Appendix

1. The real and complex number systems

In this appendix we organize some of the mathematical prerequisites for reading this book. The reader must be thoroughly informed about basic real analysis (see [Ro] and [F1]) and should know a bit of complex variable theory (see [A] and [D2]).

The real number system $\mathbf{R}$ is characterized by being a *complete ordered field*. The field axioms enable the usual operations of addition, subtraction, multiplication, and division (except by 0). These operations satisfy familiar laws. The order axioms allow us to manipulate inequalities as usual. The completeness axiom is more subtle; this crucial property distinguishes $\mathbf{R}$ from the rational number system $\mathbf{Q}$. One standard way to state the completeness axiom uses the least upper bound property:

DEFINITION 6.1. If S is a non-empty subset of $\mathbf{R}$ and S is bounded above, then S has a least upper bound α, written $\sup(S)$, and called the *supremum* of S.

Recall that a sequence of real numbers is a function $n \mapsto x_n$ from the natural numbers to $\mathbf{R}$. (Sometimes we also allow the indexing to begin with 0.) The sequence $\{x_n\}$ *converges* to the real number L if, for all $\epsilon > 0$, there is an integer N_ϵ such that $n \geq N_\epsilon$ implies $|x_n - L| < \epsilon$.

The least upper bound property enables us to prove that a bounded monotone nondecreasing sequence $\{x_n\}$ of real numbers converges to the supremum of the values of the sequence. It also enables a proof of the fundamental result of basic real analysis: a sequence of real numbers converges if and only if it is a Cauchy sequence. Recall that a sequence is *Cauchy* if, for every $\epsilon > 0$, there is an N_ϵ such that $n, m \geq N_\epsilon$ implies $|x_n - x_m| < \epsilon$. Thus a sequence has a limit L if the terms are eventually as close to L as we wish, and a sequence is Cauchy if the terms are eventually all as close to each other as we wish. The equivalence of the concepts suggests that the real number system has no gaps.

For clarity we highlight these fundamental results as a theorem. The ability to prove Theorem 6.1 should be regarded as a prerequisite for reading this book.

THEOREM 6.1. *If a sequence $\{x_n\}$ of real numbers is bounded and monotone, then $\{x_n\}$ converges. A sequence $\{x_n\}$ converges to a real number L if and only if $\{x_n\}$ is Cauchy.*

COROLLARY 6.1. *A monotone sequence converges if and only if it is bounded.*

REMARK 6.1. The first statement in Theorem 6.1 is considerably easier than the second. It is possible to prove the difficult (*if*) part of the second statement by extracting a monotone subsequence and using the first part. It is also possible to prove the second statement by using the Bolzano-Weierstrass property from Theorem 6.2 below.

The complex number system **C** is a field, but it has no ordering. As a set **C** is simply the Euclidean plane $\mathbf{R}^2$. We make this set into a field by defining addition and multiplication:

$$(x, y) + (a, b) = (x + a, y + b)$$
$$(x, y) * (a, b) = (xa - yb, xb + ya).$$

The additive identity 0 is then the ordered pair $(0,0)$ and the multiplicative identity 1 is the pair $(1,0)$. Note that $(0,1)*(0,1) = (-1,0) = -(1,0)$. As usual we denote $(0,1)$ by i and then write $x + iy$ instead of (x,y). We then drop the $*$ from the notation for multiplication, and the law becomes obvious. Namely, we expand $(x+iy)(a+ib)$ by the distributive law and set $i^2 = -1$. These operations make $\mathbf{R}^2$ into a field called **C**.

Given $z = x + iy$ we write $\overline{z} = x - iy$ and call $\overline{z}$ the complex conjugate of z. We define $|z|$ to be the Euclidean distance of z to 0; thus $|z| = \sqrt{x^2 + y^2}$ and $|z|^2 = z\overline{z}$.

The non-negative real number $|z - w|$ equals the Euclidean distance between complex numbers z and w. The following properties of distance make **C** into a complete metric space. (See the next section.)

- $|z - w| = 0$ if and only if $z = w$.
- $|z - w| \geq 0$ for all z and w.
- $|z - w| = |w - z|$ for all z and w.
- $|z - w| \leq |z - \zeta| + |\zeta - w|$ for all z, w, ζ. (the triangle inequality)

Once we know that $|z - w|$ defines a distance, we can repeat the definition of convergence.

DEFINITION 6.2. Let $\{z_n\}$ be a sequence of complex numbers, and suppose $L \in \mathbf{C}$. We say that z_n converges to L if, for all $\epsilon > 0$, there is an N_ϵ such that $n \geq N_\epsilon$ implies $|z_n - L| < \epsilon$.

Let $\{a_n\}$ be a sequence of complex numbers. We say that $\sum_{n=1}^{\infty} a_n$ *converges to L, if*

$$\lim_{N \to \infty} \sum_{n=1}^{N} a_n = L.$$

We say that $\sum_{n=1}^{\infty} a_n$ *converges absolutely* if $\sum_{n=1}^{\infty} |a_n|$ converges. It is often easy to establish absolute convergence; a series of non-negative numbers converges if and only if the sequence of partial sums is bounded. The reason is simple: if the terms of a series are non-negative, then the partial sums form a monotone sequence, and hence the sequence of partial sums converges if and only if it is bounded. See Corollary 6.1 above. We also use the following standard comparison test; we include the proof because it beautifully illustrates the Cauchy convergence criterion.

PROPOSITION 6.1. *Let $\{z_n\}$ be a sequence of complex numbers. Assume for all n that $|z_n| \leq c_n$, and that $\sum_{n=1}^{\infty} c_n$ converges. Then $\sum_{n=1}^{\infty} z_n$ converges.*

PROOF. Let S_N denote the N-th partial sum of the series $\sum z_n$, and let T_N denote the N-th partial sum of the series $\sum c_n$. For $M > N$ we have

$$|S_M - S_N| = |\sum_{N+1}^{M} z_n| \leq \sum_{N+1}^{M} |z_n| \leq \sum_{N+1}^{M} c_n = T_M - T_N. \tag{1}$$

Since $\sum c_n$ is convergent, $\{T_N\}$ is a Cauchy sequence of real numbers. By (1), $\{S_N\}$ is also Cauchy, and hence $\sum_{n=1}^{\infty} z_n$ converges by Theorem 6.1.   □

We pause to recall and discuss the notion of equivalence class, which we presume is familiar to the reader. Let S be a set. An *equivalence relation* on S is a relation $\sim$ such that, for all $a, b, c \in S$,

Reflexive property: $a \sim a$
Symmetric property: $a \sim b$ if and only if $b \sim a$
Transitive property: $a \sim b$ and $b \sim c$ implies $a \sim c$.

Given an equivalence relation on a set S, we can form a new set, sometimes written $S/\sim$, as follows. We say that a and b are *equivalent*, or lie in the same *equivalence class*, if $a \sim b$ holds. The elements of $S/\sim$ are the equivalence classes; the set $S/\sim$ is called the quotient space.

We mention three examples. The first is trivial, the second is easy but fundamental, and the third is profound.

EXAMPLE 6.1. Let S be the set of ordered pairs (a, b) of integers. We say that $(a, b) \sim (c, d)$ if $100a + b = 100c + d$. If we regard the first element of the ordered pair as the number of dollars, and the second element as the number of cents, then two pairs are equivalent if they represent the same amount of money. (Note that we allow negative money here.)

EXAMPLE 6.2. Let S be the set of ordered pairs (a, b) of integers, with $b \neq 0$. We say that $(a, b) \sim (A, B)$ if $aB = Ab$. The equivalence relation restates, without mentioning division, the condition that $\frac{a}{b}$ and $\frac{A}{B}$ define the same rational number. Then $S/\sim$ is the set of rational numbers. It becomes the system $\mathbf{Q}$ after we define addition and multiplication of equivalence classes and verify the required properties.

EXAMPLE 6.3. The real number system $\mathbf{R}$ is sometimes defined to be the *completion* of the rational number system $\mathbf{Q}$. In this definition, a real number is an equivalence class of Cauchy sequences of rational numbers. Here we define a sequence of rational numbers $\{q_n\}$ to be Cauchy if, for each positive integer K, we can find a positive integer N such that $m, n \geq N$ implies $|q_m - q_n| < \frac{1}{K}$. (The number $\frac{1}{K}$ plays the role of ϵ; we cannot use ϵ because real numbers have not yet been defined!) Two Cauchy sequences are equivalent if their difference converges to 0. Thus Cauchy sequences $\{p_n\}$ and $\{q_n\}$ of rational numbers are equivalent if, for every $M \in \mathbf{N}$, there is an $N \in \mathbf{N}$ such that $|p_n - q_n| < \frac{1}{M}$ whenever $n \geq N$. Intuitively, we can regard a real number to be the collection of all sequences of rational numbers which appear to have the same *limit*. We use the language of the next section; as a set, $\mathbf{R}$ is the metric space *completion* of $\mathbf{Q}$. As in Example 6.2, we need to define addition, multiplication, and order and establish their properties before we get the real number system $\mathbf{R}$.

We are also interested in convergence issues in higher dimensions. Let $\mathbf{R}^n$ denote real Euclidean space of dimension n and $\mathbf{C}^n$ denote complex Euclidean space of dimension n. In the next paragraph, we let $\mathbf{F}$ denote either $\mathbf{R}$ or $\mathbf{C}$.

As a set, $\mathbf{F}^n$ consists of all n-tuples of elements of the field $\mathbf{F}$. We write $z = (z_1, \ldots, z_n)$ for a point in $\mathbf{F}^n$. This set has the structure of a real or complex vector space with the usual operations of vector addition and scalar multiplication:

$$(z_1, z_2, \ldots, z_n) + (w_1, w_2, \ldots, w_n) = (z_1 + w_1, z_2 + w_2, \ldots, z_n + w_n).$$

$$c(z_1, z_2, \ldots, z_n) = (cz_1, cz_2, \ldots, cz_n)$$

DEFINITION 6.3 (norm). A *norm* on a real or complex vector space V is a function $v \mapsto ||v||$ satisfying the following three properties:

(1) $||v|| > 0$ for all nonzero v.
(2) $||cv|| = |c| \, ||v||$ for all $c \in \mathbf{C}$ and all $v \in V$.
(3) (The triangle inequality) $||v + w|| \leq ||v|| + ||w||$ for all $v, w \in V$.

We naturally say *normed vector space* for a vector space equipped with a norm. We can make a normed vector space into a metric space by defining $d(u, v) = ||u - v||$.

For us the notations $\mathbf{R}^n$ and $\mathbf{C}^n$ include the vector space structure and the *Euclidean squared norm* defined by (2):

$$||z||^2 = \langle z, z \rangle. \tag{2}$$

These norms come from the *Euclidean inner product*. In the real case we have

$$\langle x, y \rangle = \sum_{j=1}^{n} x_j y_j \tag{3.1}$$

and in the complex case we have

$$\langle z, w \rangle = \sum_{j=1}^{n} z_j \overline{w}_j. \tag{3.2}$$

In both cases $||z||^2 = \langle z, z \rangle$.

2. Metric spaces

The definitions of convergent sequence in various settings are so similar that it is natural to put these settings into one abstract framework. One such setting is *metric spaces.*

We assume that the reader is somewhat familiar with metric spaces. We recall the definition and some basic facts. Let $\mathbf{R}_+$ denote the non-negative real numbers.

DEFINITION 6.4. Let X be a set. A *distance* function on X is a function $d : X \times X \to \mathbf{R}_+$ satisfying the following properties:

(1) $d(x, y) = 0$ if and only if $x = y$.
(2) $d(x, y) = d(y, x)$ for all x, y.
(3) $d(x, z) \leq d(x, y) + d(y, z)$ for all x, y, z.

If d is a distance function on X, then the pair (X, d) is called a *metric space* and d is called the *metric.*

The real numbers, the complex numbers, real Euclidean space, and complex Euclidean space are all metric spaces under the usual Euclidean distance function. One can define other metrics, with very different properties, on these sets. For example, on any set X, the function $d : X \times X \to \mathbf{R}_+$, defined by $d(x, y) = 1$ if $x \neq y$ and $d(x, x) = 0$, is a metric. In general sets admit many different useful distance functions. When the metric is understood, one often says "Let X be a metric space". This statement is convenient but a bit imprecise.

Metric spaces provide a nice conceptual framework for convergence.

DEFINITION 6.5. Let $\{x_n\}$ be a sequence in a metric space (X, d). We say that x_n *converges* to x if, for all $\epsilon > 0$, there is an N such that $n \geq N$ implies $d(x_n, x) < \epsilon$. We say that $\{x_n\}$ is *Cauchy* if, for all $\epsilon > 0$, there is an N such that $m, n \geq N$ implies $d(x_m, x_n) < \epsilon$.

DEFINITION 6.6. A metric space (M, d) is *complete* if every Cauchy sequence converges.

If a metric space (M, d) is not complete, then we can form a new metric space called its *completion*. The idea precisely parallels the construction of $\mathbf{R}$ given $\mathbf{Q}$. The completion consists of equivalence classes of Cauchy sequences of elements of (M, d). The distance function extends to the larger set by taking limits.

Here are several additional examples of metric spaces. We omit the needed verifications of the properties of the distance function, but we mention that in some instances proving the triangle inequality requires effort.

EXAMPLE 6.4. Let X be the space of continuous functions on $[0, 1]$. Define $d(f, g) = \int_0^1 |f(x) - g(x)| dx$. Then (X, d) is a metric space. More generally, for $1 \leq p < \infty$, we define $d_p(f, g)$ by

$$d_p(f, g) = \left(\int_0^1 |f(x) - g(x)|^p dx \right)^{\frac{1}{p}}.$$

We define $d_\infty(f, g)$ by $d_\infty(f, g) = \sup |f - g|$.

Of these examples, only (X, d_∞) is complete. Completeness in this case follows because the uniform limit of a sequence of continuous functions is itself continuous.

A subset Ω of a metric space is called *open* if, whenever $p \in \Omega$, there is a positive ϵ such that $x \in \Omega$ whenever $d(p, x) < \epsilon$. In particular the empty set is open and the whole space X is open. A subset K is called *closed* if its complement is open.

PROPOSITION 6.2. *Let (X, d) be a metric space. Let $K \subseteq X$. Then K is closed if and only if, whenever $\{x_n\}$ is a sequence in K, and x_n converges to x, then $x \in K$.*

PROOF. Left to the reader. □

Let (M, d) and (M', d') be metric spaces. The natural collection of maps between them is the set of continuous functions.

DEFINITION 6.7 (Continuity). $f : (M, d) \to (M', d')$ is *continuous* if, whenever U is open in M', then $f^{-1}(U)$ is open in M.

PROPOSITION 6.3. *Suppose $f : (M, d) \to (M', d')$ is a map between metric spaces. The following are equivalent:*

(1) *f is continuous*
(2) *Whenever x_n converges to x in M, then $f(x_n)$ converges to $f(x)$ in M'.*
(3) *For all $\epsilon > 0$, there is a $\delta > 0$ such that*

$$d(x, y) < \delta \implies d'(f(x), f(y)) < \epsilon.$$

EXERCISE 6.1. Prove Propositions 6.1 and 6.2.

We next mention several standard and intuitive geometric terms. The *interior* of a set S in a metric space is the union of all open sets contained in S. The *closure*

of a set S is the intersection of all closed sets containing S. Thus a set is open if and only if it equals its interior, and a set is closed if and only if it equals its closure. The *boundary* $b\Omega$ of a set Ω consists of all points in the closure of Ω but not in the interior of Ω. Another way to define boundary is to note that $x \in b\Omega$ if and only if, for every $\epsilon > 0$, the ball of radius ϵ about x has a non-empty intersection with both Ω and its complement.

Continuity often gets used together with the notion of a *dense* subset of a metric space M. A subset S is *dense* if each $x \in M$ is the limit of a sequence of points in S. In other words, M is the closure of S. For example, the rational numbers are dense in the real numbers. If f is continuous on M, then $f(x) = \lim_n f(x_n)$, and hence f is determined by its values on a dense set.

One of the most important examples of a metric space is the collection $C(M)$ of continuous complex-valued functions on a metric space M. Several times in the book we use compactness properties in $C(M)$. We define compactness in the standard *open cover* fashion, called the Heine-Borel property. What matters most for us is the Bolzano-Weierstrass property.

We quickly review some of the most beautiful results in basic analysis.

DEFINITION 6.8. Let M be a metric space and let $K \subseteq M$. K is *compact* if, whenever K is contained in an arbitrary union $\cup A_\alpha$ of open sets, then K is contained in a finite union $\cup_{k=1}^N A_{\alpha_k}$ of these open sets. This condition is often called the *Heine-Borel property*.

This definition of compact is often stated informally "every open cover has a finite subcover", but these words are a bit imprecise.

DEFINITION 6.9. Let (M, d) be a metric space. A subset $K \subseteq M$ satisfies the *Bolzano-Weierstrass property* if, whenever $\{x_n\}$ is a sequence in K, then there is a subsequence $\{x_{n_k}\}$ converging to a limit in K.

THEOREM 6.2. *Let (M, d) be a metric space and let $K \subseteq M$. Then K is compact if and only if K satisfies the Bolzano-Weierstrass property.*

THEOREM 6.3. *A subset of Euclidean space is compact if and only if it is closed and bounded.*

EXERCISE 6.2. Prove Theorems 6.2 and 6.3.

DEFINITION 6.10 (Equicontinuity). A collection $\mathcal{K}$ of complex-valued functions on a metric space (M, d) is called *equicontinuous* if, for all x and for all $\epsilon > 0$, there is a $\delta > 0$ such that

$$d(x, y) < \delta \implies |f(x) - f(y)| < \epsilon$$

for all $f \in \mathcal{K}$.

DEFINITION 6.11 (Uniformly bounded). A collection $\mathcal{K}$ of complex-valued functions on a metric space (M, d) is called *uniformly bounded* if there is a C such that $|f(x)| \leq C$ for all $x \in M$ and for all $f \in \mathcal{K}$.

We refer to [F1] for a proof of the following major result in analysis. The statement and proof in [F1] apply in the more general context of locally compact Hausdorff topological spaces. In this book we use Theorem 6.4 to show that certain integral operators are compact. See Sections 10 and 11 of Chapter 2.

THEOREM 6.4 (Arzela-Ascoli theorem). *Let M be a compact metric space. Let $C(M)$ denote the continuous functions on M with $d(f,g) = \sup_M |f(x) - g(x)|$. Let $\mathcal{K}$ be a subset of $C(M)$. Then $\mathcal{K}$ is compact if and only if the following three items are true:*

(1) *$\mathcal{K}$ is equicontinuous.*
(2) *$\mathcal{K}$ is uniformly bounded.*
(3) *$\mathcal{K}$ is closed.*

COROLLARY 6.2. *Let $\mathcal{K}$ be a closed, uniformly bounded, and equicontinuous subset of $C(M)$. Let $\{f_n\}$ be a sequence in $\mathcal{K}$. Then $\{f_n\}$ has a convergent subsequence. That is, $\{f_{n_k}\}$ converges uniformly to an element of $\mathcal{K}$.*

PROOF. By the theorem $\mathcal{K}$ is compact; the result then follows from the Bolzano-Weierstrass characterization of compactness. $\qquad\square$

EXERCISE 6.3. Let M be a compact subset of Euclidean space. Fix $\alpha > 0$. Let H_α denote the subset of $C(M)$ satisfying the following properties:

(1) $||f||_\infty \leq 1$.
(2) $||f||_{H_\alpha} \leq 1$. Here

$$||f||_{H_\alpha} = \sup_{x \neq y} \frac{|f(x) - f(y)|}{|x - y|^\alpha}.$$

Show that H_α is compact.

A function f for which $||f||_{H_\alpha}$ is finite is said to satisfy a Hölder condition of order α. See Definition 2.13.

3. Integrals

This book presumes that the reader knows the basic theory of the Riemann-Darboux integral, which we summarize. See [Ro] among many possible texts.

Let $[a, b]$ be a closed bounded interval on $\mathbf{R}$, and suppose $f : [a, b] \to \mathbf{R}$ is a bounded function. We define $\int_a^b f(t)dt$ by a standard but somewhat complicated procedure. A partition P of $[a, b]$ is a finite collection of points p_j such that $a = p_0 < \cdots < p_j < \cdots < p_N = b$. Given f and a partition P, we define the lower and upper sums corresponding to the partition:

$$L(f, P) = \sum_{j=1}^{N} (p_j - p_{j-1}) \inf_{[p_{j-1}, p_j]} (f(x))$$

$$U(f, P) = \sum_{j=1}^{N} (p_j - p_{j-1}) \sup_{[p_{j-1}, p_j]} (f(x)).$$

DEFINITION 6.12. A bounded function $f : [a, b] \to \mathbf{R}$ is *Riemann integrable* if $\sup_P L(f, P) = \inf_P U(f, P)$. If so, we denote the common value by $\int_a^b f(t)dt$ or simply by $\int_a^b f$.

An equivalent way to state Definition 6.12 is that f is integrable if, for each $\epsilon > 0$, there is a partition P_ϵ such that $U(f, P_\epsilon) - L(f, P_\epsilon) < \epsilon$.

In case f is complex-valued, we define it to be integrable if its real and imaginary parts are integrable, and we put

$$\int_a^b f = \int_a^b u + iv = \int_a^b u + i \int_a^b v.$$

The integral satisfies the usual properties:

(1) If f, g are Riemann integrable on $[a, b]$, and c is a constant, then $f + g$ and cf are Riemann integrable and

$$\int_a^b f + g = \int_a^b f + \int_a^b g,$$

$$\int_a^b cf = c \int_a^b f.$$

(2) If f is Riemann integrable and $f(x) \geq 0$ for $x \in [a, b]$, then $\int_a^b f \geq 0$.
(3) If f is continuous on $[a, b]$, then f is Riemann integrable.
(4) If f is monotone on $[a, b]$, then f is Riemann integrable.

We assume various other basic results, such as the change of variables formula, without further mention.

The collection of complex-valued integrable functions on $[a, b]$ is a complex vector space. We would like to define the distance $\delta(f, g)$ between integrable functions f and g by

$$\delta(f, g) = ||f - g||_{L^1} = \int_a^b |f(x) - g(x)| dx,$$

but a slight problem arises. If f and g agree for example everywhere except at a single point, and each is integrable, then $\delta(f, g) = 0$ but f and g are not the same function. This point is resolved by working with equivalence classes of functions. Two functions are called equivalent if they agree except on what is called a set of measure zero. See Section 7 of Chapter 1. Even after working with equivalence classes, this vector space is not complete (in the metric space sense). One needs to use the Lebesgue integral to identify its completion.

Often one requires so-called *improper* integrals. Two possible situations arise; one is when f is unbounded on $[a, b]$, the other is when the interval is unbounded. Both situations can happen in the same example. The definitions are clear, and we state them informally. If f is unbounded at a, for example, but Riemann integrable on $[a + \epsilon, b]$ for all positive ϵ, then we define

$$\int_a^b f = \lim_{\epsilon \to 0} \int_{a+\epsilon}^b f$$

if the limit exists. If f is Riemann integrable on $[a, b]$ for all b, then we put

$$\int_a^\infty f = \lim_{b \to \infty} \int_a^b f.$$

The other possibilities are handled in a similar fashion. Here are two simple examples of improper integrals:

(1) $\int_0^1 x^\alpha dx = \frac{1}{\alpha+1}$ if $\alpha > -1$.
(2) $\int_0^\infty e^{-x} dx = 1$.

At several points in this book, whether an improper integral converges will be significant. We mention specifically Section 8 of Chapter 3, where one shows that a function has k continuous derivatives by showing that an improper integral is convergent.

The following theorem is *fundamental* to all that we do in this book.

THEOREM 6.5 (Fundamental theorem of calculus). *Assume f is continuous on $[a, b]$. For $x \in (a, b)$ put $F(x) = \int_a^x f(t)dt$. Then F is differentiable and $F'(x) = f(x)$.*

The final theorem in this section is somewhat more advanced. We state this result in Section 7 of Chapter 1, but we never use it. It is important partly because its statement is so definitive, and partly because it suggests connections between the Riemann and Lebesgue theories of integration.

THEOREM 6.6. *A function on a closed interval $[a, b]$ is Riemann integrable if and only if the set of its discontinuities has measure zero.*

EXERCISE 6.4. Establish the above properties of the Riemann integral.

EXERCISE 6.5. Verify that $\int_a^b cf = c \int_a^b f$ when c is complex and f is complex-valued. Check that $\text{Re}(\int_a^b f) = \int_a^b \text{Re}(f)$ and similarly with the imaginary part.

EXERCISE 6.6. Verify the improper integrals above.

The next three exercises involve finding sums. Doing so is generally much harder than finding integrals.

EXERCISE 6.7. Show that $\sum_{j=0}^n \binom{j}{k} = \binom{n+1}{k+1}$. Suggestion. Count the same thing in two ways.

EXERCISE 6.8. For p a nonnegative integer, consider $\sum_{j=1}^n j^p$ as a function of n. Show that it is a polynomial in n of degree $p + 1$ with leading term $\frac{n^{p+1}}{p+1}$. If you want to work harder, show that the next term is $\frac{n^p}{2}$. Comment: The previous exercise is useful in both cases.

EXERCISE 6.9. For p a positive integer, prove that $\int_0^1 t^p dt = \frac{1}{p+1}$ by using the definition of the Riemann integral. (Find upper and lower sums and use the previous exercise.)

EXERCISE 6.10. Prove the fundamental theorem of calculus. The idea of its proof recurs throughout this book.

EXERCISE 6.11. Put $f(0) = 0$ and $f(x) = x \, \text{sgn}(\sin(\frac{1}{x}))$. Here $\text{sgn}(t) = \frac{t}{|t|}$ for $t \neq 0$ and $\text{sgn}(0) = 0$.

- Sketch the graph of f.
- Determine the points where f fails to be continuous.
- Show that f is Riemann integrable on $[-1, 1]$.

4. Exponentials and trig functions

The unit circle is the set of complex numbers of unit Euclidean distance from 0, that is, the set of z with $|z| = 1$.

The complex exponential function is defined by

$$e^z = \sum_{n=0}^{\infty} \frac{z^n}{n!}.$$

The series converges absolutely for all complex z. Furthermore the resulting function satisfies $e^0 = 1$ and $e^{z+w} = e^z e^w$ for all z and w.

We define the complex trig functions by

$$\cos(z) = \frac{e^{iz} + e^{-iz}}{2}$$

$$\sin(z) = \frac{e^{iz} - e^{-iz}}{2i}.$$

When z is real these functions agree with the usual trig functions. The reader who needs convincing can express both sides as power series.

Note, by continuity of complex conjugation, we have $e^{\overline{z}} = \overline{e^z}$. Combining this property with the addition law gives (assuming t is real)

$$1 = e^0 = e^{it} e^{-it} = |e^{it}|^2.$$

Thus $z = e^{it}$ lies on the unit circle. Its real part x is given by $x = \frac{z+\overline{z}}{2}$ and its imaginary part y is given by $y = \frac{z-\overline{z}}{2i}$. Comparing with our definitions of cosine and sine, we obtain the famous Euler identity (which holds even when t is complex):

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

Complex logarithms are quite subtle. For a positive real number t we define $\log(t)$, sometimes written $\ln(t)$, by the usual formula

$$\log(t) = \int_1^t \frac{du}{u}.$$

For a nonzero complex number z, written in the form $z = |z|e^{i\theta}$, we provisionally define its logarithm by

$$\log(z) = \log(|z|) + i\theta. \tag{4}$$

The problem with this formula is that θ is defined only up to multiples of 2π. We must therefore restrict θ to an interval of length 2π. In order to define the logarithm precisely, we must choose a *branch cut*. Thus we first choose an open interval of length 2π, and then we define the logarithm only for θ in that open interval. Doing so yields a *branch* of the logarithm. For example, we often write (4) for $0 \neq z = |z|e^{i\theta}$ and $-\pi < \theta < \pi$. Combining the identity $e^{\alpha+\beta} = e^\alpha e^\beta$ with (4), we obtain $e^{\log(z)} = |z|e^{i\theta} = z$. For a second example, suppose our branch cut is the non-negative real axis; then $0 < \theta < 2\pi$. Then $\log(-1) = i\pi$, but logs of positive real numbers are not defined! To correct this difficulty, we could assume $0 \leq \theta < 2\pi$ and obtain the usual logarithm of a positive number. The logarithm, as a function on the complement of the origin in $\mathbf{C}$, is then discontinuous at points on the positive real axis.

5. Complex analytic functions

The geometric series arises throughout mathematics. Suppose that z is a complex number not equal to 1. Then we have the *finite geometric series*

$$\sum_{j=0}^{n-1} z^j = \frac{1-z^n}{1-z}.$$

When $|z| < 1$, we let $n \to \infty$ and obtain the *geometric series*

$$\sum_{j=0}^{\infty} z^j = \frac{1}{1-z}.$$

The geometric series and the exponential series lie at the foundation of complex analysis. We have seen how the exponential function informs trigonometry. The geometric series enables the proof of Theorem 6.7 below; the famous Cauchy integral formula (Theorem 6.8) combines with the geometric series to show that an arbitrary complex analytic function has a local power series expansion.

A subset Ω of $\mathbf{C}$ is called *open* if, for all $p \in \Omega$, there is an open ball about p contained in Ω. In other words, there is a positive ϵ such that $|z - p| < \epsilon$ implies $z \in \Omega$. Suppose that Ω is open and $f : \Omega \to \mathbf{C}$ is a function. We say that f is *complex analytic* on Ω if, for each $z \in \Omega$, f is complex differentiable at z. (in other words, if the limit in (5) exists).

$$\lim_{h \to 0} \frac{f(z+h) - f(z)}{h} = f'(z) \tag{5}$$

A continuously differentiable function $f : \Omega \to \mathbf{C}$ satisfies the *Cauchy-Riemann equations* if $\frac{\partial f}{\partial \bar{z}} = 0$ at all points of Ω. The complex partial derivative is defined by

$$\frac{\partial}{\partial \bar{z}} = \frac{1}{2}\left(\frac{\partial}{\partial x} + i\frac{\partial}{\partial y}\right).$$

In most elementary books on complex variables, one writes $f = u + iv$ in terms of its real and imaginary parts, and writes the Cauchy-Riemann equations as the pair of equations

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}$$

$$\frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}.$$

Perhaps the most fundamental theorem in basic complex analysis relates complex analytic functions, convergent power series, and the Cauchy-Riemann equations. Here is the precise statement:

THEOREM 6.7. *Assume that Ω is open and $f : \Omega \to \mathbf{C}$ is a function. The following are equivalent:*

(1) *f is complex analytic on Ω.*
(2) *For all p in Ω, there is a ball about p on which f is given by a convergent power series:*

$$f(z) = \sum_{n=0}^{\infty} a_n(z-p)^n.$$

(3) *f is continuously differentiable and $\frac{\partial f}{\partial \bar{z}} = 0$ on Ω.*

The key step used in establishing Theorem 6.7 is the Cauchy integral formula. Readers unfamiliar with complex line integrals should consult [A] or [D2], and should read about Green's theorem in Section 1 of Chapter 4 in this book. We mention that, in the research literature on several complex variables, the word *holomorphic* is commonly used instead of *complex analytic*.

THEOREM 6.8 (Cauchy integral theorem and Cauchy integral formula). *Let f be complex analytic on and inside a positively oriented, simple closed curve γ. Then*

$$\int_\gamma f(z)dz = 0.$$

For z in the interior of γ, we have

$$f(z) = \frac{1}{2\pi i} \int_\gamma \frac{f(\zeta)}{\zeta - z}d\zeta.$$

We close this review of complex variable theory by recalling the Fundamental Theorem of Algebra. Many proofs are known, but all of them require the methods of analysis. No purely algebraic proof can exist, because the completeness axiom for the real numbers must be used in the proof.

THEOREM 6.9 (Fundamental theorem of algebra). *Let $p(z)$ be a non-constant polynomial with complex coefficients and of degree d. Then p factors into a product of d linear factors:*

$$p(z) = c\prod_{j=1}^{d}(z - z_j),$$

where the z_j need not be distinct.

6. Probability

Many of the ideas in this book are closely connected with probability theory. We barely glimpse these connections.

We begin by briefly discussing probability densities, and we restrict our consideration to continuous densities. See a good text such as [HPS] for more information and the relationship with Fourier transforms.

Let J be a closed interval on $\mathbf{R}$; we allow the possibility of infinite endpoints. Assume that $f : J \to [0, \infty)$ is continuous. Then f is called a *continuous probability density* on J if $\int_J f = 1$. Let a denote the left-hand endpoint of J. We define the *cumulative distribution* function F by

$$F(x) = \int_a^x f(t)dt.$$

For $y < x$, we interpret $F(x) - F(y) = \int_y^x f(t)dt$ as the probability that a *random variable* lies in the interval $[x, y]$.

We do not attempt to say precisely what the phrase "Let X be a random variable" means. In our setting, we are given the continuous density function f, and we say "X is a random variable with continuous density f" to indicate the situation we have described. The intuition for the term random variable X is the following. Suppose X is a real-valued function defined on some set, and for each $x \in \mathbf{R}$, the probability that X takes on a value at most x is well-defined. We

write $F(x)$ for this probability. Thus $F(x) - F(y)$ denotes the probability that F takes on a value in the interval $(y, x]$. In the case of continuous densities, the probability that X takes on any specific value is 0. This property is sometimes taken as the *definition* of continuous random variable. Hence $F(x) - F(y)$ denotes the probability that X takes on a value in the interval $[y, x]$.

Let X denote a random variable on an interval J, with continuous density f. We say that X has *finite expectation* if

$$\int_J |t| f(t) dt < \infty.$$

We say that X has *finite variance* if

$$\int_J (t - \mu)^2 f(t) dt < \infty.$$

When these integrals are finite, we define the *mean* μ and *variance* σ^2 of X by

$$\mu = \int_J t f(t) dt$$

$$\sigma^2 = \int_J (t - \mu)^2 f(t) dt.$$

The mean is also known as the expected value. More generally, if g is any function we call $\int_J g(t) f(t) dt$ the expected value of g. Thus the variance is the expected value of $(t - \mu)^2$ and hence measures the deviation from the mean.

PROPOSITION 6.4. *The variance satisfies* $\sigma^2 = \int_J t^2 f(t) dt - \mu^2$.

PROOF. Expanding the square in the definition of the variance gives:

$$\int_J (t - \mu)^2 f(t) dt = \int_J t^2 f(t) dt - 2\mu \int_J t f(t) dt + \mu^2 \int_J f(t) dt.$$

Since $\mu = \int_J t f(t) dt$ and $1 = \int_J f(t) dt$, the last two terms combine to give $-\mu^2$. □

The computation in Proposition 6.4 arises in many contexts. It appears, for example, in the proof of the parallel axis theorem for moments of inertia. The same idea occurs in verifying the equivalence of two ways of stating Poincaré inequalities in Chapter 4. Compare also with the proof of Bessel's inequality, Proposition 2.2.

EXAMPLE 6.5 (The normal, or Gaussian, random variable). For $0 < \sigma^2 < \infty$ and $x \in \mathbf{R}$, put $g(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-x^2}{2\sigma^2}}$. See Example 1.7. Then the mean of the random variable with density g is 0 and the variance is σ^2.

EXAMPLE 6.6 (The uniform random variable). Let $f(x) = \frac{1}{b-a}$ for $a \leq x \leq b$. Then f is a probability density. Its cumulative distribution function F is given on $\mathbf{R}$ by $F(x) = 0$ if $x < a$, by $F(x) = 1$ if $x > b$, and by $F(x) = \frac{x-a}{b-a}$ for $x \in [a, b]$.

EXERCISE 6.12. Show that the mean of the uniform random variable on $[a, b]$ is $\frac{a+b}{2}$. Compute its variance.

Let X be a random variable with continuous density function f. The probability that $X \leq x$ is by definition the integral $\int_{-\infty}^x f(t) dt$. We write:

$$\text{Prob}(X \leq x) = \int_{-\infty}^x f(t) dt.$$

Let ϕ be a strictly monotone differentiable function of one real variable. We can use the fundamental theorem of calculus to compute the density of $\phi(X)$. Assuming that ϕ is increasing, we have

$$\text{Prob}(\phi(X) \leq x) = \text{Prob}(X \leq \phi^{-1}(x)) = \int_{-\infty}^{\phi^{-1}(x)} f(t)dt.$$

Differentiating and using the fundamental theorem of calculus, we see that the density of $\phi(X)$ is given by $f \circ \phi^{-1}(\phi^{-1})'$. An example of this situation gets briefly mentioned in Exercise 4.68, where X is the Gaussian and $\phi(x) = x^2$ for $x \geq 0$. In case ϕ is decreasing a similar calculation gives the answer $-f \circ \phi^{-1}(\phi^{-1})'$. Hence the answer in general is $f \circ \phi^{-1}|(\phi^{-1})'|$.

We end this appendix by glimpsing the connection between the Fourier transform and probability. Given a continuous random variable on $\mathbf{R}$ with density f, we defined above the expected value of a function g by $\int_{-\infty}^{\infty} g(t)f(t)dt$. Take $g(t) = \frac{1}{\sqrt{2\pi}}e^{-it\xi}$. Then the expected value of g is the Fourier transform of f. The terminology used in probability theory often differs from that in other branches of mathematics; for example, the expected value of e^{itX}, where X is a random variable, equals $\int_{-\infty}^{\infty} e^{itx}f(x)dx$. This function is called the *characteristic function* of the random variable X rather than (a constant times the inverse of) the Fourier transform of f.

The *central limit theorem* is one of the major results in probability theory and statistics. Most readers should have heard of the result, at least in an imprecise fashion ("everything is asymptotically normal"), and we do not state it here. See [F1] or [HPS] for precise statements of the central limit theorem. Its proof relies on several things discussed in this book: the Fourier transform is injective on an appropriate space, the Fourier transform of a Gaussian of mean zero and variance one is itself, and the Gaussian defines an approximate identity as the variance tends to 0.

EXERCISE 6.13. Show that there is a continuous probability density f on $\mathbf{R}$, with finite expectation, such that $f(n) = n$ for all positive integers n.