Appendix A:
Review of Set Theory

In this book, as in most modern mathematics, mathematical statements are couched in the language of set theory. We give here a brief descriptive summary of the parts of set theory that we use, in the form that is commonly called "naive set theory." The word naive should be understood in the same sense in which it is used by Paul Halmos in his classic text *Naive Set Theory* [Hal74]: the assumptions of set theory are to be viewed much as Euclid viewed his geometric axioms, as intuitively clear statements of fact from which reliable conclusions can be drawn.

Our description of set theory is based on the axioms of Zermelo–Fraenkel set theory together with the axiom of choice (commonly known as ZFC), augmented with a notion of *classes* (aggregations that are too large to be considered sets in ZFC), primarily for use in category theory. We do not give a formal axiomatic treatment of the theory; instead, we simply give the definitions and list the basic types of sets whose existence is guaranteed by the axioms. For more details on the subject, consult any good book on set theory, such as [Dev93, Hal74, Mon69, Sup72, Sto79]. We leave it to the set theorists to explore the deep consequences of the axioms and the relationships among different axiom systems.

Basic Concepts

A ***set*** is just a collection of objects, considered as a whole. The objects that make up the set are called its ***elements*** or its ***members***. For our purposes, the elements of sets are always "mathematical objects": integers, real numbers, complex numbers, and objects built up from them such as ordered pairs, ordered n-tuples, functions, sequences, other sets, and so on. The notation $x \in X$ means that the object x is an element of the set X. The words ***collection*** and ***family*** are synonyms for set.

Technically speaking, *set* and *element of a set* are primitive undefined terms in set theory. Instead of giving a general definition of what it means to be a set, or for an object to be an element of a set, mathematicians characterize each particular set by giving a precise definition of what it means for an object to be an element of *that*

J.M. Lee, *Introduction to Topological Manifolds*, Graduate Texts in Mathematics 202,
DOI 10.1007/978-1-4419-7940-7,

set—what might be called the set's ***membership criterion***. For example, if $\mathbb{Q}$ is the set of all rational numbers, then the membership criterion for $\mathbb{Q}$ could be expressed as follows:

$$x \in \mathbb{Q} \quad \Leftrightarrow \quad x = p/q \text{ for some integers } p \text{ and } q \text{ with } q \neq 0.$$

The essential characteristic of sets is that they are determined by their elements. Thus if X and Y are sets, to say that ***X and Y are equal*** is to say that every element of X is an element of Y, and every element of Y is an element of X. Symbolically,

$$X = Y \qquad \text{if and only if} \qquad \text{for all } x,\ x \in X \Leftrightarrow x \in Y.$$

If X and Y are sets such that every element of X is also an element of Y, then X is a ***subset of Y***, written $X \subseteq Y$. Thus

$$X \subseteq Y \qquad \text{if and only if} \qquad \text{for all } x,\ x \in X \Rightarrow x \in Y.$$

The notation $Y \supseteq X$ ("Y is a ***superset of X***") means the same as $X \subseteq Y$. It follows from the definitions that $X = Y$ if and only if $X \subseteq Y$ and $X \supseteq Y$.

If $X \subseteq Y$ but $X \neq Y$, we say that X is a ***proper subset of Y*** (or Y is a ***proper superset of X***). Some authors use the notations $X \subset Y$ and $Y \supset X$ to mean that X is a proper subset of Y; however, since other authors use the symbol "$\subset$" to mean any subset, not necessarily proper, we generally avoid using this notation, and instead say explicitly when a subset is proper.

Here are the basic types of sets whose existence is guaranteed by ZFC. In each case, the set is completely determined by its membership criterion.

- THE EMPTY SET: There is a set containing no elements, called the ***empty set*** and denoted by $\varnothing$. It is unique, because any two sets with no elements are equal by our definition of set equality, so we are justified in calling it *the* empty set.
- SETS DEFINED BY LISTS: Given any list of objects that can be explicitly named, there is a set containing those objects and no others. It is denoted by listing the objects between braces: $\{\dots\}$. For example, the set $\{0,1,2\}$ contains only the numbers 0, 1, and 2. (For now, we are defining this notation only when the objects can all be written out explicitly; a bit later, we will give a precise definition of notations such as $\{x_1,\dots,x_n\}$, in which the objects are defined implicitly with ellipses.) A set containing exactly one element is called a ***singleton***.
- SETS DEFINED BY SPECIFICATION: Given a set X and a sentence $P(x)$ that is either true or false whenever x is any particular element of X, there is a set whose elements are precisely those $x \in X$ for which $P(x)$ is true, denoted by $\{x \in X : P(x)\}$.
- UNIONS: Given any collection $\mathcal{C}$ of sets, there is a set called their ***union***, denoted by $\bigcup \mathcal{C}$, with the property that $x \in \bigcup \mathcal{C}$ if and only if $x \in X$ for some $X \in \mathcal{C}$. Other notations for unions are

$$\bigcup_{X \in \mathcal{C}} X, \qquad X_1 \cup X_2 \cup \cdots.$$

- INTERSECTIONS: Given any nonempty collection $\mathcal{C}$ of sets, there is a set called their ***intersection***, denoted by $\bigcap \mathcal{C}$, with the property that $x \in \bigcap \mathcal{C}$ if and only if $x \in X$ for every $X \in \mathcal{C}$. Other notations for intersections are

$$\bigcap_{X \in \mathcal{C}} X, \qquad X_1 \cap X_2 \cap \cdots.$$

- SET DIFFERENCES: If X and Y are sets, their ***difference***, denoted by $X \smallsetminus Y$, is the set of all elements in X that are not in Y, so $x \in X \smallsetminus Y$ if and only if $x \in X$ and $x \notin Y$. If $Y \subseteq X$, the set difference $X \smallsetminus Y$ is also called the ***complement of Y in X***.
- POWER SETS: Given any set X, there is a set $\mathcal{P}(X)$, called the ***power set of X***, whose elements are exactly the subsets of X. Thus $S \in \mathcal{P}(X)$ if and only if $S \subseteq X$.

► **Exercise A.1.** Suppose A is a set and $\mathcal{C}$ is a collection of sets. Prove the following properties of unions and intersections.

(a) DISTRIBUTIVE LAWS:

$$A \cup \left(\bigcap_{X \in \mathcal{C}} X \right) = \bigcap_{X \in \mathcal{C}} (A \cup X);$$

$$A \cap \left(\bigcup_{X \in \mathcal{C}} X \right) = \bigcup_{X \in \mathcal{C}} (A \cap X).$$

(b) DE MORGAN'S LAWS:

$$A \smallsetminus \left(\bigcap_{X \in \mathcal{C}} X \right) = \bigcup_{X \in \mathcal{C}} (A \smallsetminus X);$$

$$A \smallsetminus \left(\bigcup_{X \in \mathcal{C}} X \right) = \bigcap_{X \in \mathcal{C}} (A \smallsetminus X).$$

Note that one must be careful to start with a specific set before one can define a new set by specification. This requirement rules out the possibility of forming sets out of self-contradictory specifications such as the one discovered by Bertrand Russell and now known as "Russell's paradox": the sentence $\mathcal{C} = \{X : X \notin X\}$ looks as if it might define a set, but it does not, because each of the statements $\mathcal{C} \in \mathcal{C}$ and $\mathcal{C} \notin \mathcal{C}$ implies its own negation. Similarly, there does not exist a "set of all sets," for if there were such a set $\mathcal{S}$, we could define a set $\mathcal{C} = \{X \in \mathcal{S} : X \notin X\}$ by specification and reach the same contradiction.

There are times when we need to speak of "all sets" or other similar aggregations, primarily in the context of category theory (see Chapter 7). For this purpose, we reserve the word ***class*** to refer to any well-defined assemblage of mathematical objects that might or might not constitute a set. We treat classes informally, but there are various ways they can be axiomatized. (One such is the extension of ZFC due to von Neumann, Bernays, and Gödel, known as NBG set theory; see [Men10].) For example, we can speak of the class of all sets or the class of all vector spaces. Every set is a class, but not every class is a set. A class that is not a set is called a ***proper class***. If $\mathcal{C}$ is a class and x is a mathematical object, we use the terminology "x is an element of $\mathcal{C}$" and the notation $x \in \mathcal{C}$ to mean that x is one of the objects in $\mathcal{C}$, just as we do for sets. The main restriction on using classes is that a proper class cannot be an element of any set or class; this ensures that it is impossible to form the equivalent of Russell's paradox with classes instead of sets.

Cartesian Products, Relations, and Functions

Another primitive concept that we use without a formal definition is that of an ***ordered pair***. Think of it as a pair of objects (which could be the same or different), together with a specification of which is the first and which is the second. An ordered pair is denoted by writing the two objects in parentheses and separated by a comma, as in (a,b). The objects a and b are called the ***components*** of the ordered pair. The defining characteristic is that two ordered pairs are equal if and only if their first components are equal and their second components are equal:

$$(a,b) = (a',b') \qquad \text{if and only if} \qquad a = a' \text{ and } b = b'.$$

Given two sets, we can form a new set consisting of the ordered pairs whose components are taken one from each set in a specified order. This is another type of set whose existence is guaranteed by ZFC:

- CARTESIAN PRODUCTS: Given sets X and Y, there exists a set $X \times Y$, called their ***Cartesian product***, whose elements are precisely all the ordered pairs of the form (x,y) with $x \in X$ and $y \in Y$.

Relations

Cartesian products are used to give rigorous definitions of the most important constructions in mathematics: relations and functions. Let us begin with the simpler of these two concepts. A ***relation*** between sets X and Y is a subset of $X \times Y$. If R is a relation, it is often convenient to use the notation $x\ R\ y$ to mean $(x,y) \in R$.

An important special case arises when we consider a relation between a set X and itself, which is called a ***relation on X***. For example, both "equals" and "less than" are relations on the set of real numbers. If R is a relation on X and $Y \subseteq X$, we obtain a relation on Y, called the ***restriction of R to Y***, consisting of the set of all ordered pairs $(x,y) \in R$ such that both x and y are in Y.

Let $\sim$ denote a relation on a set X. It is said to be ***reflexive*** if $x \sim x$ for all $x \in X$, ***symmetric*** if $x \sim y$ implies $y \sim x$, and ***transitive*** if $x \sim y$ and $y \sim z$ imply $x \sim z$. A relation that is reflexive, symmetric, and transitive is called an ***equivalence relation***. The restriction of an equivalence relation to a subset $S \subseteq X$ is again an equivalence relation.

Given an equivalence relation $\sim$ on X, for each $x \in X$ the ***equivalence class of x*** is defined to be the set

$$[x] = \{y \in X : y \sim x\}.$$

(The use of the term *class* here is not meant to suggest that equivalence classes are not sets; the terminology was established before a clear distinction was made between classes and sets.) The set of all equivalence classes is denoted by $X/\sim$.

Closely related to equivalence relations is the notion of a *partition*. Given any collection $\mathcal{C}$ of sets, if $A \cap B = \varnothing$ whenever $A, B \in \mathcal{C}$ and $A \neq B$, the sets in $\mathcal{C}$ are said to be ***disjoint***. If X is a set, a ***partition of X*** is a collection $\mathcal{C}$ of disjoint nonempty subsets of X whose union is X. In this situation one also says that X is the ***disjoint union*** of the sets in $\mathcal{C}$.

► **Exercise A.2.** Given an equivalence relation $\sim$ on a set X, show that the set $X/\sim$ of equivalence classes is a partition of X. Conversely, given a partition of X, show that there is a unique equivalence relation whose set of equivalence classes is exactly the original partition.

If R is any relation on a set X, the next exercise shows that there is a "smallest" equivalence relation $\sim$ such that $x\ R\ y \Rightarrow x \sim y$. It is called the ***equivalence relation generated by R***.

► **Exercise A.3.** Let $R \subseteq X \times X$ be any relation on X, and define $\sim$ to be the intersection of all equivalence relations in $X \times X$ that contain R.

(a) Show that $\sim$ is an equivalence relation.
(b) Show that $x \sim y$ if and only if at least one of the following statements is true: $x = y$, or $x\ R'\ y$, or there is a finite sequence of elements $z_1, \dots, z_n \in X$ such that $x\ R'\ z_1\ R' \cdots R'\ z_n\ R'\ y$, where $x\ R'\ y$ means "$x\ R\ y$ or $y\ R\ x$." (See below for the formal definition of a finite sequence.)

Another particularly important type of relation is a ***partial ordering***: this is a relation $\leq$ on a set X that is reflexive, transitive, and ***antisymmetric***, which means that $x \leq y$ and $y \leq x$ together imply $x = y$. If in addition at least one of the relations $x \leq y$ or $y \leq x$ holds for each pair of elements $x, y \in X$, it is called a ***total ordering*** (or sometimes a ***linear*** or ***simple ordering***). The notation $x < y$ is defined to mean $x \leq y$ and $x \neq y$, and the notations $x > y$ and $x \geq y$ have the obvious meanings. If X is a set endowed with an ordering, one often says that X is a ***totally*** or ***partially ordered set***, with the ordering being understood from the context.

The most common examples of totally ordered sets are number systems such as the real numbers and the integers (see below). An important example of a partially ordered set is the set $\mathcal{P}(X)$ of subsets of a given set X, with the partial order relation defined by containment: $A \leq B$ if and only if $A \subseteq B$. It is easy to see that any subset of a partially ordered set is itself partially ordered with (the restriction of) the same order relation, and if the original ordering is total, then the subset is also totally ordered.

If X is a partially ordered set and $S \subseteq X$ is any subset, an element $x \in X$ is said to be an ***upper bound for S*** if $x \geq s$ for every $s \in S$. If S has an upper bound, it is said to be ***bounded above***. If x is an upper bound for S and every other upper bound x' satisfies $x' \geq x$, then x is called a ***least upper bound***. The terms ***lower bound***, ***bounded below***, and ***greatest lower bound*** are defined similarly.

An element $s \in S$ is said to be ***maximal*** if there is no $s' \in S$ such that $s' > s$, and it is the ***largest element*** of S if $s' \leq s$ for every $s' \in S$. ***Minimal*** and ***smallest*** elements are defined similarly. A largest or smallest element of S is also called a ***maximum*** or ***minimum*** of S, respectively. A largest element, if it exists, is automatically unique and maximal, and similarly for a smallest element.

Note the important difference between a maximal element and a maximum: in a subset S of a partially ordered set X, an element $s \in S$ may be maximal without being a maximum, because there might be elements in S that are neither larger nor smaller than s. On the other hand, if S is totally ordered, then a maximal element is automatically a maximum.

A totally ordered set X is said to be ***well ordered*** if every nonempty subset $S \subseteq X$ has a smallest element. For example, the set of positive integers is well ordered, but the set of all integers and the set of positive real numbers are not.

Functions

Suppose X and Y are sets. A ***function from X to Y*** is a relation $f \subseteq X \times Y$ with the property that for every $x \in X$ there is a unique $y \in Y$ such that $(x, y) \in f$. This unique element of Y is called the ***value of f at x*** and denoted by $f(x)$, so that $y = f(x)$ if and only if $(x, y) \in f$. The sets X and Y are called the ***domain*** and ***codomain of f***, respectively. We consider the domain and codomain to be part of the definition of the function, so to say that two functions are equal is to say that they have the same domain and codomain, and both give the same value when applied to each element of the domain. The words ***map*** and ***mapping*** are synonyms for function.

The notation $f : X \to Y$ means "f is a function from X to Y" (or, depending on how it is used in a sentence, "f, a function from X to Y," or "f, from X to Y"). The equation $y = f(x)$ is also sometimes written $f : x \mapsto y$ or, if the name of the function is not important, $x \mapsto y$. Note that the type of arrow ($\mapsto$) used to denote the action of a function on an element of its domain is different from the arrow ($\to$) used between the domain and codomain.

Given two functions $g : X \to Y$ and $f : Y \to Z$, their ***composition*** is the function $f \circ g : X \to Z$ defined by $(f \circ g)(x) = f(g(x))$ for each $x \in X$. It follows from the definition that composition is associative: $(f \circ g) \circ h = f \circ (g \circ h)$.

A map $f : X \to Y$ is called a ***constant map*** if there is some element $c \in Y$ such that $f(x) = c$ for every $x \in X$. This is sometimes written symbolically as $f(x) \equiv c$, and read "$f(x)$ is identically equal to c." For each set X, there exists a natural map $\mathrm{Id}_X : X \to X$ called the ***identity map of X***, defined by $\mathrm{Id}_X(x) = x$ for all $x \in X$. It satisfies $f \circ \mathrm{Id}_X = f = \mathrm{Id}_Y \circ f$ whenever $f : X \to Y$. If $S \subseteq X$ is a subset, there is a function $\iota_S : S \to X$ called the ***inclusion map of S in X***, given by $\iota_S(x) = x$ for $x \in S$. We sometimes use the notation $\iota_S : S \hookrightarrow X$ to emphasize the fact that it is an inclusion map. When the sets are understood, we sometimes denote an identity map simply by Id and an inclusion map by ι.

If $f : X \to Y$ is a function, we can obtain new functions from f by changing the domain or codomain. First consider the domain. For any subset $S \subseteq X$, there is a naturally defined function from S to Y, denoted by $f|_S : S \to Y$ and called the ***restriction of f to S***, obtained by applying f only to elements of S: $f|_S(x) = f(x)$ for all $x \in S$. In terms of ordered pairs, $f|_S$ is just the subset of $S \times Y$ consisting of ordered pairs $(x, y) \in f$ such that $x \in S$. It is immediate that $f|_S = f \circ \iota_S$, and ι_S is just the restriction of Id_X to S.

On the other hand, given $f : X \to Y$, there is no natural way to *expand* the domain of f without giving a new definition for the action of f on elements that are not in X. If W is a set that contains X, and $g : W \to Y$ is a function whose restriction to X is equal to f, we say that g is an ***extension of f***. If $W \neq X$, there are typically many possible extensions of f.

Next consider changes of codomain. Given a function $f : X \to Y$, if Z is any set that contains Y, we automatically obtain a new function $\widetilde{f} : X \to Z$, just by letting $\widetilde{f}(x) = f(x)$ for each $x \in X$. It is also sometimes possible to shrink the codomain, but this requires more care: if $T \subseteq Y$ is a subset such that $f(x) \in T$ for every $x \in X$, we get a new function $\bar{f} : X \to T$, defined by $\bar{f}(x) = f(x)$ for every $x \in X$. In terms of ordered pairs, all three functions f, $\widetilde{f}$, and $\bar{f}$ are represented by exactly the same set of ordered pairs as f itself; but it is important to observe that they are all *different functions* because they have different codomains. This observation notwithstanding, it is a common practice (which we usually follow) to denote any function obtained from f by expanding or shrinking its codomain by the same symbol as the original function. Thus in the situation above, we might have several different functions denoted by the symbol f: the original function $f : X \to Y$, a function $f : X \to Z$ obtained by expanding the codomain, and a function $f : X \to T$ obtained by restricting the codomain. In any such situation, it is important to be clear about which function is intended.

Let $f : X \to Y$ be a function. If $S \subseteq X$, the ***image of S under f***, denoted by $f(S)$, is the subset of Y defined by

$$f(S) = \{y \in Y : y = f(x) \text{ for some } x \in S\}.$$

It is common also to use the shorter notation

$$\{f(x) : x \in S\}$$

to mean the same thing. The set $f(X) \subseteq Y$, the image of the entire domain, is also called the ***image of f*** or the ***range of f***. (Warning: in some contexts—including the previous edition of this book—the word *range* is used to denote what we here call the codomain of a function. Because of this ambiguity, we avoid using the word *range* in favor of *image*.)

If T is a subset of Y, the ***preimage of T under f*** (also called the ***inverse image***) is the subset $f^{-1}(T) \subseteq X$ defined by

$$f^{-1}(T) = \{x \in X : f(x) \in T\}.$$

If $T = \{y\}$ is a singleton, it is common to use the notation $f^{-1}(y)$ in place of the more accurate but more cumbersome $f^{-1}(\{y\})$.

► **Exercise A.4.** Let $f : X \to Y$ and $g : W \to X$ be maps, and suppose $R \subseteq W$, $S, S' \subseteq X$, and $T, T' \subseteq Y$. Prove the following:

(a) $T \supseteq f(f^{-1}(T))$.
(b) $T \subseteq T' \Rightarrow f^{-1}(T) \subseteq f^{-1}(T')$.
(c) $f^{-1}(T \cup T') = f^{-1}(T) \cup f^{-1}(T')$.
(d) $f^{-1}(T \cap T') = f^{-1}(T) \cap f^{-1}(T')$.
(e) $f^{-1}(T \smallsetminus T') = f^{-1}(T) \smallsetminus f^{-1}(T')$.
(f) $S \subseteq f^{-1}(f(S))$.
(g) $S \subseteq S' \Rightarrow f(S) \subseteq f(S')$.
(h) $f(S \cup S') = f(S) \cup f(S')$.
(i) $f(S \cap S') \subseteq f(S) \cap f(S')$.
(j) $f(S \smallsetminus S') \supseteq f(S) \smallsetminus f(S')$.
(k) $f(S) \cap T = f(S \cap f^{-1}(T))$.
(l) $f(S) \cup T \supseteq f(S \cup f^{-1}(T))$.
(m) $S \cap f^{-1}(T) \subseteq f^{-1}(f(S) \cap T)$.
(n) $S \cup f^{-1}(T) \subseteq f^{-1}(f(S) \cup T)$.
(o) $(f \circ g)^{-1}(T) = g^{-1}(f^{-1}(T))$.
(p) $(f \circ g)(R) = f(g(R))$.

► **Exercise A.5.** With notation as in the previous exercise, give counterexamples to show that the following equalities do not necessarily hold true.

(a) $T = f(f^{-1}(T))$.
(b) $S = f^{-1}(f(S))$.
(c) $f(S \cap S') = f(S) \cap f(S')$.
(d) $f(S \smallsetminus S') = f(S) \smallsetminus f(S')$.

A function $f : X \to Y$ is said to be ***injective*** or ***one-to-one*** if $f(x_1) = f(x_2)$ implies $x_1 = x_2$ whenever $x_1, x_2 \in X$. It is said to be ***surjective*** or to ***map X onto Y*** if $f(X) = Y$, or in other words if every $y \in Y$ is equal to $f(x)$ for some $x \in X$. A function that is both injective and surjective is said to be ***bijective*** or a ***one-to-one correspondence***. Maps that are injective, surjective, or bijective are also called ***injections***, ***surjections***, or ***bijections***, respectively. A bijection from a set X to itself is also called a ***permutation of X***.

► **Exercise A.6.** Show that a composition of injective functions is injective, a composition of surjective functions is surjective, and a composition of bijective functions is bijective.

► **Exercise A.7.** Show that equality (a) in Exercise A.5 holds for every $T \subseteq Y$ if and only if f is surjective, and each of the equalities (b)–(d) holds for every $S, S' \subseteq X$ if and only if f is injective.

Given $f: X \to Y$, if there exists a map $g: Y \to X$ such that $f \circ g = \mathrm{Id}_Y$ and $g \circ f = \mathrm{Id}_X$, then g is said to be an ***inverse of f***. Since inverses are unique (see the next exercise), the inverse map is denoted unambiguously by f^{-1} when it exists.

► **Exercise A.8.** Let $f: X \to Y$ be a function.

(a) Show that f has an inverse if and only if it is bijective.
(b) Show that if f has an inverse, its inverse is unique.
(c) Show that if $f: X \to Y$ and $g: Y \to Z$ are both bijective, then $(g \circ f)^{-1} = f^{-1} \circ g^{-1}$.

Beware: given a function $f: X \to Y$, because the same notation f^{-1} is used for both the inverse function and the preimage of a set, it is easy to get confused. When f^{-1} is applied to a subset $T \subseteq Y$, there is no ambiguity: the notation $f^{-1}(T)$ always means the preimage. If f happens to be bijective, $f^{-1}(T)$ could also be interpreted to mean the (forward) image of T under the function f^{-1}; but a little reflection should convince you that the two interpretations yield the same result.

A little more care is required with the notation $f^{-1}(y)$ when y is an *element* of Y. If f is bijective, this generally means the value of the inverse function applied to the element y, which is an element of X. But we also sometimes use this notation to mean the preimage set $f^{-1}(\{y\})$, which makes sense regardless of whether f is bijective. In such cases, the intended meaning should be made clear in context.

Given $f: X \to Y$, a ***left inverse for f*** is a function $g: Y \to X$ that satisfies $g \circ f = \mathrm{Id}_X$. A ***right inverse for f*** is a function $g: Y \to X$ satisfying $f \circ g = \mathrm{Id}_Y$.

Lemma A.9. *If $f: X \to Y$ is a function and $X \neq \varnothing$, then f has a left inverse if and only if it is injective, and a right inverse if and only if it is surjective.*

Proof. Suppose g is a left inverse for f. If $f(x) = f(x')$, applying g to both sides implies $x = x'$, so f is injective. Similarly, if g is a right inverse and $y \in Y$ is arbitrary, then $f(g(y)) = y$, so f is surjective.

Now suppose f is injective. Choose any $x_0 \in X$, and define $g: Y \to X$ by $g(y) = x$ if $y \in f(X)$ and $y = f(x)$, and $g(y) = x_0$ if $y \notin f(X)$. The injectivity of f ensures that g is well defined, and it is immediate from the definition that $g \circ f = \mathrm{Id}_X$. The proof that surjectivity implies the existence of a right inverse requires the axiom of choice, so we postpone it until later in this appendix (Exercise A.15). □

► **Exercise A.10.** Show that if $f: X \to Y$ is bijective, then any left or right inverse for f is equal to f^{-1}.

For the purposes of category theory, it is necessary to extend some of the concepts of relations and functions to classes as well as sets. If $\mathcal{C}$ and $\mathcal{D}$ are classes, a ***relation between $\mathcal{C}$ and $\mathcal{D}$*** is just a class of ordered pairs of the form (x, y) with $x \in \mathcal{C}$ and $y \in \mathcal{D}$. A ***mapping from $\mathcal{C}$ to $\mathcal{D}$*** is a relation $\mathcal{F}$ between $\mathcal{C}$ and $\mathcal{D}$ with the property that for every $x \in \mathcal{C}$ there is a unique $y \in \mathcal{D}$ such that $(x, y) \in \mathcal{F}$. We use the same notations in this context as for relations and mappings between sets. Thus, for example, $\mathcal{F}: \mathcal{C} \to \mathcal{D}$ means that $\mathcal{F}$ is a mapping from $\mathcal{C}$ to $\mathcal{D}$, and $y = \mathcal{F}(x)$ means that $(x, y) \in \mathcal{F}$.

Number Systems and Cardinality

So far, most of the set-theoretic constructions we have introduced describe ways of obtaining new sets from already existing ones. Before the theory will have much content, we need to know that some interesting sets exist. We take the set of real numbers as our starting point. The properties that characterize it are the following:

(i) It is a ***field*** in the algebraic sense: a set with binary operations $+$ and $\times$ satisfying the usual associative, commutative, and distributive laws, containing an additive identity 0 and a multiplicative identity $1 \neq 0$, such that every element has an additive inverse and every nonzero element has a multiplicative inverse.
(ii) It is endowed with a total ordering that makes it into an ***ordered field***, which means that $y < z \Rightarrow x + y < x + z$ and $x > 0,\ y > 0 \Rightarrow xy > 0$.
(iii) It is ***complete***, meaning that every nonempty subset with an upper bound has a least upper bound.

ZFC guarantees the existence of such a set.

- EXISTENCE OF THE REAL NUMBERS: There exists a complete ordered field, called the set of ***real numbers*** and denoted by $\mathbb{R}$.

► **Exercise A.11.** Show that the real numbers are unique, in the sense that any complete ordered field admits a bijection with $\mathbb{R}$ that preserves addition, multiplication, and order.

Let $S \subseteq \mathbb{R}$ be a nonempty subset with an upper bound. The least upper bound of S is also called the ***supremum of*** $\boldsymbol{S}$, and is denoted by $\sup S$. Similarly, any nonempty set T with a lower bound has a greatest lower bound, also called its ***infimum*** and denoted by $\inf T$.

We work extensively with the usual subsets of $\mathbb{R}$:

- the set of ***natural numbers***, $\mathbb{N}$ (the positive counting numbers), defined as the smallest subset of $\mathbb{R}$ containing 1 and containing $n + 1$ whenever it contains n
- the set of ***integers***, $\mathbb{Z} = \{n \in \mathbb{R} : n = 0 \text{ or } n \in \mathbb{N} \text{ or } -n \in \mathbb{N}\}$
- the set of ***rational numbers***, $\mathbb{Q} = \{x \in \mathbb{R} : x = p/q \text{ for some } p, q \in \mathbb{Z}\}$

We consider the set $\mathbb{C}$ of ***complex numbers*** to be simply $\mathbb{R} \times \mathbb{R}$, in which the real numbers are identified with the subset $\mathbb{R} \times \{0\} \subseteq \mathbb{C}$ and i stands for the imaginary unit $(0, 1)$. Multiplication and addition of complex numbers are defined by the usual rules with $i^2 = -1$; thus $x + iy$ is another notation for (x, y).

For any pair of integers $m \le n$, we define the set $\{m, \dots, n\} \subseteq \mathbb{Z}$ by

$$\{m, \dots, n\} = \{k \in \mathbb{Z} : m \le k \le n\}.$$

For subsets of the real numbers, we use the following standard notations when $a < b$:

$$\begin{aligned}
(a,b) &= \{x \in \mathbb{R} : a < x < b\} && \text{(open interval)},\\
[a,b] &= \{x \in \mathbb{R} : a \le x \le b\} && \text{(closed interval)},\\
(a,b] &= \{x \in \mathbb{R} : a < x \le b\} && \text{(half-open interval)},\\
[a,b) &= \{x \in \mathbb{R} : a \le x < b\} && \text{(half-open interval)}.
\end{aligned}$$

(The two conflicting meanings of (a,b)—as an ordered pair or as an open interval—have to be distinguished from the context.) We also use the notations $[a,\infty)$, (a,∞), $(-\infty,b]$, $(-\infty,b)$, and $(-\infty,\infty)$, with the obvious meanings. A subset $J \subseteq \mathbb{R}$ is called an ***interval*** if it contains more than one element, and whenever $a,b \in J$, every c such that $a < c < b$ is also in J.

► **Exercise A.12.** Show that an interval must be one of the nine types of sets $[a,b]$, (a,b), $[a,b)$, $(a,b]$, $(-\infty,b]$, $(-\infty,b)$, $[a,\infty)$, (a,∞), or $(-\infty,\infty)$.

The natural numbers play a special role in set theory, as a yardstick for measuring sizes of sets. Two sets are said to ***have the same cardinality*** if there exists a bijection between them. A set is ***finite*** if it is empty or has the same cardinality as $\{1,\dots,n\}$ for some $n \in \mathbb{N}$ (in which case it is said to have ***cardinality n***), and otherwise it is ***infinite***. A set is ***countably infinite*** if it has the same cardinality as $\mathbb{N}$, ***countable*** if it is either finite or countably infinite, and ***uncountable*** otherwise. The sets $\mathbb{N}$, $\mathbb{Z}$, and $\mathbb{Q}$ are countable, but $\mathbb{R}$ and $\mathbb{C}$ are not.

► **Exercise A.13.** Prove that any subset of a countable set is countable.

► **Exercise A.14.** Prove that the Cartesian product of two countable sets is countable.

Indexed Families

Using what we have introduced so far, it is easy to extend the notion of ordered pair to more than two objects. Given a natural number n and a set S, an ***ordered n-tuple*** of elements of S is a function $x\colon \{1,\dots,n\} \to S$. It is customary to write x_i instead of $x(i)$ for the value of x at i, and the whole n-tuple is denoted by either of the notations

$$(x_1,\dots,x_n) \quad \text{or} \quad (x_i)_{i=1}^n.$$

The elements $x_i \in S$ are called the ***components of the n-tuple***. Similarly, an ***(infinite) sequence*** of elements of S is a function $x\colon \mathbb{N} \to S$, written as

$$(x_1,x_2,\dots), \quad (x_i)_{i\in\mathbb{N}}, \quad \text{or} \quad (x_i)_{i=1}^\infty.$$

A ***doubly infinite sequence*** is a function $x\colon \mathbb{Z} \to S$, written

$$(\dots,x_{-1},x_0,x_1,\dots), \quad (x_i)_{i\in\mathbb{Z}}, \quad \text{or} \quad (x_i)_{i=-\infty}^\infty.$$

An ordered n-tuple is sometimes called a ***finite sequence***. For all such sequences, we sometimes write (x_i) if the domain of the associated function ($\{1,\dots,n\}$, $\mathbb{N}$, or $\mathbb{Z}$) is understood.

It is also useful to adapt the notations for sequences to refer to the *image set* of a finite or infinite sequence, that is, the set of values $x_1, x_2, \dots$, irrespective of their order and disregarding repetitions. For this purpose we replace the parentheses by braces. Thus any of the notations

$$\{x_1,\dots,x_n\}, \quad \{x_i\}_{i=1}^n, \quad \text{or} \quad \{x_i : i = 1,\dots,n\}$$

denotes the image set of the function $x\colon \{1,\dots,n\} \to S$. Similarly,

$$\{x_1,x_2,\dots\}, \quad \{x_i\}_{i\in\mathbb{N}}, \quad \{x_i\}_{i=1}^\infty, \quad \text{or} \quad \{x_i : i \in \mathbb{N}\}$$

all represent the image set of the infinite sequence $(x_i)_{i\in\mathbb{N}}$.

A ***subsequence*** of a sequence $(x_i)_{i\in\mathbb{N}}$ is a sequence of the form $(x_{i_j})_{j\in\mathbb{N}}$, where $(i_j)_{j\in\mathbb{N}}$ is a sequence of natural numbers that is ***strictly increasing***, meaning that $j < j'$ implies $i_j < i_{j'}$.

We sometimes need to consider collections of objects that are indexed, not by the natural numbers or subsets of them, but by arbitrary sets, potentially even uncountable ones. An ***indexed family*** of elements of a set S is just a function from a set A (called the ***index set***) to S, and in this context is denoted by $(x_\alpha)_{\alpha\in A}$. (Thus a sequence is just the special case of an indexed family in which the index set is $\mathbb{N}$.) Occasionally, when the index set is understood or is irrelevant, we omit it from the notation and simply denote the family as (x_α). As in the case of sequences, we use braces to denote the image set of the function:

$$\{x_\alpha\}_{\alpha\in A} = \{x_\alpha : \alpha \in A\} = \{x \in S : x = x_\alpha \text{ for some } \alpha \in A\}.$$

Any set $\mathcal{A}$ of elements of S can be converted to an indexed family, simply by taking the index set to be $\mathcal{A}$ itself and the indexing function to be the inclusion map $\mathcal{A} \hookrightarrow S$.

If $(X_\alpha)_{\alpha\in A}$ is an indexed family of sets, $\bigcup_{\alpha\in A} X_\alpha$ is just another notation for the union of the (unindexed) collection $\{X_\alpha\}_{\alpha\in A}$. If the index set is finite, the union is usually written as $X_1 \cup \dots \cup X_n$. A similar remark applies to the intersection $\bigcap_{\alpha\in A} X_\alpha$ or $X_1 \cap \dots \cap X_n$.

The definition of Cartesian product now extends easily from two sets to arbitrarily many. If $(X_1,\dots,X_n)$ is an ordered n-tuple of sets, their Cartesian product $X_1 \times \dots \times X_n$ is the set of all ordered n-tuples $(x_1,\dots,x_n)$ such that $x_i \in X_i$ for $i = 1,\dots,n$. If $X_1 = \dots = X_n = X$, the n-fold Cartesian product $X \times \dots \times X$ is often written simply as X^n.

Every Cartesian product comes naturally equipped with ***canonical projection maps*** $\pi_i : X_1 \times \dots \times X_n \to X_i$, defined by $\pi_i(x_1,\dots,x_n) = x_i$. Each of these maps is surjective, provided the sets X_i are all nonempty. If $f : S \to X_1 \times \dots \times X_n$ is any function into a Cartesian product, the composite functions $f_i = \pi_i \circ f : S \to X_i$ are called its ***component functions***. Any such function f is completely determined by its component functions, via the formula

$$f(y) = (f_1(y), \ldots, f_n(y)).$$

More generally, the Cartesian product of an arbitrary indexed family $(X_\alpha)_{\alpha \in A}$ of sets is defined to be the set of all functions $x\colon A \to \bigcup_{\alpha \in A} X_\alpha$ such that $x_\alpha \in X_\alpha$ for each α. It is denoted by $\prod_{\alpha \in A} X_\alpha$. Just as in the case of finite products, each Cartesian product comes equipped with canonical projection maps $\pi_\beta\colon \prod_{\alpha \in A} X_\alpha \to X_\beta$, defined by $\pi_\beta(x) = x_\beta$.

Our last set-theoretic assertion from ZFC is that it is possible to choose an element from each set in an arbitrary indexed family.

- AXIOM OF CHOICE: If $(X_\alpha)_{\alpha \in A}$ is a nonempty indexed family of nonempty sets, there exists a function $c\colon A \to \bigcup_{\alpha \in A} X_\alpha$, called a ***choice function***, such that $c(\alpha) \in X_\alpha$ for each α.

In other words, the Cartesian product of a nonempty indexed family of nonempty sets is nonempty.

Here are some immediate applications of the axiom of choice.

► **Exercise A.15.** Complete the proof of Lemma A.9 by showing that every surjective function has a right inverse.

► **Exercise A.16.** Prove that if there exists a surjective map from a countable set onto S, then S is countable.

► **Exercise A.17.** Prove that the union of a countable collection of countable sets is countable.

The axiom of choice has a number of interesting equivalent reformulations; the relationships among them make fascinating reading, for example in [Hal74]. The only other formulations we make use of are the following two (the well-ordering theorem in Problem 4-6 and Zorn's lemma in Lemma 13.42).

Theorem A.18 (The Well-Ordering Theorem). *Every set can be given a total ordering with respect to which it is well ordered.*

Theorem A.19 (Zorn's Lemma). *Let X be a partially ordered set in which every totally ordered subset has an upper bound. Then X contains a maximal element.*

For proofs, see any of the set theory texts mentioned at the beginning of this appendix.

Abstract Disjoint Unions

Earlier, we mentioned that given a set X and a partition of it, X is said to be the *disjoint union* of the subsets in the partition. It sometimes happens that we are given a collection of sets, which may or may not be disjoint, but which we want to consider as disjoint subsets of a larger set. For example, we might want to form a set consisting of "five copies of $\mathbb{R}$," in which we consider the different copies to be disjoint from each other. We can accomplish this by the following trick. Suppose $(X_\alpha)_{\alpha\in A}$ is an indexed family of nonempty sets. For each α in the index set, imagine "tagging" the elements of X_α with the index α, in order to make the sets X_α and X_β disjoint when $\alpha \neq \beta$, even if they were not disjoint to begin with.

Formally, we can make sense of an element x with a tag α as an ordered pair (x,α). Thus we define the ***(abstract) disjoint union*** of the indexed family, denoted by $\coprod_{\alpha\in A} X_\alpha$, to be the set

$$\coprod_{\alpha\in A} X_\alpha = \{(x,\alpha) : \alpha \in A \text{ and } x \in X_\alpha\}.$$

If the index set is finite, the disjoint union is usually written as $X_1 \amalg \cdots \amalg X_n$.

For each index α, there is a natural map $\iota_\alpha \colon X_\alpha \to \coprod_{\alpha\in A} X_\alpha$, called the ***canonical injection of*** $\boldsymbol{X_\alpha}$, defined by $\iota_\alpha(x) = (x,\alpha)$. Each such map is injective, and its image is the set $X_\alpha^* = \{(x,\alpha) : x \in X_\alpha\}$, which we can think of as a "copy" of X_α sitting inside the disjoint union. For $\alpha \neq \beta$, the sets X_α^* and X_β^* are disjoint from each other by construction. In practice, we usually blur the distinction between X_α and X_α^*, and thus think of X_α itself as a subset of the disjoint union, and think of the canonical injection ι_α as an inclusion map. With this convention, this usage of the term *disjoint union* is consistent with our previous one.

Appendix B:
Review of Metric Spaces

Metric spaces play an indispensable role in real analysis, and their properties provide the underlying motivation for most of the basic definitions in topology. In this section we summarize the important properties of metric spaces with which you should be familiar. For a thorough treatment of the subject, see any good undergraduate real analysis text such as [Rud76] or [Apo74].

Euclidean Spaces

Most of topology, in particular manifold theory, is modeled on the behavior of Euclidean spaces and their subsets, so we begin with a quick review of their properties.

The Cartesian product $\mathbb{R}^n = \mathbb{R} \times \cdots \times \mathbb{R}$ of n copies of $\mathbb{R}$ is known as ***n*-dimensional Euclidean space**. It is the set of all ordered n-tuples of real numbers. An element of $\mathbb{R}^n$ is denoted by $(x_1, \dots, x_n)$ or simply x. The numbers x_i are called its ***components*** or ***coordinates***. Zero-dimensional Euclidean space $\mathbb{R}^0$ is, by convention, the singleton $\{0\}$.

We use without further comment the fact that $\mathbb{R}^n$ is an n-dimensional real vector space with the usual operations of scalar multiplication and vector addition. We refer to an element of $\mathbb{R}^n$ either as a ***point*** or as a ***vector***, depending on whether we wish to emphasize its location or its direction and magnitude. The geometric properties of $\mathbb{R}^n$ are derived from the ***Euclidean dot product***, defined by $x \cdot y = x_1 y_1 + \cdots + x_n y_n$. In particular, the ***norm*** or ***length*** of a vector $x \in \mathbb{R}^n$ is given by

$$|x| = (x \cdot x)^{1/2} = \left((x_1)^2 + \cdots + (x_n)^2\right)^{1/2}.$$

► **Exercise B.1.** Show that the following inequalities hold for any $x = (x_1, \dots, x_n) \in \mathbb{R}^n$:

$$\max\{|x_1|, \dots, |x_n|\} \le |x| \le \sqrt{n} \max\{|x_1|, \dots, |x_n|\}. \tag{B.1}$$

If x and y are nonzero vectors in $\mathbb{R}^n$, the ***angle between x and y*** is defined to be $\cos^{-1}\left((x \cdot y)/(|x||y|)\right)$. Given two points $x, y \in \mathbb{R}^n$, the ***line segment from x to***

y is the set $\{x + t(y - x) : 0 \le t \le 1\}$, and the ***distance between x and y*** is $|x - y|$. A ***(closed) ray*** in $\mathbb{R}^n$ is any set of the form $\{x + t(y - x) : t \ge 0\}$ for two distinct points $x, y \in \mathbb{R}^n$, and the corresponding ***open ray*** is the same set with x deleted.

Continuity and convergence in Euclidean spaces are defined in the usual ways. A map $f : U \to V$ between subsets of Euclidean spaces is ***continuous at x ∈ U*** if for any $\varepsilon > 0$ there exists $\delta > 0$ such that for all $y \in U$, $|x - y| < \delta$ implies $|f(x) - f(y)| < \varepsilon$. Such a map is said to be ***continuous*** if it is continuous at every point of its domain. A sequence (x_i) of points in $\mathbb{R}^n$ ***converges*** to $x \in \mathbb{R}^n$ if for any $\varepsilon > 0$ there exists $N \in \mathbb{N}$ such that $i \ge N$ implies $|x_i - x| < \varepsilon$. A sequence is ***bounded*** if there is some $R \in \mathbb{R}$ such that $|x_i| \le R$ for all i.

► **Exercise B.2.** Prove that if S is a nonempty subset of $\mathbb{R}$ that is bounded above and $a = \sup S$, then there is a sequence in S converging to a.

Metrics

Metric spaces are generalizations of Euclidean spaces, in which none of the vector space properties are present and only the distance function remains. Suppose M is any set. A ***metric on M*** is a function $d : M \times M \to \mathbb{R}$, also called a ***distance function***, satisfying the following three properties for all $x, y, z \in M$:

(i) SYMMETRY: $d(x, y) = d(y, x)$.
(ii) POSITIVITY: $d(x, y) \ge 0$, and $d(x, y) = 0$ if and only if $x = y$.
(iii) TRIANGLE INEQUALITY: $d(x, z) \le d(x, y) + d(y, z)$.

If M is a set and d is a metric on M, the pair (M, d) is called a ***metric space***. (Actually, unless it is important to specify which metric is being considered, one often just says "M is a metric space," with the metric being understood from the context.)

Example B.3 (Metric Spaces).

(a) If M is any subset of $\mathbb{R}^n$, the function $d(x, y) = |x - y|$ is a metric on M (see Exercise B.4 below), called the ***Euclidean metric***. Whenever we consider a subset of $\mathbb{R}^n$ as a metric space, it is always with the Euclidean metric unless we specify otherwise.
(b) Similarly, if M is any metric space and X is a subset of M, then X inherits a metric simply by restricting the distance function of M to pairs of points in X.
(c) If X is any set, define a metric on X by setting $d(x, y) = 1$ unless $x = y$, in which case $d(x, y) = 0$. This is called the ***discrete metric*** on X. //

► **Exercise B.4.** Prove that $d(x, y) = |x - y|$ is a metric on any subset of $\mathbb{R}^n$.

Here are some of the standard definitions used in metric space theory. Let M be a metric space.

- For any $x \in M$ and $r > 0$, the ***(open) ball of radius r around x*** is the set

$$B_r(x) = \{y \in M : d(y,x) < r\},$$

and the ***closed ball of radius r around x*** is

$$\overline{B}_r(x) = \{y \in M : d(y,x) \le r\}.$$

- A subset $A \subseteq M$ is said to be an ***open subset of M*** if it contains an open ball around each of its points.
- A subset $A \subseteq M$ is said to be a ***closed subset of M*** if $M \smallsetminus A$ is open.

The next two propositions summarize the most important properties of open and closed subsets of metric spaces.

Proposition B.5 (Properties of Open Subsets of a Metric Space). *Let M be a metric space.*

(*a*) *Both M and $\varnothing$ are open subsets of M.*
(*b*) *Any intersection of finitely many open subsets of M is an open subset of M.*
(*c*) *Any union of arbitrarily many open subsets of M is an open subset of M.*

Proposition B.6 (Properties of Closed Subsets of a Metric Space). *Let M be a metric space.*

(*a*) *Both M and $\varnothing$ are closed subsets of M.*
(*b*) *Any union of finitely many closed subsets of M is a closed subset of M.*
(*c*) *Any intersection of arbitrarily many closed subsets of M is a closed subset of M.*

► **Exercise B.7.** Prove the two preceding propositions.

► **Exercise B.8.** Suppose M is a metric space.

(a) Show that an open ball in M is an open subset, and a closed ball in M is a closed subset.
(b) Show that a subset of M is open if and only if it is the union of some collection of open balls.

► **Exercise B.9.** In each part below, a subset S of a metric space M is given. In each case, decide whether S is open, closed, both, or neither.

(a) $M = \mathbb{R}$, and $S = [0,1)$.
(b) $M = \mathbb{R}$, and $S = \mathbb{N}$.
(c) $M = \mathbb{Z}$, and $S = \mathbb{N}$.
(d) $M = \mathbb{R}^2$, and S is the set of points with rational coordinates.
(e) $M = \mathbb{R}^2$, and S is the unit disk $\{(x,y) \in \mathbb{R}^2 : x^2 + y^2 < 1\}$.
(f) $M = \mathbb{R}^3$, and S is the unit disk $\{(x,y,z) \in \mathbb{R}^3 : z = 0 \text{ and } x^2 + y^2 < 1\}$.
(g) $M = \{(x,y) \in \mathbb{R}^2 : x > 0 \text{ and } y > 0\}$, and $S = \{(x,y) \in M : x^2 + y^2 \le 1\}$.

► **Exercise B.10.** Suppose $A \subseteq \mathbb{R}$ is closed and nonempty. Show that if A is bounded above, then it contains its supremum, and if it is bounded below, then it contains its infimum.

Suppose M is a metric space and A is a subset of M. We say that A is ***bounded*** if there exists a positive number R such that $d(x,y) \le R$ for all $x,y \in A$. If A is a nonempty bounded subset of M, the ***diameter of A*** is the number $\operatorname{diam} A = \sup\{d(x,y) : x,y \in A\}$.

► **Exercise B.11.** Let M be a metric space and $A \subseteq M$ be any subset. Prove that the following are equivalent:

(a) A is bounded.
(b) A is contained in some closed ball.
(c) A is contained in some open ball.

Continuity and Convergence

The definition of continuity in the context of metric spaces is a straightforward generalization of the Euclidean definition. If (M_1,d_1) and (M_2,d_2) are metric spaces and x is a point in M_1, a map $f\colon M_1 \to M_2$ is said to be ***continuous at x*** if for any $\varepsilon > 0$ there exists $\delta > 0$ such that $d_1(x,y) < \delta$ implies $d_2(f(x), f(y)) < \varepsilon$ for all $y \in M_1$; and f is ***continuous*** if it is continuous at every point of M_1.

Similarly, suppose $(x_i)_{i=1}^{\infty}$ is a sequence of points in a metric space (M,d). Given $x \in M$, the sequence is said to ***converge to x***, and x is called the ***limit of the sequence***, if for any $\varepsilon > 0$ there exists $N \in \mathbb{N}$ such that $i \ge N$ implies $d(x_i,x) < \varepsilon$. If this is the case, we write $x_i \to x$ or $\lim_{i\to\infty} x_i = x$.

► **Exercise B.12.** Let M and N be metric spaces and let $f\colon M \to N$ be a map. Show that f is continuous if and only if it takes convergent sequences to convergent sequences and limits to limits, that is, if and only if $x_i \to x$ in M implies $f(x_i) \to f(x)$ in N.

► **Exercise B.13.** Suppose A is a closed subset of a metric space M, and (x_i) is a sequence of points in A that converges to a point $x \in M$. Show that $x \in A$.

A sequence $(x_i)_{i=1}^{\infty}$ in a metric space is said to be ***bounded*** if its image $\{x_i\}_{i=1}^{\infty}$ is a bounded subset of M. The sequence is said to be a ***Cauchy sequence*** if for every $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that $i,j \ge N$ implies $d(x_i,x_j) < \varepsilon$. Every convergent sequence is Cauchy (Exercise B.14), but the converse is not true in general. A metric space in which every Cauchy sequence converges is said to be ***complete***.

► **Exercise B.14.** Prove that every convergent sequence in a metric space is Cauchy, and every Cauchy sequence is bounded.

► **Exercise B.15.** Prove that every closed subset of a complete metric space is complete, when considered as a metric space in its own right.

The following criterion for continuity is frequently useful (and in fact, as is explained in Chapter 2, it is the main motivation for the definition of a topological space).

Theorem B.16 (Open Subset Criterion for Continuity). *A map $f: M_1 \to M_2$ between metric spaces is continuous if and only if the preimage of every open subset is open: whenever U is an open subset of M_2, its preimage $f^{-1}(U)$ is open in M_1.*

Proof. First assume f is continuous, and let $U \subseteq M_2$ be an open set. If x is any point in $f^{-1}(U)$, then because U is open, there is some $\varepsilon > 0$ such that $B_\varepsilon(f(x)) \subseteq U$. Continuity of f implies that there exists $\delta > 0$ such that $y \in B_\delta(x)$ implies $f(y) \in B_\varepsilon(f(x)) \subseteq U$, so $B_\delta(x) \subseteq f^{-1}(U)$. Since this is true for every point of $f^{-1}(U)$, it follows that $f^{-1}(U)$ is open.

Conversely, assume that the preimage of every open subset is open. Choose any $x \in M_1$, and let $\varepsilon > 0$ be arbitrary. Because $B_\varepsilon(f(x))$ is open in M_2, our hypothesis implies that $f^{-1}\big(B_\varepsilon(f(x))\big)$ is open in M_1. Since $x \in f^{-1}\big(B_\varepsilon(f(x))\big)$, this means there is some ball $B_\delta(x) \subseteq f^{-1}\big(B_\varepsilon(f(x))\big)$. In other words, $y \in B_\delta(x)$ implies $f(y) \in B_\varepsilon(f(x))$, so f is continuous at x. Because this is true for every $x \in X$, it follows that f is continuous. □

Appendix C:

Review of Group Theory

We assume only basic group theory such as one is likely to encounter in most undergraduate algebra courses. You can find much more detail about all of this material in, for example, [Hun97] or [Her96].

Basic Definitions

A ***group*** is a set G together with a map $G \times G \to G$, usually called ***multiplication*** and written $(g,h) \mapsto gh$, satisfying

(i) ASSOCIATIVITY: For all $g,h,k \in G$, $(gh)k = g(hk)$.
(ii) EXISTENCE OF IDENTITY: There is an element $1 \in G$ such that $1g = g1 = g$ for all $g \in G$.
(iii) EXISTENCE OF INVERSES: For each $g \in G$, there is an element $h \in G$ such that $gh = hg = 1$.

One checks easily that the identity is unique, that each element has a unique inverse (so the usual notation g^{-1} for inverses makes sense), and that $(gh)^{-1} = h^{-1}g^{-1}$. For $g \in G$ and $n \in \mathbb{Z}$, the notation g^n is defined inductively by $g^0 = 1$, $g^1 = g$, $g^{n+1} = g^n g$ for $n \in \mathbb{N}$, and $g^{-n} = (g^{-1})^n$.

The ***order*** of a group G is its cardinality as a set. The ***trivial group*** is the unique group of order 1; it is the group consisting of the identity alone. A group G is said to be ***abelian*** if $gh = hg$ for all $g,h \in G$. The group operation in an abelian group is frequently written additively, $(g,h) \mapsto g + h$, in which case the identity element is denoted by 0, the inverse of g is denoted by $-g$, and we use ng in place of g^n.

If G is a group, a subset of G that is itself a group with the same multiplication is called a ***subgroup*** of G. It follows easily from the definition that a subset of G is a subgroup if and only if it is closed under multiplication and contains the inverse of each of its elements. Thus, for example, the intersection of any family of subgroups of G is itself a subgroup of G.

If S is any subset of a group G, we let $\langle S\rangle$ denote the intersection of all subgroups of G containing S. It is a subgroup of G—in fact, the smallest subgroup of G containing S—and is called the ***subgroup generated by S***. If $S = \{g_1, \dots, g_k\}$ is a finite set, it is common to use the less cumbersome notation $\langle g_1, \dots, g_k\rangle$ for the subgroup generated by S, instead of $\langle\{g_1, \dots, g_k\}\rangle$.

► **Exercise C.1.** Suppose G is a group and S is any subset of G. Show that the subgroup generated by S is equal to the set of all finite products of integral powers of elements of S.

If $G_1, \dots, G_n$ are groups, their ***direct product*** is the set $G_1 \times \dots \times G_n$ with the group structure defined by the multiplication law

$$(g_1, \dots, g_n)(g_1', \dots, g_n') = (g_1 g_1', \dots, g_n g_n')$$

and with identity element $(1, \dots, 1)$. More generally, the direct product of an arbitrary indexed family of groups $(G_\alpha)_{\alpha\in A}$ is the Cartesian product set $\prod_{\alpha\in A} G_\alpha$ with multiplication defined componentwise: $(gg')_\alpha = g_\alpha g_\alpha'$.

If $(G_\alpha)_{\alpha\in A}$ is a family of abelian groups, we also define their ***direct sum***, denoted by $\bigoplus_\alpha G_\alpha$, to be the subgroup of the direct product $\prod_\alpha G_\alpha$ consisting of those elements $(g_\alpha)_{\alpha\in A}$ such that g_α is the identity element in G_α for all but finitely many α. The direct sum of a finite family is often written $G_1 \oplus \dots \oplus G_n$. If the family is finite (or if G_α is the trivial group for all but finitely many α), then the direct sum and the direct product are identical; but in general they are not.

A map $f: G \to H$ between groups is called a ***homomorphism*** if it preserves multiplication: $f(gh) = f(g)f(h)$. A bijective homomorphism is called an ***isomorphism***. If there exists an isomorphism between groups G and H, they are said to be ***isomorphic***, and we write $G \cong H$. A homomorphism from a group G to itself is called an ***endomorphism of G***, and an endomorphism that is also an isomorphism is called an ***automorphism of G***.

If $f: G \to H$ is a homomorphism, the ***image of f*** is the set $f(G) \subseteq H$, often written $\operatorname{Im} f$, and its ***kernel*** is the set $f^{-1}(1) \subseteq G$, denoted by $\operatorname{Ker} f$.

► **Exercise C.2.** Let $f: G \to H$ be a homomorphism.

(a) Show that f is injective if and only if $\operatorname{Ker} f = \{1\}$.
(b) Show that if f is bijective, then f^{-1} is also an isomorphism.
(c) Show that $\operatorname{Ker} f$ is a subgroup of G, and $\operatorname{Im} f$ is a subgroup of H.
(d) Show that for any subgroup $K \subseteq G$, the image set $f(K)$ is a subgroup of H.

Any element g of a group G defines a map $C_g: G \to G$ by $C_g(h) = ghg^{-1}$. This map, called ***conjugation by g***, is easily shown to be an automorphism of G, so the image under C_g of any subgroup $H \subseteq G$ (written symbolically as gHg^{-1}) is another subgroup of G. Two subgroups H, H' are ***conjugate*** if $H' = gHg^{-1}$ for some $g \in G$.

► **Exercise C.3.** Let G be a group. Show that conjugacy is an equivalence relation on the set of all subgroups of G.

The set of subgroups of G conjugate to a given subgroup H is called the ***conjugacy class of H in G***.

Cosets and Quotient Groups

Suppose G is a group. Given a subgroup $H \subseteq G$ and an element $g \in G$, the ***left coset of H determined by g*** is the set

$$gH = \{gh : h \in H\}.$$

The ***right coset*** Hg is defined similarly. The relation ***congruence modulo H*** is defined on G by declaring that $g \equiv g' \pmod{H}$ if and only if $g^{-1}g' \in H$.

► **Exercise C.4.** Show that congruence modulo H is an equivalence relation, and its equivalence classes are precisely the left cosets of H.

The set of left cosets of H in G is denoted by G/H. (This is just the partition of G defined by congruence modulo H.) The cardinality of G/H is called the ***index of H in G***.

A subgroup $K \subseteq G$ is said to be ***normal*** if it is invariant under all conjugations, that is, if $gKg^{-1} = K$ for all $g \in G$. Clearly, every subgroup of an abelian group is normal.

► **Exercise C.5.** Show that a subgroup $K \subseteq G$ is normal if and only if $gK = Kg$ for every $g \in G$.

► **Exercise C.6.** Show that the kernel of any homomorphism is a normal subgroup.

► **Exercise C.7.** If G is a group, show that the intersection of any family of normal subgroups of G is itself a normal subgroup of G.

Normal subgroups give rise to one of the most important constructions in group theory. Given a normal subgroup $K \subseteq G$, define a multiplication operator on the set G/K of left cosets by

$$(gK)(g'K) = (gg')K.$$

Theorem C.8 (Quotient Theorem for Groups). *If K is a normal subgroup of G, this multiplication is well defined on cosets and turns G/K into a group.*

Proof. First we need to show that the product does not depend on the representatives chosen for the cosets: if $gK = g'K$ and $hK = h'K$, we show that $(gh)K = (g'h')K$. From Exercise C.4, the fact that g and g' determine the same coset means that $g^{-1}g' \in K$, which is the same as saying $g' = gk$ for some $k \in K$. Similarly, $h' = hk'$ for $k' \in K$. Because K is normal, $h^{-1}kh$ is an element of K. Writing this element as k'', we have $kh = hk''$. It follows that

$$g'h' = gkhk' = ghk''k',$$

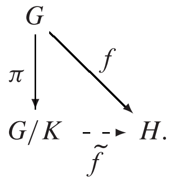
which shows that $g'h'$ and gh determine the same coset.

Now we just note that the group properties are satisfied: associativity of the multiplication in G/K follows from that of G; the element $1K = K$ of G/K acts as an identity; and $g^{-1}K$ is the inverse of gK. □

When K is a normal subgroup of G, the group G/K is called the ***quotient group of G by K***. The natural projection map $\pi : G \to G/K$ that sends each element to its coset is a surjective homomorphism whose kernel is K.

The following theorem tells how to define homomorphisms from a quotient group.

Theorem C.9. *Let G be a group and let $K \subseteq G$ be a normal subgroup. Given a homomorphism $f : G \to H$ such that $K \subseteq \operatorname{Ker} f$, there is a unique homomorphism $\widetilde{f} : G/K \to H$ such that the following diagram commutes:*



(C.1)

(A diagram such as (C.1) is said to ***commute***, or to be ***commutative***, if the maps between two sets obtained by following arrows around either side of the diagram are equal. So in this case commutativity means that $\widetilde{f} \circ \pi = f$.)

Proof. Since $\pi(g) = gK$, if such a map exists, it has to be given by the formula $\widetilde{f}(gK) = f(g)$; this proves uniqueness. To prove existence, we wish to define $\widetilde{f}$ by this formula. As long as this is well defined, it will certainly make the diagram commute. To see that it is well defined, note that if $g \equiv g' \pmod K$, then $g' = gk$ for some $k \in K$, and therefore $f(g') = f(gk) = f(g)f(k) = f(g)$. It follows from the definition of multiplication in G/K that $\widetilde{f}$ is a homomorphism. □

In the situation of the preceding theorem, we say that ***f passes to the quotient*** or ***descends to the quotient***.

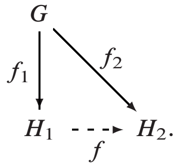
The most important fact about quotient groups is the following result, which says in essence that the projection onto a quotient group is the model for all surjective homomorphisms.

Theorem C.10 (First Isomorphism Theorem for Groups). *Suppose G and H are groups, and $f : G \to H$ is a homomorphism. Then f descends to an isomorphism from $G/\operatorname{Ker} f$ to $\operatorname{Im} f$. Thus if f is surjective, then $G/\operatorname{Ker} f$ is isomorphic to H.*

Proof. Let $K = \operatorname{Ker} f$ and $G' = \operatorname{Im} f$. From the preceding theorem, $\widetilde{f}(gK) = f(g)$ defines a homomorphism $\widetilde{f} : G/K \to G'$. Because G' is the image of f, it follows that $\widetilde{f}$ is surjective. To show that $\widetilde{f}$ is injective, suppose $1 = \widetilde{f}(gK) = f(g)$. This means that $g \in \operatorname{Ker} f = K$, so $gK = K$ is the identity element of G/K. □

► **Exercise C.11.** Suppose $f: G \to H$ is a surjective group homomorphism, and $K \subseteq G$ is a normal subgroup. Show that $f(K)$ is normal in H.

► **Exercise C.12.** Suppose $f_1: G \to H_1$ and $f_2: G \to H_2$ are group homomorphisms such that f_1 is surjective and $\operatorname{Ker} f_1 \subseteq \operatorname{Ker} f_2$. Show that there is a unique homomorphism $f: H_1 \to H_2$ such that the following diagram commutes:



Cyclic Groups

Let G be a group. If G is generated by a single element $g \in G$, then G is said to be a ***cyclic group***, and g is called a ***generator of G***. More generally, for any group G and element $g \in G$, the subgroup $\langle g \rangle = \{g^n : n \in \mathbb{Z}\} \subseteq G$ is called the ***cyclic subgroup generated by g***.

Example C.13 (Cyclic Groups).

(a) The group $\mathbb{Z}$ of integers (under addition) is an infinite cyclic group generated by 1.

(b) For any $n \in \mathbb{Z}$, the cyclic subgroup $\langle n \rangle \subseteq \mathbb{Z}$ is normal because $\mathbb{Z}$ is abelian. The quotient group $\mathbb{Z}/\langle n \rangle$ (often abbreviated $\mathbb{Z}/n$) is called the ***group of integers modulo n***. It is easily seen to be a cyclic group of order n, with the coset of 1 as a generator. //

► **Exercise C.14.** Show that every infinite cyclic group is isomorphic to $\mathbb{Z}$ and every finite cyclic group is isomorphic to $\mathbb{Z}/n$, where n is the order of the group.

► **Exercise C.15.** Show that every subgroup of a cyclic group is cyclic.

► **Exercise C.16.** Suppose G is a cyclic group and $f: G \to G$ is any homomorphism. Show there is an integer n such that $f(\gamma) = \gamma^n$ for all $\gamma \in G$. Show that if G is infinite, then n is uniquely determined by f.

References

[Apo74] Tom M. Apostol, *Mathematical Analysis*, 2nd ed., Addison-Wesley, Reading, MA, 1974.

[Bai99] René Baire, *Sur les fonctions de variables réelles*, Ann. Mat. Pur. Appl. (3) **3** (1899), 1–123.

[Bre93] Glen E. Bredon, *Topology and Geometry*, Springer-Verlag, New York, 1993.

[Con78] John B. Conway, *Functions of One Complex Variable*, 2nd ed., Springer-Verlag, New York, 1978.

[Dev93] Keith Devlin, *The Joy of Sets*, 2nd ed., Springer-Verlag, New York, 1993.

[DH07] Max Dehn and Poul Heegaard, *Analysis situs*, Enzyklopädie der Math. Wiss. **III.1.1** (1907), 153–220.

[Die89] Jean Dieudonné, *A History of Algebraic and Differential Topology: 1900–1960*, Birkhäuser Boston, Boston, 1989.

[Fre82] Michael Hartley Freedman, *The topology of four-dimensional manifolds*, J. Differ. Geom. **17** (1982), 357–453.

[GHV72] Werner Greub, Stephen Halperin, and Ray Vanstone, *Connections, Curvature, and Cohomology. Vol. I: De Rham Cohomology of Manifolds and Vector Bundles*, Academic Press, New York, 1972. Pure Appl. Math., Vol. 47.

[Hal74] Paul R. Halmos, *Naive Set Theory*, Springer-Verlag, New York, 1974.

[Hat02] Allen Hatcher, *Algebraic Topology*, Cambridge University Press, Cambridge, UK, 2002.

[Her96] I. N. Herstein, *Abstract Algebra*, 3rd ed., Prentice Hall, Upper Saddle River, NJ, 1996.

[Hun97] Thomas W. Hungerford, *Abstract Algebra: An Introduction*, 2nd ed., Saunders College Publishing, Fort Worth, TX, 1997.

[Lee02] John M. Lee, *Introduction to Smooth Manifolds*, Springer-Verlag, New York, 2002.

[Lee97] ———, *Riemannian Manifolds: An Introduction to Curvature*, Springer-Verlag, New York, 1997.

[LMM95] Silvio Levy, Delle Maxwell, and Tamara Munzer, *Outside in*, University of Minnesota Geometry Center (video), Minneapolis, MN, 1995.

[Mas77] William S. Massey, *Algebraic Topology: An Introduction*, Springer-Verlag, New York, 1977.

[Max77] Nelson Max, *Turning a Sphere Inside Out*, International Film Bureau (video), Chicago, 1977.

[Men10] Elliott Mendelson, *Introduction to Mathematical Logic*, 5th ed., Discrete Mathematics and Its Applications, CRC Press, Boca Raton, FL, 2010.

[Moi77] Edwin E. Moise, *Geometric Topology in Dimensions* 2 *and* 3, Springer-Verlag, New York, 1977.

[Mon69] J. Donald Monk, *Introduction to Set Theory*, McGraw-Hill, New York, 1969.

[MT07] John Morgan and Gang Tian, *Ricci Flow and the Poincaré Conjecture*, Clay Mathematics Monographs, vol. 3, American Mathematical Society, Providence, RI, 2007.

[Mun00] James R. Munkres, *Topology*, 2nd ed., Prentice–Hall, Upper Saddle River, NJ, 2000.

[Mun84] ______, *Elements of Algebraic Topology*, Addison–Wesley, Menlo Park, CA, 1984.

[PM07] A. O. Prishlyak and K. I. Mischenko, *Classification of noncompact surfaces with boundary*, Methods Funct. Anal. Topology **13** (2007), no. 1, 62–66.

[Rad25] Tibor Radó, *Über den Begriff der Riemannschen Fläche*, Acta Litt. Sci. Szeged. **2** (1925), 101–121.

[Ran96] A. A. Ranicki, *On the Hauptvermutung*, The Hauptvermutung Book, Kluwer Academic, Dordrecht, 1996, pp. 3–31.

[Ric63] Ian Richards, *On the classification of noncompact surfaces*, T. Am. Math. Soc. **106** (1963), 259–269.

[Rud76] Walter Rudin, *Principles of Mathematical Analysis*, 3rd ed., McGraw-Hill, New York, 1976.

[Rud79] Mary Ellen Rudin, *The undecidability of the existence of a perfectly normal nonmetrizable manifold*, Houston J. Math. **5** (1979), no. 2, 249–252.

[SFL98] John M. Sullivan, George Francis, and Stuart Levy, *The optiverse*, National Center for Supercomputing Applications (video), Champaign–Urbana, IL, 1998.

[Sie92] Allan J. Sieradski, *An Introduction to Topology and Homotopy*, PWS-KENT, Boston, 1992.

[Sma58] Stephen Smale, *A classification of immersions of the two-sphere*, T. Am. Math. Soc. **90** (1958), 281–290.

[Sma61] ______, *Generalized Poincaré conjecture in dimensions greater than 4*, Ann. Math. **74** (1961), 391–406.

[Spa81] Edwin H. Spanier, *Algebraic Topology*, Springer-Verlag, New York, 1981.

[Sti82] John Stillwell, *The word problem and the isomorphism problem for groups*, B. Am. Math. Soc. (N.S.) **6** (1982), 33–56.

[Sti93] ______, *Classical Topology and Combinatorial Group Theory*, 2nd ed., Springer-Verlag, New York, 1993.

[Sto79] Robert R. Stoll, *Set Theory and Logic*, Dover, New York, 1979.

[Sup72] Patrick Suppes, *Axiomatic Set Theory*, Dover, New York, 1972.

[Tho92] Carsten Thomassen, *The Jordan-Schönflies theorem and the classification of surfaces*, Am. Math. Mon. **99** (1992), no. 2, 116–130.

[Whi78] George W. Whitehead, *Elements of Homotopy Theory*, Springer-Verlag, New York, 1978.

Notation Index

Subject Index