{A.3.10}$$

is said to be continuous at $p \in S$ provided

$$p_j \in S, \ p_j \to p \Longrightarrow f(p_j) \to f(p). \tag{A.3.11}$$

If f is continuous at each $p \in S$, we say f is continuous on S, and write $f \in C(S)$.

Clearly $f(x) = x$ defines $f \in C(\mathbb{R})$. The following result provides an arsenal of continuous functions.

Proposition A.3.4. *Given* $S \subset \mathbb{R}$,

$$f, g \in C(S) \Longrightarrow f + g, \ fg \in C(S). \tag{A.3.12}$$

The proof is a simple application of Proposition A.2.7. As a consequence, we see that each polynomial

$$p(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_0, \quad a_j \in \mathbb{R}$$

is continuous on $\mathbb{R}$.

The following two results give important connections between continuity and compactness.

Proposition A.3.5. *If* $K \subset \mathbb{R}$ *is compact and* $f : K \to \mathbb{R}$ *is continuous, then* $f(K)$ *is compact.*

Proof. If (q_k) is an infinite sequence of points in $f(K)$, pick $p_k \in K$ such that $f(p_k) = q_k$. If K is compact, we have a subsequence $p_{k_\nu} \to p$ in K, and then $q_{k_\nu} \to f(p)$ in $\mathbb{R}$. □

This leads to the second connection.

Proposition A.3.6. *If $K \subset \mathbb{R}$ is compact and $f : K \to \mathbb{R}$ is continuous, then there exists $p \in K$ such that*

$$f(p) = \max_{x \in K} f(x), \tag{A.3.13}$$

and there exists $q \in K$ such that

$$f(q) = \min_{x \in K} f(x). \tag{A.3.14}$$

Proof. Since $f(K)$ is compact, we have well defined numbers

$$b = \max f(K), \quad a = \min f(K), \quad a, b \in f(K). \tag{A.3.15}$$

So take $p, q \in K$ such that $f(p) = b$ and $f(q) = a$. ☐

The next result is called the intermediate value theorem.

Proposition A.3.7. *Take $a, b, c \in \mathbb{R}$, $a < b$. Let $f : [a, b] \to \mathbb{R}$ be continuous. Assume*

$$f(a) < c < f(b). \tag{A.3.16}$$

Then there exists $x \in (a, b)$ such that $f(x) = c$.

Proof. Let

$$S = \{y \in [a, b] : f(y) \le c\}. \tag{A.3.17}$$

Then $a \in S$. Also, if $f(y_j) \le c$ and $y_j \to y$, then $f(y) \le c$. Hence S is a nonempty, closed (hence compact) subset of $[a, b]$. Note that $b \notin S$. Take

$$x = \max S. \tag{A.3.18}$$

Then $a < x < b$ and $f(x) \le c$. If $f(x) < c$, then there exists $\varepsilon > 0$ such that $x + \varepsilon < b$ and $f(y) < c$ for $x \le y < x + \varepsilon$. Thus $x + \varepsilon \in S$, contradicting (A.3.18). ☐

Returning to the issue of compactness, we establish some further properties of compact sets $K \subset \mathbb{R}$, leading to the important result, Proposition A.3.11 below.

Proposition A.3.8. *Let $K \subset \mathbb{R}$ be compact. Assume $X_1 \supset X_2 \supset X_3 \supset \cdots$ form a decreasing sequence of closed subsets of K. If each $X_m \ne \emptyset$, then $\cap_m X_m \ne \emptyset$.*

Proof. Pick $x_m \in X_m$. If K is compact, (x_m) has a convergent subsequence, $x_{m_k} \to y$. Since $\{x_{m_k} : k \ge \ell\} \subset X_{m_\ell}$, which is closed, we have $y \in \cap_m X_m$. ☐

Corollary A.3.9. *Let $K \subset \mathbb{R}$ be compact. Assume $U_1 \subset U_2 \subset U_3 \subset \cdots$ form an increasing sequence of open sets in $\mathbb{R}$. If $\cup_m U_m \supset K$, then $U_M \supset K$ for some M.*

Proof. Consider $X_m = K \setminus U_m$. ☐

Before getting to Proposition A.3.11, we bring in the following. Let $\mathbb{Q}$ denote the set of rational numbers. The set $\mathbb{Q} \subset \mathbb{R}$ has the following "denseness" property: given $p \in \mathbb{R}$ and $\varepsilon > 0$, there exists $q \in \mathbb{Q}$ such that $|p - q| < \varepsilon$. Let

$$\mathcal{R} = \{B_{r_j}(q_j) : q_j \in \mathbb{Q},\ r_j \in \mathbb{Q} \cap (0, \infty)\}. \tag{A.3.19}$$

Now one can show that the set $\mathbb{Q}$ is *countable*, i.e., it can be put in one-to-one correspondence with $\mathbb{N}$. Similar reasoning shows that $\mathcal{R}$ is a countable collection of open intervals. The following lemma is left as an exercise for the reader.

Lemma A.3.10. *Let $\Omega \subset \mathbb{R}$ be a nonempty open set. Then*

$$\Omega = \bigcup \{B : B \in \mathcal{R},\ B \subset \Omega\}. \tag{A.3.20}$$

To state the next result, we say that a collection $\{U_\alpha : \alpha \in \mathcal{A}\}$ covers K if $K \subset \cup_{\alpha \in \mathcal{A}} U_\alpha$. If each $U_\alpha \subset \mathbb{R}$ is open, it is called an open cover of K. If $\mathcal{B} \subset \mathcal{A}$ and $K \subset \cup_{\beta \in \mathcal{B}} U_\beta$, we say $\{U_\beta : \beta \in \mathcal{B}\}$ is a subcover. This result is part of the Heine-Borel theorem.

Proposition A.3.11. *If $K \subset \mathbb{R}$ is compact, then it has the following property.*

(A.3.21) *Every open cover $\{U_\alpha : \alpha \in \mathcal{A}\}$ of K has a finite subcover.*

Proof. By Lemma A.3.10, it suffices to prove the following.

(A.3.22) Every countable cover $\{B_j : j \in \mathbb{N}\}$ of K by open intervals has a finite subcover.

For this, we set

$$U_m = B_1 \cup \cdots \cup B_m \tag{A.3.23}$$

and apply Corollary A.3.9. ☐

Exercises

1. Consider a polynomial $p(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$. Assume each $a_j \in \mathbb{R}$ and n is *odd*. Use the intermediate value theorem to show that $p(x) = 0$ for some $x \in \mathbb{R}$.

We describe the construction of a Cantor set. Take a closed, bounded interval $[a, b] = \mathcal{C}_0$. Let $\mathcal{C}_1$ be obtained from $\mathcal{C}_0$ by deleting the open middle third interval, of length $(b - a)/3$. At the jth stage, $\mathcal{C}_j$ is a disjoint union of 2^j closed intervals, each of length $3^{-j}(b - a)$. Then $\mathcal{C}_{j+1}$ is obtained from $\mathcal{C}_j$ by deleting the open middle third of each of these 2^j intervals. We have $\mathcal{C}_0 \supset \mathcal{C}_1 \supset \cdots \supset \mathcal{C}_j \supset \cdots$, each a closed subset of $[a, b]$.

2. Show that

$$\mathcal{C} = \bigcap_{j \ge 0} \mathcal{C}_j \tag{A.3.24}$$

is nonempty, and compact. This is the Cantor set.

3. Suppose $\mathcal{C}$ is formed as above, with $[a, b] = [0, 1]$. Show that points in $\mathcal{C}$ are

precisely those of the form

$$\xi = \sum_{j=0}^{\infty} b_j \, 3^{-j}, \quad b_j \in \{0, 2\}. \tag{A.3.25}$$

4. If $p, q \in \mathcal{C}$ (and $p < q$), show that the interval $[p, q]$ must contain points not in $\mathcal{C}$. One says $\mathcal{C}$ is totally disconnected.

5. If $p \in \mathcal{C}$, $\varepsilon > 0$, show that $(p - \varepsilon, p + \varepsilon)$ contains infinitely many points in $\mathcal{C}$. Given that $\mathcal{C}$ is closed, one says $\mathcal{C}$ is *perfect*.

6. Show that $\mathrm{Card}(\mathcal{C}) = \mathrm{Card}(\mathbb{R})$.
Hint. With ξ as in (A.3.25) show that

$$\xi \mapsto \eta = \sum_{j=0}^{\infty} \Big(\frac{b_j}{2}\Big) 2^{-j}$$

maps $\mathcal{C}$ onto $[0, 1]$.

Remark. At this point, we mention the
Continuum Hypothesis. If $S \subset \mathbb{R}$ is uncountable, then $\mathrm{Card}\, S = \mathrm{Card}\, \mathbb{R}$.
This hypothesis has been shown not to be amenable to proof or disproof, from the standard axioms of set theory. See [**5**]. However, there is a large class of sets for which the conclusion holds. For example, it holds whenever $S \subset \mathbb{R}$ is uncountable and compact. See Chapter 2 of [**15**] for further results along this line.

7. Show that Proposition A.3.7 implies the existence of kth roots of each element of $\mathbb{R}^+$.

8. In the setting of Proposition A.3.7 (the intermediate value theorem), in which $f : [a, b] \to \mathbb{R}$ is continuous and $f(a) < c < f(b)$, consider the following.

(a) Divide $I = [a, b]$ into two equal intervals I_ℓ and I_r, meeting at the midpoint $\alpha_0 = (a + b)/2$. Select $I_1 = I_\ell$ if $f(\alpha_0) \geq c$, $I_1 = I_r$ if $f(\alpha_0) < c$. Say $I_1 = [x_1, y_1]$. Note that $f(x_1) < c,\ f(y_1) \geq c$.

(b) Divide I_1 into two equal intervals $I_{1\ell}$ and I_{1r}, meeting at the midpoint $(x_1 + y_1)/2 = \alpha_1$. Select $I_2 = I_{1\ell}$ if $f(\alpha_1) \geq c$, $I_2 = I_{1r}$ if $f(\alpha_1) < c$. Say $I_2 = [x_2, y_2]$. Note that $f(x_2) < c,\ f(y_2) \geq c$.

(c) Continue. Having $I_k = [x_k, y_k]$, of length $2^{-k}(b - a)$, with $f(x_k) < c,\ f(y_k) \geq c$, divide I_k into two equal intervals $I_{k\ell}$ and I_{kr}, meeting at the midpoint $\alpha_k = (x_k + y_k)/2$. Select $I_{k+1} = I_{k\ell}$ if $f(\alpha_k) \geq c$, $I_{k+1} = I_{kr}$ if $f(\alpha_k) < c$. Again, $I_{k+1} = [x_{k+1}, y_{k+1}]$ with $f(x_{k+1}) < c$ and $f(y_{k+1}) \geq c$.

(d) Show that there exists $x \in (a, b)$ such that

$$x_k \nearrow x, \quad y_k \searrow x, \quad \text{and} \quad f(x) = c.$$

This method of approximating a solution to $f(x) = c$ is called the *bisection method*.

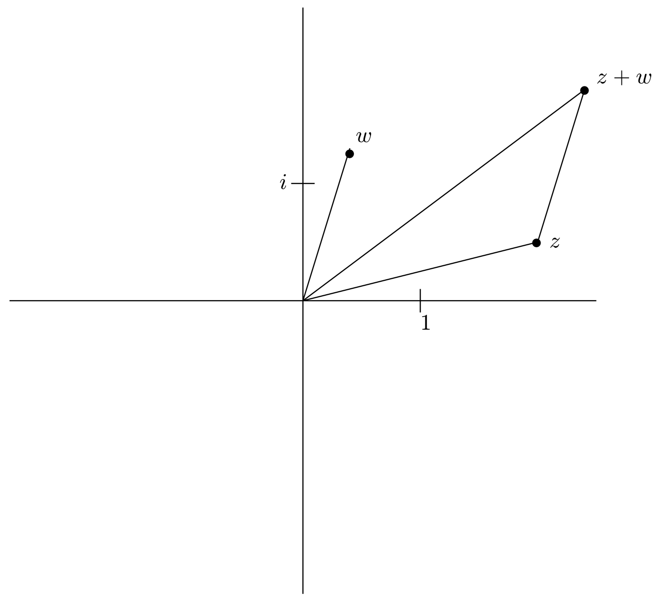

Figure A.4.1. Addition in the complex plane

A.4. Complex numbers

A complex number is a number of the form

$$z = x + iy, \quad x, y \in \mathbb{R}, \tag{A.4.1}$$

where the new object i has the property

$$i^2 = -1. \tag{A.4.2}$$

We denote the set of complex numbers by $\mathbb{C}$. We have $\mathbb{R} \hookrightarrow \mathbb{C}$, identifying $x \in \mathbb{R}$ with $x + i0 \in \mathbb{C}$.

We define addition and multiplication in $\mathbb{C}$ as follows. Suppose $w = a + ib$, $a, b \in \mathbb{R}$. We set

$$\begin{aligned} z + w &= (x + a) + i(y + b), \\ zw &= (xa - yb) + i(xb + ya). \end{aligned} \tag{A.4.3}$$

See Figures A.4.1 and A.4.2 for illustrations of these operations.

It is routine to verify various commutative, associative, and distributive laws of arithmetic. If $z \neq 0$, i.e., either $x \neq 0$ or $y \neq 0$, we can set

$$z^{-1} = \frac{1}{z} = \frac{x}{x^2 + y^2} - i\frac{y}{x^2 + y^2}, \tag{A.4.4}$$

and verify that $zz^{-1} = 1$.

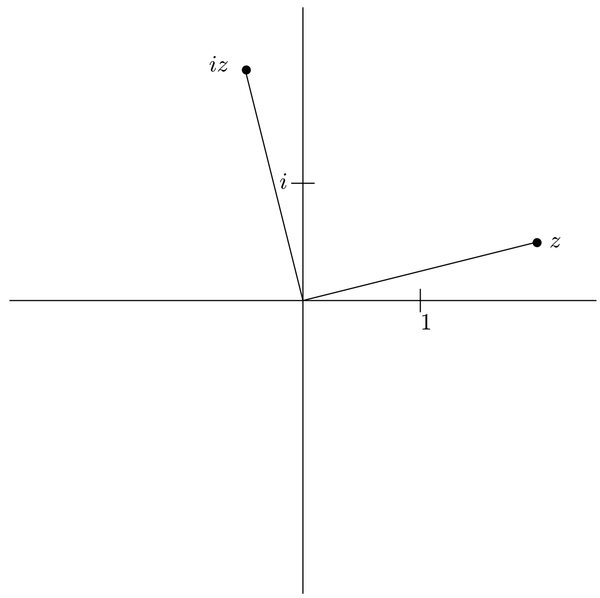

Figure A.4.2. Multiplication by i in $\mathbb{C}$

For some more notation, for $z \in \mathbb{C}$ of the form (A.4.1), we set

$$\overline{z} = x - iy, \quad \operatorname{Re} z = x, \quad \operatorname{Im} z = y. \tag{A.4.5}$$

We say $\overline{z}$ is the complex conjugate of z, $\operatorname{Re} z$ is the real part of z, and $\operatorname{Im} z$ is the imaginary part of z.

We next discuss the concept of the magnitude (or absolute value) of an element $z \in \mathbb{C}$. If z has the form (A.4.1), we take a cue from the Pythagorean theorem, giving the Euclidean distance from z to 0, and set

$$|z| = \sqrt{x^2 + y^2}. \tag{A.4.6}$$

Note that

$$|z|^2 = z\,\overline{z}. \tag{A.4.7}$$

With this notation, (A.4.4) takes the compact (and clear) form

$$z^{-1} = \frac{\overline{z}}{|z|^2}. \tag{A.4.8}$$

We have

$$|zw| = |z| \cdot |w|, \tag{A.4.9}$$

for $z, w \in \mathbb{C}$, as a consequence of the identity (readily verified from the definition (A.4.5))

$$\overline{zw} = \overline{z} \cdot \overline{w}. \tag{A.4.10}$$

In fact, $|zw|^2 = (zw)(\overline{zw}) = z\,w\,\overline{z}\,\overline{w} = z\overline{z}w\overline{w} = |z|^2|w|^2$. This extends the first part of (A.2.23) from $\mathbb{R}$ to $\mathbb{C}$. The extension of the second part also holds, but it requires a little more work. The following is the triangle inequality in $\mathbb{C}$.

Proposition A.4.1. *Given $z, w \in \mathbb{C}$,*

$$|z + w| \le |z| + |w|. \tag{A.4.11}$$

Proof. We compare the squares of each side of (A.4.11). First,

$$\begin{aligned} |z+w|^2 &= (z+w)(\overline{z} + \overline{w}) \\ &= |z|^2 + |w|^2 + w\overline{z} + z\overline{w} \\ &= |z|^2 + |w|^2 + 2\operatorname{Re} z\overline{w}. \end{aligned} \tag{A.4.12}$$

Now, for any $\zeta \in \mathbb{C}$, $\operatorname{Re}\zeta \le |\zeta|$, so $\operatorname{Re} z\overline{w} \le |z\overline{w}| = |z| \cdot |w|$, so (A.4.12) is

$$\le |z|^2 + |w|^2 + 2|z| \cdot |w| = (|z| + |w|)^2, \tag{A.4.13}$$

and we have (A.4.11). □

We now discuss matters related to convergence in $\mathbb{C}$. Parallel to the real case, we say a sequence (z_j) in $\mathbb{C}$ converges to a limit $z \in \mathbb{C}$ (and write $z_j \to z$) if and only if for each $\varepsilon > 0$ there exists N such that

$$j \ge N \Longrightarrow |z_j - z| < \varepsilon. \tag{A.4.14}$$

Equivalently,

$$z_j \to z \Longleftrightarrow |z_j - z| \to 0. \tag{A.4.15}$$

It is easily seen that

$$z_j \to z \Longleftrightarrow \operatorname{Re} z_j \to \operatorname{Re} z \quad \text{and} \quad \operatorname{Im} z_j \to \operatorname{Im} z. \tag{A.4.16}$$

The set $\mathbb{C}$ also has the completeness property, given as follows. A sequence (z_j) in $\mathbb{C}$ is said to be a Cauchy sequence if and only if

$$|z_j - z_k| \to 0, \quad \text{as} \quad j, k \to \infty. \tag{A.4.17}$$

It is easy to see (using the triangle inequality) that if $z_j \to z$ for some $z \in \mathbb{C}$, then (A.4.17) holds. Here is the converse:

Proposition A.4.2. *If (z_j) is a Cauchy sequence in $\mathbb{C}$, then it has a limit.*

Proof. If (z_j) is Cauchy in $\mathbb{C}$, then $(\operatorname{Re} z_j)$ and $(\operatorname{Im} z_j)$ are Cauchy in $\mathbb{R}$, so, by Theorem A.2.9, they have limits. □

We turn to infinite series $\sum_{k=0}^{\infty} a_k$, with $a_k \in \mathbb{C}$. We say this converges if and only if the sequence of partial sums

$$S_n = \sum_{k=0}^{n} a_k \tag{A.4.18}$$

converges:

$$\sum_{k=0}^{\infty} a_k = A \Longleftrightarrow S_n \to A \quad \text{as} \quad n \to \infty. \tag{A.4.19}$$

The following is a useful condition guaranteeing convergence. Compare Proposition A.2.13.

Proposition A.4.3. *The infinite series $\sum_{k=0}^{\infty} a_k$ converges provided*

$$\sum_{k=0}^{\infty} |a_k| < \infty, \tag{A.4.20}$$

i.e., there exists $B < \infty$ such that $\sum_{k=0}^{n} |a_k| \le B$ for all n.

Proof. The triangle inequality gives, for $\ell \ge 1$,

$$\begin{aligned} |S_{n+\ell} - S_n| &= \Bigl| \sum_{k=n+1}^{n+\ell} a_k \Bigr| \\ &\le \sum_{k=n+1}^{n+\ell} |a_k|, \end{aligned} \tag{A.4.21}$$

which tends to 0 as $n \to \infty$, uniformly in $\ell \ge 1$, provided (A.4.20) holds (cf. (A.2.45)–(A.2.46)). Hence (A.4.20) $\Rightarrow$ (S_n) is Cauchy. Convergence then follows, by Proposition A.4.2. $\square$

As in the real case, if (A.4.20) holds, we say the infinite series $\sum_{k=0}^{\infty} a_k$ is *absolutely convergent.*

An example to which Proposition A.4.3 applies is the following power series, giving the exponential function e^z:

$$e^z = \sum_{k=0}^{\infty} \frac{z^k}{k!}, \quad z \in \mathbb{C}. \tag{A.4.22}$$

Compare Exercise 13 of §A.2. The exponential function is explored in depth in §3.2 of Chapter 3.

We turn to a discussion of polar coordinates on $\mathbb{C}$. Given a nonzero $z \in \mathbb{C}$, we can write

$$z = r\omega, \quad r = |z|, \ \omega = \frac{z}{|z|}. \tag{A.4.23}$$

Then ω has unit distance from 0. If the ray from 0 to ω makes an angle θ with the positive real axis, we have

$$\operatorname{Re} \omega = \cos\theta, \quad \operatorname{Im} \omega = \sin\theta, \tag{A.4.24}$$

by definition of the trigonometric functions cos and sin. Hence

$$z = r \ \operatorname{cis}\theta, \tag{A.4.25}$$

where

$$\operatorname{cis}\theta = \cos\theta + i\sin\theta. \tag{A.4.26}$$

If also

$$w = \rho \operatorname{cis} \varphi, \quad \rho = |w|, \tag{A.4.27}$$

then

$$zw = r\rho \operatorname{cis}(\theta + \varphi), \tag{A.4.28}$$

as a consequence of the identity

$$\operatorname{cis}(\theta + \varphi) = (\operatorname{cis} \theta)(\operatorname{cis} \varphi), \tag{A.4.29}$$

which in turn is equivalent to the pair of trigonometric identities

$$\begin{aligned} \cos(\theta + \varphi) &= \cos\theta \, \cos\varphi - \sin\theta \, \sin\varphi, \\ \sin(\theta + \varphi) &= \cos\theta \, \sin\varphi + \sin\theta \, \cos\varphi. \end{aligned} \tag{A.4.30}$$

There is another way to write (A.4.25), using the classical Euler identity

$$e^{i\theta} = \cos\theta + i \sin\theta. \tag{A.4.31}$$

Then (A.4.25) becomes

$$z = r \, e^{i\theta}. \tag{A.4.32}$$

The identity (A.4.29) is equivalent to

$$e^{i(\theta+\varphi)} = e^{i\theta} e^{i\varphi}. \tag{A.4.33}$$

We give a self-contained derivation of (A.4.31) (and also of (A.4.30) and (A.4.33)) in Chapter 3, §§3.1–3.2. The analysis there includes a precise description of what "angle θ" means.

We next define closed and open subsets of $\mathbb{C}$, and discuss the notion of compactness. A set $S \subset \mathbb{C}$ is said to be closed if and only if

$$z_j \in S, \ z_j \to z \Longrightarrow z \in S. \tag{A.4.34}$$

The complement $\mathbb{C} \setminus S$ of a closed set S is open. Alternatively, $\Omega \subset \mathbb{C}$ is open if and only if, given $q \in \Omega$, there exists $\varepsilon > 0$ such that $B_\varepsilon(q) \subset \Omega$, where

$$B_\varepsilon(q) = \{z \in \mathbb{C} : |z - q| < \varepsilon\}, \tag{A.4.35}$$

so q cannot be a limit of a sequence of points in $\mathbb{C} \setminus \Omega$. We define the closure $\overline{S}$ of a set $S \subset \mathbb{C}$ to consist of all points $p \in \mathbb{C}$ such that $B_\varepsilon(p) \cap S \neq \emptyset$ for all $\varepsilon > 0$. Equivalently, $p \in \overline{S}$ if and only if there exists an infinite sequence (p_j) of points in S such that $p_j \to p$.

Parallel to (A.3.7), we say a nonempty set $K \subset \mathbb{C}$ is compact if and only if the following property holds.

(A.4.36) Each infinite sequence (p_j) in K has a subsequence that converges to a point in K.

As in §A.3, if $K \subset \mathbb{C}$ is compact, it must be closed and bounded. Parallel to Theorem A.3.2, we have the converse.

Proposition A.4.4. *If a nonempty $K \subset \mathbb{C}$ is closed and bounded, then it is compact.*

Proof. Let (z_j) be a sequence in K. Then $(\operatorname{Re} z_j)$ and $(\operatorname{Im} z_j)$ are bounded, so Theorem A.2.10 implies the existence of a subsequence such that $\operatorname{Re} z_{j_\nu}$ and $\operatorname{Im} z_{j_\nu}$ converge. Hence the subsequence (z_{j_ν}) converges in $\mathbb{C}$. Since K is closed, the limit must belong to K. □

If $S \subset \mathbb{C}$, a function

$$f : S \longrightarrow \mathbb{C} \tag{A.4.37}$$

is said to be continuous at $p \in S$ provided

$$p_j \in S, \ p_j \to p \Longrightarrow f(p_j) \to f(p). \tag{A.4.38}$$

If f is continuous at each $p \in S$, we say f is continuous on S. The following result has the same proof as Proposition A.3.5.

Proposition A.4.5. *If $K \subset \mathbb{C}$ is compact and $f : K \to \mathbb{C}$ is continuous, then $f(K)$ is compact.*

Then the following variant of Proposition A.3.6 is straightforward.

Proposition A.4.6. *If $K \subset \mathbb{C}$ is compact and $f : K \to \mathbb{C}$ is continuous, then there exists $p \in K$ such that*

$$|f(p)| = \max_{z \in K} |f(z)|, \tag{A.4.39}$$

and there exists $q \in K$ such that

$$|f(q)| = \min_{z \in K} |f(z)|. \tag{A.4.40}$$

There are also straightforward extensions to $K \subset \mathbb{C}$ of Propositions A.3.8–A.3.11. We omit the details.

Exercises

We define π as the smallest positive number such that

$$\operatorname{cis} \pi = -1.$$

See Chapter 4, §§3.1–3.2 for more on this matter.

1. Show that

$$\omega = \operatorname{cis} \frac{2\pi}{n} \Longrightarrow \omega^n = 1.$$

For this, use (A.4.29). In conjunction with (A.4.25)–(A.4.28) and the existence of nth roots of positive real numbers, use this to prove the following:

Given $a \in \mathbb{C}, \ a \neq 0, \ n \in \mathbb{N}$, there exist $z_1, \dots, z_n \in \mathbb{C}$ such that $z_j^n = a$.

2. Compute

$$\Big(\frac{1}{2} + \frac{\sqrt{3}}{2} i\Big)^3,$$

and verify that

$$\cos\frac{\pi}{3} = \frac{1}{2}, \quad \sin\frac{\pi}{3} = \frac{\sqrt{3}}{2}. \tag{A.4.41}$$

3. Find $z_1, \dots, z_n$ such that

$$z_j^n = 1, \tag{A.4.42}$$

explicitly in the form $a + ib$ (not simply as $\mathrm{cis}(2\pi j/n)$), in case

$$n = 3, 4, 6, 8. \tag{A.4.43}$$

Hint. Use (A.4.41), and also the fact that the equation $u_j^2 = i$ has solutions

$$u_1 = \frac{1}{\sqrt{2}} + \frac{i}{\sqrt{2}}, \quad u_2 = -u_1. \tag{A.4.44}$$

4. Take the following path to finding the 5 solutions to

$$z_j^5 = 1. \tag{A.4.45}$$

One solution is $z_1 = 1$. Since $z^5 - 1 = (z-1)(z^4 + z^3 + z^2 + z + 1)$, we need to find 4 solutions to $z^4 + z^3 + z^2 + z + 1 = 0$. Write this as

$$z^2 + z + 1 + \frac{1}{z} + \frac{1}{z^2} = 0, \tag{A.4.46}$$

which, for

$$w = z + \frac{1}{z}, \tag{A.4.47}$$

becomes

$$w^2 + w - 1 = 0. \tag{A.4.48}$$

Use the quadratic formula to find 2 solutions to (A.4.48). Then solve (A.4.47), i.e., $z^2 - wz + 1 = 0$, for z. Use these calculations to show that

$$\cos\frac{2\pi}{5} = \frac{\sqrt{5}-1}{4}.$$

The roots z_j of (A.4.45) form the vertices of a regular pentagon. See Figure A.4.3.

5. Take the following path to explicitly finding the real and imaginary parts of a solution to

$$z^2 = a + ib.$$

Namely, with $x = \operatorname{Re} z,\ y = \operatorname{Im} z$, we have

$$x^2 - y^2 = a, \quad 2xy = b,$$

and also

$$x^2 + y^2 = \rho = \sqrt{a^2 + b^2},$$

hence

$$x = \sqrt{\frac{\rho + a}{2}}, \quad y = \frac{b}{2x},$$

as long as $a + ib \neq -|a|$.

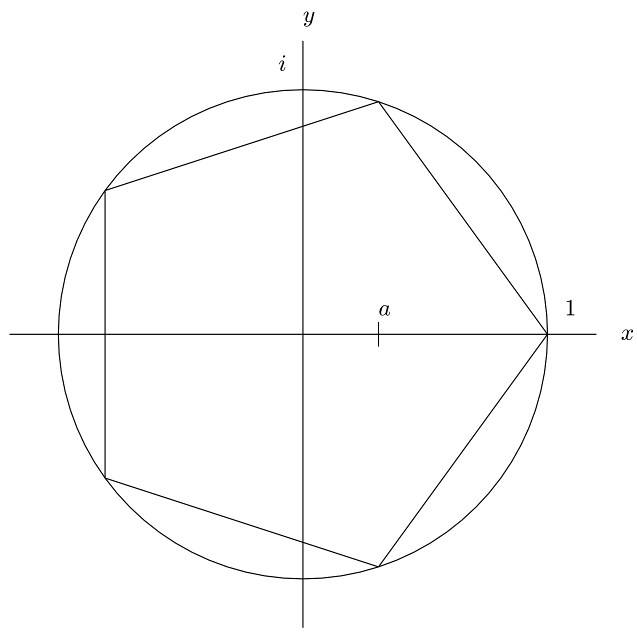

Figure A.4.3. Regular pentagon, $a = (\sqrt{5} - 1)/4$.

6. Taking a cue from Exercise 4 of §A.2, show that

$$\frac{1}{1-z} = \sum_{k=0}^{\infty} z^k, \quad \text{for } \ z \in \mathbb{C}, \ |z| < 1. \tag{A.4.49}$$

7. Show that

$$\frac{1}{1-z^2} = \sum_{k=0}^{\infty} z^{2k}, \quad \text{for } \ z \in \mathbb{C}, \ |z| < 1.$$

8. Produce a power series series expansion in z, valid for $|z| < 1$, for

$$\frac{1}{1+z^2}.$$

9. Consider the unit circle $S^1 = \{z \in \mathbb{C} : |z| = 1\}$. Show that

$$\varphi(x) = \frac{x+i}{x-i}$$

defines

$$\varphi : \mathbb{R} \longrightarrow S^1 \setminus \{1\}, \quad \text{one-one and onto,}$$

with inverse ψ, given by

$$\psi(\omega) = \frac{\omega + 1}{\omega - 1}\, i, \quad \omega \in S^1 \setminus \{1\}.$$

10. Set

$$\begin{aligned} \mathbb{Q}[i] &= \{x + iy : x, y \in \mathbb{Q}\}, \\ S^1_{\mathbb{Q}} &= S^1 \cap \mathbb{Q}[i] = \{x + iy : x, y \in \mathbb{Q},\ x^2 + y^2 = 1\}. \end{aligned}$$

Show that

$$\varphi : \mathbb{Q} \longrightarrow S^1_{\mathbb{Q}} \setminus \{1\}, \quad \text{one-one and onto,}$$

with inverse ψ, as in Exercise 9.

11. A triple (j, k, ℓ) of positive integers is called a Pythagorean triple if

$$j^2 + k^2 = \ell^2.$$

Show that for each such triple, there is a unique $m/n \in \mathbb{Q}$ such that

$$\varphi\Bigl(\frac{m}{n}\Bigr) = \frac{j}{\ell} + \frac{k}{\ell} i.$$

Use this to produce a formula that yields all Pythagorean triples.

Appendix B

Sequences and series of continuous functions

Here we consider sequences of functions

$$f_j : X \longrightarrow \mathbb{R}^n, \tag{B.0.1}$$

where X is a subset of $\mathbb{R}^k$, and produce results on convergence of such sequences, and on series

$$\sum_{k=0}^{\infty} f_k. \tag{B.0.2}$$

We pay particular attention to sequences and series of continuous functions. These results are useful in the development of calculus, for example in results on the Riemann integral and on power series.

Section B.1 gives basic information on continuous functions $f : X \to \mathbb{R}^n$. Some of this extends material from §A.3, which took $n = 1$ and $X \subset \mathbb{R}$. In addition, we define the notion of uniform continuity, and show that each continuous function f is uniformly continuous when X is compact.

In §B.2 we discuss convergence $f_j \to f$ of functions on X, with emphasis on the notion of uniform convergence. We show that if each f_j is continuous and $f_j \to f$ uniformly, then f is continuous.

In §B.3 we consider infinite series (B.0.2), and establish a sufficient condition for uniform convergence known as the Weierstrass M-test. It follows from §B.2 that, if this condition is satisfied and each f_j is continuous, so is the sum.

B.1. Continuous functions

Here we discuss continuity, extending the treatment in in §A.3. Let $X \subset \mathbb{R}^k$. A function

$$f : X \longrightarrow \mathbb{R}^n \tag{B.1.1}$$

is said to be continuous at $p \in X$ provided

$$p_j \in X,\ p_j \to p \Longrightarrow f(p_j) \to f(p). \tag{B.1.2}$$

An equivalent condition is the following: given $\varepsilon > 0$, there exists $\delta = \delta(p) > 0$ such that

$$x \in X,\ |x - p| < \delta \Longrightarrow |f(x) - f(p)| < \varepsilon. \tag{B.1.3}$$

If f is continuous at each point $p \in X$, we say f is continuous on X, and write $f \in C(X, \mathbb{R}^n)$.

There are some important connections between continuity and compactness. The following two results extend Propositions A.3.5–A.3.6.

Proposition B.1.1. *If $X \subset \mathbb{R}^k$ is compact and $f : X \to \mathbb{R}^n$ is continuous, then $f(X)$ is compact.*

Proposition B.1.2. *If $X \subset \mathbb{R}^k$ is compact and $f : X \to \mathbb{R}^n$ is continuous, then there exists $p \in X$ such that*

$$|f(p)| = \max_{x \in X} |f(x)|, \tag{B.1.4}$$

and there exists $q \in X$ such that

$$|f(q)| = \min_{x \in K} |f(x)|. \tag{B.1.5}$$

The proofs are similar to their analogues in §A.3.

Going further, we say $f : X \to \mathbb{R}^n$ is *uniformly continuous* provided that, given $\varepsilon > 0$, there exists $\delta > 0$ (independent of p) such that

$$x, p \in X,\ |x - p| < \delta \Longrightarrow |f(x) - f(p)| < \varepsilon. \tag{B.1.6}$$

Uniform continuity is a very important concept, useful for example in the study of the Riemann integral. An example of a bounded continuous function that is not uniformly continuous is

$$f : \Big(0, \frac{1}{4}\Big] \longrightarrow \mathbb{R}, \quad f(x) = \sin \frac{1}{x}. \tag{B.1.7}$$

See Figure B.1.1. In light of this, it is useful to have the following.

Proposition B.1.3. *If $X \subset \mathbb{R}^k$ is compact and $f : X \to \mathbb{R}^n$ is continuous, then f is uniformly continuous.*

Proof. If f is not uniformly continuous, then there exists $\varepsilon_0 > 0$ such that, for each $\ell \in \mathbb{N}$, there are

$$\begin{aligned} &x_\ell, y_\ell \in X \text{ such that } |x_\ell - y_\ell| < 2^{-\ell}, \text{ but} \\ &|f(x_\ell) - f(y_\ell)| \geq \varepsilon_0. \end{aligned} \tag{B.1.8}$$

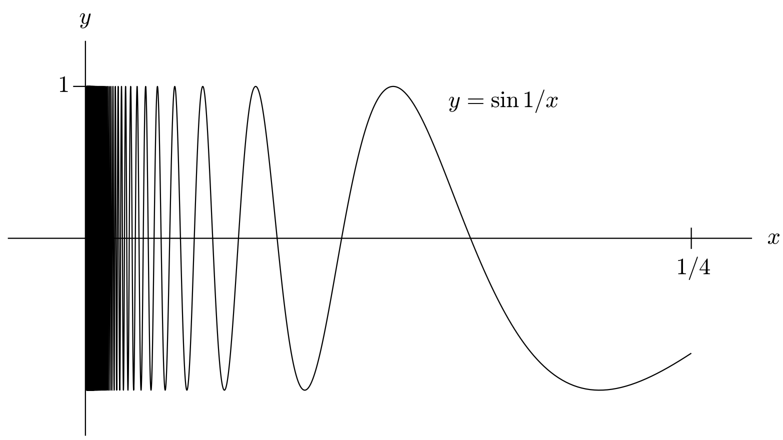

Figure B.1.1. Graph of $y = \sin 1/x$

Since X is compact, (x_ℓ) and (y_ℓ) have convergent subsequences. We hence have

$$x_{\ell_\nu} \to x,\ y_{\ell_\nu} \to y, \quad \text{and} \quad x = y. \tag{B.1.9}$$

The continuity of f then implies

$$f(x_{\ell_\nu}) \to f(x), \quad f(y_{\ell_\nu}) \to f(y), \tag{B.1.10}$$

and then (B.1.8) gives

$$|f(x) - f(y)| \geq \varepsilon_0, \tag{B.1.11}$$

contradicting the fact that $x = y$. □

B.2. Sequences of functions: uniform convergence

Let $X \subset \mathbb{R}^k$ and suppose $f_j, f : X \to \mathbb{R}^n$. We say $f_j \to f$ pointwise on X provided $f_j(x) \to f(x)$ as $j \to \infty$ for each $x \in X$. A stronger type of convergence is *uniform convergence*. We say $f_j \to f$ uniformly on X provided

$$\sup_{x\in X} |f_j(x) - f(x)| \longrightarrow 0, \quad \text{as} \quad j \to \infty. \tag{B.2.1}$$

An equivalent characterization is that, for each $\varepsilon > 0$, there exists $K \in \mathbb{N}$ such that

$$j \geq K \Longrightarrow |f_j(x) - f(x)| \leq \varepsilon, \quad \forall\, x \in X. \tag{B.2.2}$$

A significant property of uniform convergence is that passing to the limit preserves continuity.

Proposition B.2.1. *If $f_j : X \to \mathbb{R}^n$ is continuous for each j and $f_j \to f$ uniformly, then $f : X \to \mathbb{R}^n$ is continuous.*

Proof. Fix $p \in X$ and take $\varepsilon > 0$. Pick $K \in \mathbb{N}$ such that (B.2.2) holds. Then pick $\delta > 0$ such that

$$|x - p| < \delta \Longrightarrow |f_K(x) - f_K(p)| < \varepsilon, \tag{B.2.3}$$

which can be done since $f_K : X \to \mathbb{R}^n$ is continuous. Together, (B.2.2) and (B.2.3) imply

$$\begin{aligned} |x - p| < \delta \Rightarrow |f(x) - f(p)| & \\ &\leq |f(x) - f_K(x)| + |f_K(x) - f_K(p)| + |f_K(p) - f(p)| \\ &\leq 3\varepsilon. \end{aligned} \tag{B.2.4}$$

Thus f is continuous at p, for each $p \in X$. □

We next consider Cauchy sequences of functions $f_j : X \to \mathbb{R}^n$. To say (f_j) is Cauchy at x is simply to say that $(f_j(x))$ is a Cauchy sequence in $\mathbb{R}^n$. We say (f_j) is *uniformly Cauchy* provided

$$\sup_{x\in X} |f_j(x) - f_k(x)| \longrightarrow 0, \quad \text{as} \quad j, k \to \infty. \tag{B.2.5}$$

An equivalent characterization is that, for each $\varepsilon > 0$, there exists $K \in \mathbb{N}$ such that

$$j, k \geq K \Longrightarrow |f_j(x) - f_k(x)| \leq \varepsilon, \quad \forall\, x \in X. \tag{B.2.6}$$

Since, as seen in Chapter 2, $\mathbb{R}^n$ is complete, each Cauchy sequence (f_j) will have a limit $f : X \to \mathbb{R}^n$. We have the following.

Proposition B.2.2. *Assume $f_j : X \to \mathbb{R}^n$. If (f_j) is uniformly Cauchy, then (f_j) converges uniformly to a limit $f : X \to \mathbb{R}^n$.*

Proof. We have already seen that there exists $f : X \to \mathbb{R}^n$ such that $f_j(x) \to f(x)$ for each $x \in X$. To finish the proof, take $\varepsilon > 0$ and pick $K \in \mathbb{N}$ such that (B.2.6) holds. Then taking $k \to \infty$ yields

$$j \geq K \Longrightarrow |f_j(x) - f(x)| \leq \varepsilon, \quad \forall\, x \in X, \tag{B.2.7}$$

yielding the uniform convergence. □

If, in addition, each $f_j : X \to \mathbb{R}^n$ is continuous, we can put Propositions B.2.1 and B.2.2 together. We leave this to the reader.

B.3. Series of functions: the Weierstrass M-test

We move from sequences to series. Again we assume $X \subset \mathbb{R}^k$ and

$$f_j : X \longrightarrow \mathbb{R}^n, \tag{B.3.1}$$

for some $n \in \mathbb{N}$. We look at the infinite series

$$\sum_{k=0}^{\infty} f_k(x), \tag{B.3.2}$$

and seek conditions for convergence, which is the same as convergence of the sequence of partial sums

$$S_j(x) = \sum_{k=0}^{j} f_k(x). \tag{B.3.3}$$

We have convergence at $x \in X$ provided

$$\sum_{k=0}^{\infty} |f_k(x)| < \infty, \tag{B.3.4}$$

i.e., provided there exists $B_x < \infty$ such that

$$\sum_{k=0}^{j} |f_k(x)| \le B_x, \quad \forall\, j \in \mathbb{N}. \tag{B.3.5}$$

In such a case, we say the series (B.3.2) converges *absolutely* at x. We say (B.3.2) converges uniformly on X if and only if (S_j) converges uniformly on X. The following sufficient condition for uniform convergence is called the Weierstrass M-test.

Proposition B.3.1. *Assume there exist M_k such that $|f_k(x)| \le M_k$, for all $x \in X$, and*

$$\sum_{k=0}^{\infty} M_k < \infty. \tag{B.3.6}$$

Then the series (B.3.2) *converges uniformly on X, to a limit $S : X \to \mathbb{R}^n$.*

Proof. This proof is similar to that of Proposition A.2.13, but we review it. We have

$$\begin{aligned} |S_{m+\ell}(x) - S_m(x)| &\le \Big| \sum_{k=m+1}^{m+\ell} f_k(x) \Big| \\ &\le \sum_{k=m+1}^{m+\ell} |f_k(x)| \\ &\le \sum_{k=m+1}^{m+\ell} M_k. \end{aligned} \tag{B.3.7}$$

Now (B.3.6) implies $\sigma_m = \sum_{k=0}^{m} M_k$ is uniformly bounded, so (by Proposition A.2.11), $\sigma_m \nearrow \beta$ for some $\beta \in \mathbb{R}^+$. Hence

$$|S_{m+\ell}(x) - S_m(x)| \le \sigma_{m+\ell} - \sigma_m \le \beta - \sigma_m \to 0, \quad \text{as } m \to \infty, \tag{B.3.8}$$

independent of $\ell \in \mathbb{N}$ and $x \in X$. Thus (S_j) is uniformly Cauchy on X, and uniform convergence follows by Proposition B.2.2. □

Bringing in Proposition B.2.1, we have the following.

Corollary B.3.2. *In the setting of Proposition B.3.1, if also each* $f_k : X \to \mathbb{R}^n$ *is continuous, so is the limit* S*.*

Appendix C

Supplementary material on linear algebra

Chapter 2 introduced some topics in linear algebra needed for the subsequent development of multivariable calculus, starting with Euclidean space $\mathbb{R}^n$ and proceeding to more general vector spaces, and then to linear transformations and determinants. Here we provide some complementary material that is also of occasional use in the text.

Section C.1 deals with inner product spaces, of which $\mathbb{R}^n$ equipped with the dot product is a standard example. We consider both real and complex inner product spaces. Contact with Euclidean space is made through the existence of orthonormal bases (via the Gramm-Schmidt construction). We define the adjoint of a linear map between inner product spaces, $T : V \to W$, and use this to define self-adjoint and unitary transformations, and investigate some of their properties. These results will play a role in §C.2.

Section C.2 deals with eigenvalues and eigenvectors of a linear transformation $T : V \to V$. It includes results on the existence of an orthonormal basis of eigenvectors when T is self adjoint or unitary, or more generally normal. These results have a number of uses in the text, including the discussion of various types of critical points (local max., local min., saddles) of real-valued smooth functions.

Section C.3 deals with matrix norms, filling out material introduced on §2.4. We define an inner product

$$\langle A, B\rangle = \operatorname{Tr} AB^*,$$

for $A, B \in M(n, \mathbb{F})$, $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$, where if $B = (b_{jk})$, then $B^* = (\bar{b}_{kj})$. More generally, this inner product is defined for $A, B \in \mathcal{L}(V, W)$, where V and W are finite-dimensional inner product spaces. The associated norm on A, denoted $\|A\|_{\mathrm{HS}}$, is the *Hilbert-Schmidt norm.* There is also an *operator norm*,

$$\|A\| = \sup\{\|Tv\| : \|v\| \le 1\},$$

and we discuss significant interplays between these two norms.

Section C.4 deals with the matrix exponential,

$$e^{tA} = \sum_{k=0}^{\infty} \frac{t^k}{k!} A^k, \quad A \in M(n, \mathbb{C}),$$

which generalizes the exponential of complex numbers from §3.2. This arises in the treatment of curvature equations, in §§3.3–3.4, and has further roles in subsequent chapters. Results on matrix norms from §C.3 allow us to establish convergence of the defining series for e^{tA}.

C.1. Inner product spaces

Here we look at norms and inner products on vector spaces other than $\mathbb{R}^n$. Generally, as discussed in §2.2, a complex vector space V is a set on which there are operations of vector addition:

$$f, g \in V \Longrightarrow f + g \in V, \tag{C.1.1}$$

and multiplication by an element of $\mathbb{C}$ (called scalar multiplication):

$$a \in \mathbb{C},\ f \in V \Longrightarrow af \in V, \tag{C.1.2}$$

satisfying the following properties. For vector addition, we have

$$f + g = g + f,\ (f + g) + h = f + (g + h),\ f + 0 = f,\ f + (-f) = 0. \tag{C.1.3}$$

For multiplication by scalars, we have

$$a(bf) = (ab)f, \quad 1 \cdot f = f. \tag{C.1.4}$$

Furthermore, we have two distributive laws:

$$a(f + g) = af + ag, \quad (a + b)f = af + bf. \tag{C.1.5}$$

These properties are readily verified for the function spaces mentioned above.

An inner product on a complex vector space V assigns to elements $f, g \in V$ the quantity $(f, g) \in \mathbb{C}$, in a fashion that obeys the following three rules:

$$\begin{gathered}(a_1 f_1 + a_2 f_2, g) = a_1(f_1, g) + a_2(f_2, g), \\ (f, g) = \overline{(g, f)}, \\ (f, f) > 0 \quad \text{unless} \quad f = 0.\end{gathered} \tag{C.1.6}$$

A vector space equipped with an inner product is called an inner product space. For example,

$$(f, g) = \frac{1}{2\pi} \int\limits_{S^1} f(\theta)\overline{g(\theta)}\, d\theta \tag{C.1.7}$$

defines an inner product on $C(S^1)$, and also on $\mathcal{R}(S^1)$, where we identify two functions that differ only on a set of upper content zero. Similarly,

$$(f, g) = \int_{-\infty}^{\infty} f(x)\overline{g(x)}\, dx \tag{C.1.8}$$

defines an inner product on $\mathcal{R}(\mathbb{R})$ (where, again, we identify two functions that differ only on a set of upper content zero).

As another example, in we define ℓ^2 to consist of sequences $(a_k)_{k\in\mathbb{Z}}$ such that

$$\sum_{k=-\infty}^{\infty} |a_k|^2 < \infty. \tag{C.1.9}$$

An inner product on ℓ^2 is given by

$$\big((a_k), (b_k)\big) = \sum_{k=-\infty}^{\infty} a_k \overline{b_k}. \tag{C.1.10}$$

Given an inner product on V, one says the object $\|f\|$ defined by

$$\|f\| = \sqrt{(f,f)} \tag{C.1.11}$$

is the *norm* on V associated with the inner product. Generally, a norm on V is a function $f \mapsto \|f\|$ satisfying

$$\|af\| = |a| \cdot \|f\|, \quad a \in \mathbb{C}, \ f \in V, \tag{C.1.12}$$

$$\|f\| > 0 \quad \text{unless} \quad f = 0, \tag{C.1.13}$$

$$\|f+g\| \le \|f\| + \|g\|. \tag{C.1.14}$$

The property (C.1.14) is called the triangle inequality. A vector space equipped with a norm is called a normed vector space. We can define a distance function on such a space by

$$d(f,g) = \|f-g\|. \tag{C.1.15}$$

If $\|f\|$ is given by (C.1.11), from an inner product satisfying (C.1.6), it is clear that (C.1.12)–(C.1.13) hold, but (C.1.14) requires a demonstration. Note that

$$\begin{aligned}\|f+g\|^2 &= (f+g, f+g) \\ &= \|f\|^2 + (f,g) + (g,f) + \|g\|^2 \\ &= \|f\|^2 + 2\operatorname{Re}(f,g) + \|g\|^2,\end{aligned} \tag{C.1.16}$$

while

$$(\|f\| + \|g\|)^2 = \|f\|^2 + 2\|f\| \cdot \|g\| + \|g\|^2. \tag{C.1.17}$$

Thus to establish (C.1.17) it suffices to prove the following, known as Cauchy's inequality.

Proposition C.1.1. *For any inner product on a vector space V, with $\|f\|$ defined by* (C.1.11),

$$|(f,g)| \le \|f\| \cdot \|g\|, \quad \forall\, f,g \in V. \tag{C.1.18}$$

Proof. We start with

$$0 \le \|f-g\|^2 = \|f\|^2 - 2\operatorname{Re}(f,g) + \|g\|^2, \tag{C.1.19}$$

which implies

$$2\operatorname{Re}(f,g) \le \|f\|^2 + \|g\|^2, \quad \forall\, f,g \in V. \tag{C.1.20}$$

Replacing f by af for arbitrary $a \in \mathbb{C}$ of absolute velue 1 yields $2\operatorname{Re} a(f,g) \le \|f\|^2 + \|g\|^2$, for all such a, hence

$$2|(f,g)| \le \|f\|^2 + \|g\|^2, \quad \forall\, f,g \in V.$$

Replacing f by tf and g by $t^{-1}g$ for arbitrary $t \in (0,\infty)$, we have

$$2|(f,g)| \le t^2\|f\|^2 + t^{-2}\|g\|^2, \quad \forall\, f,g \in V, \ t \in (0,\infty). \tag{C.1.21}$$

If we take $t^2 = \|g\|/\|f\|$, we obtain the desired inequality (C.1.18). This assumes f and g are both nonzero, but (C.1.18) is trivial if f or g is 0. $\square$

An inner product space V is called a *Hilbert space* if it is a complete metric space, i.e., if every Cauchy sequence (f_ν) in V has a limit in V. The space ℓ^2 has this completeness property, but $C(S^1)$, with inner product (C.1.7), does not, nor does $\mathcal{R}(S^1)$. Appendix A.2 describes a process of constructing the completion of the space $\mathbb{Q}$. When applied to an incomplete inner product space, it produces a Hilbert space. When this process is applied to $C(S^1)$, the completion is the space $L^2(S^1)$. An alternative construction of $L^2(S^1)$ uses the Lebesgue integral. For this approach, one can consult Chapter 4 of [**16**].

For the rest of this appendix, we confine attention to finite-dimensional inner product spaces.

If V is a finite-dimensional inner product space, a basis $\{u_1, \dots, u_n\}$ of V is called an *orthonormal basis* of V provided

$$(u_j, u_k) = \delta_{jk}, \quad 1 \le j, k \le n, \tag{C.1.22}$$

i.e.,

$$\|u_j\| = 1, \quad j \ne k \Rightarrow (u_j, u_k) = 0. \tag{C.1.23}$$

In such a case we see that

$$\begin{aligned} &v = a_1 u_1 + \cdots + a_n u_n, \ w = b_1 u_1 + \cdots + b_n u_n \\ &\Longrightarrow (v, w) = a_1 \overline{b}_1 + \cdots + a_n \overline{b}_n. \end{aligned} \tag{C.1.24}$$

It is often useful to construct orthonormal bases. The construction we now describe is called the Gramm-Schmidt construction.

Proposition C.1.2. *Let $\{v_1, \dots, v_n\}$ be a basis of V, an inner product space. Then there is an orthonormal basis $\{u_1, \dots, u_n\}$ of V such that*

$$\operatorname{Span}\{u_j : j \le \ell\} = \operatorname{Span}\{v_j : j \le \ell\}, \quad 1 \le \ell \le n. \tag{C.1.25}$$

Proof. To begin, take

$$u_1 = \frac{1}{\|v_1\|} v_1. \tag{C.1.26}$$

Now define the linear transformation $P_1 : V \to V$ by $P_1 v = (v, u_1) u_1$ and set

$$\tilde{v}_2 = v_2 - P_1 v_2 = v_2 - (v_2, u_1) u_1.$$

We see that $(\tilde{v}_2, u_1) = (v_2, u_1) - (v_2, u_1) = 0$. Also $\tilde{v}_2 \ne 0$ since u_1 and v_2 are linearly independent. Hence we set

$$u_2 = \frac{1}{\|\tilde{v}_2\|} \tilde{v}_2. \tag{C.1.27}$$

Inductively, suppose we have an orthonormal set $\{u_1, \dots, u_m\}$ with $m < n$ and (C.1.25) holding for $1 \le \ell \le m$. Then define $P_m : V \to V$ by

$$P_m v = (v, u_1) u_1 + \cdots + (v, u_m) u_m, \tag{C.1.28}$$

and set

$$\begin{aligned} \tilde{v}_{m+1} &= v_{m+1} - P_m v_{m+1} \\ &= v_{m+1} - (v_{m+1}, u_1) u_1 - \cdots - (v_{m+1}, u_m) u_m. \end{aligned} \tag{C.1.29}$$

We see that

$$j \le m \Rightarrow (\tilde{v}_{m+1}, u_j) = (v_{m+1}, u_j) - (v_{m+1}, u_j) = 0. \tag{C.1.30}$$

Also, since $v_{m+1} \notin \operatorname{Span}\{v_1, \dots, v_m\} = \operatorname{Span}\{u_1, \dots, u_m\}$, it follows that $\tilde{v}_{m+1} \ne 0$. Hence we set

$$u_{m+1} = \frac{1}{\|\tilde{v}_{m+1}\|} \tilde{v}_{m+1}. \tag{C.1.31}$$

This completes the construction. □

EXAMPLE. Take $V = \mathcal{P}_2$, with basis $\{1, x, x^2\}$, and inner product given by

$$(p, q) = \int_{-1}^{1} p(x)\overline{q(x)}\, dx. \tag{C.1.32}$$

The Gramm-Schmidt construction gives first

$$u_1(x) = \frac{1}{\sqrt{2}}. \tag{C.1.33}$$

Then

$$\tilde{v}_2(x) = x,$$

since by symmetry $(x, u_1) = 0$. Now $\int_{-1}^{1} x^2\, dx = 2/3$, so we take

$$u_2(x) = \sqrt{\frac{3}{2}} x. \tag{C.1.34}$$

Next

$$\tilde{v}_3(x) = x^2 - (x^2, u_1)u_1 = x^2 - \frac{1}{3},$$

since by symmetry $(x^2, u_2) = 0$. Now $\int_{-1}^{1} (x^2 - 1/3)^2\, dx = 8/45$, so we take

$$u_3(x) = \sqrt{\frac{45}{8}} \Big(x^2 - \frac{1}{3}\Big). \tag{C.1.35}$$

Let V be an n-dimensional inner product space, $W \subset V$ an m-dimensional linear subspace. By Proposition C.1.2, W has an orthonormal basis

$$\{w_1, \dots, w_m\}.$$

We know from §2.2 that V has a basis of the form

$$\{w_1, \dots, w_m, v_1, \dots, v_\ell\}, \quad \ell + m = n. \tag{C.1.36}$$

Applying Proposition C.1.2 again gives the following.

Proposition C.1.3. *If V is an n-dimensional inner product space and $W \subset V$ an m-dimensional linear subspace, with orthonormal basis $\{w_1, \dots, w_m\}$, then V has an orthonormal basis of the form*

$$\{w_1, \dots, w_m, u_1, \dots, u_\ell\}, \quad \ell + m = n. \tag{C.1.37}$$

We see that, if we define the orthogonal complement of W in V as

$$W^\perp = \{v \in V : (v, w) = 0, \ \forall\, w \in W\}, \tag{C.1.38}$$

then

$$W^\perp = \text{Span}\{u_1, \dots, u_\ell\}. \tag{C.1.39}$$

In particular,

$$\dim W + \dim W^\perp = \dim V. \tag{C.1.40}$$

In the setting of Proposition C.1.3, we can define $P_W \in \mathcal{L}(V)$ by

$$P_W v = \sum_{j=1}^{m} (v, w_j) w_j, \quad \text{for} \ \ v \in V, \tag{C.1.41}$$

and see that P_W is uniquely defined by the properties

$$P_W w = w, \ \ \forall\, w \in W, \quad P_W u = 0, \ \ \forall\, u \in W^\perp. \tag{C.1.42}$$

We call P_W the orthogonal projection of V onto W. Note the appearance of such orthogonal projections in the proof of Proposition C.1.2, namely in (C.1.28).

Another object that arises in the setting of inner product spaces is the *adjoint*, defined as follows. If V and W are finite-dimensional inner product spaces and $T \in \mathcal{L}(V, W)$, we define the adjoint

$$T^* \in \mathcal{L}(W, V), \quad (v, T^* w) = (Tv, w). \tag{C.1.43}$$

If V and W are real vector spaces, we also use the notation T^t for the adjoint, and call it the transpose. In case $V = W$ and $T \in \mathcal{L}(V)$, we say

$$T \text{ is self-adjoint } \Longleftrightarrow T^* = T, \tag{C.1.44}$$

and

$$\begin{aligned} T \text{ is unitary (if } \mathbb{F} = \mathbb{C}\text{), or orthogonal (if } \mathbb{F} = \mathbb{R}\text{)} \\ \Longleftrightarrow T^* = T^{-1}. \end{aligned} \tag{C.1.45}$$

The following gives a significant connection between adjoints and orthogonal complements.

Proposition C.1.4. *Let V be an n-dimensional inner product space, $W \subset V$ a linear subspace. Take $T \in \mathcal{L}(V)$. Then*

$$T : W \to W \Longrightarrow T^* : W^\perp \to W^\perp. \tag{C.1.46}$$

Proof. Note that

$$(w, T^* u) = (Tw, u) = 0, \quad \forall\, w \in W, \ u \in W^\perp, \tag{C.1.47}$$

if $T : W \to W$. This shows that $T^* u \perp W$ for all $u \in W^\perp$, and we have (C.1.46). $\square$

In particular,

$$T = T^*, \ T : W \to W \Longrightarrow T : W^\perp \to W^\perp. \tag{C.1.48}$$

C.2. Eigenvalues and eigenvectors

Let $T : V \to V$ be linear. If there is a nonzero $v \in V$ such that

$$Tv = \lambda_j v, \tag{C.2.1}$$

for some $\lambda_j \in \mathbb{F}$, we say λ_j is an eigenvalue of T, and v is an eigenvector. Let $\mathcal{E}(T, \lambda_j)$ denote the set of vectors $v \in V$ such that (C.2.1) holds. It is clear that $\mathcal{E}(T, \lambda_j)$ (the λ_j-eigenspace of T) is a linear subspace of V and

$$T : \mathcal{E}(T, \lambda_j) \longrightarrow \mathcal{E}(T, \lambda_j). \tag{C.2.2}$$

The set of $\lambda_j \in \mathbb{F}$ such that $\mathcal{E}(T, \lambda_j) \neq 0$ is denoted $\operatorname{Spec}(T)$. Clearly $\lambda_j \in \operatorname{Spec}(T)$ if and only if $T - \lambda_j I$ is not injective, so, if V is finite dimensional,

$$\lambda_j \in \operatorname{Spec}(T) \Longleftrightarrow \det(\lambda_j I - T) = 0. \tag{C.2.3}$$

We call $K_T(\lambda) = \det(\lambda I - T)$ the *characteristic polynomial* of T.

If $\mathbb{F} = \mathbb{C}$, we can use the *fundamental theorem of algebra*, which says every non-constant polynomial with complex coefficients has at least one complex root. (See Appendix E for a proof of this result.) This proves the following.

Proposition C.2.1. *If V is a finite-dimensional complex vector space and $T \in \mathcal{L}(V)$, then T has at least one eigenvector in V.*

Remark. If V is real and $K_T(\lambda)$ does have a real root λ_j, then there is a real λ_j-eigenvector.

Sometimes a linear transformation has only one eigenvector, up to a scalar multiple. Consider the transformation $A : \mathbb{C}^3 \to \mathbb{C}^3$ given by

$$A = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{pmatrix}. \tag{C.2.4}$$

We see that $\det(\lambda I - A) = (\lambda - 2)^3$, so $\lambda = 2$ is a triple root. It is clear that

$$\mathcal{E}(A, 2) = \operatorname{Span}\{e_1\}, \tag{C.2.5}$$

where $e_1 = (1, 0, 0)^t$ is the first standard basis vector of $\mathbb{C}^3$.

If one is given $T \in \mathcal{L}(V)$, it is of interest to know whether V has a basis of eigenvectors of T. The following result is useful.

Proposition C.2.2. *Assume that the characteristic polynomial of $T \in \mathcal{L}(V)$ has k distinct roots, $\lambda_1, \dots, \lambda_k$, with eigenvectors $v_j \in \mathcal{E}(T, \lambda_j)$, $1 \le j \le k$. Then $\{v_1, \dots, v_k\}$ is linearly independent. In particular, if $k = \dim V$, these vectors form a basis of V.*

Proof. We argue by contradiction. If $\{v_1, \dots, v_k\}$ is linearly dependent, take a minimal subset that is linearly dependent and (reordering if necessary) say this set is $\{v_1, \dots, v_m\}$, with $Tv_j = \lambda_j v_j$, and

$$c_1 v_1 + \cdots + c_m v_m = 0, \tag{C.2.6}$$

with $c_j \neq 0$ for each $j \in \{1, \dots, m\}$. Applying $T - \lambda_m I$ to (C.2.6) gives

$$c_1(\lambda_1 - \lambda_m)v_1 + \cdots + c_{m-1}(\lambda_{m-1} - \lambda_m)v_{m-1} = 0, \tag{C.2.7}$$

a linear dependence relation on the smaller set $\{v_1, \dots, v_{m-1}\}$. This contradiction proves the proposition. □

Here is another important class of transformations that have a full complement of eigenvectors.

Proposition C.2.3. *Let V be an n-dimensional inner product space, $T \in \mathcal{L}(V)$. Assume T is self-adjoint, i.e., $T = T^*$. The V has an orthonormal basis of eigenvectors of T.*

Proof. First, assume V is a complex vector space ($\mathbb{F} = \mathbb{C}$). Proposition C.2.1 implies that there exists an eigenvector v_1 of T. Let $W = \operatorname{Span}\{v_1\}$. Then Proposition C.1.4 gives

$$T : W^\perp \longrightarrow W^\perp, \tag{C.2.8}$$

and $\dim W^\perp = n - 1$. The proposition then follows by induction on n. □

If V is a real vector space ($\mathbb{F} = \mathbb{R}$), then the characteristic polynomial $\det(\lambda I - T)$ has a complex root, say $\lambda_1 \in \mathbb{C}$. Denote by $\widetilde{V}$ the complexification of V. The transformation T extends to $T \in \mathcal{L}(\widetilde{V})$, as a self-adjoint transformation on this complex inner product space. Hence there exists nonzero $v_1 \in \widetilde{V}$ such that $Tv_1 = \lambda_1 v_1$. We now take note of the following.

Proposition C.2.4. *If $T = T^*$, every eigenvalue of T is real.*

Proof. Say $Tv_1 = \lambda_1 v_1,\ v_1 \neq 0$. Then

$$\begin{aligned} \lambda_1 \|v_1\|^2 &= (\lambda_1 v_1, v_1) = (Tv_1, v_1) \\ &= (v_1, Tv_1) = (v_1, \lambda v_1) = \overline{\lambda}_1 \|v_1\|^2. \end{aligned} \tag{C.2.9}$$

Hence $\lambda_1 = \overline{\lambda}_1$, so λ_1 is real. □

Returning to the proof of Proposition C.2.3 when V is a real inner product space, we see that the (complex) root λ_1 of $\det(\lambda I - T)$ must in fact be real. Hence $\lambda_1 I - T : V \to V$ is not injective, so there exists a λ_1-eigenvector $v_1 \in V$. Induction on n, as in the argument above, finishes the proof.

Here is a useful general result on orthogonality of eigenvectors.

Proposition C.2.5. *Let V be an inner product space, $T \in \mathcal{L}(V)$. If*

$$Tu = \lambda u, \quad T^* v = \overline{\mu} v, \quad \lambda \neq \mu, \tag{C.2.10}$$

then

$$u \perp v. \tag{C.2.11}$$

Proof. We have

$$\lambda(u, v) = (Tu, v) = (u, T^* v) = \mu(u, v). \tag{C.2.12}$$

□

As a corollary, it $T = T^*$, then

$$Tu = \lambda u, \quad Tv = \mu v, \quad \lambda \neq \mu \Rightarrow u \perp v.$$

Our next goal is to extend Proposition C.2.3 to a broader class of transformations. Given $T \in \mathcal{L}(V)$, where V is an n-dimensional complex inner product space, we say T is *normal* if T and T^* commute, i.e., $TT^* = T^*T$. Equivalently, taking

$$T = A + iB, \quad A = A^*, \ B = B^*, \tag{C.2.13}$$

we have

$$T \text{ normal } \Longleftrightarrow AB = BA. \tag{C.2.14}$$

Generally, for $A, B \in \mathcal{L}(V)$, we see that

$$BA = AB \Longrightarrow B : \mathcal{E}(A, \lambda_j) \to \mathcal{E}(A, \lambda_j). \tag{C.2.15}$$

Thus, in the setting of (C.2.13), we can find an orthonormal basis of each space $\mathcal{E}(A, \lambda)$, $\lambda \in \operatorname{Spec} A$, consisting of eigenvectors of B, to get an orthonormal basis of V consisting of vectors that are simultaneously eigenvectors of A and B, hence eigenvectors of T. This establishes the following.

Proposition C.2.6. *Let V be an n-dimensional complex inner product space, $T \in \mathcal{L}(V)$ a normal transformation. Then V has an orthonormal basis of eigenvectore of T.*

Note that if T has the form (C.2.13)–(C.2.14) and $\lambda = a + ib$, $a, b \in \mathbb{R}$, then

$$\begin{aligned} \mathcal{E}(T, \lambda) &= \mathcal{E}(A, a) \cap \mathcal{E}(B, b) \\ &= \mathcal{E}(T^*, \overline{\lambda}). \end{aligned} \tag{C.2.16}$$

We deduce from Proposition C.2.5 the following.

Proposition C.2.7. *In the setting of Proposition C.2.6, with T normal,*

$$\lambda \neq \mu \Longrightarrow \mathcal{E}(T, \lambda) \perp \mathcal{E}(T, \mu). \tag{C.2.17}$$

An important class of normal operators is the class of unitary operators, defined in §C.1. We recall that if V is an inner product space and $T \in \mathcal{L}(V)$, then

$$T \text{ is unitary } \Longleftrightarrow T^* = T^{-1}. \tag{C.2.18}$$

We write $T \in U(V)$, if V is a complex inner product space. We see from (C.2.16) (or directly) that

$$\begin{aligned} T \in U(V), \ \lambda \in \operatorname{Spec} T &\Longrightarrow \overline{\lambda} = \lambda^{-1} \\ &\Longrightarrow |\lambda| = 1. \end{aligned} \tag{C.2.19}$$

We deduce that if $T \in U(V)$, then V has an orthonormal basis of eigenvectors of T, each eigenvalue being a complex number of absolute value 1.

If V is a *real* n-dimensional inner product space and (C.2.18) holds, we say T is an *orthogonal transformation*, and write $T \in O(V)$. In such a case, V typically does not have an orthonormal basis of eigenvectors of T. However, V does have an orthonormal basis with respect to which such an orthogonal transformation has a special structure, as we proceed to show. To get it, we construct the *complexification* of V,

$$V_{\mathbb{C}} = \{u + iv : u, v \in V\}, \tag{C.2.20}$$

which has a natural structure of a complex n-dimensional vector space, with a Hermitian inner product. A transformation $T \in O(V)$ has a unique $\mathbb{C}$-linear extension to a transformation on $V_{\mathbb{C}}$, which we continue to denote by T, and this extended transformation is unitary on $V_{\mathbb{C}}$. Hence $V_{\mathbb{C}}$ has an orthonormal basis of eigenvectors of T. Say $u + iv \in V_{\mathbb{C}}$ is such an eigenvector,

$$T(u + iv) = e^{-i\theta}(u + iv), \quad e^{i\theta} \notin \{1, -1\}. \tag{C.2.21}$$

Writing $e^{i\theta} = c + is,\ c, s \in \mathbb{R}$, we have

$$\begin{aligned} Tu + iTv &= (c - is)(u + iv) \\ &= cu + sv + i(-su + cv), \end{aligned} \tag{C.2.22}$$

hence

$$\begin{aligned} Tu &= cu + sv, \\ Tv &= -su + cv. \end{aligned} \tag{C.2.23}$$

In such a case, applying complex conjugation to (C.2.21) yields

$$T(u - iv) = e^{i\theta}(u - iv),$$

and $e^{i\theta} \neq e^{-i\theta}$ if $e^{i\theta} \notin \{1, -1\}$, so Proposition C.2.7 yields

$$u + iv \perp u - iv, \tag{C.2.24}$$

hence

$$\begin{aligned} 0 &= (u + iv, u - iv) \\ &= (u, u) - (v, v) + i(v, u) + i(u, v) \\ &= |u|^2 - |v|^2 + 2i(u, v), \end{aligned} \tag{C.2.25}$$

or equivalently

$$|u| = |v| \ \text{ and } \ u \perp v. \tag{C.2.26}$$

Now

$$\operatorname{Span}\{u, v\} \subset V$$

has an $(n-2)$-dimensional orthogonal complement, on which T acts, and an inductive argument gives the following.

Proposition C.2.8. *Let V be a n-dimensional real inner product space, $T : V \to V$ an orthogonal transformation. Then V has an orthonormal basis in which the matrix representation of T consists of blocks*

$$\begin{pmatrix} c_j & -s_j \\ s_j & c_j \end{pmatrix}, \quad c_j^2 + s_j^2 = 1, \tag{C.2.27}$$

plus perhaps an identity matrix block if $1 \in \operatorname{Spec} T$, and a block that is $-I$ if $-1 \in \operatorname{Spec} T$.

This result has the following consequence, advertised in Exercise 14 of §6.1.

Corollary C.2.9. *For each integer $n \geq 2$,*

$$\text{Exp} : \text{Skew}(n) \longrightarrow SO(n) \quad \textit{is onto.} \tag{C.2.28}$$

As in §6.1 we leave the proof as an exercise for the reader. The key is to use the Euler-type identity

$$J = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \Longrightarrow e^{\theta J} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}. \tag{C.2.29}$$

In cases when T is a linear transform on an n-dimensional complex vector space V, and V does not have a basis of eigenvectors of T, it is useful to have the concept of a generalized eigenspace, defined as

$$\mathcal{GE}(T, \lambda_j) = \{v \in V : (t - \lambda_j I)^k v = 0 \text{ for some } k\}. \tag{C.2.30}$$

If λ_j is an eigenvalue of T, nonzero elements of $\mathcal{GE}(T, \lambda_j)$ are called generalized eigenvectors. Clearly $\mathcal{E}(T, \lambda_j) \subset \mathcal{GE}(T, \lambda_j)$. Also $T : \mathcal{GE}(T, \lambda_j) \to \mathcal{GE}(T, \lambda_j)$. Furthermore, one has the following.

Proposition C.2.10. *If $\mu \neq \lambda_j$, then*

$$T - \mu I : \mathcal{GE}(T, \lambda_j) \xrightarrow{\approx} \mathcal{GE}(T, \lambda_j). \tag{C.2.31}$$

It is useful to know the following.

Proposition C.2.11. *If W is an n-dimensional complex vector space, and $T \in \mathcal{L}(V)$, then W has a basis of generalized eigenvectors of T.*

We will not give a proof of this result here. A proof can be found in Chapter 2, §7 of [**19**], and also in [**20**].

C.3. Matrix norms

Let V and W be inner product spaces, of dimension n and m, respectively. They can be real or complex. If $T \in \mathcal{L}(V, W)$, we define

$$\|T\| = \sup\{\|Tv\| : \|v\| \le 1\}. \tag{C.3.1}$$

Equivalently, $\|T\|$ is the smallest quantity K such that

$$\|Tv\| \le K\|v\|, \quad \forall\, v \in V. \tag{C.3.2}$$

To see the equivalence, note that (C.3.2) holds if and only if $\|Tv\| \le K$ for all v such that $\|v\| = 1$. We call $\|T\|$ the *operator norm* of T. The fact that the unit ball in V is compact guarantees that $\|T\|$ is well defined. We will make some explicit estimates below.

If also $S : W \to X$, another inner product space, then

$$\|STv\| \le \|S\|\,\|TV\| \le \|S\|\,\|T\|\,\|v\|, \quad \forall\, v \in V, \tag{C.3.3}$$

and hence

$$\|ST\| \le \|S\|\,\|T\|. \tag{C.3.4}$$

In particular, we have by induction that

$$T \in \mathcal{L}(V) \Longrightarrow \|T^k\| \le \|T\|^k, \quad \forall\, k \in \mathbb{N}. \tag{C.3.5}$$

This will be useful when we discuss the matrix exponential, in §C.4.

We turn to the notion of the trace of a transformation $T \in \mathcal{L}(V)$. We start with $A = (a_{jk}) \in M(n, \mathbb{F})$, and as in §2.4 we set

$$\operatorname{Tr} A = \sum_{j=1}^{n} a_{jj}. \tag{C.3.6}$$

Note that is also $B = (b_{jk}) \in M(n, \mathbb{F})$, then

$$\begin{aligned} AB = C = (c_{jk}), &\quad c_{jk} = \sum_{\ell} a_{j\ell} b_{\ell k}, \\ BA = D = (d_{jk}), &\quad d_{jk} = \sum_{\ell} b_{j\ell} a_{\ell k}, \end{aligned} \tag{C.3.7}$$

and hence

$$\operatorname{Tr} AB = \sum_{j,\ell} a_{j\ell} b_{\ell j} = \operatorname{Tr} BA. \tag{C.3.8}$$

Hence, if B is invertible,

$$\operatorname{Tr} B^{-1}AB = \operatorname{Tr} ABB^{-1} = \operatorname{Tr} A. \tag{C.3.9}$$

Now, if $T \in \mathcal{L}(V)$, we can choose a basis $S = \{v_1, \dots, v_n\}$ of V, and set up an isomorphism $J_S : \mathbb{F}^n \to V$, and define

$$\operatorname{Tr} T = \operatorname{Tr} A, \quad A = J_S^{-1} T J_S. \tag{C.3.10}$$

It follows from (C.3.9) that this is independent of the choice of basis of V.

Next we recall from §C.1 the notion of the adjoint of $T \in \mathcal{L}(V, W)$, the map $T^* \in \mathcal{L}(W, V)$ satisfying

$$(Tv, w) = (v, T^*w), \quad \forall\, v \in V,\ w \in W. \tag{C.3.11}$$

If $\{v_1, \dots, v_n\}$ is an orthonormal basis of V and $\{w_1, \dots, w_m\}$ an orthonormal basis of W, then

$$A = (a_{ij}), \quad a_{ij} = (Tv_j, w_i) \tag{C.3.12}$$

is the matrix representation of T, and the matrix representation of T^* is

$$A^* = (\overline{a}_{ji}). \tag{C.3.13}$$

Now we define the *Hilbert-Schmidt norm* of $T \in \mathcal{L}(V, W)$ when V and W are finite-dimensional inner product spaces. Namely, we set

$$\|T\|_{\mathrm{HS}}^2 = \operatorname{Tr} T^*T = \operatorname{Tr} TT^*. \tag{C.3.14}$$

In terms of the matrix representation (C.3.12) of T, we have

$$T^*T = (b_{jk}), \quad b_{jk} = \sum_\ell \overline{a}_{\ell j} a_{\ell k}, \tag{C.3.15}$$

hence

$$\|T\|_{\mathrm{HS}}^2 = \sum_j b_{jj} = \sum_{j,k} |a_{jk}|^2. \tag{C.3.16}$$

Equivalently, using an arbitrary orthonormal basis $\{v_1, \dots, v_n\}$ of V, we have

$$\|T\|_{\mathrm{HS}}^2 = \sum_{j=1}^n \|Tv_j\|^2. \tag{C.3.17}$$

If also $\{w_1, \dots, w_m\}$ is an orthonormal basis of W, then

$$\begin{aligned} \|T\|_{\mathrm{HS}}^2 &= \sum_{j,k} |(Tv_j, w_k)|^2 = \sum_{j,k} |(v_j, T^*w_k)|^2 \\ &= \sum_k \|T^*w_k\|^2. \end{aligned} \tag{C.3.18}$$

This gives $\|T\|_{\mathrm{HS}} = \|T^*\|_{\mathrm{HS}}$. Also the right side of (C.3.18) is clearly independent of the choice of orthonormal basis $\{v_1, \dots, v_n\}$ of V. Of course, we already know that the right side of (C.3.14) is independent of such a choice of basis.

Using (C.3.17), we can show that the operator norm of $\|T\|$ is dominated by the Hilbert-Schmidt norm:

$$\|T\| \le \|T\|_{\mathrm{HS}}. \tag{C.3.19}$$

In fact, pick a unit $v_1 \in V$ such that $\|Tv_1\|$ is maximized over $\{v : \|v\| \le 1\}$, extend this to an orthonormal basis $\{v_1, \dots, v_n\}$, and use

$$\|T\|^2 = \|Tv_1\|^2 \le \sum_{j=1}^n \|Tv_j\|^2 = \|T\|_{\mathrm{HS}}^2. \tag{C.3.20}$$

Also we can dominate each term on the right side of (C.3.17) by $\|T\|^2$, so

$$\|T\|_{\mathrm{HS}} \le \sqrt{n}\, \|T\|, \quad n = \dim V. \tag{C.3.21}$$

Another consequence of (C.3.17)–(C.3.19) is

$$\|ST\|_{\mathrm{HS}} \le \|S\|\,\|T\|_{\mathrm{HS}} \le \|S\|_{\mathrm{HS}}\|T\|_{\mathrm{HS}}, \tag{C.3.22}$$

for S as in (C.3.3). In particular, parallel to (C.3.5), we have

$$T \in \mathcal{L}(V) \Longrightarrow \|T^k\|_{\mathrm{HS}} \le \|T\|_{\mathrm{HS}}^k, \quad \forall\, k \in \mathbb{N}. \tag{C.3.23}$$

Exercises

Here V and W are finite-dimensional inner product spaces.

1. Let $S, T \in \mathcal{L}(V, W)$. Show that
$$\|S + T\| \le \|S\| + \|T\|, \quad \|S + T\|_{\mathrm{HS}} \le \|S\|_{\mathrm{HS}} + \|T\|_{\mathrm{HS}}.$$

2. Show that, if $T \in \mathcal{L}(V),\ k \in \mathbb{N}$,
$$\|T^k\|_{\mathrm{HS}} \le \|T\|^{k-1}\|T\|_{\mathrm{HS}}.$$

3. Suppose $A \in \mathcal{L}(V)$ and $\|A\| < 1$. Show that
$$(I - A)^{-1} = I + A + A^2 + \cdots + A^k + \cdots,$$
a convergent series.

4. Show that, for any real θ, the matrix
$$A = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$
has operator norm 1. Compute its Hilbert-Schmidt norm.

5. Show that, for $T \in \mathcal{L}(V)$,
$$\|T\| = \sup\{|(Tu, v)| : \|u\|, \|v\| \le 1\}.$$
Show that
$$\|T^*\| = \|T\| \ \text{ and } \ \|T^*T\| = \|T\|^2.$$

C.4. The matrix exponential

Given $A \in M(n, \mathbb{R})$ or $M(n, \mathbb{C})$, the matrix exponential $\operatorname{Exp}(tA) = e^{tA}$ is constructed to solve the differential equation

$$\frac{d}{dt} e^{tA} = A e^{tA}, \quad e^{0A} = I. \tag{C.4.1}$$

Trying a power series representation and arguing as in §3.2 yields a solution which we take to define e^{tA}:

$$e^{tA} = \sum_{k=0}^{\infty} \frac{t^k}{k!} A^k. \tag{C.4.2}$$

Note that we then also have

$$\frac{d}{dt} e^{tA} = e^{tA} A. \tag{C.4.3}$$

We claim that this solution is unique, i.e., if $X : \mathbb{R} \to M(n, \mathbb{C})$ solves

$$X'(t) = AX(t), \quad X(0) = I, \tag{C.4.4}$$

then $X(t) = e^{tA}$. To see this, note that

$$\begin{aligned} \frac{d}{dt} e^{-tA} X(t) &= e^{-tA} X'(t) - e^{-tA} A X(t) \\ &= e^{-tA} A X(t) - e^{-tA} A X(t) \\ &= 0, \end{aligned} \tag{C.4.5}$$

so $e^{-tA} X(t)$ is independent of t. Taking $t = 0$ gives

$$e^{-tA} X(t) = I, \quad \forall\, t \in \mathbb{R}. \tag{C.4.6}$$

This holds for each solution to (C.4.4), in particular for e^{tA}, so

$$e^{-tA} e^{tA} = I, \quad \forall\, t \in \mathbb{R}. \tag{C.4.7}$$

Hence we can multiply each side of (C.4.6) on the left by e^{tA}, to get the desired result.

Continuing along this line, we can compute

$$\frac{d}{dt} e^{(s+t)A} e^{-tA} = 0, \tag{C.4.8}$$

to get $e^{(s+t)A} e^{-tA} = e^{sA}$, hence

$$e^{(s+t)A} = e^{sA} e^{tA}, \quad \forall\, s, t \in \mathbb{R},\ A \in M(n, \mathbb{C}). \tag{C.4.9}$$

A related identity is

$$e^{t(A+B)} = e^{tA} e^{tB}, \quad \text{provided} \quad AB = BA, \tag{C.4.10}$$

given $A, B \in M(n, \mathbb{C})$. To see this, we compute that

$$\frac{d}{dt} e^{t(A+B)} e^{-tB} e^{-tA} = 0, \tag{C.4.11}$$

provided

$$e^{-tB} A = A e^{-tB}, \tag{C.4.12}$$

which holds provided $AB = BA$. The desired identity (C.4.10) then follows from (C.4.11).

The matrix exponential plays an important role in the study of $n \times n$ linear systems of differential equations. For more on this, see Chapter 3 of [**19**].

If $A(t)$ is a smooth function of t with values in $M(n, \mathbb{C})$, one does not always have an explicit formula for solutions to

$$\frac{d}{dt} X(t) = A(t)X(t), \quad X(0) = I, \tag{C.4.13}$$

though there is a body of results on this sort of system of ODE, which can be found in [**19**]. On the other hand, one does have a neat formula in the special case that

$$A(t_1)A(t_2) = A(t_2)A(t_1), \quad \forall\, t_j. \tag{C.4.14}$$

We can get this by observing that if $B(t)$ is a smooth function of t with values in $M(n, \mathbb{C})$, and if

$$B(t_1)B(t_2) = B(t_2)B(t_1), \quad \forall\, t_j, \tag{C.4.15}$$

then, thanks to (C.4.10), we have

$$\frac{d}{dt} e^{B(t)} = B'(t) e^{B(t)}. \tag{C.4.16}$$

Hence a solution to (C.4.13) is given by

$$X(t) = e^{B(t)}, \quad B(t) = \int_0^t A(\tau)\, d\tau, \tag{C.4.17}$$

provided (C.4.14) holds.

Exercises

1. Use results of §C.3 to show that, for $A \in M(n, \mathbb{F})$,

$$\Big\| \sum_{k=m}^{m+n} \frac{t^k}{k!} A^k \Big\| \le \sum_{k=m}^{m+n} \frac{|t|^k}{k!} \|A\|^k.$$

Use the ratio test and the Weierstrass M-test to show that the infinite series (C.4.2) converges for all $t \in \mathbb{R}$, uniformly on $|t| \le R$, for each $R < \infty$.

2. Show that (C.4.15) implies (C.4.16).
Hint. Start with

$$e^{B(t+h)} - e^{B(t)} = \Big[e^{B(t+h)-B(t)} - I \Big] e^{B(t)},$$

and plug in the power series for

$$e^Y, \quad Y = B(t+h) - B(t) = hB'(t) + o(h).$$

3. Show that, for $A \in M(n, \mathbb{C}),\ \lambda \in \mathbb{C}$,
$$Av = \lambda v \Longrightarrow e^{tA}v = e^{t\lambda}v.$$

4. Show that
$$(A - \lambda I)^2 v = 0 \Longrightarrow e^{tA}v = e^{t\lambda}\Big[I + t(A - \lambda I)\Big]v.$$
Extend this calculation to the setting where
$$(A - \lambda I)^k v = 0.$$
Hint. Start by showing that $e^{tA} = e^{t(\lambda I + (A - \lambda I))}v = e^{t\lambda}e^{t(A - \lambda I)}$, via (C.4.10).

Green's theorem and complex differentiable functions

Let f be a complex valued C^1 function on a region $\Omega \subset \mathbb{R}^2$. We identify $\mathbb{R}^2$ and $\mathbb{C}$, via $z = x + iy$, and write $f(z) = f(x, y)$. We say f is *holomorphic* on Ω provided it is complex differentiable, in the sense that

$$\lim_{h \to 0} \frac{1}{h} \big[f(z+h) - f(z) \big] \ \text{ exists,} \tag{D.0.1}$$

for each $z \in \Omega$. When this limit exists, we denote it $f'(x)$, or df/dz. An equivalent condition (given $f \in C^1$) is that f satisfies the *Cauchy-Riemann equation*

$$\frac{\partial f}{\partial x} = \frac{1}{i} \frac{\partial f}{\partial y}. \tag{D.0.2}$$

In such a case,

$$f'(z) = \frac{\partial f}{\partial x}(z) = \frac{1}{i} \frac{\partial f}{\partial y}(z). \tag{D.0.3}$$

Note that $f(z) = z$ has this property, but $f(z) = \overline{z}$ does not. The following is a convenient tool for producing more holomorphic functions.

Lemma D.0.1. *If f and g are holomorphic on Ω, so is fg.*

Proof. We have

$$\frac{\partial}{\partial x}(fg) = \frac{\partial f}{\partial x} g + f \frac{\partial g}{\partial x}, \quad \frac{\partial}{\partial y}(fg) = \frac{\partial f}{\partial y} g + f \frac{\partial g}{\partial y}, \tag{D.0.4}$$

so if f and g satisfy the Cauchy-Riemann equation, so does fg. □

Note that

$$\frac{d}{dz}(fg) = f'(z) g(z) + f(z) g'(z). \tag{D.0.5}$$

Using this lemma, we can inductively show that if $k \in \mathbb{N}$, then z^k is holomorphic on $\mathbb{C}$, and

$$\frac{d}{dz} z^k = k z^{k-1}. \tag{D.0.6}$$

We can also treat $1/z$:

$$\frac{d}{dz}\frac{1}{z} = \lim_{h \to 0} \frac{1}{h}\Bigl(\frac{1}{z+h} - \frac{1}{z}\Bigr) = \lim_{h \to 0} \frac{1}{h}\,\frac{-h}{z(z+h)} = -\frac{1}{z^2}. \tag{D.0.7}$$

Then we can verify (D.0.6) for all $k \in \mathbb{Z}$ ($z \in \mathbb{C} \setminus 0$ if $k < 0$).

Here is another important example.

Lemma D.0.2. *The exponential function e^z is holomorphic on $\mathbb{C}$, and*

$$\frac{d}{dz} e^z = e^z. \tag{D.0.8}$$

Proof. Write $e^z = e^{x+iy} = e^x e^{iy}$. Then

$$\frac{\partial}{\partial x} e^x e^{iy} = e^x e^{iy}, \quad \frac{\partial}{\partial y} e^x e^{iy} = i e^x e^{iy}, \tag{D.0.9}$$

so the Cauchy-Riemann equation holds and we have (D.0.8). □

Our goal in this appendix is to show how Green's theorem can be used to establish results about holomorphic functions on domains in $\mathbb{C}$. The first result is the Cauchy integral theorem, established in §D.1. This is followed in §D.2 by the Cauchy integral formula, and in §D.3 by Liouville's theorem, which will be applied in Appendix E.

Material here gives a taste of results in the important area of complex function theory. For more on this, the reader can look at [**17**].

D.1. The Cauchy integral theorem

Let $\Omega \subset \mathbb{C}$ be a smoothly bounded open set, with closure $\overline{\Omega}$. The Cauchy integral theorem says the following.

Theorem D.1.1. *Assume $f : \overline{\Omega} \to \mathbb{C}$ is C^1, and holomorphic on Ω. Then*

$$\int_{\partial\Omega} f(z)\,dz = 0. \tag{D.1.1}$$

Proof. Here $dz = dx + i dy$, so

$$\int_{\partial\Omega} f(z)\,dz = \int_{\partial\Omega} f\,dx + if\,dy. \tag{D.1.2}$$

Recall that Green's theorem gives

$$\int_{\partial\Omega} f\,dx + g\,dy = \int_{\Omega} \Big(\frac{\partial g}{\partial x} - \frac{\partial f}{\partial y}\Big)\,dx\,dy. \tag{D.1.3}$$

We apply this with $g = if$. We see that (D.1.2) is equal to the left side of (D.1.3), with $g = if$. In this case, the right side of (D.1.3) is equal to

$$\int_{\Omega} \Big(i\frac{\partial f}{\partial x} - \frac{\partial f}{\partial y}\Big)\,dx\,dy = 0, \tag{D.1.4}$$

by the Cauchy-Riemann equation for f. □

D.2. The Cauchy integral formula

As in §D.1, let $\Omega \subset \mathbb{C}$ be a smoothly bounded open set, with closure $\overline{\Omega}$. The Cauchy integral formula is the following.

Theorem D.2.1. *Assume $f : \overline{\Omega} \to \mathbb{C}$ is C^1, and holomorphic on Ω. Then*

$$f(z_0) = \frac{1}{2\pi i} \int_{\partial\Omega} \frac{f(z)}{z - z_0}\, dz, \tag{D.2.1}$$

for each $z_0 \in \Omega$.

Proof. Pick $\varepsilon_0 > 0$ so small that $D_{\varepsilon_0}(z_0) \subset \Omega$. For $\varepsilon \in (0, \varepsilon_0)$, let $\Omega_\varepsilon = \Omega \setminus \overline{D_\varepsilon(z_0)}$. Then

$$g(z) = \frac{f(z)}{z - z_0} \quad \text{is holomorphic on} \quad \Omega_\varepsilon, \tag{D.2.2}$$

and C^1 on $\overline{\Omega}_\varepsilon$, so Theorem D.1.1 implies

$$\int_{\partial\Omega_\varepsilon} g(z)\, dz = 0, \tag{D.2.3}$$

hence

$$\int_{\partial\Omega} \frac{f(z)}{z - z_0}\, dz = \int_{\partial D_\varepsilon(z_0)} \frac{f(z)}{z - z_0}\, dz. \tag{D.2.4}$$

Parametrizing $\partial D_\varepsilon(z_0)$ by $\gamma(t) = z_0 + \varepsilon e^{it}$, so $\gamma'(t) = i\varepsilon e^{it}$, we see that the right side of (D.2.4) is equal to

$$\int_0^{2\pi} \frac{f(z_0 + \varepsilon e^{it})}{\varepsilon e^{it}}\, i\varepsilon e^{it}\, dt = i \int_0^{2\pi} f(z_0 + \varepsilon e^{it})\, dt, \tag{D.2.5}$$

and taking the limit $\varepsilon \to 0$ gives (D.2.1). □

It is useful to change labels, and write (D.2.1) as

$$f(z) = \frac{1}{2\pi i} \int_{\partial\Omega} \frac{f(\zeta)}{z - \zeta}\, d\zeta, \quad z \in \Omega. \tag{D.2.6}$$

Now it is natural to regard z as a variable and take the z-derivative. Parallel to (D.0.7), we have

$$\frac{d}{dz} \frac{1}{\zeta - z} = \frac{1}{(\zeta - z)^2}, \tag{D.2.7}$$

and (D.2.6) yields

$$f'(z) = \frac{1}{2\pi i} \int_{\partial\Omega} \frac{f(\zeta)}{(\zeta - z)^2}\, d\zeta, \quad z \in \Omega. \tag{D.2.8}$$

We see from (D.2.8) that f' is holomorphic on Ω. We can keep this up, obtaining inductively that

$$f^{(n)}(z) = \frac{n!}{2\pi i} \int_{\partial\Omega} \frac{f(\zeta)}{(\zeta - z)^{n+1}}\, d\zeta, \quad z \in \Omega. \tag{D.2.9}$$

D.3. Liouville's theorem

The following consequence of the Cauchy integral formula is known as Liouville's theorem.

Theorem D.3.1. *Let $f : \mathbb{C} \to \mathbb{C}$ be holomorphic. If f is bounded, i.e.,*

$$|f(z)| \le M < \infty, \quad \forall\, z \in \mathbb{C}, \tag{D.3.1}$$

then f is constant.

Proof. We will show that the hypothesis (D.3.1) implies that f' is identically zero, which implies that f is constant. To see this, apply (D.2.8) with $\Omega = D_R(z)$. Parametrizing $\partial D_R(z)$ by $\gamma(t) = z + Re^{it}$, we have

$$\begin{aligned} f'(z) &= \frac{1}{2\pi i} \int\limits_{\partial D_R(z)} \frac{f(\zeta)}{(\zeta - z)^2}\, d\zeta \\ &= \frac{1}{2\pi R} \int_0^{2\pi} f(z + Re^{it})\, dt, \end{aligned} \tag{D.3.2}$$

hence

$$|f'(z)| \le \frac{M}{R}. \tag{D.3.3}$$

Taking $R \to \infty$ yields

$$|f'(z)| = 0, \quad \forall\, z \in \mathbb{C}, \tag{D.3.4}$$

and we are done. □

Appendix E

Polynomials and the fundamental theorem of algebra

The polynomial

$$p(x) = x^2 + 1, \quad x \in \mathbb{R}, \tag{E.0.1}$$

clearly has no real root. The complex number $i = \sqrt{-1}$ was introduced to provide such a root. Then there is a factorization

$$z^2 + 1 = (z + i)(z - i), \quad z \in \mathbb{C}. \tag{E.0.2}$$

It was then established that one need go no further to produce roots of polynomials. The fundamental theorem of algebra asserts that every nonconstant polynomial

$$p(z) = a_n z^n + a_{n-1} z^{n-1} + \cdots + a_0, \tag{E.0.3}$$

with $a_j \in \mathbb{C}$, $n \geq 1$, $a_n \neq 0$, vanishes for some $z \in \mathbb{C}$. Furthermore, such a polynomial has a factorization into linear factors. This result is of use in Appendix C, to produce eigenvalues of matrices, which in turn is useful for the study of Hessian matrices in Chapter 3.

We give two proofs of the fundamental theorem of algebra, one in §E.1 that is elementary, in the sense that it does not use Green's theorem, and a second in §E.2, which uses Liouville's theorem, established in §D.3 as a consequence of the Cauchy integral theorem.

E.1. Elementary proof of the fundamental theorem of algebra

The following result is the fundamental theorem of algebra.

Theorem E.1.1. *If $P(z)$ is a nonconstant polynomial (with complex coefficients), then $p(z)$ must have a complex root.*

Proof. We have, for some $n \ge 1,\ a_n \neq 0$,

$$\begin{aligned} p(z) &= a_n z^n + \cdots + a_1 z + a_0 \\ &= a_n z^n \bigl(1 + O(z^{-1})\bigr), \quad |z| \to \infty, \end{aligned} \tag{E.1.1}$$

which implies

$$\lim_{|z|\to\infty|} |p(z)| = \infty. \tag{E.1.2}$$

Picking $R \in (0,\infty)$ such that

$$\inf_{|z|\ge R} |p(z)| > |p(0)|, \tag{E.1.3}$$

we deduce that

$$\inf_{|z|\le R} |p(z)| = \inf_{z\in\mathbb{C}} |p(z)|. \tag{E.1.4}$$

Since $\overline{D}_R = \{z : |z| \le R\}$ is compact and p is continuous, there exists $z_0 \in \overline{D}_R$ such that

$$|p(z_0)| = \inf_{z\in\mathbb{C}} |p(z)|. \tag{E.1.5}$$

The theorem hence follows from the following result. ☐

Lemma E.1.2. *If $p(z)$ is a nonconstant polynomial and* (E.1.5) *holds, then $p(z_0) = 0$.*

Proof. Suppose to the contrary that

$$p(z_0) = a \neq 0. \tag{E.1.6}$$

We can write

$$p(z_0 + \zeta) = a + q(\zeta), \tag{E.1.7}$$

where $q(\zeta)$ is a (nonconstant) polynomial in ζ, satisfying $q(0) = 0$. Hence, for some $k \ge 1$ and $b \neq 0$, we have $q(\zeta) = b\zeta^k + \cdots + b_n\zeta^n$, i.e.,

$$q(\zeta) = b\zeta^k + O(\zeta^{k+1}), \quad \zeta \to 0, \tag{E.1.8}$$

so, uniformly on $S^1 = \{\omega \in \mathbb{C} : |\omega = 1\}$,

$$p(z_0 + \varepsilon\omega) = a + b\omega^k\varepsilon^k + O(\varepsilon^{k+1}), \quad \varepsilon \searrow 0. \tag{E.1.9}$$

Pick $\omega \in S^1$ such that

$$\frac{b}{|b|}\omega^k = -\frac{a}{|a|}, \tag{E.1.10}$$

which is possible since $a \neq 0$ and $b \neq 0$. In more detail, since $-(a/|a|)(|b|, b) \in S^1$, Euler's identity implies

$$-\frac{a}{|a|}\,\frac{|b|}{b} = e^{i\theta},$$

for some $\theta \in \mathbb{R}$, so we can take

$$\omega = e^{i\theta/k}.$$

Given (E.1.10),

$$p(z_0 + \varepsilon\omega) = a\Big(1 - \Big|\frac{b}{a}\Big|\varepsilon^k\Big) + O(\varepsilon^{k+1}), \tag{E.1.11}$$

which contradicts (E.1.5) for $\varepsilon > 0$ small enough. Thus (E.1.6) is impossible. This proves Lemma E.1.2, hence Theorem E.1.1. □

Now that we have shown that $p(z)$ in (E.1.1) must have one root, we can show that it has n roots (counting multiplicity).

Proposition E.1.3. *For a polynomial $p(z)$ of degree n as in* (E.1.1)*, there exist $r_1, \dots, r_n \in \mathbb{C}$ such that*

$$p(z) = a_n(z - r_1) \cdots (z - r_n). \tag{E.1.12}$$

Proof. We have shown that $p(z)$ has one root, call it r_1. Dividing $p(z)$ by $z - r_1$, we have

$$p(z) = (z - r_1)\tilde{p}(z) + q, \tag{E.1.13}$$

where $\tilde{p}(z) = az^{n-1} + \cdots + \tilde{a}_0$ and q is a polynomial of degree < 1, i.e., a constant. Setting $z = r_1$ in (E.1.13) yields $q = 0$, so

$$p(z) = (z - r_1)\tilde{p}(z). \tag{E.1.14}$$

Since $\tilde{p}(z)$ is apolynomial of degree $n - 1$, the result (E.1.12) follows by induction on n. □

The numbers r_j, $1 \le j \le n$ in (E.1.12) are called the roots of $p(z)$. If k of them coincide (say with r_ℓ) we say r_ℓ is a root of multiplicity k. If r_ℓ is distinct from r_j for all $j \neq \ell$, we say r_ℓ is a simple root.

E.2. Proof via Liouville's theorem

Here we use Liouville's theorem to give a second proof of the fundamental theorem of algebra. So take a polynomial

$$p(z) = a_n z^n + \cdots + a_1 z + a_0, \quad a_n \neq 0,\ n \geq 1. \tag{E.2.1}$$

We will continue to make use of (E.1.2), i.e.,

$$\lim_{|z|\to\infty} |p(z)| = \infty. \tag{E.2.2}$$

Now suppose

$$p(z) \neq 0, \quad \forall\, z \in \mathbb{C}. \tag{E.2.3}$$

Set

$$f(z) = \frac{1}{p(z)}, \quad z \in \mathbb{C}. \tag{E.2.4}$$

Then $f(z)$ is holomorphic on $\mathbb{C}$; one checks that

$$f'(z) = -\frac{p'(z)}{p(z)^2}. \tag{E.2.5}$$

However, (E.2.2) implies $|f(z)| \to 0$ as $|z| \to \infty$, hence f is bounded. Then Liouville's theorem implies f is constant. This is clearly not possible, so we have a contradiction to (E.2.3). This completes the second proof of Theorem E.1.1.

Bibliography

[1] J. Arndt and C. Haenel, π Unleashed, Springer-Verlag, New York, 2001.

[2] L. Baez-Duarte, Brouwer's fixed point theorem and a generalization of the formula for change of variables in multiple integrals, J. Math. Anal. Appl. 177 (1993), 412-414.

[3] R. Bartle and D. Sherbert, Introduction to Real Analysis, J. Wiley, New York, 1992.

[4] P. Beckmann, A History of π, St. Martin's Press, New York, 1971.

[5] P. Cohen, Set Theory and the Continuum Hypothesis, Dover, New York, 2008.

[6] H. Federer, Geometric Measure Theory, Springer, New York, 1969.

[7] G. Folland, Real Analysis: Modern Techniques and Applications, Wiley-Interscience, New York, 1984.

[8] J. Kitchen, Calculus of One Variable, Addison-Wesley, New York, 1968.

[9] S. Lang, Algebra, Addison-Wesley, Reading MA, 1965.

[10] S. Lang, Short Calculus, Springer-Verlag, New York, 2002.

[11] P. Lax, Change of variables in multiple integrals, Amer. Math. Monthly 106 (1999), 497–501.

[12] I. Niven, A simple proof that π is irrational, Bull. AMS 53 (1947), 509.

[13] K. Smith, Primer of Modern Analysis, Springer-Verlag, New York, 1983.

[14] M. Taylor, Measure Theory and Integration, American Mathematical Society, Providence RI, 2006.

[15] M. Taylor, Introduction to Analysis in One Variable, Undergraduate text #47, American Mathematical Society, Providence RI, 2020.

[16] M. Taylor, Introduction to Analysis in Several Variables (Advanced Calculus), Undergraduate text #46, American Mathematical Society, Providence RI, 2020.

[17] M. Taylor, Introduction to Complex Analysis, GSM #202, American Mathematical Society, Providence RI, 2019.

[18] M. Taylor, Partial Differential Equations, Vols. 1–3, Springer-Verlag, New York, 1996 (2nd ed., 2011).

[19] M. Taylor, Introduction to Differential Equations, American Mathematical Society, Providence RI, 2011.

[20] M. Taylor, Linear Algebra, Undergraduate text #45, American Mathematical Society, Providence RI, 2020.

[21] G. Thomas, Calculus and Analytic Geometry, 3rd ed., Addison-Wesley, Reading MA, 1961.

[22] O. Toeplitz, The Calculus, a Genetic Approach, Univ. of Chicago Press, 1963 (translated from German original).

Index