Table of Contents

Chapter 1

Set theory and terminology

The principle purpose of this chapter is to introduce the mathematical notation and language that will be used in the remainder of these volumes. Much of this notation is standard, or at least the notation we use is generally among a collection of standard possibilities. In this respect, the chapter is a simple one. However, we also wish to introduce the reader to some elementary, although somewhat abstract, mathematics. The secondary objective behind this has three components.

1. We aim to provide a somewhat rigorous foundation for what follows. This means being fairly clear about defining the (usually) somewhat simple concepts that arise in the chapter. Thus "intuitively clear" concepts like sets, subsets, maps, etc., are given a fairly systematic and detailed discussion. It is at least interesting to know that this can be done. And, if it is not of interest, it can be sidestepped at a first reading.
2. This chapter contains some results, and many of these require very simple proofs. We hope that these simple proofs might be useful to readers who are new to the world where everything is proved. Proofs in other chapters in these volumes may not be so useful for achieving this objective.
3. The material is standard mathematical material, and should be known by anyone purporting to love mathematics.

Do I need to read this chapter? Readers who are familiar with standard mathematical notation (e.g., who understand the symbols $\in$, $\subset$, $\cup$, $\cap$, $\times$, $f\colon S \to T$, $\mathbb{N}$, and $\mathbb{Z}$) can simply skip this chapter in its entirety. Some ideas (e.g., relations, orders, Zorn's Lemma) may need to be referred to during the course of later chapters, but this is easily done.

Readers not familiar with the above standard mathematical notation will have some work to do. They should certainly read Sections 1.1, 1.2, and 1.3 closely enough that they understand the language, notation, and main ideas. And they should read enough of Section 1.4 that they know what objects, familiar to them from their being human, the symbols $\mathbb{N}$ and $\mathbb{Z}$ refer to. The remainder of the material can be overlooked until it is needed later. •

Contents

Section 1.1

Sets

The basic ingredient in modern mathematics is the set. The idea of a set is familiar to everyone at least in the form of "a collection of objects." In this section, we shall not really give a definition of a set that excels that intuitive one. Rather we shall accept this intuitive idea of a set, and move forward from there. This way of dealing with sets is called ***naïve set theory***. There are some problems with naïve set theory, as described in Section 1.8.1, and these lead to a more formal notion of a set as an object that satisfies certain axioms, those given in Section 1.8.2. However, these matters will not concern us much at the moment.

Do I need to read this section? Readers familiar with basic set theoretic notation can skip this section. Other readers should read it, since it contains language, notation, and ideas that are absolutely commonplace in these volumes. •

1.1.1 Definitions and examples

First let us give our working definition of a set. A ***set*** is, for us, a well-defined collection of objects. Thus one can speak of everyday things like "the set of red-haired ladies who own yellow cars." Or one can speak of mathematical things like "the set of even prime numbers." Sets are therefore defined by describing their ***members*** or ***elements***, i.e., those objects that are in the set. When we are feeling less formal, we may refer to an element of a set as a ***point*** in that set. The set with no members is the ***empty set***, and is denoted by $\emptyset$. If S is a set with member x, then we write $x \in S$. If an object x is *not* in a set S, then we write $x \notin S$.

1.1.1 Examples (Sets)

1. If S is the set of even prime numbers, then $2 \in S$.
2. If S is the set of even prime numbers greater than 3, then S is the empty set.
3. If S is the set of red-haired ladies who own yellow cars and if $x =$ Ghandi, then $x \notin S$. •

If it is possible to write the members of a set, then they are usually written between braces { }. For example, the set of prime numbers less that 10 is written as $\{2, 3, 5, 7\}$ and the set of physicists to have won a Fields Prize as of 2005 is {Edward Witten}.

A set S is a ***subset*** of a set T if $x \in S$ implies that $x \in T$. We shall write $S \subset T$, or equivalently $T \supset S$, in this case. If $x \in S$, then the set $\{x\} \subset S$ with one element, namely x, is a ***singleton***. Note that x and $\{x\}$ are different things. For example, $x \in S$ and $\{x\} \subset S$. If $S \subset T$ and if $T \subset S$, then the sets S and T are ***equal***, and we write $S = T$. If two sets are not equal, then we write $S \neq T$. If $S \subset T$ and if $S \neq T$, then S is a ***proper*** subset of T, and we write $S \subsetneq T$ if we wish to emphasise this fact. Some of the following examples may not be perfectly obvious, so may require sorting through the definitions.

1.1.2 Examples (Subsets)

1. For any set S, $\emptyset \subset S$ (see Exercise 1.1.1).
2. $\{1, 2\} \subset \{1, 2, 3\}$.
3. $\{1, 2\} \subsetneq \{1, 2, 3\}$.
4. $\{1, 2\} = \{2, 1\}$.

5. $\{1,2\} = \{2,1,2,1,1,2\}$. •

A common means of defining a set is to define it as the subset of an existing set that satisfies conditions. Let us be slightly precise about this. A ***one-variable predicate*** is a statement which, in order that its truth be evaluated needs a single argument to be specified. For example, $P(x)$ = "x is blue" needs the single argument x in order that it be decided whether it is true or not. We then use the notation

$$\{x \in S \mid \ P(x)\}$$

to denote the members x of S for which the predicate P is true when evaluated at x. This is read as something like, "the set of x's in S such that $P(x)$ holds."

For sets S and T, the ***relative complement*** of T in S is the set

$$S - T = \{x \in S \mid \ x \notin T\}.$$

Note that for this to make sense, we do not require that T be a subset of S. It is a common occurrence when dealing with complements that one set be a subset of another. We use different language and notation to deal with this. If S is a set and if $T \subset S$, then $S \setminus T$ denotes the ***absolute complement*** of T in S, and is defined by

$$S \setminus T = \{x \in S \mid \ x \notin T\}.$$

Note that, if we forget that T is a subset of S, then we have $S \setminus T = S - T$. Thus $S - T$ is the more general notation. Of course, if $A \subset T \subset S$, one needs to be careful when using the words "absolute complement of A," since one must say whether one is taking the complement in T or the larger complement in S. For this reason, we prefer the notation we use rather the commonly encountered notation A^{C} or A' to refer to the absolute complement. Note that one should not talk about the absolute complement to a set, without saying within which subset the complement is being taken. To do so would imply the existence of "a set containing all sets," an object that leads one to certain paradoxes (see Section 1.8).

A useful set associated with every set S is its ***power set***, by which we mean the set

$$\mathbf{2}^S = \{A \mid \ A \subset S\}.$$

The reader can investigate the origins of the peculiar notation in Exercise 1.1.2.

1.1.2 Unions and intersections

In this section we indicate how to construct new sets from existing ones.

Given two sets S and T, the ***union*** of S and T is the set $S \cup T$ whose members are members of S *or* T. The ***intersection*** of S and T is the set $S \cap T$ whose members are members of S *and* T. If two sets S and T have the property that $S \cap T = \emptyset$, then S and T are said to be ***disjoint***. For sets S and T their ***symmetric complement*** is the set

$$S \triangle T = (S - T) \cup (T - S).$$

Thus $S \triangle T$ is the set of objects in union $S \cup T$ that do not lie in the intersection $S \cap T$. The symmetric complement is so named because $S \triangle T = T \triangle S$. In Figure 1.1 we give Venn diagrams describing union, intersection, and symmetric complement.

The following result gives some simple properties of pairwise unions and intersections of sets. We leave the straightforward verification of some or all of these to the reader as Exercise 1.1.3.

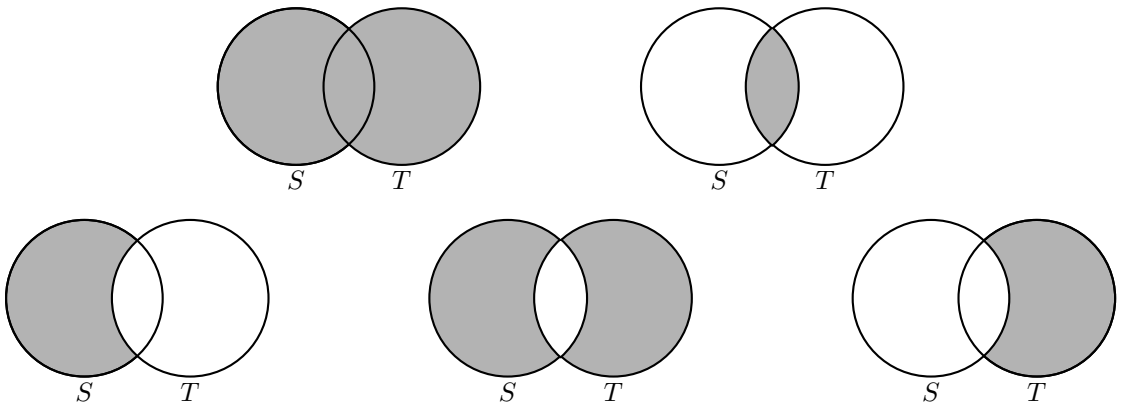


Figure 1.1 $S \cup T$ (top left), $S \cap T$ (top right), $S - T$ (bottom left), $S \triangle T$ (bottom middle), and $T - S$ (bottom right)

1.1.3 Proposition (Properties of unions and intersections) *For sets* S *and* T*, the following statements hold:*

(i) $\mathrm{S} \cup \emptyset = \mathrm{S}$*;*

(ii) $\mathrm{S} \cap \emptyset = \emptyset$*;*

(iii) $\mathrm{S} \cup \mathrm{S} = \mathrm{S}$*;*

(iv) $\mathrm{S} \cap \mathrm{S} = \mathrm{S}$*;*

(v) $\mathrm{S} \cup \mathrm{T} = \mathrm{T} \cup \mathrm{S}$ *(**commutativity**);*

(vi) $\mathrm{S} \cap \mathrm{T} = \mathrm{T} \cap \mathrm{S}$ *(**commutativity**);*

(vii) $\mathrm{S} \subset \mathrm{S} \cup \mathrm{T}$*;*

(viii) $\mathrm{S} \cap \mathrm{T} \subset \mathrm{S}$*;*

(ix) $\mathrm{S} \cup (\mathrm{T} \cup \mathrm{U}) = (\mathrm{S} \cup \mathrm{T}) \cup \mathrm{U}$ *(**associativity**);*

(x) $\mathrm{S} \cap (\mathrm{T} \cap \mathrm{U}) = (\mathrm{S} \cap \mathrm{T}) \cap \mathrm{U}$ *(**associativity**);*

(xi) $\mathrm{S} \cap (\mathrm{T} \cup \mathrm{U}) = (\mathrm{S} \cap \mathrm{T}) \cup (\mathrm{S} \cap \mathrm{U})$ *(**distributivity**);*

(xii) $\mathrm{S} \cup (\mathrm{T} \cap \mathrm{U}) = (\mathrm{S} \cup \mathrm{T}) \cap (\mathrm{S} \cup \mathrm{U})$ *(**distributivity**).*

We may more generally consider not just two sets, but an arbitrary collection $\mathscr{S}$ of sets. In this case we *posit* the existence of a set, called the ***union*** of the sets $\mathscr{S}$, with the property that it contains each element of each set $S \in \mathscr{S}$. Moreover, one can specify the subset of this big set to *only* contain members of sets from $\mathscr{S}$. This set we will denote by $\cup_{S \in \mathscr{S}} S$. We can also perform a similar construction with intersections of an arbitrary collection $\mathscr{S}$ of sets. Thus we denote by $\cap_{S \in \mathscr{S}} S$ the set, called the ***intersection*** of the sets $\mathscr{S}$, having the property that $x \in \cap_{S \in \mathscr{S}} S$ if $x \in S$ for every $S \in \mathscr{S}$. Note that we do not need to posit the existence of the intersection.

Let us give some properties of general unions and intersections as they relate to complements.

1.1.4 Proposition (De Morgan's[1] Laws) *Let* T *be a set and let* $\mathscr{S}$ *be a collection of subsets of* T*. Then the following statements hold:*

(i) $\mathrm{T} \setminus (\cup_{\mathrm{S} \in \mathscr{S}} \mathrm{S}) = \cap_{\mathrm{S} \in \mathscr{S}} (\mathrm{T} \setminus \mathrm{S})$*;*

[1]Augustus De Morgan (1806–1871) was a British mathematician whose principal mathematical contributions were to analysis and algebra

(ii) $\mathrm{T} \setminus (\cap_{\mathrm{S} \in \mathscr{S}} \mathrm{S}) = \cup_{\mathrm{S} \in \mathscr{S}} (\mathrm{T} \setminus \mathrm{S})$.

Proof (i) Let $x \in T \setminus (\cup_{S \in \mathscr{S}})$. Then, for each $S \in \mathscr{S}$, $x \notin S$, or $x \in T \setminus S$. Thus $x \in \cap_{S \in \mathscr{S}} (T \setminus S)$. Therefore, $T \setminus (\cup_{S \in \mathscr{S}}) \supset \cap_{S \in \mathscr{S}} (T \setminus S)$. Conversely, if $x \in \cap_{S \in \mathscr{S}} (T \setminus S)$, then, for each $S \in \mathscr{S}$, $x \notin S$. Therefore, $x \notin \cup_{S \in \mathscr{S}}$. Therefore, $x \in T \setminus (\cup_{S \in \mathscr{S}})$, thus showing that $\cap_{S \in \mathscr{S}} (T \setminus S) \subset T \setminus (\cup_{S \in \mathscr{S}})$. It follows that $T \setminus (\cup_{S \in \mathscr{S}}) = \cap_{S \in \mathscr{S}} (T \setminus S)$.

(ii) This follows in much the same manner as part (i), and we leave the details to the reader. ■

1.1.5 Remark (Showing two sets are equal) Note that in proving part (i) of the preceding result, we proved two things. First we showed that $T \setminus (\cup_{S \in \mathscr{S}}) \subset \cap_{S \in \mathscr{S}} (T \setminus S)$ and then we showed that $\cap_{S \in \mathscr{S}} (T \setminus S) \subset T \setminus (\cup_{S \in \mathscr{S}})$. This is the standard means of showing that two sets are equal; first show that one is a subset of the other, and then show that the other is a subset of the one. •

For general unions and intersections, we also have the following generalisation of the distributive laws for unions and intersections. We leave the straightforward proof to the reader (Exercise 1.1.4)

1.1.6 Proposition (Distributivity laws for general unions and intersections) *Let* T *be a set and let* $\mathscr{S}$ *be a collection of sets. Then the following statements hold:*

(i) $\mathrm{T} \cap (\cup_{\mathrm{S} \in \mathscr{S}} \mathrm{S}) = \cup_{\mathrm{S} \in \mathrm{S}} (\mathrm{T} \cap \mathrm{S})$*;*

(ii) $\mathrm{T} \cup (\cap_{\mathrm{S} \in \mathscr{S}} \mathrm{S}) = \cap_{\mathrm{S} \in \mathrm{S}} (\mathrm{T} \cup \mathrm{S})$.

There is an alternative notion of the union of sets, one that retains the notion of membership in the original set. The issue that arises is this. If $S = \{1, 2\}$ and $T = \{2, 3\}$, then $S \cup T = \{1, 2, 3\}$. Note that we lose with the usual union the fact that 1 is an element of S only, but that 2 is an element of both S and T. Sometimes it is useful to retain these sorts of distinctions, and for this we have the following definition.

1.1.7 Definition (Disjoint union) For sets S and T, their ***disjoint union*** is the set

$$S \mathring{\cup} T = \{ (S, x) \mid \ x \in S \} \cup \{ (T, y) \mid \ y \in T \}. \qquad \bullet$$

Let us see how the disjoint union differs from the usual union.

1.1.8 Example (Disjoint union) Let us again take the simple example $S = \{1, 2\}$ and $T = \{2, 3\}$. Then $S \cup T = \{1, 2, 3\}$ and

$$S \mathring{\cup} T = \{(S, 1), (S, 2), (T, 2), (T, 3)\}.$$

We see that the idea behind writing an element in the disjoint union as an ordered pair is that the first entry in the ordered pair simply keeps track of the set from which the element in the disjoint union was taken. In this way, if $S \cap T \neq \emptyset$, we are guaranteed that there will be no "collapsing" when the disjoint union is formed. •

1.1.3 Finite Cartesian products

As we have seen, if S is a set and if $x_1, x_2 \in S$, then $\{x_1, x_2\} = \{x_2, x_1\}$. There are times, however, when we wish to keep track of the order of elements in a set. To accomplish this and other objectives, we introduce the notion of an ordered pair. First, however, in order to make sure that we understand the distinction between ordered and unordered pairs, we make the following definition.

1.1.9 Definition (Unordered pair) If S is a set, an ***unordered pair*** from S is any subset of S with two elements. The collection of unordered pairs from S is denoted by $S^{(2)}$. •

Obviously one can talk about unordered collections of more than two elements of a set, and the collection of subsets of a set S comprised of k elements is denoted by $S^{(k)}$ and called the set of ***unordered* k-*tuples***.

With the simple idea of an unordered pair, the notion of an ordered pair is more distinct.

1.1.10 Definition (Ordered pair and Cartesian product) Let S and T be sets, and let $x \in S$ and $y \in T$. The ***ordered pair*** of x and y is the set $(x, y) = \{\{x\}, \{x, y\}\}$. The ***Cartesian product*** of S and T is the set

$$S \times T = \{\, (x, y) \mid \ x \in S,\ y \in T \}\,. \qquad \bullet$$

The definition of the ordered pair seems odd at first. However, it is as it is to secure the objective that if two ordered pairs (x_1, y_1) and (x_2, y_2) are equal, then $x_1 = x_2$ and $y_1 = y_2$. The reader can check in Exercise 1.1.6 that this objective is in fact achieved by the definition. It is also worth noting that the form of the ordered pair as given in the definition is seldom used after its initial introduction.

Clearly one can define the Cartesian product of any finite number of sets $S_1, \ldots, S_k$ inductively. Thus, for example, $S_1 \times S_2 \times S_3 = (S_1 \times S_2) \times S_3$. Note that, according to the notation in the definition, an element of $S_1 \times S_2 \times S_3$ should be written as $((x_1, x_2), x_3)$. However, it is immaterial that we define $S_1 \times S_2 \times S_3$ as we did, or as $S_1 \times S_2 \times S_3 = S_1 \times (S_2 \times S_3)$. Thus we simply write elements in $S_1 \times S_2 \times S_3$ as (x_1, x_2, x_3), and similarly for a Cartesian product $S_1 \times \cdots \times S_k$. The Cartesian product of a set with itself k-times is denoted by S^k. That is,

$$S^k = \underbrace{S \times \cdots \times S}_{k\text{-times}}.$$

In Section 1.6.1 we shall indicate how to define Cartesian products of more than finite collections of sets.

Let us give some simple examples.

1.1.11 Examples (Cartesian products)

1. If S is a set then note that $S \times \emptyset = \emptyset$. This is because there are no ordered pairs from S and $\emptyset$. It is just as clear that $\emptyset \times S = \emptyset$. It is also clear that, if $S \times T = \emptyset$, then either $S = \emptyset$ or $T = \emptyset$.
2. If $S = \{1, 2\}$ and $T = \{2, 3\}$, then

$$S \times T = \{(1, 2), (1, 3), (2, 2), (2, 3)\}. \qquad \bullet$$

Cartesian products have the following properties.

1.1.12 Proposition (Properties of Cartesian product) *For sets* S, T, U, *and* V, *the following statements hold:*

(i) $(\mathrm{S} \cup \mathrm{T}) \times \mathrm{U} = (\mathrm{S} \times \mathrm{U}) \cup (\mathrm{T} \times \mathrm{U})$;

(ii) $(\mathrm{S} \cap \mathrm{U}) \times (\mathrm{T} \cap \mathrm{V}) = (\mathrm{S} \times \mathrm{T}) \cap (\mathrm{U} \times \mathrm{V})$;

(iii) $(\mathrm{S} - \mathrm{T}) \times \mathrm{U} = (\mathrm{S} \times \mathrm{U}) - (\mathrm{T} \times \mathrm{U})$.

Proof Let us prove only the first identity, leaving the remaining two to the reader. Let $(x, u) \in (S \cup T) \times U$. Then $x \in S \cup T$ and $u \in U$. Therefore, x is an element of at least one of S and T. Without loss of generality, suppose that $x \in S$. Then $(x, u) \in S \times U$ and so $(x, u) \in (S \times U) \cup (T \times U)$. Therefore, $(S \cup T) \times U = (S \times U) \cup (T \times U)$. Conversely, suppose that $(x, u) \in (S \times U) \cup (T \times U)$. Without loss of generality, suppose that $(x, u) \in S \times U$. Then $x \in S \subset S \cup T$ and $u \in U$. Therefore, $(x, u) \in (S \cup T) \times U$. Thus $(S \times U) \cup (T \times U) \subset (S \cup T) \times U$, giving the result. ■

1.1.13 Remark ("Without loss of generality") In the preceding proof, we twice employed the expression "without loss of generality." This is a commonly encountered expression, and is frequently used in one of the following two contexts. The first, as above, indicates that one is making an arbitrary selection, but that were another arbitrary selection to have been made, the same argument holds. This is a more or less straightforward use of "without loss of generality." A more sophisticated use of the expression might indicate that one is making a simplifying assumption, and that this is okay, because it can be shown that the general case follows easily from the simpler one. The trick is to then understand *how* the general case follows from the simpler one, and this can sometimes be nontrivial, depending on the willingness of the writer to describe this process. •

Exercises

1.1.1 Prove that the empty set is a subset of every set.
Hint: *Assume the converse, and arrive at an absurdity.*

1.1.2 If S is a set with n members, show that $\mathbf{2}^S$ is a set with 2^n members.

1.1.3 Prove as many parts of Proposition 1.1.3 as you wish.

1.1.4 Prove Proposition 1.1.6.

1.1.5 Let S be a set with n members and let T be a set with m members. Show that $S \mathbin{\mathring{\cup}} T$ is a set with nm members.

1.1.6 Let S and T be sets, let $x_1, x_2 \in S$, and let $y_1, y_2 \in T$. Show that $(x_1, y_1) = (x_2, y_2)$ if and only if $x_1 = x_2$ and $y_1 = y_2$.

Section 1.2

Relations

Relations are a fundamental ingredient in the description of many mathematical ideas. One of the most valuable features of relations is that they allow many useful constructions to be explicitly made only using elementary ideas from set theory.

Do I need to read this section? The ideas in this section will appear in many places in the series, so this material should be regarded as basic. However, readers looking to proceed with minimal background can skip the section, referring back to it when needed. •

1.2.1 Definitions

We shall describe in this section "binary relations," or relations between elements of two sets. It is possible to define more general sorts of relations where more sets are involved. However, these will not come up for us.

1.2.1 Definition (Relation) A ***binary relation from*** **S** ***to*** **T** (or simply a ***relation from*** **S** ***to*** **T** is a subset of $S \times T$. If $R \subset S \times T$ and if $(x, y) \in R$, then we shall write $x\,R\,y$, meaning that x and y are related by R. A relation from S to S is a ***relation in*** **S**. •

The definition is simple. Let us give some example to give it a little texture.

1.2.2 Examples (Relations)

1. Let S be the set of husbands and let T be the set of wives. Define a relation R from S to T by asking that $(x, y) \in R$ if x is married to y. Thus, to say that x and y are related in this case means to say that x is married to y.
2. Let S be a set and consider the relation R in the power set $\mathbf{2}^S$ of S given by

$$R = \{(A, B) \mid \ A \subset B\}.$$

Thus A is related to B if A is a subset of B.

3. Let S be a set and define a relation R in S by

$$R = \{(x, x) \mid \ x \in S\}.$$

Thus, under this relation, two members in S are related if and only if they are equal.

4. Let S be the set of integers, let k be a positive integer, and define a relation R_k in S by

$$R_k = \{(n_1, n_2) \mid \ n_1 - n_2 = k\}.$$

Thus, if $n \in S$, then all integers of the form $n + mk$ for an integer m are related to n. •

1.2.3 Remark ("If" versus "if and only if") In part 3 of the preceding example we used the expression "if and only if" for the first time. It is, therefore, worth saying a few words about this commonly used terminology. One says that statement A holds "if and only if" statement B holds to mean that statements A and B are exactly equivalent. Typically, this language arises in theorem statements. In proving such theorems, it is important to note that one must prove *both* that statement A implies statement B *and* that statement B implies statement A.

To confuse matters, when stating a definition, the convention is to use "if" rather than "if and only if," even though it is the case, by the very meaning of the word "definition," that "if and only if" is also proper. However, it is usually taken as understood in a definition. •

In the next section we will encounter the notion of the inverse of a function; this idea is perhaps known to the reader. However, the notion of inverse also applies to the more general setting of relations.

1.2.4 Definition (Inverse of a relation) If $R \subset S \times T$ is a relation from S to T, then the ***inverse*** of R is the relation R^{-1} from T to S defined by

$$R^{-1} = \{\, (y, x) \in T \times S \mid \ (x, y) \in R \}\,. \qquad \bullet$$

There are a variety of properties that can be bestowed upon relations to ensure they have certain useful attributes. The following is a partial list of such properties.

1.2.5 Definition (Properties of relations) Let S be a set and let R be a relation in S. The relation R is:

(i) ***reflexive*** if $(x, x) \in R$ for each $x \in S$;

(ii) ***irreflexive*** if $(x, x) \notin R$ for each $x \in S$;

(iii) ***symmetric*** if $(x_1, x_2) \in R$ implies that $(x_2, x_1) \in R$;

(iv) ***antisymmetric*** if $(x_1, x_2) \in R$ and $(x_2, x_1) \in R$ implies that $x_1 = x_2$;

(v) ***transitive*** if $(x_1, x_2) \in R$ and $(x_2, x_3) \in R$ implies that $(x_1, x_3) \in R$. •

1.2.6 Examples (Example 1.2.2 cont'd)

1. The relation of inclusion in the power set $\mathbf{2}^S$ of a set S is reflexive, antisymmetric, and transitive.
2. The relation of equality in a set S is reflexive, symmetric, antisymmetric, and transitive.
3. The relation R_k in the set S of integers is reflexive, symmetric, and transitive. •

1.2.2 Equivalence relations

In this section we turn our attention to an important class of relations, and we indicate why these are important by giving them a characterisation in terms of a decomposition of a set.

1.2.7 Definition (Equivalence relation, equivalence class) An ***equivalence relation*** in a set S is a relation R that is reflexive, symmetric, and transitive. For $x \in S$, the set of elements of S related to x is denoted by $[x]$, and is the ***equivalence class*** of x with respect to R. An element x' is an equivalence class $[x]$ is a ***representative*** of that equivalence class. The set of equivalence classes is denoted by S/R (typically pronounced as **S** ***modulo*** **R**). •

It is common to denote that two elements $x_1, x_2 \in S$ are related by an equivalence relation by writing $x_1 \sim x_2$. Of the relations defined in Example 1.2.2, we see that those in parts 3 and 4 are equivalence relations, but that in part 2 is not.

Let us now characterise equivalence relations in a more descriptive manner. We begin by defining a (perhaps seemingly unrelated) notion concerning subsets of a set.

1.2.8 Definition (Partition of a set) A ***partition*** of a set S is a collection $\mathscr{A}$ of subsets of S having the properties that

(i) two distinct subsets in $\mathscr{A}$ are disjoint and

(ii) $S = \cup_{A \in \mathscr{A}} A$. •

We now prove that there is an exact correspondence between equivalence classes associated to an equivalence relation.

1.2.9 Proposition *Let* S *be a set and let* R *be an equivalence relation in* S. *Then the set of equivalence classes with respect to* R *is a partition of* S.

Conversely, if $\mathscr{A}$ is a partition of S, *then the relation*

$$\{(x_1, x_2) \mid \ x_1, x_2 \in A \text{ for some } A \in \mathscr{A}\}$$

is an equivalence relation in S.

Proof We first claim that two distinct equivalence classes are disjoint. Thus we let $x_1, x_2 \in S$ and suppose that $[x_1] \neq [x_2]$. Suppose that $x \in [x_1] \cap [x_2]$. Then $x \sim x_1$ and $x \sim x_2$, or, by transitivity of R, $x_1 \sim x$ and $x \sim x_2$. By transitivity of R, $x_1 \sim x_2$, contradicting the fact that $[x_1] \neq [x_2]$. To show that S is the union of its equivalence classes, merely note that, for each $x \in S$, $x \in [x]$ by reflexivity of R.

Now let $\mathscr{A}$ be a partition and defined R as in the statement of the proposition. Let $x \in S$ and let A be the element of $\mathscr{A}$ that contains x. Then clearly we see that $(x, x) \in R$ since $x \in A$. Thus R is reflexive. Next let $(x_1, x_2) \in R$ and let A be the element of $\mathscr{A}$ such that $x_1, x_2 \in A$. Clearly then, $(x_2, x_1) \in R$, so R is symmetric. Finally, let $(x_1, x_2), (x_2, x_3 \in R$. Then there are elements $A_{12}, A_{23} \in \mathscr{A}$ such that $x_1, x_2 \in A_{12}$ and such that $x_2, x_3 \in A_{23}$. Since A_{12} and A_{23} have the point x_2 in common, we must have $A_{12} = A_{23}$. Thus $(x_1, x_3 \in A_{12} = A_{23}$, giving transitivity of R. ■

Exercises

1.2.1 In a set S define a relation $R = \{(x, y) \in S \times S \mid \ x = y\}$.

(a) Show that R is an equivalence relation.

(b) Show that $S/R = S$.

Section 1.3

Maps

Another basic concept in all of mathematics is that of a map between sets. Indeed, many of the interesting objects in mathematics are maps of some sort. In this section we review the notation associated with maps, and give some simple properties of maps.

Do I need to read this section? The material in this section is basic, and will be used constantly throughout the series. Unless you are familiar already with maps and the notation associated to them, this section is essential reading. •

1.3.1 Definitions and notation

We begin with the definition.

1.3.1 Definition (Map) For sets S and T, a ***map*** from S to T is a relation R from S to T having the property that, for each $x \in S$, there exists a unique $y \in T$ such that $(x, y) \in R$. The set S is the ***domain*** of the map and the set T is the ***codomain*** of the map. The set of maps from S to T is denoted by T^S.[2] •

By definition, a map is a relation. This is not how one most commonly thinks about a map, although the definition serves to render the concept of a map in terms of concepts we already know. Suppose one has a map from S to T defined by a relation R. Then, given $x \in S$, there is a single $y \in T$ such that x and y are related. Denote this element of T by $f(x)$, since it is defined by x. When one refers to a map, one more typically refers to the assignment of the element $f(x) \in T$ to $x \in S$. Thus one refers to the map as f, leaving aside the baggage of the relation as in the definition. Indeed, this is how we from now on will think of maps. The definition above does, however, have some use, although we alter our language, since we are now thinking of a map as an "assignment." We call the set

$$\text{graph}(f) = \{(x, f(x)) \mid \ x \in S\}$$

(which we originally called the map in Definition 1.3.1) the ***graph*** of the map $f\colon S \to T$.

If one wishes to indicate a map f with domain S and codomain T, one typically writes $f\colon S \to T$ to compactly express this. If one wishes to *define* a map by saying what it does, the notation

$$\begin{aligned} f\colon\ & S \to T \\ & x \mapsto \text{what } x \text{ gets mapped to} \end{aligned}$$

is sometimes helpful. Sometimes we shall write this in the text as $f\colon x \mapsto$ "what x gets mapped to". Note the distinct uses of the symbols "$\to$" and "$\mapsto$".

1.3.2 Notation (f versus f(x)) Note that a map is denoted by "f". It is quite common to see the expression "consider the map $f(x)$". Taken literally, these words are difficult to comprehend. First of all, x is unspecified. Second of all, even if x were specified, $f(x)$ is an element of T, not a map. Thus it is considered bad form mathematically to use an

[2] The idea behind this notation is the following. A map from S to T assigns to each point in S a point in T. If S and T are finite sets with k and l elements, respectively, then there are l possible values that can be assigned to each of the k elements of S. Thus the set of maps has l^k elements.

expression like "consider the map $f(x)$". However, there are times when it is quite convenient to use this poor notation, with an understanding that some compromises are being made. For instance, in this volume, we will be frequently dealing simultaneously with functions of both time (typically denoted by t) and frequency (typically denoted by ν). Thus it would be convenient to write "consider the map $f(t)$" when we wish to write a map that we are considering as a function of time, and similarly for frequency. Nonetheless, we shall refrain from doing this, and shall consistently use the mathematically precise language "consider the map f". •

The following is a collection of examples of maps. Some of these examples are not just illustrative, but also define concepts and notation that we will use throughout the series.

1.3.3 Examples (Maps)

1. There are no maps having $\emptyset$ as a domain or codomain since there are no elements in the empty set.
2. If S is a set and if $T \subset S$, then the map $i_T \colon T \to S$ defined by $i_T(x) = x$ is called the ***inclusion*** of T in S.
3. The inclusion map $i_S \colon S \to S$ of a set S into itself (since $S \subset S$) is the ***identity map***, and we denote it by id_S.
4. If $f \colon S \to T$ is a map and if $A \subset S$, then the map from A to T which assigns to $x \in A$ the value $f(x) \in T$ is called the ***restriction*** of f to A, and is denoted by $f|A \colon A \to T$.
5. If S is a set with $A \subset S$, then the map χ_A from S to the integers defined by
$$\chi_A(x) = \begin{cases} 1, & x \in A, \\ 0, & x \notin A, \end{cases}$$
is the ***characteristic function*** of A.
6. If $S_1, \ldots, S_k$ are sets, if $S_1 \times \cdots \times S_k$ is the Cartesian product, and if $j \in \{1, \ldots, k\}$, then the map
$$\begin{aligned} \mathrm{pr}_j \colon\ & S_1 \times \cdots \times S_j \times \cdots \times S_k \to\ S_j \\ & (x_1, \ldots, x_j, \ldots, x_k) \mapsto\ x_j \end{aligned}$$
is the ***projection onto the jth factor***.
7. If R is an equivalence relation in a set S, then the map $\pi_R \colon S \to S/R$ defined by $\pi_R(x) = [x]$ is called the ***canonical projection*** associated to R.
8. If S, T, and U are sets and if $f \colon S \to T$ and $g \colon T \to U$ are maps, then we define a map $g \circ f \colon S \to U$ by $g \circ f(x) = g(f(x))$. This is the ***composition*** of f and g. •

Next we introduce the notions of images and preimages of points and sets.

1.3.4 Definition (Image and preimage) Let S and T be sets and let $f \colon S \to T$ be a map.

(i) If $A \subset S$, then $f(A) = \{ f(x) \mid\ x \in A \}$.

(ii) The ***image*** of f is the set $\mathrm{image}(f) = f(S) \subset T$.

(iii) If $B \subset T$, then $f^{-1}(B) = \{ x \in S \mid\ f(x) \in B \}$ is the ***preimage*** of B under f. If $B = \{y\}$ for some $y \in T$, then we shall often write $f^{-1}(y)$ rather that $f^{-1}(\{y\})$. •

Note that one can think of f as being a map from $\mathbf{2}^S$ to $\mathbf{2}^T$ and of f^{-1} as being a map from $\mathbf{2}^T$ to $\mathbf{2}^S$. Here are some elementary properties of f and f^{-1} thought of in this way.

1.3.5 Proposition (Properties of images and preimages) *Let* S *and* T *be sets, let* $\mathrm{f}\colon \mathrm{S} \to \mathrm{T}$ *be a map, let* $\mathrm{A} \subset \mathrm{S}$ *and* $\mathrm{B} \subset \mathrm{T}$*, and let* $\mathscr{A}$ *and* $\mathscr{B}$ *be collections of subsets of* S *and* T*, respectively. Then the following statements hold:*

(i) $\mathrm{A} \subset \mathrm{f}^{-1}(\mathrm{f}(\mathrm{A}))$*;*

(ii) $\mathrm{f}(\mathrm{f}^{-1}(\mathrm{B})) \subset \mathrm{B}$*;*

(iii) $\cup_{\mathrm{A}\in\mathscr{A}}\mathrm{f}(\mathrm{A}) = \mathrm{f}(\cup_{\mathrm{A}\in\mathscr{A}}\mathrm{A})$*;*

(iv) $\cup_{\mathrm{B}\in\mathscr{B}}\mathrm{f}^{-1}(\mathrm{B}) = \mathrm{f}^{-1}(\cup_{\mathrm{B}\in\mathscr{B}}\mathrm{B})$*;*

(v) $\cap_{\mathrm{A}\in\mathscr{A}}\mathrm{f}(\mathrm{A}) = \mathrm{f}(\cap_{\mathrm{A}\in\mathscr{A}}\mathrm{A})$*;*

(vi) $\cap_{\mathrm{B}\in\mathscr{B}}\mathrm{f}^{-1}(\mathrm{B}) = \mathrm{f}^{-1}(\cap_{\mathrm{B}\in\mathscr{B}}\mathrm{B})$.

Proof We shall prove only some of these, leaving the remainder for the reader to complete.

(i) Let $x \in A$. Then $x \in f^{-1}(f(x))$ since $f(x) = f(x)$.

(iii) Let $y \in \cup_{A\in\mathscr{A}} f(A)$. Then $y = f(x)$ for some $x \in \cup_{A\in\mathscr{A}} A$. Thus $y \in f(\cup_{A\in\mathscr{A}} A)$. Conversely, let $y \in f(\cup_{A\in\mathscr{A}} A)$. Then, again, $y = f(x)$ for some $x \in \cup_{A\in\mathscr{A}} A$, and so $y \in \cup_{A\in\mathscr{A}} f(A)$.

(vi) Let $x \in \cap_{B\in\mathscr{B}} f^{-1}(B)$. Then, for each $B \in \mathscr{B}$, $x \in f^{-1}(B)$. Thus $f(x) \in B$ for all $B \in \mathscr{B}$ and so $f(x) \in \cap_{B\in\mathscr{B}} B$. Thus $x \in f^{-1}(\cap_{B\in\mathscr{B}} B)$. Conversely, if $x \in f^{-1}(\cap_{B\in\mathscr{B}} B)$, then $f(x) \in B$ for each $B \in \mathscr{B}$. Thus $x \in f^{-1}(B)$ for each $B \in \mathscr{B}$, or $x \in \cap_{B\in\mathscr{B}} f^{-1}(B)$. ■

1.3.2 Properties of maps

Certain basic features of maps will be of great interest.

1.3.6 Definition (Injection, surjection, bijection) Let S and T be sets. A map $f\colon S \to T$ is:

(i) ***injective***, or an ***injection***, if $f(x) = f(y)$ implies that $x = y$;

(ii) ***surjective***, or a ***surjection***, if $f(S) = T$;

(iii) ***bijective***, or a ***bijection***, if it is both injective and surjective. •

1.3.7 Remarks (One-to-one, onto, 1–1 correspondence)

1. It is not uncommon for an injective map to be said to be ***1–1*** or ***one-to-one***, and that a surjective map be said to be ***onto***. In this series, we shall exclusively use the terms injective and surjective, however. These words appear to have been given prominence by their adoption by Bourbaki (see footnote on page **??**).
2. If there exists a bijection $f\colon S \to T$ between sets S and T, it is common to say that there is a ***1–1 correspondence*** between S and T. This can be confusing if one is familiar with the expression "1–1" as referring to an injective map. The words "1–1 correspondence" mean that there is a bijection, not an injection. In case S and T are in 1–1 correspondence, we shall also say that S and T are ***equivalent***. •

Closely related to the above concepts, although not immediately obviously so, are the following notions of inverse.

1.3.8 Definition (Left-inverse, right-inverse, inverse) Let S and T be sets, and let $f\colon S \to T$ be a map. A map $g\colon T \to S$ is:

(i) a ***left-inverse*** of f if $g \circ f = \mathrm{id}_S$;

(ii) a ***right-inverse*** of f if $f \circ g = \mathrm{id}_T$;

(iii) an ***inverse*** of f if it is both a left- and a right-inverse. •

In Definition 1.2.4 we gave the notion of the inverse of a relation. Functions, being relations, also possess inverses in the sense of relations. We ask the reader to explore the relationships between the two concepts of inverse in Exercise 1.3.5.

The following result relates these various notions of inverse to the properties of injective, surjective, and bijective.

1.3.9 Proposition (Characterisation of various inverses) *Let* S *and* T *be sets and let* f: S → T *be a map. Then the following statements hold:*

(i) f *is injective if and only if it possesses a left-inverse;*

(ii) f *is surjective if and only if it possess a right-inverse;*

(iii) f *is bijective if and only if it possesses an inverse;*

(iv) there is at most one inverse for f*;*

(v) if f *possesses a left-inverse and a right-inverse, then these necessarily agree.*

Proof (i) Suppose that f is injective. For $y \in \text{image}(f)$, define $g(y) = x$ where $f^{-1}(y) = \{x\}$, this being well-defined since f is injective. For $y \notin \text{image}(f)$, define $g(y) = x_0$ for some $x_0 \in S$. The map g so defined is readily verified to satisfy $g \circ f = \text{id}_S$, and so is a left-inverse. Conversely, suppose that f possesses a left-inverse g, and let $x_1, x_2 \in S$ satisfy $f(x_1) = f(x_2)$. Then $g \circ f(x_1) = g \circ f(x_2)$, or $x_1 = x_2$. Thus f is injective.

(ii) Suppose that f is surjective. For $y \in T$ let $x \in f^{-1}(y)$ and define $g(y) = x$.[3] With g so defined it is easy to see that $f \circ g = \text{id}_T$, so that g is a right-inverse. Conversely, suppose that f possesses a right-inverse g. Now let $y \in T$ and take $x = g(y)$. Then $f(x) = f \circ g(y) = y$, so that f is surjective.

(iii) Since f is bijective, it possesses a left-inverse g_L and a right-inverse g_R. We claim that these are equal, and each is actually an inverse of f. We have

$$g_L = g_L \circ \text{id}_T == g_L \circ f \circ g_R = \text{id}_S \circ g_R = g_R,$$

showing equality of g_L and g_R. Thus each is a left- and a right-inverse, and therefore an inverse for f.

(iv) Let g_1 and g_2 be inverses for f. Then, just as in part (iii),

$$g_1 = g_1 \circ \text{id}_T = g_1 \circ f \circ g_2 = \text{id}_S \circ g_2 = g_2.$$

(v) This follows from the proof of part (iv), noting that there we only used the facts that g_1 is a left-inverse and that g_2 is a right-inverse. ■

In Figure 1.2 we depict maps that have various of the properties of injectivity, surjectivity, or bijectivity. From these cartoons, the reader may develop some intuition for Proposition 1.3.9. In the case that $f: S \to T$ is a bijection, we denote its unique inverse by $f^{-1}: T \to S$. The confluence of the notation f^{-1} introduced when discussing preimages is not a problem, in practice.

It is worth mentioning at this point that the characterisation of left- and right-inverses in Proposition 1.3.9 is not usually very helpful. Normally, in a given setting, one will want these inverses to have certain properties. For vector spaces, for example, one may want left- or right-inverses to be linear (see), and for topological spaces, for another example, one may want a left- or right-inverse to be continuous (see Chapter II-2). what

[3] Note that the ability to choose an x from each set $f^{-1}(y)$ requires the Axiom of Choice (see Section 1.8.3).

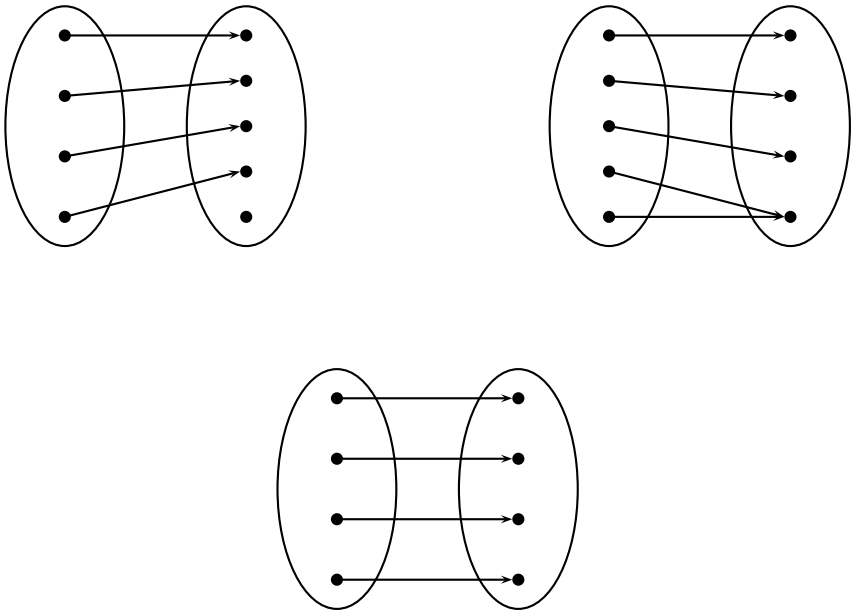


Figure 1.2 A depiction of maps that are injective but not surjective (top left), surjective but not injective (top right), and bijective (bottom)

1.3.3 Graphs and commutative diagrams

Often it is useful to be able to understand the relationship between a number of maps by representing them together in a diagram. We shall be somewhat precise about what we mean by a diagram by making it a special instance of a graph. We shall encounter graphs
where? in , although for the present purposes we merely use them as a means of making precise the notion of a commutative diagram.

First the definitions for graphs.

1.3.10 Definition (Graph) A ***graph*** is a pair (V, E) where V is a set an element of which is called a ***vertex*** and E is a subset of the set $V^{(2)}$ of unordered pairs from V an element of which is called an ***edge***. If $\{v_1, v_2\} \in E$ is an edge, then the vertices v_1 and v_2 are the ***endvertices*** of this edge. •

In a graph, it is the way that vertices and edges are related that is of interest. To capture this structure, the following language is useful.

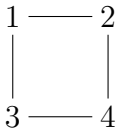
1.3.11 Definition (Adjacent and incident) Let (V, E) be a graph. Two vertices $v_1, v_2 \in V$ are ***adjacent*** if $\{v_1, v_2\} \in E$ and a vertex $v \in V$ and an edge $e \in E$ are ***incident*** if there exists $v' \in V$ such that $e = \{v, v'\}$. •

One typically represents a graph by placing the vertices in some sort of array on the page, and then drawing a line connecting two vertices if there is a corresponding edge associated with the two vertices. Some examples make this process clear.

1.3.12 Examples (Graphs)

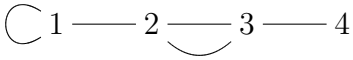
1. Consider the graph (V, E) with $V = \{1, 2, 3, 4\}$ and $E = \{\{1, 2\}, \{1, 3\}, \{2, 4\}, \{3, 4\}\}$. There are many ways one can lay out the vertices on the page, but for this diagram,

it is most convenient to arrange them in a square. Doing so gives rise to the following representation of the graph:



The vertices 1 and 2 are adjacent, but the vertices 1 and 4 are not. The vertex 1 and the edge $\{1, 2\}$ are incident, but the vertex 1 and the edge $\{3, 4\}$ are not.

2. For the graph (V, E) with $V = \{1, 2, 3, 4\}$ and $E = \{\{1, 2\}, \{2, 3\}, \{2, 3\}, \{3, 4\}\}$ we have the representation



Note that we allow the same edge to appear twice, and we allow for an edge to connect a vertex to itself. We observe that the vertices 2 and 3 are adjacent, but the vertices 1 and 3 are not. Also, the vertex 3 and the edge $\{2, 3\}$ are incident, but the vertex 4 and the edge $\{1, 2\}$ are not. •

Often one wishes to attach "direction" to vertices. This is done with the following notion.

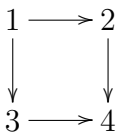
1.3.13 Definition (Directed graph) A ***directed graph***, or ***digraph***, is a pair (V, E) where V is a set an element of which is called a ***vertex*** and E is a subset of the set $V \times V$ of ordered pairs from V an element of which is called an ***edge***. If $e = (v_1, v_2) \in E$ is an edge, then v_1 is the ***source*** for e and v_2 is the ***target*** for e. •

Note that every directed graph is certainly also a graph, since one can assign an unordered pair to every ordered pair of vertices.

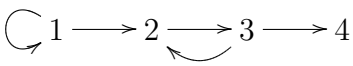
The examples above of graphs are easily turned into directed graphs, and we see that to represent a directed graph one needs only to put a "direction" on an edge, typically via an arrow.

1.3.14 Examples (Directed graphs)

1. Consider the directed graph (V, E) with $V = \{1, 2, 3, 4\}$ and $E = \{(1, 2), (1, 3), (2, 4), (3, 4)\}$. A convenient representation of this directed graph is as follows:



2. For the directed graph (V, E) with $V = \{1, 2, 3, 4\}$ and $E = \{(1, 1), (1, 2), (2, 3), (2, 3), (3, 4)\}$ we have the representation



•

Of interest in graph theory is the notion of connecting two, perhaps nonadjacent, vertices with a sequence of edges. This is made precise as follows.

1.3.15 Definition (Path)

(i) If (V, E) is a graph, a ***path*** in the graph is a sequence $\{a_j\}_{j \in \{1, \dots, k\}}$ in $V \cup E$ with the following properties:

(a) $a_1, a_k \in V$;
(b) for $j \in \{1, \dots, k-1\}$, if $a_j \in V$ (resp. $a_j \in E$), then $a_{j+1} \in E$ (resp. $a_{j+1} \in V$).

(ii) If (V, E) is a directed graph, a ***path*** in the graph is a sequence $\{a_j\}_{j\in\{1,\dots,k\}}$ in $V \cup E$ with the following properties:
(a) $\{a_j\}_{j\in\{1,\dots,k\}}$ is a path in the graph associated to (V, E);
(b) for $j \in \{2, \dots, k-1\}$, if $a_j \in E$, then $a_j = (a_{j-1}, a_{j+1})$.

(iii) If $\{a_j\}_{j\in\{1,\dots,k\}}$ is a path, the ***length*** of the path is the number of edges in the path.

(iv) For a path $\{a_j\}_{j\in\{1,\dots,k\}}$, the ***source*** is the vertex a_1 and the ***target*** is the vertex a_k. •

Let us give some examples of paths for graphs and for directed graphs.

1.3.16 Examples (Paths)

1. For the graph (V, E) with $V = \{1, 2, 3, 4\}$ and $E = \{\{1, 2\}, \{1, 3\}, \{2, 4\}, \{3, 4\}\}$, there are an infinite number of paths. Let us list a few:
   (a) $\{1\}$, $\{2\}$, $\{3\}$, and $\{4\}$;
   (b) $\{4, \{3, 4\}, 3, \{1, 3\}, 1\}$;
   (c) $\{1, \{1, 2\}, 2, \{2, 4\}, 4, \{3, 4\}, 3, \{1, 3\}, 1\}$;
   (d) $\{1, \{1, 2\}, 2, \{1, 2\}, 1, \{1, 2\}, 2, \{1, 2\}, 1\}$.

   Note that for this graph there are infinitely many paths.
2. For the directed graph (V, E) with $V = \{1, 2, 3, 4\}$ and $E = \{(1, 2), (1, 3), (2, 4), (3, 4)\}$, there are a finite number of paths:
   (a) $\{1\}$, $\{2\}$, $\{3\}$, and $\{4\}$;
   (b) $\{1, (1, 2), 2\}$;
   (c) $\{1, (1, 2), 2, (2, 4), 4\}$;
   (d) $\{1, (1, 3), 3\}$;
   (e) $\{1, (1, 3), 3, (2, 4), 4\}$;
   (f) $\{2, (2, 4)\}$;
   (g) $\{3, (3, 4), 4\}$.
3. For the graph (V, E) with $V = \{1, 2, 3, 4\}$ and $E = \{\{1, 2\}, \{2, 3\}, \{2, 3\}, \{3, 4\}\}$ some examples of paths are:
   (a) $\{1\}$, $\{2\}$, $\{3\}$, and $\{4\}$;
   (b) $\{1, \{1, 2\}, 2, \{2, 3\}, 3, \{2, 3\}, 2, \{1, 2\}, 1\}$;
   (c) $\{4, \{3, 4\}, 3\}$.

   There are an infinite number of paths for this graph.
4. For the directed graph (V, E) with $V = \{1, 2, 3, 4\}$ and $E = \{(1, 1), (1, 2), (2, 3), (2, 3), (3, 4)\}$ some paths include:
   (a) $\{1\}$, $\{2\}$, $\{3\}$, and $\{4\}$;
   (b) $\{1, (1, 2), 2, (2, 3), 3, (3, 2), 2, (2, 3), 3, (3, 4), 4\}$;
   (c) $\{3, (3, 4), 4\}$.

   This directed graph has an infinite number of paths by virtue of the fact that the path $\{2, (2, 3), 3, (3, 2), 2\}$ can be repeated an infinite number of times. •

1.3.17 Notation (Notation for paths of nonzero length) For paths which contain at least one edge, i.e., which have length at least 1, the vertices in the path are actually redundant. For this reason we will often simply write a path as the sequence of edges contained in the path, since the vertices can be obviously deduced. •

There is a great deal one can say about graphs, a little of which we will say in . However, where for our present purposes of defining diagrams, the notions at hand are sufficient. In the definition we employ Notation 1.3.17.

1.3.18 Definition (Diagram, commutative diagram) Let (V, E) be a directed graph.

(i) A ***diagram*** on (V, E) is a collection $\{S_v\}_{v\in V}$ of sets associated with each vertex and a collection $\{f_e\}_{e\in E}$ of maps associated with each edge such that, if $e = (v_1, v_2)$, then f_e has domain S_{v_1} and codomain S_{v_2}.

(ii) If $P = \{e_j\}_{j\in\{1,\dots,k\}} \subset E$ is a path of nonzero length in a diagram on (V, E), the ***composition*** along P is the map $f_{e_k} \circ \cdots \circ f_{e_1}$.

(iii) A diagram is ***commutative*** if, for every two vertices $v_1, v_2 \in V$ and any two paths P_1 and P_2 with source v_1 and target v_2, the composition along P_1 is equal to the composition along P_2. •

The notion of a diagram, and in particular a commutative diagram is straightforward.

1.3.19 Examples (Diagrams and commutative diagrams)

1. Let S_1, S_2, S_3, and S_4 be sets and consider maps $f_{21}\colon S_1 \to S_2$, $f_{31}\colon S_1 \to S_3$, $f_{42}\colon S_2 \to S_4$, and $f_{43}\colon S_3 \to S_4$.[4] Note that if we assign set S_j to j for each $j \in \{1, 2, 3, 4\}$, then where? this gives a diagram on (V, E) where$V = \{1, 2, 3, 4\}$ and $E = \{(1, 2), (1, 3), (2, 4), (3, 4)\}$. This diagram can be represented by

$$\begin{array}{ccc} S_1 & \xrightarrow{f_{21}} & S_2 \\ {\scriptstyle f_{31}}\downarrow & & \downarrow{\scriptstyle f_{42}} \\ S_3 & \xrightarrow[f_{43}]{} & 4 \end{array}$$

The diagram is commutative if and only if $f_{42} \circ f_{21} = f_{43} \circ f_{31}$.

2. Let S_1, S_2, S_3, and S_4 be sets and let $f_{11}\colon S_1 \to S_1$, $f_{21}\colon S_1 \to S_2$, $f_{32}\colon S_2 \to S_3$, $f_{23}\colon S_3 \to S_2$, and $f_{43}\colon S_3 \to S_4$ be maps. This data then represents a commutative diagram on the directed graph (V, E) where $V = \{1, 2, 3, 4\}$ and $E = \{(1, 1), (1, 2), (2, 3), (2, 3), (3, 4)\}$. The diagram is represented as

$$f_{11}\circlearrowleft S_1 \xrightarrow{f_{21}} S_2 \underset{f_{23}}{\overset{f_{32}}{\rightleftarrows}} S_3 \xrightarrow{f_{43}} S_4$$

While it is possible to write down conditions for this diagram to be commutative, there will be infinitely many such conditions. In practice, one encounters commutative diagrams with only finitely many paths with a given source and target. This example, therefore, is not so interesting as a commutative diagram, but is more interesting as a signal flow graph, as we shall see . • where

[4] It might seem more natural to write, for example, $f_{12}\colon S_1 \to S_2$ to properly represent the normal order of the domain and codomain. However, we instead write $f_{21}\colon S_1 \to S_2$ for reasons having to do with conventions that will become convenient in .

Exercises

1.3.1 Let S, T, U, and V be sets, and let $f\colon S \to T$, $g\colon T \to U$, and $h\colon U \to V$ be maps. Show that $h \circ (g \circ f) = (h \circ g) \circ f$.

1.3.2 If S, T, and U are sets and if $f\colon S \to T$ and $g\colon T \to U$ are bijections, then show that $(g \circ f)^{-1} = f^{-1} \circ g^{-1}$.

1.3.3 Let S, T and U be sets and let $f\colon S \to T$ and $g\colon T \to U$ be maps.

(a) Show that if f and g are injective, then so too is $g \circ f$.

(b) Show that if f and g are surjective, then so too is $g \circ f$.

1.3.4 Let S and T be sets, let $f\colon S \to T$ be a map, and let $A \subset S$ and $B \subset T$. Do the following:

(a) show that if f is injective then $A = f^{-1}(f(A))$;

(b) show that if f is surjective then $f(f^{-1}(B)) = B$.

1.3.5 Let S and T be sets and let $f\colon S \to T$ be a map.

(a) Show that if f is invertible as a map, then "the relation of its inverse is the inverse of its relation." (Part of the question is to precisely understand the statement in quotes.)

(b) Show that the inverse of the relation defined by f is itself the relation associated to a function if and only if f is invertible.

1.3.6 Show that equivalence of sets, as in Remark 1.3.7–2, is an "equivalence relation"[5] on collection of all sets.

[5]The quotes are present because the notion of equivalence relation, as we have defined it, applies to sets. However, there is no set containing all sets; see Section 1.8.1

Section 1.4

Construction of the integers

It can be supposed that the reader has some idea of what the set of integers is. In this section we actually give the set of integers a *definition*. As will be seen, this is not overly difficult to do. Moreover, the construction has little bearing on what we do. We merely present it so that the reader can be comfortable with the fact that the integers, and so subsequently the rational numbers and the real numbers (see Section 2.1), have a formal definition.

Do I need to read this section? Much of this section is not of importance in the remainder of this series. The reader should certainly know what the sets $\mathbb{N}$ and $\mathbb{Z}$ are. However, the details of their construction should be read only when the inclination strikes. •

1.4.1 Construction of the natural numbers

The natural numbers are the numbers 1, 2, 3, and so on, i.e., the "counting numbers." As such, we are all quite familiar with them in that we can recognise, in the absence of trickery, when we are presented with 4 of something. However, what is 4? This is what we endeavour to define in this section.

The important concept in defining the natural numbers is the following.

1.4.1 Definition (Successor) Let S be a set. The ***successor*** of S is the set $S^+ = S \cup \{S\}$. •

Thus the successor is a set whose elements are the elements of S, plus an additional element which is the set S itself. This seems, and indeed is, a simple enough idea. However, it does make possible the following definition.

1.4.2 Definition (0, 1, 2, etc.)

(i) The number ***zero***, denoted by 0, is the set $\emptyset$.

(ii) The number ***one***, denoted by 1, is the set 0^+.

(iii) The number ***two***, denoted by 2, is the set 1^+.

(iv) The number ***three***, denoted by 3, is the set 2^+.

(v) The number ***four***, denoted by 4, is the set 3^+.

This procedure can be inductively continued to define any finite nonnegative integer. •

The procedure above is well-defined, and so gives meaning to the symbol "k" where k is any nonnegative finite number. Let us give the various explicit ways of writing the first few numbers:

$$\begin{aligned}
&0 = \emptyset,\\
&1 = 0^+ = \{0\} &&= \{\emptyset\},\\
&2 = 1^+ = \{0, 1\} &&= \{\emptyset, \{\emptyset\}\},\\
&3 = 2^+ = \{0, 1, 2\} &&= \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\},\\
&4 = 3^+ = \{0, 1, 2, 3\} &&= \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}, \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\}\}.
\end{aligned}$$

This settles the matter of defining any desired number. We now need to indicate how to talk about the *set* of numbers. This necessitates an assumption. As we shall see in Section 1.8.2, this assumption is framed as an axiom in axiomatic set theory.

1.4.3 Assumption There exists a set containing $\emptyset$ and all subsequent successors. •

We are now almost done. The remaining problem is that the set guaranteed by the assumption may contain more than what we want. However, this is easily remedied as follows. Let S be the set whose existence is guaranteed by Assumption 1.4.3. Define a collection $\mathscr{A}$ of subsets of S by

$$\mathscr{A} = \left\{ A \subset S \mid \ \emptyset \in A \text{ and } n^+ \in A \text{ if } n \in A \right\}.$$

Note that $S \in \mathscr{A}$ so that $\mathscr{A}$ is nonempty. The following simple result is now useful.

1.4.4 Lemma *If $\mathscr{B} \subset \mathscr{A}$, then $(\cap_{\mathrm{B}\in\mathscr{B}}\mathrm{B}) \in \mathscr{A}$.*

Proof For each $B \in \mathscr{B}$, $\emptyset \in B$. Thus $\emptyset \in \cap_{B\in\mathscr{B}}B$. Also let $n \in \cap_{B\in\mathscr{B}}B$. Since $n^+ \in B$ for each $B \in \mathscr{B}$, $xn^+ \in \cap_{B\in\mathscr{B}}B$. Thus $(\cap_{B\in\mathscr{B}}B) \in \mathscr{A}$, as desired. ■

The lemma shows that $\cap_{A\in\mathscr{A}}A \in \mathscr{A}$. Now we have the following definition of the *set* of numbers.

1.4.5 Definition (Natural numbers) Let S and $\mathscr{A}$ be as defined above.

(i) The set $\cap_{A\in\mathscr{A}}A$ is denoted by $\mathbb{N}_0$, and is the set of ***nonnegative integers***.

(ii) The set $\mathbb{N}_0 \setminus \{0\}$ is denoted by $\mathbb{N}$, and is the set of ***natural numbers***. •

1.4.6 Remark (Convention concerning $\mathbb{N}$) It is not uncommon to see the set that we denote by $\mathbb{N}_0$ called the natural numbers, and denoted, therefore, by $\mathbb{N}$. This is a matter of convention, so the reader should be aware that both conventions are in use. One of the uncomfortable things about the convention we use appears in the preceding definition. Namely, we declared $\mathbb{N}_0$ to be the set of "nonnegative integers." This seems an odd definition since we have not given meaning to either of the words "nonnegative" or "integer." While the reader may think they already know what these things are, they are well advised to forget that for now, and just thing of "nonnegative integer" as a pair of randomly chosen words whose meaning will be justified shortly. •

Next we turn to the definition of the usual operations of arithmetic with the set $\mathbb{N}_0$. That is to say, we indicate how to "add" and "multiply." First we consider addition.

1.4.7 Definition (Addition in $\mathbb{N}_0$) For $k \in \mathbb{N}_0$, inductively define a map $a_k \colon \mathbb{N}_0 \to \mathbb{N}_0$, called ***addition by*** **k**, by

(i) $a_k(0) = k$;

(ii) $a_k(j^+) = (a_k(j))^+$, $j \in \mathbb{N}$.

We denote $a_k(j) = k + j$. •

Upon a moments reflection, it is easy to convince yourself that this formal definition of addition agrees with our established intuition. Roughly speaking, one defines $k + (j + 1) = (k+j)+1$, where, by definition, the operation of adding 1 means taking the successor. With these definitions it is straightforward to verify such commonplace assertions as "$1 + 1 = 2$."

Now we define multiplication.

1.4.8 Definition (Multiplication in $\mathbb{N}_0$) For $k \in \mathbb{N}_0$, inductively define a map $m_k \colon \mathbb{N}_0 \to \mathbb{N}_0$, called ***multiplication by*** **k**, by

(i) $m_k(0) = 0$;

(ii) $m_k(j^+) = m_k(j) + k$.

We denote $m_k(j) = k \cdot j$, or simply jk where no confusion can arise. •

Again, this definition of multiplication is in concert with our intuition. The definition says that $k \cdot (j+1) = k \cdot j + k$. For $k, m \in \mathbb{N}_0$, define k^m recursively by $k^0 = 1$, and $k^{m^+} = k^m \cdot k$. The element $k^m \in \mathbb{N}_0$ is the mth ***power*** of k.

Let us verify that addition and multiplication in $\mathbb{N}_0$ have the expected properties. In stating the properties, we use the usual order of operation rules one learns in high school; in this case, operations are done with the following precedence: (1) operations enclosed in parentheses, (2) multiplication, then (3) addition.

1.4.9 Proposition (Properties of arithmetic in $\mathbb{N}_0$) *Addition and multiplication in* $\mathbb{N}_0$ *satisfy the following rules:*

(i) $\mathrm{k}_1 + \mathrm{k}_2 = \mathrm{k}_2 + \mathrm{k}_1$, $\mathrm{k}_1, \mathrm{k}_2 \in \mathbb{N}_0$ *(**commutativity** of addition);*

(ii) $(\mathrm{k}_1 + \mathrm{k}_2) + \mathrm{k}_3 = \mathrm{k}_1 + (\mathrm{k}_2 + \mathrm{k}_3)$, $\mathrm{k}_1, \mathrm{k}_2, \mathrm{k}_3 \in \mathbb{N}_0$ *(**associativity** of addition);*

(iii) $\mathrm{k} + 0 = \mathrm{k}$, $\mathrm{k} \in \mathbb{N}_0$ *(**additive identity**);*

(iv) $\mathrm{k}_1 \cdot \mathrm{k}_2 = \mathrm{k}_2 \cdot \mathrm{k}_1$, $\mathrm{k}_1, \mathrm{k}_2 \in \mathbb{N}_0$ *(**commutativity** of multiplication);*

(v) $(\mathrm{k}_1 \cdot \mathrm{k}_2) \cdot \mathrm{k}_3 = \mathrm{k}_1 \cdot (\mathrm{k}_2 \cdot \mathrm{k}_3)$, $\mathrm{k}_1, \mathrm{k}_2, \mathrm{k}_3 \in \mathbb{N}_0$ *(**associativity** of multiplication);*

(vi) $\mathrm{k} \cdot 1 = \mathrm{k}$, $\mathrm{k} \in \mathbb{N}_0$ *(**multiplicative identity**);*

(vii) $\mathrm{j} \cdot (\mathrm{k}_1 + \mathrm{k}_2) = \mathrm{j} \cdot \mathrm{k}_1 + \mathrm{j} \cdot \mathrm{k}_2$, $\mathrm{j}, \mathrm{k}_1, \mathrm{k}_2 \in \mathbb{N}_0$ *(**distributivity**);*

(viii) $\mathrm{j}^{\mathrm{k}_1} \cdot \mathrm{j}^{\mathrm{k}_2} = \mathrm{j}^{\mathrm{k}_1 + \mathrm{k}_2}$, $\mathrm{j}, \mathrm{k}_1, \mathrm{k}_2 \in \mathbb{N}_0$.

Proof We shall prove these in logical sequence, rather than the sequence in which they are stated.

(ii) We prove this by induction on k_3. For $k_3 = 0$ we have $(k_1 + k_2) + 0 = k_1 + k_2$ and $k_1 + (k_2 + 0) = k_1 + k_2$, giving the result in this case. Now suppose that $(k_1 + k_2) + j = k_1 + (k_2 + j)$ for $j \in \{0, 1, \dots, k_3\}$. Then

$$(k_1 + k_2) + k_3^+ = ((k_1 + k_2) + k_3)^+ = (k_1 + (k_2 + k_3))^+ = k_1 + (k_2 + k_3)^+ = k_1 + (k_2 + k_3^+),$$

where we have used the definition of addition, the induction hypothesis, and then twice used the definition of addition.

(i) We first claim that $0 + k = k$ for all $k \in \mathbb{N}_0$. It is certainly true, by definition, that $0 + 0 = 0$. Now suppose that $0 + j = j$ for $j \in \{0, 1, \dots, k\}$. Then

$$0 + k^+ = 0 + (k + 1) = (0 + k) + 1 = k + 1 = k^+.$$

We next claim that $k_1^+ + k_2 = (k_1 + k_2)^+$ for $k_1, k_2 \in \mathbb{N}_0$. We prove this by induction on k_2. For $k_2 = 0$ we have $k_1^+ + 0 = k_1^+$ and $(k_1 + 0)^+ = k_1^+$, using the definition of addition. This gives the claim for $k_2 = 0$. Now suppose that $k_1^+ + j = (k_1 + j)^+$ for $j \in \{0, 1, \dots, k_2\}$. Then

$$k_1^+ + k_2^+ = k_1^+ + (k_2 + 1) = (k_1^+ + k_2) + 1 = (k_1^+ + k_2)^+,$$

as desired.

We now complete the proof of this part of the result by induction on k_1. For $k_1 = 0$ we have $0 + k_2 = k_2 = k_2 + 0$, using the first of our claims above and the definition of addition. Now suppose that $j + k_2 = k_2 + j$ for $j \in \{0, 1, \dots, k_1\}$. Then

$$k_1^+ + k_2 = (k_1 + k_2)^+ = (k_2 + k_1)^+ = k_2 + k_1^+,$$

using the second or our claims above and the definition of addition.

(iii) This is part of the definition of addition.

(vii) We prove the this by induction on k_2. First note that for $k_2 = 0$ we have $j \cdot (k_1 + 0) = j \cdot k_1$ and $j \cdot k_1 + j \cdot 0 = j \cdot k_1 + 0 = j \cdot k_1$, so the result holds when $k_2 = 0$. Now suppose that $j \cdot (k_1 + k) = j \cdot k_1 + j \cdot k$ for $k \in \{0, 1, \dots, k_2\}$. Then we have

$$\begin{aligned} j \cdot (k_1 + k_2^+) &= j \cdot (k_1 + k_2)^+ = j \cdot (k_1 + k_2) + j \\ &= (j \cdot k_1 + j \cdot k_2) + j = j \cdot k_1 + (j \cdot k_2 + j) \\ &= j \cdot k_1 + j \cdot k_2^+, \end{aligned}$$

as desired, where we have used, in sequence, the definition of addition, the definition of multiplication, the induction hypothesis, the associativity of addition, and the definition of multiplication.

(iv) We first prove by induction on k that $0 \cdot k = 0$ for $k \in \mathbb{N}_0$. For $k = 0$ the claim holds by definition of multiplication. So suppose that $0 \cdot j = 0$ for $j \in \{0, 1, \dots, k\}$ and then compute $0 \cdot k^+ = 0 \cdot k + 0 = 0$, as desired.

We now prove the result by induction on k_2. For $k_2 = 0$ we have $k_1 \cdot 0 = 0$ by definition of multiplication. We also have $k_2 \cdot 0 = 0$ by the first part of the proof. So now suppose that $k_1 \cdot j = j \cdot k$ for $j \in \{0, 1, \dots, k_2\}$. We then have

$$k_1 \cdot k_2^+ = k_1 \cdot k_2 + k_1 = k_2 \cdot k_1 + k_1 = k_1 + k_2 \cdot k_1 = (1 + k_2) \cdot k_1 = k_2^+ \cdot k_1,$$

where we have used, in sequence, the definition of multiplication, the induction hypothesis, commutativity of addition, distributivity, commutativity of addition, and the definition of addition.

(v) We prove this part of the result by induction on k_3. For $k_3 = 0$ we have $(k_1 \cdot k_2) \cdot 0 = 0$ and $k_1 \cdot (k_2 \cdot 0) = k_1 \cdot 0 = 0$. Thus the result is true when $k_3 = 0$. Now suppose that $(k_1 \cdot k_2) \cdot j = k_1 \cdot (k_2 \cdot j)$ for $j \in \{0, 1, \dots, k_3\}$. Then

$$(k_1 \cdot k_2) \cdot k_3^+ = (k_1 \cdot k_2) \cdot k_3 + k_1 \cdot k_2 = k_1 \cdot (k_2 \cdot k_3) + k_1 \cdot k_2 = k_1 \cdot (k_2 \cdot k_3 + k_2) = k_1 \cdot (k_2 \cdot k_3^+),$$

where we have used, in sequence, the definition of multiplication, the induction hypothesis, distributivity, and the definition of multiplication.

(vi) This follows from the definition of multiplication.

(viii) We prove the result by induction on k_1. The result is obviously true for $k_2 = 0$, so suppose that $j^{k_1 + l} = j^{k_1} \cdot j^l$ for $l \in \{1, \dots, k_2\}$. Then

$$j^{k_1 + k_2^+} = j^{(k_1 + k_2)^+} = j^{k_1 + k_2} \cdot j = j^{k_1} \cdot j^{k_2} \cdot j = j^{k_1} \cdot j^{k_2^+},$$

as desired. ■

1.4.2 Two relations on $\mathbb{N}_0$

Another property of the naturals that we would all agree they ought to have is an "order." Thus we should have a means of saying when one natural number is less than another. To get started at this, we have the following result.

1.4.10 Lemma *For* $\mathrm{j}, \mathrm{k} \in \mathbb{N}_0$*, exactly one of the following possibilities holds:*

(i) $\mathrm{j} \subset \mathrm{k}$*;*

(ii) $\mathrm{k} \subset \mathrm{j}$*;*

(iii) $\mathrm{j} = \mathrm{k}$*.*

Proof For $k \in \mathbb{N}_0$ define

$$S(k) = \{ j \in \mathbb{N} \mid \ j \subset k,\, k \subset j, \text{ or } j = k \}.$$

We shall prove by induction that $S(k) = \mathbb{N}_0$ for each $k \in \mathbb{N}_0$.

First take the case of $k = 0$. Since $\emptyset$ is a subset of every set, $0 \in S(0)$. Now suppose that $j \in S(0)$ for $j \in \mathbb{N}_0$. We have the following cases.

1. $j \in 0$: This is impossible since 0 is the empty set.
2. $0 \in j$: In this case $0 \in j^+$.
3. $0 = j$: In this case $0 \in j^+$.

Thus $j \in S(0)$ implies that $j^+ \in S(0)$, and so $S(0) = \mathbb{N}_0$.

Now suppose that $S(m) = \mathbb{N}_0$ for $m \in \{0, 1, \ldots, k\}$. We will show that $S(k^+) = \mathbb{N}_0$. Clearly $0 \in S(k^+)$. So suppose that $j \in S(k^+)$. We again have three cases.

1. $j \in k^+$: We have the following two subcases.
   (a) $j = k$: Here we have $j^+ = k^+$.
   (b) $j \in k$: Since $j^+ \in S(k)$ by the induction hypothesis, we have the following three cases.
      i. $k \in j^+$: This is impossible since $j \in k$.
      ii. $j^+ \in k$: Here $j^+ \in k^+$.
      iii. $j^+ = k$: Here again, $j^+ \in n^+$.
2. $k^+ \in j$: In this case $k^+ \in j^+$.
3. $k^+ = j$: In this case $k^+ \in j^+$.

In all cases we conclude that $j^+ \in S(k^+)$, and this completes the proof. ■

It is easy to show that $j \in k$ if and only if $j \subset k$, and that, if $j \in k$ but $j \neq k$, then $j \subsetneq k$ (see Exercise 1.4.2). With this result, it is now comparatively easy to prove the following.

1.4.11 Proposition (Order[6] on $\mathbb{N}_0$) *On $\mathbb{N}_0$ define two relations $<$ and $\leq$ by*

$$\begin{aligned} \mathrm{j} < \mathrm{k} \quad &\Longleftrightarrow \quad \mathrm{j} \subsetneq \mathrm{k}, \\ \mathrm{j} \leq \mathrm{k} \quad &\Longleftrightarrow \quad \mathrm{j} \subset \mathrm{k}. \end{aligned}$$

Then

(i) $<$ *and* $\leq$ *are transitive,*

(ii) $<$ *is irreflexive;*

(iii) $\leq$ *is reflexive and antisymmetric.*

Furthermore, for any $\mathrm{j}, \mathrm{k} \in \mathbb{N}_0$, *either* $\mathrm{j} \leq \mathrm{k}$ *or* $\mathrm{k} \leq \mathrm{j}$.

The following rewording of the final part of the result is distinguished.

1.4.12 Corollary (Trichotomy Law for $\mathbb{N}_0$) *For* $\mathrm{j}, \mathrm{k} \in \mathbb{N}_0$, *exactly one of the following possibilities holds:*

(i) $\mathrm{j} < \mathrm{k}$;

(ii) $\mathrm{k} < \mathrm{j}$;

(iii) $\mathrm{j} = \mathrm{k}$.

[6] We have not introduced the notion of order yet, but refer the reader to Section 1.5.

Of course, the symbols "$<$" and "$\leq$" have their usual meaning, which is "less than" and "less than or equal to," respectively. We shall explore such matters in more depth and generality in Section 1.5.

We shall also sometimes write "$j > k$" (resp. "$j \geq k$") for "$k < j$" (resp. "$k \leq j$"). The symbols "$>$" and "$\geq$" then have their usual meaning as "greater than" and "greater than or equal to," respectively.

The relations $<$ and $\leq$ satisfy some natural properties with respect to addition and multiplication in $\mathbb{N}_0$. Let us record these, leaving their proof as Exercise 1.4.3.

1.4.13 Proposition (Relation between addition and multiplication and $<$) *For* $\mathrm{j}, \mathrm{k}, \mathrm{m} \in \mathbb{N}_0$*, the following statements hold:*

(i) if $\mathrm{j} < \mathrm{k}$ *then* $\mathrm{j} + \mathrm{m} < \mathrm{k} + \mathrm{m}$*;*

(ii) if $\mathrm{j} < \mathrm{k}$ *and if* $\mathrm{m} \neq 0$ *then* $\mathrm{m} \cdot \mathrm{j} < \mathrm{m} \cdot \mathrm{k}$.

1.4.3 Construction of the integers from the natural numbers

Next we construct negative numbers to arrive at a definition of the integers. The construction renders the integers as the set of equivalence classes under a prescribed equivalence relation in $\mathbb{N}_0 \times \mathbb{N}_0$. The equivalence relation is defined formally as follows:

$$(j_1, k_1) \sim (j_2, k_2) \quad \iff \quad j_1 + k_2 = k_1 + j_2. \tag{1.1}$$

It is a simple exercise to check that this is indeed an equivalence relation.

We now define the integers.

1.4.14 Definition (Integers) The set of ***integers*** is the set $\mathbb{Z} = (\mathbb{N}_0 \times \mathbb{N}_0)/\sim$, where $\sim$ is the equivalence relation in (1.1). •

Now let us try to understand this definition by understanding the equivalence classes under the relation of (1.1). Key to this is the following result.

1.4.15 Lemma *Let* Z *be the subset of* $\mathbb{N}_0 \times \mathbb{N}_0$ *defined by*

$$\mathrm{Z} = \{(\mathrm{k}, 0) \mid \ \mathrm{k} \in \mathbb{N}\} \cup \{(0, \mathrm{k}) \mid \ \mathrm{k} \in \mathbb{N}\} \cup \{(0, 0)\},$$

and define a map $\mathrm{f_Z} \colon \mathrm{Z} \to \mathbb{Z}$ *by* $\mathrm{f_Z}(\mathrm{j}, \mathrm{k}) = [(\mathrm{j}, \mathrm{k})]$*. Then* $\mathrm{f_Z}$ *is a bijection.*

Proof First we show that f_Z is injective. Suppose that $f_Z(j_1, k_1) = f_Z(j_2, k_2)$. This means that $(j_1, k_1) \sim (j_2, k_2)$, or that $j_1 + k_2 = k_1 + j_2$. If $(j_1, k_1) = (0, 0)$, then this means that $k_2 = j_2$, which means that $(j_2, k_2) = (0, 0)$ since this is the only element of Z whose entries agree. If $j_1 = 0$ and $k_1 > 0$, then we have $k_2 = k_1 + j_2$. Since at least one of j_2 and k_2 must be zero, we then deduce that it must be that j_2 is zero (or else the equality $k_2 = k_1 + j_2$) cannot hold. This then also gives $k_2 = k_1$. A similar argument holds if $j_1 > 0$ and $k_1 = 0$. This shows injectivity of f_Z.

Next we show that f_Z is surjective. Let $[(j, k)] \in \mathbb{Z}$. By the Trichotomy Law, we have three cases.

1. $j = k$: We claim that $[(j, j)] = f_Z(0, 0)$. Indeed, we need only note that $(0, 0) \sim (j, j)$ since $0 + j = 0 + j$.
2. $j < k$: Let $m \in \mathbb{N}$ be defined such that $j + m = k$. (Why can this be done?) We then claim that $f_Z(0, m) = [(j, k)]$. Indeed, since $0 + k = m + j$, this is so.
3. $k < j$: Here we let $m \in \mathbb{N}$ satisfy $k + m = j$, and, as in the previous case, we can easily check that $f_Z(m, 0) = [(j, k)]$. ■

With this in mind, we introduce the following notation to denote an integer.

1.4.16 Notation (Notation for integers) Let $[(j,k)] \in \mathbb{Z}$.

(i) If $f_Z^{-1}[(j,k)] = [(0,0)]$ then we write $[(j,k)] = 0$.

(ii) If $[(j,k)] = [(m,0)]$, $m > 0$, then we write $[(j,k)] = m$. Such integers are ***positive***.

(iii) If $[(j,k)] = [(0,m)]$, $m > 0$, then we write $[(j,k)] = -m$. Such integers are ***negative***.

An integer is ***nonnegative*** if it is either positive or zero, and an integer is ***nonpositive*** if it is either negative or zero. •

This then relates the equivalence class definition of integers to the notion we are more familiar with: positive and negative numbers. We can also define the familiar operations of addition and multiplication of integers.

1.4.17 Definition (Addition and multiplication in $\mathbb{Z}$) Define the operations of ***addition*** and ***multiplication*** in $\mathbb{Z}$ by

(i) $[(j_1,k_1)] + [(j_2,k_2)] = [(j_1 + j_2, k_1 + k_2)]$ and

(ii) $[(j_1,k_1)] \cdot [(j_2,k_2)] = [(j_1 \cdot j_2 + k_1 \cdot k_2, j_1 \cdot k_2 + k_1 \cdot j_2)]$,

respectively, for $[(j_1,k_1)], [(j_2,k_2)] \in \mathbb{Z}$. As with multiplication in $\mathbb{N}_0$, we shall sometimes omit the "$\cdot$". •

These definitions do not *a priori* make sense; this needs to be verified.

1.4.18 Lemma *The definitions for addition and multiplication in $\mathbb{Z}$ a well-defined in that they do not depend on the choice of representative.*

Proof Let $(j_1,k_1) \sim (\tilde{j}_1,\tilde{k}_1)$ and $(j_2,k_2) \sim (\tilde{j}_2,\tilde{k}_2)$. Thus

$$j_1 + \tilde{k}_1 = k_1 + \tilde{j}_1, \quad j_2 + \tilde{k}_2 = k_2 + \tilde{j}_2.$$

It therefore follows that

$$(\tilde{j}_1 + \tilde{j}_2) + (k_1 + k_2) = (\tilde{k}_1 + \tilde{k}_2) + (j_1 + j_2),$$

which gives the independence of addition on representative. One may also directly verify that

$$(\tilde{j_1} \cdot \tilde{j}_2 + \tilde{k}_1 \cdot \tilde{k}_2) + (j_1 \cdot k_2 + k_1 \cdot j_2) = (\tilde{j}_1 \cdot \tilde{k}_2 + \tilde{k}_1 \cdot \tilde{j}_2) + (j_1 \cdot j_2 + k_1 \cdot k_2),$$

which gives the independence of multiplication on representative. ■

As with elements of $\mathbb{N}_0$, we can define powers for integers. Let $k \in \mathbb{Z}$ and $m \in \mathbb{N}_0$. We define k^m recursively as follows. We take $k^0 = 1$ and define $k^{m^+} = k^m \cdot k$. We call k^m the mth ***power*** of k. Note that, at this point, k^m only makes sense for $m \in \mathbb{N}_0$.

Finally, we give the properties of addition and multiplication in $\mathbb{Z}$. Some of these properties are as for $\mathbb{N}_0$. However, there is a useful new feature that arises in $\mathbb{Z}$ that mirrors our experience with negative numbers. In the statement of the result, it is convenient to denote an integer as in Notation 1.4.16, rather than as in the definition.

1.4.19 Proposition (Properties of addition and multiplication in $\mathbb{Z}$) *Addition and multiplication in $\mathbb{Z}$ satisfy the following rules:*

(i) $\mathrm{k}_1 + \mathrm{k}_2 = \mathrm{k}_2 + \mathrm{k}_1$, $\mathrm{k}_1, \mathrm{k}_2 \in \mathbb{Z}$ *(**commutativity** of addition);*

(ii) $(\mathrm{k}_1 + \mathrm{k}_2) + \mathrm{k}_3 = \mathrm{k}_1 + (\mathrm{k}_2 + \mathrm{k}_3)$, $\mathrm{k}_1, \mathrm{k}_2, \mathrm{k}_3 \in \mathbb{Z}$ *(**associativity** of addition);*

(iii) $\mathrm{k} + 0 = \mathrm{k}$, $\mathrm{k} \in \mathbb{Z}$ *(**additive identity**);*

(iv) $\mathrm{k} + (-1 \cdot \mathrm{k}) = 0$, $\mathrm{k} \in \mathbb{Z}$ ***(additive inverse)***;

(v) $\mathrm{k}_1 \cdot \mathrm{k}_2 = \mathrm{k}_2 \cdot \mathrm{k}_1$, $\mathrm{k}_1, \mathrm{k}_2 \in \mathbb{Z}$ ***(commutativity** of multiplication)*;

(vi) $(\mathrm{k}_1 \cdot \mathrm{k}_2) \cdot \mathrm{k}_3 = \mathrm{k}_1 \cdot (\mathrm{k}_2 \cdot \mathrm{k}_3)$, $\mathrm{k}_1, \mathrm{k}_2, \mathrm{k}_3 \in \mathbb{Z}$ ***(associativity** of multiplication)*;

(vii) $\mathrm{k} \cdot 1 = \mathrm{k}$, $\mathrm{k} \in \mathbb{Z}$ ***(multiplicative identity)***;

(viii) $\mathrm{j} \cdot (\mathrm{k}_1 + \mathrm{k}_2) = \mathrm{j} \cdot \mathrm{k}_1 + \mathrm{j} \cdot \mathrm{k}_2$, $\mathrm{j}, \mathrm{k}_1, \mathrm{k}_2 \in \mathbb{Z}$ ***(distributivity)***;

(ix) $\mathrm{j}^{\mathrm{k}_1} \cdot \mathrm{j}^{\mathrm{k}_2} = \mathrm{j}^{\mathrm{k}_1 + \mathrm{k}_2}$, $\mathrm{j} \in \mathbb{Z}$, $\mathrm{k}_1, \mathrm{k}_2 \in \mathbb{N}_0$.

Moreover, if we define $\mathrm{i}_{\mathbb{N}_0} \colon \mathbb{N}_0 \to \mathbb{Z}$ *by* $\mathrm{i}_{\mathbb{N}_0}(\mathrm{k}) = [(\mathrm{k}, 0)]$, *then addition and multiplication in* $\mathbb{Z}$ *agrees with that in* $\mathbb{N}_0$*:*

$$\mathrm{i}_{\mathbb{N}_0}(\mathrm{k}_1) + \mathrm{i}_{\mathbb{N}_0}(\mathrm{k}_2) = \mathrm{i}_{\mathbb{N}_0}(\mathrm{k}_1 + \mathrm{k}_2), \quad \mathrm{i}_{\mathbb{N}_0}(\mathrm{k}_1) \cdot \mathrm{i}_{\mathbb{N}_0}(\mathrm{k}_2) = \mathrm{i}_{\mathbb{N}_0}(\mathrm{k}_1 \cdot \mathrm{k}_2).$$

Proof These follow easily from the definitions of addition and multiplication, using the fact that the corresponding properties hold for $\mathbb{N}_0$. We leave the details to the reader as Exercise 1.4.4. We therefore only prove the new property (iv). For this, we suppose without loss of generality that $k \in \mathbb{N}_0$, i.e., $k = [(k, 0)]$. Then $-k = [(0, k)]$ so that

$$k + (-k) = [(k + 0, 0 + k)] = [(k, k)] = [(0, 0)] = 0,$$

as claimed. ■

We shall make the convention that $-1 \cdot k$ be written as $-k$, whether k be positive or negative. We shall also, particularly as we move along to things of more substance, think of $\mathbb{N}_0$ as a subset of $\mathbb{Z}$, without making explicit reference to the map $i_{\mathbb{N}_0}$.

1.4.4 Two relations in $\mathbb{Z}$

Finally we introduce in $\mathbb{Z}$ two relations that extend the relations $<$ and $\leq$ for $\mathbb{N}_0$. The following result is the analogue of Proposition 1.4.11.

1.4.20 Proposition (Order on $\mathbb{Z}$) *On* $\mathbb{Z}$ *define two relations* $<$ *and* $\leq$ *by*

$$[(\mathrm{j}_1, \mathrm{k}_1)] < [(\mathrm{j}_2, \mathrm{k}_2)] \iff \mathrm{j}_1 + \mathrm{k}_2 < \mathrm{k}_1 + \mathrm{j}_2,$$
$$[(\mathrm{j}_1, \mathrm{k}_1)] \leq [(\mathrm{j}_2, \mathrm{k}_2)] \iff \mathrm{j}_1 + \mathrm{k}_2 \leq \mathrm{k}_1 + \mathrm{j}_2.$$

sindex $\iff$ $\Longrightarrow$ *Then*

(i) $<$ *and* $\leq$ *are transitive,*

(ii) $<$ *is irreflexive, and*

(iii) $\leq$ *is reflexive.*

Furthermore, for any $\mathrm{j}, \mathrm{k} \in \mathbb{Z}$, *either* $\mathrm{j} \leq \mathrm{k}$ *or* $\mathrm{k} \leq \mathrm{j}$.

Proof First one must show that the relations are well-defined in that they do not depend on the choice of representative. Thus let $[(j_1, k_1)] \sim [(\tilde{j}_1, \tilde{k}_1)]$ and $[(j_2, k_2)] \sim [(\tilde{j}_2, \tilde{k}_2)]$, so that

$$j_1 + \tilde{k}_1 = k_1 + \tilde{j}_1, \quad j_2 + \tilde{k}_2 = k_2 + \tilde{j}_2.$$

Now suppose that the relation $j_1 + k_2 < k_1 + j_2$ holds. Now perform the following steps:

1. add $\tilde{j}_1 + k_1 + j_2 + \tilde{k}_2 + j_1 + \tilde{k}_1 + k_2 + \tilde{j}_2$ to both sides of the relation;
2. observe that $j_1 + k_2 + k_1 + j_2$ appears on both sides of the relation;
3. observe that $j_1 + \tilde{k}_1$ appears on one side of the relation and that $\tilde{j}_1 + k_1$ appears on the other;

4. observe that $k_2 + \tilde{j}_2$ appears on one side of the relation and that $j_2 + \tilde{k}_2$ appears on the other.

After simplification using the above observations, and using Proposition 1.4.13, we note that the relation $\tilde{j}_1 + \tilde{k}_2 < \tilde{k}_1 + \tilde{j}_2$ holds, which gives independence of the definition of $<$ on the choice of representative. The same argument works for the relation $\leq$.

The remainder of the proof follows in a fairly straightforward manner from the corresponding assertions for $\mathbb{N}_0$, and we leave the details to the reader as Exercise 1.4.6. ■

As with the natural numbers, the last assertion of the previous result has a standard restatement.

1.4.21 Corollary (Trichotomy Law for $\mathbb{Z}$) *For $\mathrm{j}, \mathrm{k} \in \mathbb{Z}$, exactly one of the following possibilities holds:*

(i) $\mathrm{j} < \mathrm{k}$;

(ii) $\mathrm{k} < \mathrm{j}$;

(iii) $\mathrm{j} = \mathrm{k}$.

Similarly with $\mathbb{N}_0$, we shall also write "$j > k$" for "$k < j$" and "$j \geq k$" for "$k \leq j$. It is also easy to directly verify that the relations $<$ and $\leq$ have the expected properties with respect to positive and negative integers. These are given in Exercise 1.4.7, for the interested reader.

We also have the following extension of Proposition 1.4.13 that relates addition and multiplication to the relations $<$ and $\leq$. We again leave these to the reader to verify in Exercise 1.4.8.

1.4.22 Proposition (Relation between addition and multiplication and $<$) *For $\mathrm{j}, \mathrm{k}, \mathrm{m} \in \mathbb{Z}$, the following statements hold:*

(i) if $\mathrm{j} < \mathrm{k}$ *then* $\mathrm{j} + \mathrm{m} < \mathrm{k} + \mathrm{m}$;

(ii) if $\mathrm{j} < \mathrm{k}$ *and if* $\mathrm{m} > 0$ *then* $\mathrm{m} \cdot \mathrm{j} < \mathrm{m} \cdot \mathrm{k}$;

(iii) if $\mathrm{j} < \mathrm{k}$ *and if* $\mathrm{m} < 0$ *then* $\mathrm{m} \cdot \mathrm{k} < \mathrm{m} \cdot \mathrm{j}$;

(iv) if $0 < \mathrm{j}, \mathrm{k}$ *then* $0 < \mathrm{j} \cdot \mathrm{k}$.

1.4.5 The absolute value function

On the set of integers there is an important map that assigns a nonnegative integer to each integer.

1.4.23 Definition (Integer absolute value function) The ***absolute value function*** on $\mathbb{Z}$ is the map from $\mathbb{Z}$ to $\mathbb{N}_0$, denoted by $k \mapsto |k|$, defined by

$$|k| = \begin{cases} k, & 0 < k, \\ 0, & k = 0, \\ -k, & k < 0. \end{cases}$$

•

The absolute value has the following properties.

1.4.24 Proposition (Properties of absolute value on $\mathbb{Z}$) *The following statements hold:*

(i) $|\mathrm{k}| \geq 0$ *for all* $\mathrm{k} \in \mathbb{Z}$;

(ii) $|\mathrm{k}| = 0$ *if and only if* $\mathrm{k} = 0$;

(iii) $|\mathrm{j}\cdot\mathrm{k}| = |\mathrm{j}|\cdot|\mathrm{k}|$ *for all* $\mathrm{j},\mathrm{k}\in\mathbb{Z}$*;*

(iv) $|\mathrm{j}+\mathrm{k}| \le |\mathrm{j}|+|\mathrm{k}|$ *for all* $\mathrm{j},\mathrm{k}\in\mathbb{Z}$ *(**triangle inequality**).*

Proof Parts (i) and (ii) follow directly from the definition of $|\cdot|$.

(iii) We first note that $|-k| = |k|$ for all $k\in\mathbb{Z}$. Now, if $0\le j,k$, then the result is clear. If $j<0$ and $k\ge 0$, then

$$|j\cdot k| = |-1\cdot(-j)\cdot k| = |(-j)\cdot k| = |-j|\cdot|k| = |j|\cdot|k|\,.$$

A similar argument hold when $k<0$ and $j\ge 0$.

(iv) We consider various cases.

1. $|j|\le|k|$:
   - (a) $0\ge j,k$: Here $|j+k| = j+k$, and $|j| = j$ and $|k| = k$. So the result is obvious.
   - (b) $j<0$, $k\ge 0$: Here one can easily argue, using the definition of addition, that $0<j+k$. From Proposition 1.4.22 we have $j+k<0+k=k$. Therefore, $|j+k|<|k|<|j|+|k|$, again by Proposition 1.4.22.
   - (c) $k<0$, $j\ge 0$: This follows as in the preceding case, swapping j and k.
   - (d) $j,k<0$: Here $|j+k| = |-j+(-k)| = |-(j+k)| = -(j+k)$, and $|j| = -j$ and $|k| = -k$, so the result follows immediately.
2. $|k|\le|j|$: The argument here is the same as the preceding one, but swapping j and k. ■

Exercises

1.4.1 Let $k\in\mathbb{N}$. Show that $k\subset\mathbb{N}$; thus k is both an element of $\mathbb{N}$ and a subset of $\mathbb{N}$.

1.4.2 Let $j,k\in\mathbb{N}_0$. Do the following:
- (a) show that $j\in k$ if and only if $j\subset k$;
- (b) show that if $j\subsetneq k$, then $k\notin j$ (and so $j\in k$ by the Trichotomy Law).

1.4.3 Prove Proposition 1.4.13.

1.4.4 Complete the proof of Proposition 1.4.19.

1.4.5 For $j_1,j_2,k\in\mathbb{Z}$, prove the distributive rule $(j_1+j_2)\cdot k = j_1\cdot k + j_2\cdot k$.

1.4.6 Complete the proof of Proposition 1.4.20.

1.4.7 Show that the relations $<$ and $\le$ on $\mathbb{Z}$ have the following properties:
1. $[(0,j)] < [(0,0)]$ for all $j\in\mathbb{N}$;
2. $[(0,j)] < [(k,0)]$ for all $j,k\in\mathbb{N}$;
3. $[(0,j)] < [(0,k)]$, $j,k,\in\mathbb{N}_0$, if and only if $k<j$;
4. $[(0,0)] < [(j,0)]$ for all $j\in\mathbb{N}$;
5. $[(j,0)] < [(k,0)]$, $j,k\in\mathbb{N}_0$, if and only if $j<k$;
6. $[(0,j)] \le [(0,0)]$ for all $j\in\mathbb{N}_0$;
7. $[(0,j)] \le [(k,0)]$ for all $j,k\in\mathbb{N}_0$;
8. $[(0,j)] \le [(0,k)]$, $j,k,\in\mathbb{N}_0$, if and only if $k\le j$;
9. $[(0,0)] \le [(j,0)]$ for all $j\in\mathbb{N}_0$;
10. $[(j,0)] \le [(k,0)]$, $j,k\in\mathbb{N}_0$, if and only if $j\le k$.

1.4.8 Prove Proposition 1.4.22.

Section 1.5

Orders of various sorts

In Section 1.4 we defined two relations, denoted by $<$ and $\leq$, on both $\mathbb{N}_0$ and $\mathbb{Z}$. Here we see that these relations have additional properties that fall into a general class of relations called orders. There are various classes or orders, having varying degrees of "strictness," as we shall see.

Do I need to read this section? Much of the material in this section is not used widely in the series, so perhaps can be overlooked until it is needed. •

1.5.1 Definitions

Let us begin by defining the various types of orders we consider.

1.5.1 Definition (Partial order, total order, well order) Let S be a set and let R be a relation in S.

(i) R is a ***partial order*** in S if it is reflexive, transitive, and antisymmetric.

(ii) A ***partially ordered set*** is a pair (S, R) where R is a partial order in S.

(iii) R is a ***strict partial order*** in S if it is irreflexive and transitive.

(iv) A ***strictly partially ordered set*** is a pair (S, R) where R is a strict partial order in S.

(v) R is a ***total order*** in S if it is a partial order and if, for each $x_1, x_2 \in S$, either $(x_1, x_2) \in R$ or $(x_2, x_1) \in R$.

(vi) A ***totally ordered set*** is a pair (S, R) where R is a total order in S.

(vii) R is a ***well order*** in S if it is a partial order and if, for every nonempty subset $A \subset S$, there exists an element $x \in A$ such that $(x, x') \in R$ for every $x' \in A$.

(viii) A ***well ordered set*** is a pair (S, R) where R is a well order in S. •

1.5.2 Remark (Mathematical structures as ordered pairs) In the preceding definitions we see four instances of an "X set," where X is some property, e.g., a partial order. In such cases, it is common practice to do as we have done and write the object as an ordered pair, in the cases above, as (S, R). The practice dictates that the first element in the ordered pair be the name of the set, and that the second specifies the structure.

In many cases one simply wishes to refer to the set, with the structure being understood. For example, one might say, "Consider the partially ordered set $S\ldots$" and not make explicit reference to the partial order. Both pieces of language are in common use by mathematicians, and in mathematical texts. •

Let us consider some simple examples of partial and strict partial orders.

1.5.3 Examples (Partial orders)

1. Consider the relation $R = \{(k_1, k_2) \mid k_1 \leq k_2\}$ in either $\mathbb{N}_0$ or $\mathbb{Z}$. Then one verifies that R is a partial order. In fact, it is both a total order and a well order.
2. Consider the relation $R = \{(k_1, k_2) \mid k_1 \leq k_2\}$ in either $\mathbb{N}_0$ or $\mathbb{Z}$. Here one can verify that R is a strict partial order.

3. Let S be a set and consider the relation R in $\mathbf{2}^S$ defined by $R = \{(A, B) \mid A \subset B\}$. Here one can see that R is a partial order, but it is generally neither a total order nor a well order (cf. Exercise 1.5.2).
4. Let S be a set and consider the relation R in $\mathbf{2}^S$ defined by $R = \{(A, B) \mid A \subsetneq B\}$. In this case R can be verified to be a strict partial order.
5. A well order R is a total order. Indeed, for $(x_1, x_2) \in R$, there exists an element $x \in \{x_1, x_2\}$ such that $(x, x') \in R$ for every $x' \in \{x_1, x_2\}$. But this implies that either $(x_1, x_2) \in R$ or $(x_2, x_1) \in R$, meaning that R is a total order. •

Motivated by the first and second of these examples, we utilise the following more or less commonplace notation for partial orders.

1.5.4 Notation ($\preceq$ and $\prec$) If R is a partial order in S, we shall normally write $x_1 \preceq x_2$ for $(x_1, x_2) \in R$, and shall refer to $\preceq$ as the partial order. In like manner, if R is a strict partial order in S, we shall write $x_1 \prec x_2$ for $(x_1, x_2) \in R$. We shall also use $x_1 \succeq x_2$ and $x_1 \succ x_2$ to stand for $x_2 \preceq x_1$ and $x_2 \prec x_1$, respectively. •

There is a natural way of associating to every partial order a strict partial order, and vice versa.

1.5.5 Proposition (Relationship between partial and strict partial orders) *Let* S *be a set.*

(i) If $\preceq$ is a partial order in S*, then the relation $\prec$ defined by*

$$\mathrm{x}_1 \prec \mathrm{x}_2 \quad \Longleftrightarrow \quad \mathrm{x}_1 \preceq \mathrm{x}_2 \text{ and } \mathrm{x}_1 \neq \mathrm{x}_2$$

is a strict partial order in S*.*

(ii) If $\prec$ is a strict partial order in S*, then the relation $\preceq$ defined by*

$$\mathrm{x}_1 \preceq \mathrm{x}_2 \quad \Longleftrightarrow \quad \mathrm{x}_1 \prec \mathrm{x}_2 \text{ or } \mathrm{x}_1 = \mathrm{x}_2$$

is a partial order in S*.*

Proof This is a straightforward matter of verifying that the definitions are satisfied. ■

When talking about a partial order $\preceq$, the symbol $\prec$ will always refer to the strict partial order as in part (i) of the preceding result. Similarly, given a strict partial order $\prec$, the symbol $\preceq$ will always refer to the partial order as in part (ii) of the preceding result.

1.5.6 Examples (Example 1.5.3 cont'd)

1. One can readily verify that $<$ is the strict partial order associated with the partial order $\leq$ in either $\mathbb{N}_0$ or $\mathbb{Z}$, and that $\leq$ is the partial order associated to $<$.
2. It is also easy to verify that, for a set S, $\subsetneq$ is the strict partial order in $\mathbf{2}^S$ associated to the partial order $\subset$, and that $\subset$ is the partial order associated to $\subsetneq$. •

1.5.2 Subsets of partially ordered sets

Surrounding subsets of a partially ordered set $(S, \preceq)$ there is some useful language. For the following definition, it is helpful to think of an order, be it partial, strictly partial, or whatever, as a relation, and to use the notation of a relation. Thus we refer to an order as R, and not as $\preceq$.

1.5.7 Definition (Restriction of an order) Let S be a set and let R be a partial order, (resp. strict partial order, total order, well order) in S. For a subset $T \subset S$, the ***restriction*** of R to T is the partial order (resp. strict partial order, total order, well order) in T defined by

$$R|T = R \cap \{ (x_1, x_2) \in S \times S \mid \ x_1, x_2 \in T \} . \qquad \bullet$$

It is a trivial matter to see that if R is an order, then its restriction to T is an order having the same properties as R, as is tacitly assumed in the definition. The notion of the restriction of an order allows us to talk unambiguously about the order on a subset of a given set, and we shall do this freely in this section.

Since most of this section is language, let us begin with some simple language associated with points.

1.5.8 Definition (Comparing elements in a partially ordered set) Let $(S, \preceq)$ be a partially ordered set.

(i) A point $x_1 \in S$ is ***less*** than or ***smaller*** than x_2, or equivalently is a ***predecessor*** of x_2, if $x_1 \preceq x_2$.

(ii) A point $x_1 \in S$ is ***greater*** than or ***larger*** than x_2, or equivalently is a ***successor*** of x_2, if $x_1 \succeq x_2$.

(iii) A point x' is ***between*** x_1 and x_2 if $x_1 \preceq x'$ and if $x' \preceq x_2$.

Similarly, let $(S, \prec)$ be a strictly partially ordered set.

(iv) A point $x_1 \in S$ is ***strictly less*** than or ***strictly smaller*** than x_2, or equivalently is a ***strict predecessor*** of x_2, if $x_1 \prec x_2$.

(v) A point $x_1 \in S$ is ***strictly greater*** than or ***strictly larger*** than x_2, or equivalently is a ***strict successor*** of x_2, if $x_1 \succ x_2$.

(vi) A point x' is ***strictly between*** x_1 and x_2 if $x_1 \prec x'$ and if $x' \prec x_2$.

(vii) If $x_1 < x_2$ and there exists no $x' \in S$ that is strictly between x_1 and x_2, then x_1 is the ***immediate predecessor*** of x_2. $\bullet$

Next we talk about some language attached to subsets of a partially ordered set.

1.5.9 Definition (Segment, least, greatest, minimal, maximal) Let $(S, \preceq)$ be a partially ordered set.

(i) The ***initial segment*** determined by $x \in S$ is the set $\underline{\text{seg}}(x) = \{ x' \in S \mid \ x' \preceq S \}$.

(ii) A ***least***, ***smallest***, or ***first*** element in S is an element $x \in S$ with the property that $x \preceq x'$ for every $x' \in S$.

(iii) A ***greatest***, ***largest***, or ***last*** element in S is an element $x \in S$ with the property that $x' \preceq x$ for every $x' \in S$.

(iv) A ***minimal*** element of S is an element $x \in S$ with the property that $x \preceq x'$ implies that $x' = x$.

(v) A ***maximal*** element of S is an element $x \in S$ with the property that $x \prec x'$ implies that $x' = x$.

Now let $(S, \preceq)$ be a partially ordered set.

(vi) The ***strict initial segment*** determined by $x \in S$ is the set $\text{seg}(x) = \{ x' \in S \mid \ x' \prec S \}$. $\bullet$

The least and greatest elements of a set, if they exist, are unique. This is easy to prove (Exercise 1.5.4).

Let us give an example that distinguishes between least and minimal.

1.5.10 Example (Least and minimal are different) Let S be a set and consider the partially ordered set $(\mathbf{2}^S \setminus \emptyset, \subset)$. Then any singleton is a minimal element of $\mathbf{2}^S \setminus \emptyset$. However, unless S is itself a set with only one member, then $\mathbf{2}^S$ has no least element, i.e., there is no subset which is contained in every other subset. •

Next we turn to two important concepts related to partial orders.

1.5.11 Definition (Greatest lower bound and least upper bound) Let $(S, \preceq)$ be a partially ordered set and let $A \subset S$.

(i) An element $x \in S$ is a ***lower bound*** for A if $x \preceq x'$ for every $x' \in A$.

(ii) An element $x \in S$ is an ***upper bound*** for A if $x' \preceq x$ for every $x' \in A$.

(iii) If, in the set of lower bounds for A, there is a greatest element, this is the ***greatest lower bound***, or the ***infimum***, of E. This is denoted by $\inf(A)$.

(iv) If, in the set of upper bounds for A, there is a least element, this is the ***least upper bound***, or the ***supremum***, of E. This is denoted by $\sup(A)$.

Now let $(S, \prec)$ be a strictly partially ordered set and let $A \subset S$.

(v) An element $x \in S$ is a ***strict lower bound*** for A if $x \prec x'$ for every $x' \in A$.

(vi) An element $x \in S$ is a ***strict upper bound*** for A if $x' \prec x$ for every $x' \in A$. •

Let us give some examples that illustrate the various possibilities arising from the preceding definitions. The examples will be given for lower bounds, but similar examples can be conjured to give similar conclusions for upper bounds.

1.5.12 Examples (Greatest lower bounds)

1. A subset $A \subset S$ may have no lower bounds. For example, the set of negative integers has no lower bound if we use the standard partial order in $\mathbb{Z}$.
2. A subset $A \subset S$ may have a greatest lower bound in A. For example, the set of nonnegative integers has as lower bounds all nonpositive integers. The greatest of these lower bounds is 0, which is itself a nonnegative integer.
3. A subset $A \subset S$ may have a greatest lower bound that is not an element of A. To see this, let S be the set of nonpositive integers, let A be the set of negative integers, and define a partial order $\preceq$ in S by

$$k_1 \preceq k_2 \iff \begin{cases} k_1 \leq k_2, \ k_1, k_2 \in A, & \text{or} \\ k_1 = k_2 = 0, & \text{or} \\ k_1 = 0, \ k_2 \in A. \end{cases}$$

Thus this is the usual partial order in $A \subset S$, and one declares 0 to be less than all elements of A. In this case, 0 is the only lower bound for A, and so is, therefore, the greatest lower bound. But $0 \notin A$. •

1.5.3 Zorn's Lemma

Zorn's[7] Lemma comes up frequently in mathematics during the course of nonconstructive existence proofs. Since some of these proofs appear in this series and are important, we state Zorn's Lemma.

1.5.13 Theorem (Zorn's Lemma) *Every partially ordered set* $(\mathrm{S}, \preceq)$ *in which every totally ordered subset has an upper bound contains at least one maximal member.*

Proof Suppose that every totally ordered subset has an upper bound, but that S has no maximal member. By assumption, if $A \subset S$ is a totally ordered subset, then there exists an upper bound x for A. Since S has no maximal element, there exists $x' \in S$ such that $x < x'$. Therefore, x' is a strict upper bound for A. Thus we have shown that every totally ordered subset possesses a strict upper bound. Let b be a function from the collection of totally ordered subsets into S having the property that $b(A)$ is a strict upper bound for A.[8]

A **b-*set*** is a subset B of S that is well ordered and has the property that, for every $x \in B$, we have $x = b(\mathrm{seg}_B(x))$, where $\mathrm{seg}_B(x)$ denotes the strict initial segment of x in B.

1 Lemma *If* B_1 *and* B_2 *are unequal* b*-sets, then one of the following statements holds:*

(i) there exists $\mathrm{x}_1 \in \mathrm{B}_1$ *such that* $\mathrm{B}_2 = \mathrm{seg}_{\mathrm{B}_1}(\mathrm{x}_1)$*;*

(ii) there exists $\mathrm{x}_2 \in \mathrm{B}_2$ *such that* $\mathrm{B}_1 = \mathrm{seg}_{\mathrm{B}_2}(\mathrm{x}_2)$*.*

Proof If $B_2 \subsetneq B_1$, then we claim that (i) holds. Take x_1 to be the least member of $B_1 - B_2$. We claim that $B_2 = \mathrm{seg}_{B_1}(x_1)$. First of all, if $x \in B_2$, then $x < x_1$ since x_1 is the least member of $B_1 - B_2$. Therefore, $B_2 \subset \mathrm{seg}_{B_1}(x_1)$. Now suppose that $\mathrm{seg}_{B_1}(x_1) - B_2 \neq \emptyset$, and let x be the least member of this set. Note that for any $x' \in B_2$ we therefore have $x' < x$, contradicting the fact that x_1 is the least member of $B_1 - B_2$. Thus we must have $\mathrm{seg}_{B_1}(x_1) - B_2 = \emptyset$, and so $B_2 = \mathrm{seg}_{B_1}(x_1)$.

We now suppose that $B_2 - B_1 \neq \emptyset$. Let x_2 be the least member of $B_2 - B_1$. If $x \in \mathrm{seg}_{B_2}(x_2)$ then $x < x_2$ and x must therefore be an element of B_1, or else this contradicts the definition of x_2. Now suppose that $B_1 \setminus \mathrm{seg}_{B_2}(x_2) \neq \emptyset$ and let y_1 be the least member of this set. If $y \in \mathrm{seg}_{B_1}(y_1)$ and $y' \in B_2$ satisfies $y' < y$, then $y' \in \mathrm{seg}_{B_1}(y_1)$. If z is the least member of $B_2 \setminus \mathrm{seg}_{B_1}(y_1)$, we then have $\mathrm{seg}_{B_2}(z) = \mathrm{seg}_{B_1}(y_1)$. Therefore

$$z = b(\mathrm{seg}_{B_2}(z)) = b(\mathrm{seg}_{B_1}(y_1)) = y_1.$$

Since $y_1 \in B_1$, $z = y_1 \neq x_2$. Since $z \leq x_2$, it follows that $z < x_2$. Thus $y_1 = z \in \underline{\mathrm{seg}}_{B_2}(x_2)$. This, however, contradicts the choice of y_1, so we conclude that $B_1 \setminus \mathrm{seg}_{B_2}(x_2) = \emptyset$, and so that $B_1 = \mathrm{seg}_{B_2}(x_2)$. Thus (ii) holds.

A swapping of the rôles of B_1 and B_2 will complete the proof. ▼

2 Lemma *The union of all* b*-sets is a* b*-set.*

Proof Let U denote the union of all b-sets. First we must show that U is well ordered. Let $A \subset U$ and let $x \in A$. Then there is a b-set B such that $x \in B$. We claim that $\mathrm{seg}_A(x) \subset B$. Indeed, if $x' < x$ then, by Lemma 1, either $x' \in B$ or x' does not lie in any b-set. Since A lies in the union of all b-sets, it must be the case that $x' \in B$. Thus $\mathrm{seg}_A(x)$ is a subset of the well ordered set B, and as such has a least element x_0. This is clearly also a least element for A, so U is well ordered.

Next, let $x \in U$ and let B be a b-set such that $x \in B$. Our above argument shows that $\mathrm{seg}_U(x) \subset B$ so that $\mathrm{seg}_U(x) = \mathrm{seg}_B(x)$. Therefore, $x = b(\mathrm{seg}_B(x)) = b(\mathrm{seg}_U(x))$. This completes the proof. ▼

[7] Max August Zorn (1906–1993) was a German mathematician who did work in the areas of set theory, algebra, and topology.

[8] The existence of the function b relies on the Axiom of Choice (see Section 1.8.3).

To complete the proof, let U be the union of all b-sets and let $x = b(U)$. Then we claim that $U \cup \{x\}$ is a b-set. That $U \cup \{x\}$ is well ordered follows since U is well ordered and since x is an upper bound for U. Since U is the union of all b-sets, it must hold that $x \in U$. However, this contradicts the fact that x is a strict upper bound for U. ■

1.5.4 Induction and recursion

In some of the proofs we have given in this section, and in our definition of $\mathbb{N}_0$, we have used the idea of induction. This idea is an eminently reasonable one. One starts with a fact or a definition that applies to the element $0 \in \mathbb{N}_0$, and a rule for extending this from the jth number to the $(j+1)$st number, and then asserts that the fact or definition applies to all elements of $\mathbb{N}_0$. In this section we formulate this principle in a more general setting that the set $\mathbb{N}_0$, namely for a well ordered set.

Since the result will have to do with a property being true for the elements of a well ordered set, let us formally say that a ***property*** defined in a set S is a map $P\colon S \to \{\text{true}, \text{false}\}$. A property is ***true***, or ***holds***, at x if $P(x) = \text{true}$.

1.5.14 Theorem (Principle of Transfinite Induction) *Let* $(\mathrm{W}, \preceq)$ *be a well ordered set and let* P *be a property defined in* W*. Suppose that, for every* $\mathrm{w} \in \mathrm{W}$*, the fact that* $\mathrm{P}(\mathrm{w}')$ *is true for every* $\mathrm{w}' \prec \mathrm{w}$ *implies that* $\mathrm{P}(\mathrm{w})$ *is true. Then* $\mathrm{P}(\mathrm{w})$ *is true for every* $\mathrm{w} \in \mathrm{W}$*.*

Proof Suppose that the hypothesis is true, but the conclusion is false. Then

$$F = \{\, w \in W \mid \ P(w) = \text{false}\} \neq \emptyset.$$

Let w be the least element of F Therefore, for $w' < w$ it must hold that $P(w') = \text{true}$. But then the hypotheses imply that $P(w) = \text{true}$, so that $w \in W \setminus F$. This is a contradiction. ■

Next we turn to the process of defining something using recursion. As we did for induction, let us first consider doing this for $\mathbb{N}_0$. What we wish to define is a map $f\colon \mathbb{N}_0 \to S$. The idea for doing this is that, if, for each $k \in \mathbb{N}_0$, one knows the value of f on the first k elements of $\mathbb{N}_0$, and if one knows a rule for then giving the value of f at $k+1$, then the f extends uniquely to a function on all of $\mathbb{N}_0$. To give a concrete example, if $S = \mathbb{Z}$ and if we define $f(k+1) = 2 \cdot f(k)$, then the resulting function $f\colon \mathbb{N}_0 \to \mathbb{Z}$ is determined by its value at 0: $f(k) = 2^k \cdot f(0)$.

To state the general theorem requires some notation. We let W be a well ordered set and let S be a set. For $w \in W$, we let $\text{seq}_S(w)$ be the set of maps from $\text{seg}(w)$ into S. We then let $\text{Seq}_S(W)$ be the set of all maps of the form $g\colon \text{seq}_S(w) \to S$. The idea is that an element of $\mathrm{S}_S(W)$ tells us how to extend a map from $\text{seg}(w)$ to give its value at w.

The desired result is now the following.

1.5.15 Theorem (Transfinite recursion) *Let* $(\mathrm{W}, \preceq)$ *be a well ordered set and let* S *be a set. Given a member* $\mathrm{g} \in \mathrm{Seq}_{\mathrm{S}}(\mathrm{W})$*, there exists a unique map* $\mathrm{f_g}\colon \mathrm{W} \to \mathrm{S}$ *such that* $\mathrm{f_g(w)} = \mathrm{g(f|\,seg(w))}$*.*

Proof That there can be only one map f_g as in the theorem statement follows from the Principle of Transfinite Induction (take $P(w) = \text{true}$ if and only if $f_g(w) = g(f_g|\,\text{seg}(w))$).

So we shall prove the existence of f_g. Define

$$\mathscr{C}_g = \{A \subset W \times S| \\ w \in W,\, h \in \text{seq}_S(w),\, (w', h(w')) \in A \text{ for all } w' \in \text{seg}(w) \implies (w, g(h)) \in A\}.$$

Note that $W \times S \in \mathscr{C}_g$, so that C_g is not empty. It is easy to check that the intersection of members of $\mathscr{C}_g$ is also a member of $\mathscr{C}_g$. Therefore we let $F_g = \cap_{A \in \mathscr{C}_g} A$, and note that $F_g \in \mathscr{C}_g$.

We shall show that F_g is the graph of a function f_g that satisfies the conditions in the theorem statement.

First we need to show that, for each $w \in W$, there exists exactly one $x \in S$ such that $(w, x) \in F_g$. Define

$$A_g = \{ w \in W \mid \text{ there exists exactly one } x \in S \text{ such that } (w, x) \in F_g \}.$$

For $w \in W$, we claim that if $\text{seg}(w) \subset A_g$, then $w \in A_g$. Indeed, if $\text{seg}(w) \subset A_g$, define $h \in \text{seq}_S(w)$ by $h(w') = x'$ where $x' \in S$ is the unique element such that $(w', x') \in A_g$. Since $F_g \in \mathscr{C}_g$, there exists some $x \in S$ such that $(w, x) \in F_g$. Suppose that $x \neq g(h)$. We claim that $F_g - \{(w, x)\} \in \mathscr{C}_g$. Let $w' \in W$ and let $h' \in \text{seg}_S(w')$ satisfy $(w'', h'(w'')) \in F_g - \{(w, x)\}$ for all $w'' \in \text{seg}(w')$. If $w' = w$ then $h' = h$ by the uniqueness assertion of the theorem, and therefore $(w', g(h')) \in F_g - \{(w, x)\}$ since $x \neq g(h) = g(h')$. On the other hand, if $w' \neq w$ then $(w', g(h')) \in F_g - \{(w, x)\}$ since $F_g \in \mathscr{C}_g$. Thus, indeed, $F_g - \{(w, x)\} \in \mathscr{C}_g$, contradicting the fact that F_g is the intersection of all sets in $\mathscr{C}_g$. Thus we can conclude that $x = g(h)$, and therefore that there is exactly one $x \in S$ such that $(w, x) \in F_g$. By the Principle of Transfinite Induction, we can then conclude that for *every* $w \in W$, there is exactly one $x \in S$ such that $(w, x) \in F_g$. Thus F_g is the graph of a map $f_g \colon W \to S$.

It remains to verify that $f_g(w) = g(f_g|\,\text{seg}(w))$. This, however, follows easily from the definition of F_g. ■

One of the features of transfinite induction and transfinite recursion that requires some getting used to is that, unlike the usual induction with natural numbers as the well ordered set, one does not begin the induction or recursion by starting at 0 (or, in the case of a well ordered set, the least element), and proceeding element by element. Rather, one deals with initial segments. The reason for this is that in a well ordered set one may not have an immediate predecessor for every element, so that cannot be part of the induction/recursion; so the initial segment serves this purpose instead.

1.5.5 Zermelo's Well Ordering Theorem

The final topic in this section is a somewhat counterintuitive one. It says that every set possesses as well order.

1.5.16 Theorem (Zermelo's[9] Well Ordering Theorem) *For every set* S*, there is a well order in* S*.*

Proof Define

$$\mathscr{W} = \{ (W, \preceq_W) \mid \ W \subset S \text{ and } \preceq_W \text{ is a well order on } W \}.$$

Since $\emptyset \in \mathscr{W}$, $\mathscr{W}$ is nonempty. Define a partial order $\preceq$ on $\mathscr{W}$ by

$$W_1 \preceq W_2 \quad \Longleftrightarrow \quad W_2 \text{ is similar to a segment of } W_1.$$

Suppose that $\mathscr{T}$ is a totally ordered subset of $\mathscr{W}$.

1 Lemma *The set* $\cup_{\mathrm{A} \in \mathscr{T}} \mathrm{A}$ *has a unique well ordering, denoted by* $\lesssim$*, such that* $\mathrm{A}' \lesssim \cup_{\mathrm{A} \in \mathscr{T}}$ *for all* $\mathrm{A}' \in \mathscr{T}$*.*

Proof Let $x_1, x_2 \in \cup_{A \in \mathscr{T}} A$, and let $W_1, W_2 \in \mathscr{T}$ have the property that $x_1 \in W_1$ and $x_2 \in W_2$. Note that since either $W_1 = W_2$, $W_1 \preceq W_2$, or $W_2 \preceq W_1$, it must be the case that x_1 and x_2 lie in the same set from $\mathscr{C}$, let us call this W. The order in $\cup_{A \in \mathscr{T}} A$ is then defined by giving to the points x_1 and x_2 their order in W. This is unambiguous since $\mathscr{T}$ is totally ordered. It is then a simple exercise, left to the reader, that this is a well order. ▼

[9] Ernst Friedrich Ferdinand Zermelo (1871–1953) was a German mathematician whose mathematical contributions were mainly in the area of set theory.

The lemma ensures that the hypotheses of Zorn's Lemma apply to the totally ordered subsets of $\mathscr{W}$, and therefore the conclusions of Zorn's Lemma ensure that there is a maximal element W in $\mathscr{W}$. We claim that this maximal element is S. Suppose this is not the case, and that $x \in S - W$. We claim that $W \cup \{x\} \in \mathscr{W}$. To see this, simply define a well order on $W \cup \{x\}$ by asking that points in W have their usual order, and that x be greater that all points in W. The result is easily verified to be a well order on $W \cup \{x\}$, so contradiction the maximality of W. This completes the proof. ■

It might be surprising that it should be possible to well order any set. A well order can be thought of as allowing an arranging of the elements in a set, starting from the least element, and moving upwards in order:

$$x_0 < x_1 < x_2 < \cdots .$$

The complicated thing to understand here are the "$\cdots$," since they only mean "and so on" with an appropriate interpretation of these words (this is entirely related to the idea of ordinal numbers discussed in Section 1.7.1). As an example, the reader might want to imagine trying to order the real numbers (which we define in Section 2.1). It might seem absurd that it is possible to well order the real numbers. However, this is one of the many counterintuitive consequences arising from set theory, in this case directly related to the Axiom of Choice (Section 1.8.3).

1.5.6 Similarity

Between partially ordered sets, there are classes of maps that are distinguished by their preserving of the order relation. In this section we look into these and some of their properties, particularly with respect to well orders.

1.5.17 Definition (Similarity) If $(S, \preceq_S)$ and $(T, \preceq_T)$ are partially ordered sets, a bijection $f\colon S \to T$ is a ***similarity***, and $(S, \preceq_S)$ and $(T, \preceq_T)$ are said to be ***similar***, if $f(x_1) \preceq_T f(x_2)$ if and only if $x_1 \preceq_S x_2$. •

Now we prove a few results relating to similarities between well ordered sets. These shall be useful in our discussion or ordinal numbers in Section 1.7.1.

1.5.18 Proposition (Similarities of a well ordered set with itself) *If* $(\mathrm{S}, \preceq)$ *is a well ordered set and if* $\mathrm{f}\colon \mathrm{S} \to \mathrm{S}$ *is a similarity, then* $\mathrm{x} \preceq \mathrm{f(x)}$ *for each* $\mathrm{x} \in \mathrm{S}$.

Proof Define $A = \{x \in S \mid f(x) \prec x\}$ and let x be the least element of A. Then, for any $x' < x$, we have $x' \not\preceq f(x')$. In particular, $f(x) \preceq f \circ f(x)$. But $f(x) < x$ implies that $f \circ f(x) < f(x)$, giving a contradiction. Thus $A = \emptyset$. ■

1.5.19 Proposition (Well ordered sets are similar in at most one way) *If* $\mathrm{f, g}\colon \mathrm{S} \to \mathrm{T}$ *are similarities between well ordered sets* $(\mathrm{S}, \preceq_{\mathrm{S}})$ *and* $(\mathrm{T}, \preceq_{\mathrm{T}})$, *then* $\mathrm{f} = \mathrm{g}$.

Proof Let $h = f^{-1} \circ g$, and note that h is a similarity from S to itself. By Proposition 1.5.18 this implies that $x \preceq_S h(x)$ for each $x \in S$. Thus

$$\begin{aligned} & x \preceq_S f^{-1} \circ g(x), \qquad x \in S \\ \implies \quad & f(x) \preceq_T g(x), \qquad x \in S. \end{aligned}$$

Reversing the argument gives $g(x) \preceq_T f(x)$ for every $x \in S$. This gives the result. ■

1.5.20 Proposition (Well ordered sets are not similar to their segments) *If* $(\mathrm{S}, \prec)$ *is a well ordered set and if* $\mathrm{x} \in \mathrm{S}$, *then* S *is not similar to* $\mathrm{seg(x)}$.

Proof If $f(x) \in \mathrm{seg}(x)$ then $f(x) < x$, contradiction Proposition 1.5.18. ■

The final result is the deepest of the results we give here, because it gives a rather simple structure to the collection of all well ordered sets.

1.5.21 Proposition (Comparing well ordered sets) *If* $(\mathrm{S}, \preceq_{\mathrm{S}})$ *and* $(\mathrm{T}, \preceq_{\mathrm{T}})$ *are well ordered sets, then one of the following statements holds:*

(i) S *and* T *are similar;*

(ii) there exists $\mathrm{x} \in \mathrm{S}$ *such that* seg(x) *and* T *are similar;*

(iii) there exists $\mathrm{y} \in \mathrm{T}$ *such that* seg(y) *and* S *are similar.*

Proof Define

$$S_0 = \{\, x \in S \mid \text{ there exists } y \in T \text{ such that } \mathrm{seg}(x) \text{ is similar to } \mathrm{seg}(y) \}\,,$$

noting that S_0 is nonempty, since the segment of the least element in S is similar to the segment of the least element in T. Define $f\colon S_0 \to T$ by $f(x) = y$ where $\mathrm{seg}(x)$ is similar to $\mathrm{seg}(y)$. Note that this uniquely defines f by Propositions 1.5.19 and 1.5.20. We then take $T_0 = \mathrm{image}(f)$. If $S_0 = S$, then the result immediately follows. If $S_0 \subsetneq S$, then we claim that $S_0 = \mathrm{seg}(x_0)$ for some $x_0 \in S$. Indeed, we simply take x_0 to be the least strict upper bound for S_0, and then apply the definition of S_0 to see that $S_0 = \mathrm{seg}(x_0)$. We next claim that $T_0 = T$. Indeed, suppose that $T_0 \subsetneq T$, let y_0 be the least strict upper bound for T_0, and let x_0 be the least strict upper bound for S_0. We claim that $\mathrm{seg}(x_0)$ is similar to $\mathrm{seg}(y_0)$. Indeed, if this is not the case, then there exists $y < y_0$ such that $\mathrm{seg}(y)$ is not similar to a segment in S. However, this contradicts the definition of T_0. ■

Exercises

1.5.1 Show that any set S possesses a partial order.

1.5.2 Give conditions on S under which the partial order $\subset$ on $\mathbf{2}^S$ is

(a) a total order or

(b) a well-order.

1.5.3 Given two partially ordered sets $(S, \preceq_S)$ and $(T, \preceq_T)$, we define a relation $\preceq_{S\times T}$ in $S \times T$ by

$$(x_1, y_1) \preceq_{S\times T} (x_2, y_2) \quad \Longleftrightarrow \quad (x_1 \prec_S x_2) \text{ or } (x_1 = x_2 \text{ and } y_1 \preceq_T y_2).$$

This is called the ***lexicographic order*** on $S \times T$. Show the following:

(a) the lexicographic order is a partial order;

(b) if $\preceq_S$ and $\preceq_T$ are total orders, then the lexicographic order is a total order.

1.5.4 Show that a partially ordered set $(S, \preceq)$ possesses at most one least element and/or at most one greatest element.

Section 1.6

Families of sets and elements of sets

In this section we discuss general collections of sets, and general collections of members of sets. In Section 1.1.3 we considered Cartesian products of a finite collection of sets. In this section, we wish to extend this to allow for an arbitrary collection of sets. The often used idea of an index set is introduced here, and will come up on many occasions in the text.

Do I need to read this section? The idea of a general family of sets, and notions related to it, do not arise in a lot of places in these volumes. But they do arise. The ideas here are simple, and so perhaps can be read through. But the reader in a rush can skip the material, knowing they can look back on it if necessary. •

1.6.1 General Cartesian products

Before giving general definitions, it pays to revisit the idea of the Cartesian product $S_1 \times S_2$ of sets S_1 and S_2 as defined in Section 1.1.3 (the reason for our change from S and T to S_1 and S_2 will become clear shortly). Let $A = \{1, 2\}$, and let $f\colon A \to S_1 \cup S_2$ be a map satisfying $f(1) \in S_1$ and $f(2) \in S_2$. Then $(f(1), f(2)) \in S_1 \times S_2$. Conversely, given a point $(x_1, x_2) \in S_1 \times S_2$, we define a map $f\colon A \to S_1 \cup S_2$ by $f(1) = x_1$ and $f(2) = x_2$, noting that $f(1) \in S_1$ and $f(2) \in S_2$.

The punchline is that, for a pair of sets S_1 and S_2, their Cartesian product is in 1–1 correspondence with maps f from $A = \{1, 2\}$ to $S_1 \cup S_1$ having the property that $f(x_1) \in S_1$ and $f(x_2) \in S_2$. There are two things to note here: (1) the use of the set A to label the sets S_1 and S_2 and (2) the alternative characterisation of the Cartesian product.

The preceding discussion motivates the following definitions.

1.6.1 Definition (Family of sets) Let A be a set. A ***family of sets*** with ***index set*** A is a collection of sets, one associated to each member of A. The set associated to $a \in A$ is typically denoted by S_a, and the collection is denoted by $\{S_a\}_{a\in A}$. •

Thus a family of sets is nothing more than what we have been referring to previously as a "collection" of sets. The difference now is that we are using a separate set, the index set, to label the sets. In practice, this is typically convenient. In this case, one also uses the notation $\cup_{a\in A} S_a$ and $\cap_{a\in A} S_a$ to denote the union and intersection of a family of sets indexed by A. Similarly, when considering the disjoint union of a family of sets indexed by A, we define this to be

$$\mathring{\cup}_{a\in A} S_a = \cup_{a\in A}(\{a\} \times S_a).$$

Thus an element in the disjoint union has the form (a, x) where $x \in S_a$. Just as with the disjoint union of a pair of sets, the disjoint union of a family of sets keeps track of the set that element belongs to, now labelled by the index set A, along with the element. A family of sets $\{S_a\}_{a\in A}$ is ***pairwise disjoint*** if, for every distinct $a_1, a_2 \in A$, $S_{a_1} \cap S_{a_2} = \emptyset$.

Often when one writes $\{S_a\}_{a\in A}$, one omits saying that the family is "indexed by A," this being understood from the notation. Moreover, many authors will say things like, "Consider the family of sets $\{S_a\}$," so omitting any reference to the index set. In such cases, the index set is usually understood (often it is $\mathbb{N}$). However, we shall not use this notation, and will always give a symbol for the index set.

Next we generalise the Cartesian product to families of sets.

1.6.2 Definition (Cartesian product) The ***Cartesian product*** of a family of sets $\{S_a\}_{a\in A}$ is the set

$$\prod_{a\in A} S_a = \{\, f\colon A \to \cup_{a\in A} S_a \mid \ \ f(a) \in S_a\}\,. \qquad \bullet$$

Note that the analogue to the ordered pair in a general Cartesian product is simply the set $f(A)$ for some $f \in \prod_{a\in A} S_a$. The reader should convince themselves that this is indeed the appropriate generalisation.

1.6.2 Sequences and generalisations

In this section we fix a set S and we consider collections of elements in S.

1.6.3 Definition (Families of elements of a set, sequence, subsequence) Let A and S be sets.

(i) A ***family of elements*** of S with ***index set*** A is a map $f\colon A \to S$. A family $f\colon A \to S$ of elements of S is typically denoted by $\{f(a)\}_{a\in A}$, or simply by $\{x_a\}_{a\in A}$.

(ii) If $A = \mathbb{N}$, a family $\{x_j\}_{j\in\mathbb{N}}$ is a ***sequence*** in S.

(iii) A ***subsequence*** of a sequence $\{x_j\}_{j\in\mathbb{N}}$ in S is a map $f\colon A \to S$ where

(a) $A \subset \mathbb{N}$ is a nonempty set with no upper bound and

(b) $f(k) = x_k$ for all $k \in A$.

If the elements in the set A are ordered as $j_1 < j_2 < j_3 < \cdots$, then the subsequence may be written as $\{x_{j_k}\}_{k\in\mathbb{N}}$. $\bullet$

As with families of sets, it is not uncommon to see a family of elements of a set simply denoted by $\{x_a\}$, with no reference to the index set. Again, we shall not adopt this notation, preferring to always name the index set.

Exercises

1.6.1

Section 1.7

Ordinal numbers, cardinal numbers, cardinality

The notion of cardinality has to do with the "size" of a set. For sets with finite numbers of elements, there is no problem with "size." For example, it is clear what it means for one set with a finite number of elements to be "larger" or "smaller" than another set with a finite number of elements. However, for sets with infinite numbers of elements, can one be larger than another? If so, how can this be decided? In this section we see that there is a set, called the ***cardinal numbers***, which exactly characterises the "size" of all sets, just as natural numbers characterise the "size" if finite sets.

Do I need to read this section? The material in this section is used only slightly, so it can be thought of as "cultural," and hopefully interesting. Certainly the details of constructing the ordinal numbers, and then the cardinal numbers, plays no essential rôle in these volumes. The idea of cardinality comes up, but only in the simple sense of Theorem 1.7.12. •

1.7.1 Ordinal numbers

Ordinal numbers generalise the natural numbers. Recall from Section 1.4.1 that a natural number is a set, and moreover, from Section 1.4.2, a well ordered set. Indeed, the number $k \in \mathbb{N}_0$ is, by definition,

$$k = \{0, 1, \dots, k-1\}.$$

Moreover, note that, for every $j \in k$, $j = \operatorname{seg}(j)$. This motivates our definition of the ordinal numbers.

1.7.1 Definition (Ordinal number) An ***ordinal number*** is a well ordered set $(o, \leq)$ with the property that, for each $x \in o$, $x = \operatorname{seg}(x)$. •

Let us give some examples of ordinal numbers. The examples we give are all of "small" ordinals. We begin our constructions in a fairly detailed way, and then we omit the details as we move on, since the idea becomes clear after the initial constructions.

1.7.2 Examples (Ordinal numbers)

1. As we saw before we stated Definition 1.7.1, each nonnegative integer is an ordinal number.
2. The set $\mathbb{N}_0$ is an ordinal number. This is easily verified, but discomforting. We are saying that the set of numbers is itself a new kind of number, an ordinal number. Let us call this ordinal number ω. Pressing on...
3. The successor $\mathbb{N}_0^+ = \mathbb{N}_0 \cup \{\mathbb{N}_0\}$ is also an ordinal number, in just the same manner as a natural number is an ordinal number. This ordinal number is denoted by $\omega + 1$.
4. One carries on in this way defining ordinal numbers $\omega + (k+1) = (\omega + k)^+$.
5. Next we assume that there is a set containing ω and all of its successors. In axiomatic set theory, this follows from a construction like that justifying Assumption 1.4.3, along with another axiom (the Axiom of Substitution; see Section 1.8.2) saying, essentially, that we can repeat the process. Just as we did with the definition of $\mathbb{N}_0$, we take the smallest of these sets of successors to arrive at a net set that is to ω as ω is to 0. As was $\omega = \mathbb{N}_0$,

we well order this set by the partial order $\subset$. This set is then clearly an ordinal number, and is denoted by $\omega 2$.

6. One now proceeds to construct the successors $\omega 2+1 = \omega 2^+$, $\omega 2+2 = (\omega 2+1)^+$, and so on. These new sets are also ordinal numbers.
7. The preceding process yields ordinal numbers $\omega, \omega 2, \omega 3$, and so on.
8. We now again apply the same procedure to define an ordinal number that is contains ω, $\omega 2$, etc. This set we denote by ω^2.
9. One then defines $\omega^2+1 = (\omega^2)^+$, $\omega^2+2 = (\omega^2+1)^+$, etc., noting that these two are all ordinal numbers.
10. Next comes $\omega^2+\omega$, which is the set containing all ordinal numbers ω^2+1, ω^2+2, etc.
11. Then comes $\omega^2+\omega+1$, $\omega^2+\omega+2$, etc.
12. Following these is $\omega^2+\omega 2$, $\omega^2+\omega 2+1$, and so on.
13. Then comes $\omega^2+\omega 3$, $\omega^2+\omega 3+1$, and so on.
14. After ω^2, $\omega^2+\omega$, $\omega^2+\omega 2$, and so on, we arrive at $\omega^2 2$.
15. One then arrives at $\omega^2 2+1, \ldots, \omega^2 2+\omega, \ldots, \omega^2 2+\omega 2$, etc.
16. After $\omega^2 2$, $\omega^2 3$, and so on comes ω^3.
17. After ω, ω^2, ω^3, etc., comes ω^ω.
18. After ω, ω^ω, ω^{ω^ω}, etc., comes ϵ_0. The entire construction starts again from ϵ_0. Thus we get to ϵ_0+1, ϵ_0+2, and so on reproducing all of the above steps with an ϵ_0 in front of everything.
19. Then we get $\epsilon_0 2$, $\epsilon_0 3$, and so on up to $\epsilon_0\omega$.
20. These are followed by $\epsilon_0\omega^2$, $\epsilon_0\omega^3$ and so on up to $\epsilon_0\omega^\omega$.
21. Then comes $\epsilon_0\omega^{\omega^\omega}$, etc.
22. These are followed by ϵ_0^2.
23. We hope the reader is getting the point of these constructions, and can produce more such ordinals derived from the natural numbers. •

The above constructions of examples of ordinal numbers suggests that there are a lot of them. However, the concrete constructions do not really do justice to the number of ordinals. The ordinals that are elements of $\mathbb{N}_0$ are called ***finite*** ordinals, and all other ordinals are ***transfinite***. All of the ordinals we have named above are called "countable" (see Definition 1.7.13). There are many other ordinals not included in the above list, but before we can appreciate this, we first have to describe some properties of ordinals.

First we note that ordinals are exactly defined by similarity. More precisely, we have the following result.

1.7.3 Proposition (Similar ordinals are equal) *If* o_1 *and* o_2 *are similar ordinal numbers then* $\mathrm{o}_1 = \mathrm{o}_2$.

Proof Let $f\colon o_1 \to o_2$ be a similarity and define

$$S = \{x \in o_1 \mid \ f(x) = x\}.$$

We wish to show that $S = o_1$. Suppose that $\mathrm{seg}(x) \subset S$ for $x \in o_1$. Then x is the least element of $\mathrm{seg}(x)$ and, since f is a similarity, $f(x)$ is the least element of $f(\mathrm{seg}(x))$. Therefore, x and $f(x)$ both have $\mathrm{seg}(x)$ as their strict initial segment, by definition of S. Thus, by the definition of ordinal numbers, $x = f(x)$. The result now follows by the Principle of Transfinite Induction. ■

The next result gives a rather rigid structure to any set of ordinal numbers.

1.7.4 Proposition (Sets of ordinals are always well ordered) *If* O *is a set of ordinal numbers, then this set is well ordered by* $\subset$.

Proof First we claim that O is totally ordered. Let $o_1, o_2 \in O$ and note that these are both well ordered sets. Therefore, by Proposition 1.5.21, either $o_1 = o_2$, o_1 is similar to a strict initial segment in o_2, or o_2 is similar to a strict initial segment in o_1. In either of the last two cases, it follows from Proposition 1.7.3 that either o_1 is *equal* to a strict initial segment in o_2, or vice versa. Thus, either $o_1 \leq o_2$ or $o_2 \leq o_1$. Thus O is totally ordered, a fact we shall assume in the remainder of the proof.

Let $o \in O$. If $o \leq o'$ for every $o' \in O$, then o is the least member of O, and so O has a least member, namely o. If o is not the least member of O, then there exists $o' \in O$ such that $o' < o$. Thus $o' \in o$ and so the set $o \cap E$ is nonempty. Let o_0 be the least element of o. We claim that o_0 is also the least element of O. Indeed, let $o' \in O$. If $o' < o$ then $o' \in o \cap E$ and so $o_0 \leq o'$. If $o \leq o'$ then $o_0 < o'$, so showing that o_0 is indeed the least element of O. ■

Our constructions in Example 1.7.2, and indeed the definition of an ordinal number, suggest the true fact that every ordinal number has a successor that is an ordinal number. However, it may not be the case that an ordinal number has an immediate predecessor. For example, each of the ordinals that are natural numbers has an immediate predecessor, but the ordinal ω does not have an immediate predecessor. That is to say, there is no largest ordinal number strictly less ω.

Recall that the set $\mathbb{N}_0$ was defined by being the smallest set, having a certain property, that contains all nonnegative integers. One can then ask, "Is there a set containing all ordinal numbers?" It turns out the definition of the ordinal numbers prohibits this.

1.7.5 Proposition (Burali-Forti Paradox) *There is no set* $\mathbb{O}$ *having the property that, if* o *is an ordinal number, then* $\mathrm{o} \in \mathbb{O}$.

Proof Suppose that such a set $\mathbb{O}$ exists. We claim that $\operatorname{supp} \mathbb{O}$ exists and is an ordinal number. Indeed, we claim that $\operatorname{supp} \mathbb{O} = \cup_{o \in \mathbb{O}} o$. Note that the set $\cup_{o \in \mathbb{O}} o$ is well ordered by inclusion by Proposition 1.7.4. Clearly, $\cup_{o \in \mathbb{O}}$ is the smallest such set containing each $o \in \mathbb{O}$. Moreover, it is also clear from Proposition 1.7.4 that if $o' \in \cup_{o \in \mathbb{O}}$, then $o' = \operatorname{seg}(o')$. Thus $\operatorname{supp} \mathbb{O}$ exists, and is an ordinal number. Moreover, this order number is greater than all those in $\mathbb{O}$, thus showing that $\mathbb{O}$ cannot exist. ■

For our purposes, the most useful feature of the ordinal numbers is the following.

1.7.6 Theorem (Ordinal numbers can count the size of a set) *If* $(\mathrm{S}, \preceq)$ *is a well ordered set, then there exists a unique ordinal number* o_S *with the property that* S *and* o_S *are similar.*

Proof The uniqueness follows from Proposition 1.7.3. Let $x_0 \in S$ have the property that if $x < x_0$ then $\operatorname{seg}(x)$ is similar to some (necessarily unique) ordinal. (Why does x_0 exist?) Now let $P(x, o)$ be the proposition "o is an ordinal number similar to $\operatorname{seg}(x)$". Then define the set of ordinal numbers

$$o_0 = \{\, o \mid \text{ for each } x \in \operatorname{seg}(x_0), \text{ there exists } o \text{ such that } P(x, o) \text{ holds} \}\,.$$

One can easily verify that o_0 is itself an ordinal number that is similar to $\operatorname{seg}(x_0)$. Therefore, the Principle of Transfinite Induction can be applied to show that S is similar to an ordinal number. ■

This theorem is important, because it tells us that the ordinal numbers are the same, essentially, as the well ordered sets. Thus one can use the two concepts interchangeably; this is not obvious from the definition of an ordinal number.

It is also possible to define addition and multiplication of ordinal numbers. Since we will not make use of this, let us merely sketch how this goes. For ordinal numbers o_1 and o_2, let $(S_1, \preceq_1)$ and $(S_2, \preceq_2)$ be well ordered sets similar to o_1 and o_2, respectively. Define a partial order in $S_1 \mathring{\cup} S_2$ by

$$(i_1, x_1) \preceq_+ (i_2, x_2) \quad \iff \quad \begin{cases} i_1 = i_2,\ x_1 \preceq_{i_1}, & \text{or} \\ i_1 < i_2. \end{cases}$$

One may verify that this is a well order. Then define $o_1 + o_2$ as the unique ordinal number equivalent to the well ordered set $(S_1 \mathring{\cup} S_2, \preceq_+)$. To define product of o_1 and o_2, on the Cartesian product $S_1 \times S_2$ consider the partial order

$$(x_1, x_2) \preceq_\times (y_1, y_2) \quad \iff \begin{cases} x_2 \prec_2 y_2, & \text{or} \\ x_2 = y_2,\ x_1 \prec_1 y_1. \end{cases}$$

Again, this is verifiable as being a well order. One then defines $o_1 \cdot o_2$ to be the unique ordinal number similar to the well ordered set $(S_1 \times S_2, \preceq_\times)$. One must exercise care when dealing with addition and multiplication of ordinals, since, for example, neither addition not multiplication are commutative. For example, $1 + \omega \neq \omega + 1$ (why?). However, since we do not make use of this arithmetic, we shall not explore this further. It is worth noting that the notation in Example 1.7.2 is derived from ordinal arithmetic. Thus, for example, $\omega 2 = \omega \cdot 2$, etc.

1.7.2 Cardinal numbers

The cardinal numbers, as mentioned at the beginning of this section, are intended to be measures of the size of a set. If one combines the Zermelo's Well Ordering Theorem (Theorem 1.5.16) and Theorem 1.7.6, one might be inclined to say that the ordinal numbers are suited to this task. Indeed, simply place a well order on the set of interest by Theorem 1.5.16, and then use the associated ordinal number, given by Theorem 1.7.6, to define "size." The problem with this construction is that this notion of the "size" of a set would depend on the choice of well ordering. As an example, let us take the set $\mathbb{N}_0$. We place two well orderings on $\mathbb{N}_0$, one being the natural well ordering $\leq$ and the other being defined by

$$k_1 \preceq k_2 \quad \iff \quad \begin{cases} k_1 \leq k_2,\ k_1, k_2 \in \mathbb{N}, & \text{or} \\ k_1 = k_2 = 0, & \text{or} \\ k_1 = 0,\ k_2 \in \mathbb{N}. \end{cases}$$

Thus, for the partial order $\preceq$, one places 0 after all other natural numbers. One then verifies that $(\mathbb{N}_0, \leq)$ is similar to the ordinal number ω and that $(\mathbb{N}_0, \preceq)$ is similar to the ordinal number $\omega + 1$. Thus, even in a fairly simple example of a non-finite set, we see that the well order can change the size, if we go with size being determined by ordinals.

Therefore, we introduce a special subset of ordinals.

1.7.7 Definition (Cardinal number) A ***cardinal number*** is an ordinal number c with the property that, for all ordinal numbers o for which there exists a bijection from c to o, we have $c \leq o$. •

In other words, a cardinal number is the least ordinal number in a collection of ordinal numbers that are equivalent. Note that finite ordinals are only equivalent with a single

ordinal, namely themselves. However, transfinite ordinals may be equivalent to different transfinite ordinals. The following example illustrates this.

1.7.8 Example (Equivalent transfinite ordinals) We claim that there is a 1–1 correspondence between ω and $\omega + 1$. We can establish this correspondence explicitly by defining a map $f\colon \omega \to \omega + 1$ by

$$f(x) = \begin{cases} \omega, & x = 0, \\ x - 1, & x \in \mathbb{N}, \end{cases}$$

where $x - 1$ denotes the immediate predecessor of $x \in \mathbb{N}$.

One can actually check that *all* of the ordinal numbers presented in Example 1.7.2 are equivalent to ω! This is a consequence of Proposition 1.7.16 below. Accepting this as fact for the moment, we see that the only ordinals from Example 1.7.2 that are cardinal numbers are the elements of $\mathbb{N}_0$ along with ω. •

Certain of the facts about ordinal numbers translate directly to equivalent facts about cardinal numbers. Let us record these

1.7.9 Proposition (Properties of cardinal numbers) *The following statements hold:*

(i) *if* c_1 *and* c_2 *are similar cardinal numbers then* $\mathrm{c}_1 = \mathrm{c}_2$*;*

(ii) *if* C *is a set of cardinal numbers, then this set is well ordered by* $\subset$*;*

(iii) *there is no set* $\mathbb{C}$ *having the property that, if* c *is an cardinal number, then* $\mathrm{c} \in \mathbb{C}$ *(**Cantor's paradox**).*[10]

Proof The only thing that does not follow immediately from the corresponding results for ordinal numbers is Cantor's Paradox. The proof of this part of the result goes exactly as does that of Proposition 1.7.5. One only needs to verify that, if $\mathbb{C}$ is any set of cardinal numbers, then there exists a cardinal number greater or equal to $\operatorname{supp}\mathbb{C}$. This, however, is clear since $\operatorname{supp}\mathbb{C}$ is an ordinal number strictly greater than any element of $\mathbb{C}$, meaning that there is a corresponding cardinal number c equivalent to $\operatorname{supp}\mathbb{C}$. Thus $c \geq \operatorname{supp}\mathbb{C}$. ■

1.7.3 Cardinality

Cardinality is the measure of the "size" of a set that we have been after. The following result sets the stage for the definition.

1.7.10 Lemma *For a set* S *there exists a unique cardinal number* card(S) *such that* S *and* card(S) *are equivalent.*

Proof By Theorem 1.7.6 there exists an ordinal number o_S that is similar to S, and therefore equivalent to S. Any ordinal equivalent to o_S is therefore also equivalent to S, since equivalence of sets is an "equivalence relation" (Exercise 1.3.6). Therefore, the result follows by choosing the unique least element in the set of ordinals equivalent to o_S. ■

With this fact at hand, the following definition makes sense.

1.7.11 Definition (Cardinality) The ***cardinality*** of a set S is the unique cardinal number $\operatorname{card}(S)$ that is equivalent to S. •

The next result indicates how one often deals with cardinality in practice. The important thing to note is that, provided one is interested only in *comparing* cardinalities of sets, then one need not deal with the complication of cardinal numbers.

details

[10]

1.7.12 Theorem (Cantor–Schröder–Bernstein[11] Theorem) *For sets* S *and* T*, the following statements are equivalent:*

(i) $\mathrm{card}(\mathrm{S}) = \mathrm{card}(\mathrm{T})$*;*

(ii) there exists a bijection $\mathrm{f}\colon \mathrm{S} \to \mathrm{T}$*;*

(iii) there exists injections $\mathrm{f}\colon \mathrm{S} \to \mathrm{T}$ *and* $\mathrm{g}\colon \mathrm{T} \to \mathrm{S}$*;*

(iv) there exists surjections $\mathrm{f}\colon \mathrm{S} \to \mathrm{T}$ *and* $\mathrm{g}\colon \mathrm{T} \to \mathrm{S}$*.*

Proof It is clear from Lemma 1.7.10 that (i) and (ii) are equivalent. It is also clear that (ii) implies both (iii) and (iv).

(iii) $\Longrightarrow$ (ii) We start with a lemma.

1 Lemma *If* $\mathrm{A} \subset \mathrm{S}$ *and if there exists an injection* $\mathrm{f}\colon \mathrm{S} \to \mathrm{A}$*, then there exists a bijection* $\mathrm{g}\colon \mathrm{S} \to \mathrm{A}$*.*

Proof Define $B_0 = S \setminus A$ and then inductively define B_j, $j \in \mathbb{N}$, by $B_{j+1} = f(B_j)$. We claim that the sets $\{B_j\}_{j\in\mathbb{N}_0}$ are pairwise disjoint. Suppose not and let $(j,k) \in \mathbb{N}_0 \times \mathbb{N}_0$ be the least pair, with respect to the lexicographic ordering (see Exercise 1.5.3), for which $B_j \cap B_k \neq \emptyset$. Since clearly $B_0 \cap B_j = \emptyset$ for $j \in \mathbb{N}$, we can assume that $j = \tilde{j} + 1$ and $k = \tilde{k} + 1$ for $\tilde{j}, \tilde{k} \in \mathbb{N}_0$, and so therefore that $B_j = f(B_{\tilde{j}})$ and $B_k = f(B_{\tilde{k}})$. Thus $f(B_{\tilde{j}} \cap B_{\tilde{k}}) \neq \emptyset$ by Proposition 1.3.5, and so $B_{\tilde{j}} \cap B_{\tilde{k}} \neq \emptyset$. Since $(\tilde{j}, \tilde{k})$ is less that (j,k) with respect to the lexicographic order, we have a contradiction.

Now let $B = \cup_{j\in\mathbb{N}_0} B_j$ and define $g\colon S \to A$ by

$$g(x) = \begin{cases} f(x), & x \in B, \\ x, & x \notin B. \end{cases}$$

For $x \in B$, $g(x) = f(x) \in A$. For $x \notin B$, we have $x \in A$ by definition of B_0, so that g indeed takes values in A. By definition g is injective. Also, let $x \in A$. If $x \notin B$ then $g(x) = x$. If $x \in B$ then $x \in B_{j+1}$ for some $j \in \mathbb{N}_0$. Since $B_{j+1} = f(B_j)$, $x \in \mathrm{image}(g)$, so showing that g is surjective. ▼

We now continue with the proof of this part of the theorem. Note that $g \circ f\colon S \to g(T)$ is injective (cf. Exercise 1.3.3). Therefore, by the preceding lemma, there exists a bijection $h\colon S \to g(T)$. Since g is injective, $g\colon T \to g(T)$ is bijective, and let us denote the inverse by, abusing notation, $g^{-1}\colon g(T) \to T$. We then define $b\colon S \to T$ by $b = g^{-1} \circ h$, and leave it to the reader to perform the easy verification that b is a bijection.

(iv) $\Longrightarrow$ (iii) Since f is surjective, by Proposition 1.3.9 there exists a right inverse $f_R\colon T \to S$. Thus $f \circ f_R = \mathrm{id}_T$. Thus f is a *left*-inverse for f_R, implying that f_R is injective, again by Proposition 1.3.9. In like manner, g being surjective implies that there is an injective map from S to T, namely a right-inverse for g. ■

Distinguished names are given to certain kinds of sets, based on their cardinality. Recall that ω is the cardinal number corresponding to the set of natural numbers.

1.7.13 Definition (Finite, countable, uncountable) A set S is:

(i) ***finite*** if $\mathrm{card}(S) \in \mathbb{N}_0$;

[11] Georg Ferdinand Ludwig Philipp Cantor (1845–1918) was a Russian mathematician who made many contributions to the foundations of mathematics, and is regarded as the founder of set theory as we now know it. Friedrich Wilhelm Karl Ernst Schröder (1814–1902) was a German mathematician whose work was in the area of mathematical logic. Felix Bernstein (1878–1956) was born in Germany. Despite his name being attached to a basic result in set theory, Bernstein's main contributions were in the areas of statistics, mathematical biology, and actuarial mathematics.

(ii) ***infinite*** if $\text{card}(S) \geq \omega$;

(iii) ***countable*** if $\text{card}(S) \in \mathbb{N}_0$ or if $\text{card}(S) = \omega$;

(iv) ***countably infinite*** if $\text{card}(S) = \omega$;

(v) ***uncountable***, or ***uncountably infinite***, if $\text{card}(S) > \omega$. •

Let us give some examples illustrating the distinctions between the various notions of set size.

1.7.14 Examples (Cardinality)

1. All elements of $\mathbb{N}_0$ are, of course, finite sets.
2. The set $\mathbb{N}_0$ is countably infinite. Indeed, $\text{card}(\mathbb{N}_0) = \omega$.
3. We claim that $\mathbf{2}^{\mathbb{N}_0}$ is uncountable. More generally, we claim that, for any set S, $\text{card}(S) < \text{card}(\mathbf{2}^S)$. To see this, we shall show that any map $f\colon S \to \mathbf{2}^S$ is not surjective. For such a map, let
$$A_f = \{\, x \in S \mid \ x \notin f(x)\}\,.$$
We claim that $A_f \notin \text{image}(f)$. Indeed, suppose that $A_f = f(x)$. If $x \in A_f$ then $x \notin f(x) = A_f$ by definition of A_f; a contradiction. On the other hand, if $x \notin A_f$, then $x \in f(x) = A_f$; again a contradiction. We thus conclude that $A_f \notin \text{image}(f)$.
Thus there is no surjective map from S to $\mathbf{2}^S$. There is, however, a surjective map from $\mathbf{2}^S$ to S; for example, for any $x_0 \in S$, the map
$$g(A) = \begin{cases} x, & A = \{x\}, \\ x_0, & \text{otherwise} \end{cases}$$
is surjective. Thus S is "smaller than" $\mathbf{2}^S$, or $\text{card}(S) < \text{card}(\mathbf{2}^S)$. •

1.7.15 Remark (Uncountable sets exist, Continuum Hypothesis) A consequence of the last of the preceding examples is that fact that uncountable sets exist since $\mathbf{2}^{\mathbb{N}_0}$ has a cardinality strictly greater than that of $\mathbb{N}_0$.

It is usual to denote the countable ordinal by $\aleph_0$ (pronounced "aleph zero" or "aleph naught"). The smallest uncountable ordinal is then denoted by $\aleph_1$. An easy way to characterise $\aleph_1$ is as follows. Note that the cardinal $\aleph_0$ has the property that each of its initial segments is finite. In like manner, $\aleph_1$ has the property that each of its segments is countable. This does not *define* $\aleph_1$, but perhaps gives the reader some idea what it is.

It is conjectured that there are no cardinal numbers between $\aleph_0$ and $\aleph_1$; this conjecture is called the ***Continuum Hypothesis***. For readers prepared to accept the existence of the real numbers (or to look ahead to Section 2.1), we comment that $\text{card}(\mathbb{R}) = \text{card}(\mathbf{2}^{\mathbb{N}_0})$ (see Exercise 1.7.5). From this follows a slightly more concrete statement of the Continuum Hypothesis, namely the conjecture that $\text{card}(\mathbb{R}) = \aleph_1$. Said yet otherwise, the Continuum Hypothesis is the conjecture that, among the subsets of $\mathbb{R}$, the only possibilities are (1) countable sets and (2) sets having the same cardinality as $\mathbb{R}$. •

It is clear the finite union of finite sets is finite. The following result, however, is less clearly true.

1.7.16 Proposition (Countable unions of countable sets are countable) *Let $\{\mathrm{S_j}\}_{\mathrm{j}\in\mathbb{N}_0}$ be a family of sets, each of which is countable. Then $\cup_{\mathrm{j}\in\mathbb{N}_0}\mathrm{S_j}$ is countable.*

Proof Let us explicitly enumerate the elements in the sets S_j, $j \in \mathbb{N}_0$. Thus we write $S_j = \{x_{jk}\}_{k\in\mathbb{N}_0}$. We now indicate how one constructs a surjective map f from $\mathbb{N}_0$ to $\cup_{j\in\mathbb{N}_0} S_j$:

$$f(0) = x_{00},\ f(1) = x_{01},\ f(2) = x_{10},\ f(3) = x_{02},\ f(4) = x_{11},\ f(5) = x_{20},$$
$$f(6) = x_{03},\ f(7) = x_{12},\ f(8) = x_{21},\ f(9) = x_{30},\ f(10) = x_{04}, \ldots.$$

We leave it to the reader to examine this definition and convince themselves that, if it were continued indefinitely, it would include every element of the set $\cup_{j\in\mathbb{N}} S_j$ in the domain of f. ■

For cardinal numbers one can define arithmetic in a manner similar to, but not the same as, that for ordinal numbers. Given cardinal numbers c_1 and c_2 we let S_1 and S_2 be sets equivalent to (not necessarily similar to, note) c_1 and c_2, respectively. We then define $c_1 + c_2 = \text{card}(S_1 \mathring{\cup} S_2)$ and $c_1 \cdot c_2 = \text{card}(S_1 \times S_2)$. Note that cardinal number arithmetic is not just ordinal number arithmetic restricted to the cardinal numbers. That is to say, for example, the sum of two cardinal numbers is *not* the ordinal sum of the cardinal numbers thought of as ordinal numbers. It is easy to see this with an example. If S and T are two countably infinite sets, then so too is $S \mathring{\cup} T$ a countably infinite set (this is Proposition 1.7.16). Therefore, $\text{card}(S) + \text{card}(T) = \text{card}(S \mathring{\cup} T) = \omega = \text{card}(S) = \text{card}(T)$.

The only result that we shall care about concerning cardinal arithmetic is the following.

1.7.17 Theorem (Sums and products of infinite cardinal number) *If* c *is an infinite cardinal number then*

(i) $\mathrm{c} + \mathrm{k} = \mathrm{c}$ *for every finite cardinal number* k*,*

(ii) $\mathrm{c} = \mathrm{c} + \mathrm{c}$*, and*

(iii) $\mathrm{c} = \mathrm{c} \cdot \mathrm{c}$.

Proof (i) Let S and T be disjoint sets such that $\text{card}(S) = c$ and $\text{card}(T) = k$. Let $g\colon T \to \{1, \ldots, k\}$ be a bijection. Since S is infinite, we may suppose that S contains $\mathbb{N}$ as a subset. Define $f\colon S \cup T \to S$ by

$$f(x) = \begin{cases} g(x), & x \in T, \\ x + k, x \in \mathbb{N} \subset S, & \\ x, & x \in S \setminus \mathbb{N}. \end{cases}$$

This is readily seen to be a bijection, and so gives the result by definition of cardinal addition.

(ii) Let S be a set such that $\text{card}(S) = c$ and define

$$G(S) = \{\, (f, A) \mid\ A \subset S,\ f\colon A \times \{0,1\} \to A \text{ is a bijection}\}\,.$$

If $A \subset S$ is countably infinite, then $\text{card}(A \times \{0,1\}) = \text{card}(A)$, and so $G(S)$ is not empty. Place a partial order $\preceq$ on $G(S)$ by $(f_1, A_1) \preceq (f_2, A_2)$ if $A_1 \subset A_2$ and if $f_2|A_1 = f_1$. This is readily verified to be a partial order. Moreover, if $\{(f_j, A_j)\}_{j\in J}$ is a totally ordered subset, then we define an upper bound (f, A) as follows. We take $A = \cup_{j\in J} A_j$ and $f(x,k) = f_j(x,k)$ where $j \in J$ is defined such that $x \in A_j$. One can now use Zorn's Lemma to assert the existence of a maximal element of $G(S)$ which we denote by (f, A). We claim that $S \setminus A$ is finite. Indeed, if $S \setminus A$ is infinite, then there exists a countably infinite subset B of $S \setminus A$. Let g be a bijection from $B \times \{0,1\}$ to B and note that the map $f \times g\colon (A \cup B) \times \{0,1\} \to A \cup B$ defined by

$$f \times g(x,k) = \begin{cases} f(x,k), & x \in A, \\ g(x,k), & x \in B \end{cases}$$

if then a bijection, thus contradicting the maximality of (f, A). Thus $S \setminus A$ is indeed finite. Finally, since $(f, A) \in G(S)$, we have $\text{card}(A) + \text{card}(A) = \text{card}(A)$. Also, $\text{card}(S) = \text{card}(A) + \text{card}(A \setminus S)$. Since $\text{card}(S \setminus A)$ is finite, by part (i) this part of the theorem follows.

(iii) Let S be a set such that $\text{card}(S) = c$ and define

$$F(S) = \{ (f, A) \mid \ A \subset S,\ f \colon A \times A \to Af \text{ is a bijection} \}.$$

If $A \subset S$ is countably infinite, then $\text{card}(A \times A) = \text{card}(A)$ and so there exists a bijection from $A \times A$ to A. Thus $F(S)$ is not empty. Place a partial order $\preceq$ on $F(S)$ by asking that $(f_1, A_1) \preceq (f_2, A_2)$ if $A_1 \subset A_2$ and $f_2|A_1 \times A_1 = f_1$; we leave to the reader the straightforward verification that this is a partial order. Moreover, if $\{(f_j, A_j)\}_{j \in J}$ is a totally ordered subset, it is easy to define an upper bound (f, A) for this set as follows. Take $A = \cup_{j \in J} A_j$ and define $f(x, y) = f_j(x, y)$ where $j \in J$ is defined such that $(x, y) \in A_j \times A_j$. Thus, by Zorn's Lemma, there exists a maximal element (f, A) of $F(S)$. By definition of $F(S)$ we have $\text{card}(A)\,\text{card}(A) = \text{card}(A)$. We now show that $\text{card}(A) = \text{card}(S)$.

Clearly $\text{card}(A) \leq \text{card}(S)$ since $A \subset S$. Thus suppose that $\text{card}(A) < \text{card}(S)$. We now use a lemma.

1 Lemma *If* c_1 *and* c_2 *are cardinal numbers at least one of which is infinite, and if* c_3 *is the larger of* c_1 *and* c_2*, then* $\mathrm{c}_1 + \mathrm{c}_2 = \mathrm{c}_3$.

Proof Let S_1 and S_2 be disjoint sets such that $\text{card}(S_1) = c_1$ and $\text{card}(S_2) = c_2$. Since $c_1 \leq c_3$ and $c_2 \leq c_3$ it follows that $c_1 + c_2 = c_3 + c_3$. Also, $\text{card}(c_3) \leq \text{card}(c_1) + \text{card}(c_2)$. The lemma now follows from part (ii). ▼

From the lemma we know that $\text{card}(S)$ is the larger of $\text{card}(A)$ and $\text{card}(S \setminus A)$, i.e., that $\text{card}(S) = \text{card}(S \setminus A)$. Therefore $\text{card}(A) < \text{card}(S \setminus A)$. Thus there exists a subset $B \subset (S \setminus A)$ such that $\text{card}(B) = \text{card}(A)$. Therefore,

$$\text{card}(A \times B) = \text{card}(B \times A) = \text{card}(B \times B) = \text{card}(A) = \text{card}(B).$$

Therefore,

$$\text{card}((A \times B) \cup (B \times A) \cup (B \times B)) = \text{card}(B)$$

by part (ii). Therefore, there exists a bijection g from $(A \times B) \cup (B \times A) \cup (B \times B)$ to B. Thus we can define a bijection $f \times g$ from

$$(A \cup B) \times (A \cup B) = (A \times A) \cup (A \times B) \cup (B \times A) \cup (B \times B)$$

to $A \cup B$ by

$$f \times g(x, y) = \begin{cases} f(x, y), & (x, y) \in A \times A, \\ g(x, y), & \text{otherwise.} \end{cases}$$

Since $A \subset (A \cup B)$ and since $f \times g|(A \times A) = f$, this contradicts the maximality of (f, A). Thus our assumption that $\text{card}(A) < \text{card}(S)$ is invalid. ■

The following corollary will be particularly useful.

1.7.18 Corollary (Sum and product of a countable cardinal and an infinite cardinal) *If* c *is an infinite cardinal number then*

(i) $\mathrm{c} \leq \mathrm{c} + \text{card}(\mathbb{N})$ *and*

(ii) $\mathrm{c} \leq \mathrm{c} \cdot \text{card}(\mathbb{N})$.

Proof This follows from Theorem 1.7.17 since $\text{card}(\mathbb{N})$ is the smallest infinite cardinal number, and so $\text{card}(\mathbb{N}) \leq c$. ■

Exercises

1.7.1 Show that every element of an ordinal number is an ordinal number.

1.7.2 Show that any finite union of finite sets is finite.

1.7.3 Show that the Cartesian product of a finite number of countable sets is countable.

1.7.4 For a set S, as per Definition 1.3.1, let 2^S denote the collection of maps from the set S to the set 2. Show that $\mathrm{card}(2^S) = \mathrm{card}(\mathbf{2}^S)$, so justifying the notation $\mathbf{2}^S$ as the collection of subsets of S.
Hint: *Given a subset* A $\subset$ S*, think of a natural way of assigning a map from* S *to* 2*.*

In the next exercise you will show that $\mathrm{card}(\mathbb{R}) = \mathrm{card}(\mathbf{2}^{\mathbb{N}})$. We refer to Section 2.1 for the definition of the real numbers. There the reader can also find the definition of the rational numbers, as these are also used in the next exercise.

1.7.5 Show that $\mathrm{card}(\mathbb{R}) = \mathrm{card}(\mathbf{2}^{\mathbb{N}})$ by answering the following questions.
Define $f_1 \colon \mathbb{R} \to \mathbf{2}^{\mathbb{Q}}$ by

$$f_1(x) = \{\, q \in \mathbb{Q} \mid \ q \leq x \,\}\,.$$

(a) Show that f_1 is injective to conclude that $\mathrm{card}(\mathbb{R}) \leq \mathrm{card}(\mathbf{2}^{\mathbb{Q}})$.
(b) Show that $\mathrm{card}(\mathbf{2}^{\mathbb{Q}}) = \mathrm{card}(\mathbf{2}^{\mathbb{N}})$, and conclude that $\mathrm{card}(\mathbb{R}) \leq \mathrm{card}(\mathbf{2}^{\mathbb{N}})$.

Let $\{0,2\}^{\mathbb{N}}$ be the set of maps from $\mathbb{N}$ to $\{0,2\}$, and regard $\{0,2\}^{\mathbb{N}}$ as a subset of $[0,1]$ by thinking of $\{0,2\}^{\mathbb{N}}$ as being a sequence representing a decimal expansion in base 3. That is, to $f \colon \mathbb{N} \to \{0,2\}$ assign the real number

$$f_2(f) = \sum_{j=1}^{\infty} \frac{f(j)}{3^j}.$$

Thus f_2 is a map from $\{0,2\}^{\mathbb{N}}$ to $[0,1]$.

(c) Show that f_2 is injective so that $\mathrm{card}(\{0,2\}^{\mathbb{N}}) \leq \mathrm{card}([0,1])$.
(d) Show that $\mathrm{card}([0,1]) \leq \mathrm{card}(\mathbb{R})$.
(e) Show that $\mathrm{card}(\{0,2\}^{\mathbb{N}}) = \mathrm{card}(\mathbf{2}^{\mathbb{N}})$, and conclude that $\mathrm{card}(\mathbf{2}^{\mathbb{N}}) \leq \mathrm{card}(\mathbb{R})$.
Hint: *Use Exercise 1.7.4.*

This shows that $\mathrm{card}(\mathbb{R}) = \mathrm{card}(\mathbf{2}^{\mathbb{N}})$, as desired.

Section 1.8

Some words on axiomatic set theory

The account of set theory in this chapter is, as we said at the beginning of Section 1.1, called "naïve set theory." It turns out that the lack of care in saying what a set *is* in naïve set theory causes some problems. We indicate the nature of these problems in Section 1.8.1. To get around these problems, the presently accepted technique is the define a set as an element of a collection of objects satisfying certain axioms. This is called ***axiomatic set theory***, and we refer the reader to [Suppes 1960] for a detailed discussion. The most commonly used such axioms are those of Zermelo–Fränkel set theory, and we give these in Section 1.8.2. There are alternative collections of axioms, some equivalent to the Zermelo–Fränkel axioms, and some not. We shall not discuss this here. An axiom commonly, although not incontroversially, accepted is the Axiom of Choice, which we discuss in Section 1.8.3. We also discuss the Peano Axioms in Section 1.8.4, as these are the axioms of arithmetic. We close with a discussion of some of the issues in set theory, since these are of at least cultural interest.

Do I need to read this section? The material in this section is used exactly nowhere else in the texts. However, we hope the reader will find the informal presentation, and historical slant, interesting. •

1.8.1 Russell's Paradox

Russell's Paradox[12] is the following. Let S be the set of all sets that are not members of themselves. For example, the set P of prime numbers is not in S since the set of prime numbers is not a prime number. However, the set N of all things that are not prime numbers is in S since the set of all things that are not prime numbers is not a prime number. Now argue as follows. If $T \in S$ then $T \notin S$ by definition of S. On the other hand, if $T \notin S$ then $T \in S$, again by definition of S. This is clearly absurd, so the set S cannot exist, although there seems to be nothing wrong with its definition. That a contradiction can be derived from the naïve version of set theory means that it is ***inconsistent***.

A consequence of Russell's Paradox is that there is no set containing all sets. Indeed, let S be any set. Then define

$$T = \{ x \in S \mid \ x \notin x \} .$$

We claim that $T \notin S$. Indeed, suppose that $T \in S$. Then either $T \in T$ or $T \notin T$. In the first instance, since $T \in S$, $T \notin T$. In the second instance, again since $T \in S$, we have $T \notin T$. This is clearly a contradiction, and so we have concluded that, for every set S, there exists something that is not in A. Thus there can be no set of subsets.

Another consequence of Russell's Paradox is the ridiculous conclusion that everything is true. This is a simply logical consequence of the fact that, if a contradiction holds, then all statements hold. Here a contradiction means that a proposition P and its negation $\neg P$ both hold. The argument is as follows. Consider a proposition P'. Then P or P' holds, since P holds. However, since $\neg P$ holds and either P or P' holds, it must be the case that P' holds, no matter what P' is!

[12]So named for Bertrand Arthur William Russell (1872–1970), who was a British philosopher and mathematician. Russell received a Nobel prize for literature in recognition of his popular writings on philosophy.

Thus the contradiction arising from Russell's Paradox is unsettling since it now calls into question any conclusions that might arise from our discussion of set theory. Various attempts were made to eliminate the eliminate the inconsistency in the naïve version of set theory. The presently most widely accepted of these attempts is the collection of axioms forming Zermelo–Fränkel set theory.

1.8.2 The axioms of Zermelo–Fränkel set theory

The axioms we give here are the culmination of the work of Ernst Friedrich Ferdinand Zermelo (1871–1953) and Adolf Abraham Halevi Fränkel (1891–1965).[13] The axioms were constructed in an attempt to arrive at a basis for set theory that was free of inconsistencies. At present, it is unknown whether the axioms of Zermelo–Fränkel set theory, abbreviated ***ZF***, are consistent.

Here we shall state the axioms, give a slight discussion of them, and indicate some of the places in the chapter where the axioms were employed.

The first axiom merely says that two sets are equal if they have the same elements. This is not controversial, and we have used this axiom out of hand throughout the chapter.

Axiom of Extension For sets S and T, if $x \in S$ if and only if $x \in T$, then $A = B$. •

The next axiom indicates that one can form the set of elements for which a certain property holds. Again, this is not controversial, and is an axiom we have used throughout the chapter.

Axiom of Separation For a set S and a property P defined in S, there exists a set A such that $x \in A$ if and only if $x \in S$ and $P(x) = \text{true}$. •

We also have an axiom which says that one can extract two members from two sets, and think of these as members of another set. This is another uncontroversial axiom that we have used without much fuss.

Axiom of the Unordered Pair For sets S_1 and S_2 and for $x_1 \in S_1$ and $x_2 \in S_2$, there exists a set T such that $x \in T$ if and only if $x = x_1$ or $x = x_2$. •

To form the union of two sets, one needs an axiom asserting that the union exists. This is natural, and we have used it whenever we use the notion of union, i.e., frequently.

Axiom of Union For sets S_1 and S_2 there exists a set T such that $x \in T$ if and only if $x \in S_1$ or $x \in S_2$. •

The existence of the power set is also included in the axioms. It is natural and we have used it frequently.

Axiom of the Power Set For a set S there exists a set T such that $A \in T$ if and only if $A \subset S$. •

When we constructed the set of natural numbers, we needed an axiom to ensure that this set existed (cf. Assumption 1.4.3). This axiom is the following.

Axiom of Infinity There exists a set S such that

(i) $\emptyset \in S$ and

[13] Fränkel was a German mathematician who worked primarily in the areas of set theory and mathematical logic.

(ii) for each $x \in S$, $x^+ \in S$. •

When we constructed a large number of ordinal numbers in Example 1.7.2, we repeatedly used an axiom, the essence of which was, "The same principle used to assert the existence of $\mathbb{N}_0$ can be applied to this more general setting." Let us now state this idea more formally.

Axiom of Substitution For a set S, if for all $x \in S$ there exists a unique y such that $P(x, y)$ holds, then there exists a set T and a map $f\colon S \to T$ such that $f(x) = y$ where $P(x, y) = \text{true}$. •

The idea is that, for each $x \in S$, the collection of objects y for which $P(x, y)$ holds forms a set. Let us illustrate how the Axiom of Substitution can be used to define the ordinal number $\omega 2$, as in Example 1.7.2. For $k \in \mathbb{N}_0$ we define

$$P(k, y) = \begin{cases} \text{true}, & y = \omega + k, \\ \text{false}, & \text{otherwise.} \end{cases}$$

The Axiom of Substitution then says that there is a set T and a map $f\colon \mathbb{N}_0 \to T$ such that $f(k) = \omega + k$. The ordinal number $\omega 2$ is then simply the image of the map f.

The final axiom in ZF is the one whose primary purpose is to eliminate inconsistencies such as those arising from Russell's Paradox.

Axiom of Regularity For each nonempty set S there exists $x \in S$ such that $x \cap S = \emptyset$. •

The Axiom of Regularity rules out sets like $S = \{S\}$ whose only members are themselves. It is no great loss having to live without such sets.

1.8.3 The Axiom of Choice

The Axiom of Choice has its origins in Zermelo's proof of his theorem that every set can be well ordered. In order to prove the theorem, he had to introduce a new axiom in addition to those accepted at the time to characterise sets. The new axiom is the following.

Axiom of Choice For each family $\{S_a\}_{a \in A}$ of nonempty sets, there exists a function, $f\colon A \to \cup_{a \in A} S_a$, called a ***choice function***, having the property that $f(a) \in S_a$. •

The combination of the axioms of ZF with the Axiom of Choice is sometimes called ***ZF with Choice***, or ***ZFC***. Work of Cohen [1963][14] shows that the Axiom of Choice is independent of the axioms of ZF. Thus, when one adopts ZFC, the Axiom of Choice is really something additional that one is adding to one's list of assumptions of set theory.

At first glance, the Axiom of Choice, at least in the form we give it, does not seem startling. It merely says that, from any collection of sets, it is possible to select an element from each set. A trivial rephrasing of the Axiom of Choice is that, for any family $\{S_a\}_{a \in A}$ of nonempty sets, the Cartesian product $\prod_{a \in A} S_a$ is nonempty.

What is less settling about the Axiom of Choice is that it can lead to some nonintuitive conclusions. For example, as mentioned above, Zermelo's Well Ordering Theorem follows from the Axiom of Choice. Indeed, the two are equivalent. Let us, in fact, list the equivalence of the Axiom of Choice with two other important results from the chapter, one of which is Zermelo's Well Ordering Theorem.

[14] Paul Joseph Cohen was born in the United States in 1934, and has made outstanding contributions to the foundations of mathematics and set theory.

1.8.1 Theorem (Equivalents of the Axiom of Choice) *If the axioms of ZF hold, then the following statements are equivalent:*

(i) the Axiom of Choice holds;

(ii) Zorn's Lemma holds;

(iii) Zermelo's Well Ordering Theorem holds.

Proof Let us suppose that the proofs we give of Theorems 1.5.13 and 1.5.16 are valid using the axioms of ZF. This is true, and can be verified, if tediously. One only needs to check that no constructions, other than those allowed by the axioms of ZF were used in the proofs. Assuming this, the implications (i) $\Longrightarrow$ (ii) and (ii) $\Longrightarrow$ (iii) hold, since these are what is used in the proofs of Theorems 1.5.13 and 1.5.16. It only remains to prove the implication (iii) $\Longrightarrow$ (i). However, this is straightforward. Let $\{S_a\}_{a\in A}$ be a family of sets. By Zermelo's Well Ordering Theorem, well order each of these sets, and then define a choice function by assigning to $a \in A$ the least member of S_a. ■

There are, in fact, many statements that are equivalent to the Axiom of Choice. For example, the fact that a surjective map possesses a right-inverse is equivalent to the Axiom of Choice. A discussion of such matters may be found in the book of Moore [1982]. In Exercise 1.8.1 we give a few of the more easily proved equivalents of the Axiom of Choice. At the time of its introduction, the equivalence of the Axiom of Choice with Zermelo's Well Ordering Theorem led many mathematicians to reject the validity of the Axiom of Choice. Zermelo, however, countered that many mathematicians implicitly used the Axiom of Choice without saying so. This then led to much activity in mathematics along the lines of deciding which results *required* the Axiom of Choice for their proof. Results can then be divided into three groups, in ascending order of "goodness," where the Axiom of Choice is deemed "bad":

1. results that are equivalent to the Axiom of Choice;
2. results that are not equivalent to the Axiom of Choice, but can be shown to require it for their proof;
3. results that are true, whether or not the Axiom of Choice holds.

Somewhat more startling is that, if one accepts the Axiom of Choice, then it is possible to derive results which seem absurd. Perhaps the most famous of these is the ***Banach–Tarski Paradox***,[15] which says, very roughly, that it is possible to divide a sphere into a finite number of pieces and then reassemble them, while maintaining their shape, into two spheres of equal volume. Said in this way, the result seems impossible. However, if one looks at the result carefully, the nature of the pieces into which the sphere is divided is, obviously, extremely complicated. In the language of Chapter II-1, they are nonmeasurable sets. Such sets correspond poorly with our intuition, and indeed require the Axiom of Choice to assert their existence.

On the flip side of this is the fact that there are statements that seem like they *must* be true, and that are equivalent to the Axiom of Choice. One such statement is the Trichotomy Law for the real numbers, which says that, given two real numbers x and y, either $x < y$, $y < x$, or $x = y$. If rejecting the Axiom of Choice means rejecting the Trichotomy Law for real numbers, then many mathematicians would have to rethink the way they do mathematics!

Indeed, there is a branch of mathematics that is dedicated to just this sort of rethinking, and this is called ***constructivism***; see [Bridges and Richman 1987], for example. The

[15] Stefan Banach (1892–1945) was a well-known Polish mathematician who made significant and foundational contributions to functional analysis. Alfred Tarski (1902–1983) was also Polish, and his main contributions were to set theory and mathematical logic.

genesis of this branch of mathematics is the dissatisfaction, often arising from applications of the Axiom of Choice, with nonconstructive proofs in mathematics (for example, our proof that a surjective map possesses a right-inverse).

In this book, we will unabashedly assume the validity of the Axiom of Choice. In doing so, we follow in the mainstream of contemporary mathematics.

1.8.4 Peano's axioms

Peano's axioms[16] were derived in order to establish a basis for arithmetic. They essentially give those properties of the set of "numbers" that allow the establishment of the usual laws for addition and multiplication of natural numbers. ***Peano's axioms*** are these:

1. $0 = \emptyset$ is a number;
2. if k is a number, the successor of k is a number;
3. there is no number for which 0 is a successor;
4. if $j^+ = k^+$ then $j = k$ for all numbers j and k;
5. if S is a set of numbers containing 0 and having the property that the successor of every element of S is in S, then S contains the set of numbers.

Peano's axioms, since they led to the integers, and so there to the rational and real numbers (as in Section 2.1), were once considered as the basic ingredient from which all the rest of mathematics stemmed. This idea, however, received a blow with the publication of a paper by Kurt Gödel [1931][17]. Gödel showed that in any logical system sufficiently general to include the Peano axioms, there exist statements whose truth cannot be validated within the axioms of the system. Thus, this showed that any system built on arithmetic could not possibly be self-contained.

1.8.5 Discussion of the status of set theory

In this section, we have painted a picture of set theory that suggests it is something of a morass of questionable assumptions and possibly unverifiable statements. There is some validity in this, in the sense that there are many fundamental questions unanswered. However, we shall not worry much about these matters as we proceed onto more concrete topics.

Exercises

1.8.1 Prove the following result.

Theorem *If the axioms of ZF hold, then the following statements are equivalent:*

(i) the Axiom of Choice holds;

(ii) for any family $\{S_a\}_{a\in A}$ of sets, the Cartesian product $\prod_{a\in A} S_a$ is nonempty;

(iii) every surjective map possesses a right inverse.

[16]Named after Giuseppe Peano (1858–1932), an Italian mathematician who did work with differential equations and set theory.

[17]Kurt Gödel (1906–1978) was born in a part of the Austro-Hungarian Empire that is now Czechoslovakia. He made outstanding contributions to the subject of mathematical logic.

Section 1.9

Some words about proving things

Rigour is an important part of the presentation in this series, and if you are so unfortunate as to be using these books as a text, then hopefully you will be asked to prove some things, for example, from the exercises. In this section we say a few (almost uselessly) general things about techniques for proving things. We also say some things about poor proof technique, much (but not all) of which is delivered with tongue in cheek. The fact of the matter is that the best way to become proficient at proving things is to (1) read a lot of (needless to say, good) proofs, and (2) most importantly, get lots of practice. What is certainly true is that it much easier to begin your theorem-proving career by proving simple things. In this respect, the proofs and exercises in this chapter are good ones. Similarly, many of the proofs and exercises in Chapters 4 and 5 provide a good basis for honing one's theorem-proving skills. By contrast, some of the results in Chapter 2 are a little more sophisticated, while still not difficult. As we progress through the preparatory material, we shall increasingly encounter material that is quite challenging, and so proofs that are quite elaborate. The neophyte should not be so ambitious as to tackle these early on in their mathematical development.

Do I need to read this section? Go ahead, read it. It will be fun. •

1.9.1 Legitimate proof techniques

The techniques here are the principle ones use in proving simple results. For very complicated results, many of which appear in this series, one is unlikely to get much help from this list.

1. *Proof by definition:* Show that the desired proposition follows directly from the given definitions and assumptions. Theorems that have already been proven to follow from the definitions and assumptions may also be used. Proofs of this sort are often abbreviated by "This is obvious." While this may well be true, it is better to replace this hopelessly vague assertion with something more meaningful like "This follows directly from the definition."
2. *Proof by contradiction:* Assume that the hypotheses of the desired proposition hold, but that the conclusions are false, and make no other assumption. Show that this leads to an impossible conclusion. This implies that the assumption must be false, meaning the desired proposition is true.
3. *Proof by induction:* In this method one wishes to prove a proposition for an enumerable number of cases, say $1, 2, \ldots, n, \ldots$. One first proves the proposition for case 1. Then one proves that, if the proposition is true for the nth case, it is true for the $(n+1)$st case.
4. *Proof by exhaustion:* One proves the desired proposition to be true for all cases. This method only applies when there is a *finite* number of cases.
5. *Proof by contrapositive:* To show that proposition A implies proposition B, one shows that proposition B *not* being true implies that proposition A is *not* true. It is common to see newcomers get proof by contrapositive and proof by contradiction confused.
6. *Proof by counterexample:* This sort of proof is typically useful in showing that some general assertion *does not* hold. That is to say, one wishes to show that a certain

conclusion does not follow from certain hypotheses. To show this, it suffices to come up with a single example for which the hypotheses hold, but the conclusion does not. Such an example is called a ***counterexample***.

1.9.2 Improper proof techniques

Many of these seem so simple that a first reaction is, "Who would be dumb enough to do something so obviously incorrect." However, it is easy, and sometimes tempting, to hide one of these incorrect arguments inside something complicated.

1. *Proof by reverse implication:* To prove that A implies B, shows that B implies A.
2. *Proof by half proof:* One is required to show that A and B are equivalent, but one only shows that A implies B. Note that the appearance of "if and only if" means that you have two implications to prove!
3. *Proof by example:* Show only a single case among many. Assume that only a single case is sufficient (when it is not) or suggest that the proof of this case contains most of the ideas of the general proof.
4. *Proof by picture:* A more convincing form of proof by example. Pictures can provide nice illustrations, but suffice in no part of a rigorous argument.
5. *Proof by special methods:* You are allowed to divide by zero, take wrong square roots, manipulate divergent series, etc.
6. *Proof by convergent irrelevancies:* Prove a lot of things related to the desired result.
7. *Proof by semantic shift:* Some standard but inconvenient definitions are changed for the statement of the result.
8. *Proof by limited definition:* Define (or implicitly assume) a set S, for which all of whose elements the desired result is true, then announce that in the future only members of the set S will be considered.
9. *Proof by circular cross-reference:* Delay the proof of a lemma until many theorems have been derived from it. Use one or more of these theorems in the proof of the lemma.
10. *Proof by appeal to intuition:* Cloud-shaped drawings frequently help here.
11. *Proof by elimination of counterexample:* Assume the hypothesis is true. Then show that a counterexample cannot exist. (This is really just a well-disguised proof by reverse implication.) A common variation, known as "begging the question" involves getting deep into the proof and then using a step that assumes the hypothesis.
12. *Proof by obfuscation:* A long plotless sequence of true and/or meaningless syntactically related statements.
13. *Proof by cumbersome notation:* Best done with access to at least four alphabets and special symbols. Can help make proofs by special methods look more convincing.
14. *Proof by cosmology:* The negation of a proposition is unimaginable or meaningless.
15. *Proof by reduction to the wrong problem:* To show that the result is true, compare (reduce/translate) the problem (in)to another problem. This is valid if the other problem is then solvable. The error lies in comparing to an unsolvable problem.

Exercises

1.9.1 Find the flaw in the following inductive "proof" of the fact that, in any class, if one selects a subset of students, they will have received the same grade.

Suppose that we have a class with students $S = \{S_1, \ldots, S_m\}$. We shall prove by induction on the size of the subset that any subset of students receive the same grade. For a subset $\{S_{j_1}\}$, the assertion is clearly true. Now suppose that the assertion holds for all subsets of S with k students with $k \in \{1, \ldots, l\}$, and suppose we have a subset $\{S_{j_1}, \ldots, S_{j_l}, S_{j_{l+1}}\}$ of $l+1$ students. By the induction hypothesis, the students from the set $\{S_{j_1}, \ldots, S_{j_l}\}$ all receive the same grade. Also by the induction hypothesis, the students from the set $\{S_2, \ldots, S_{j_l}, S_{j_{l+1}}\}$ all receive the same grade. In particular, the grade received by student $S_{j_{l+1}}$ is the same as the grade received by student S_{j_l}. But this is the same as the grade received by students $S_{j_1}, \ldots, S_{j_{l-1}}$, and so, by induction, we have proved that all students receive the same grade.

In the next exercise you will consider one of Zeno's paradoxes. Zeno[18] is best known for having developed a collection of paradoxes, some of which touch surprisingly deeply on mathematical ideas that were not perhaps fully appreciated until the 19th century. Many of his paradoxes have a flavour similar to the one we give here, which may be the most commonly encountered during dinnertime conversations.

1.9.2 Consider the classical problem of the Achilles chasing the tortoise. A tortoise starts off a race T seconds before Achilles. Achilles, of course, is faster than the tortoise, but we shall argue that, despite this, Achilles will actually never overtake the tortoise.

At time T when Achilles starts after the tortoise, the tortoise will be some distance d_1 ahead of Achilles. Achilles will reach this point after some time t_1. But, during the time it took Achilles to travel distance d_1, the tortoise will have moved along to some point d_2 ahead of d_1. Achilles will then take a time t_2 to travel the distance d_2. But by then the tortoise will have travelled another distance d_3. This clearly will continue, and when Achilles reaches the point where the tortoise was at some moment before, the tortoise will have moved inexorably ahead. Thus Achilles will never actually catch up to the tortoise.

What is the flaw in the argument?

[18] Zeno of Elea (∼490BC–∼425BC) was an Italian born philosopher of the Greek school.

Bibliography

Bridges, D. S. and Richman, F. [1987] *Varieties of Constructive Mathematics*, number 97 in London Mathematical Society Lecture Note Series, Cambridge University Press, New York/Port Chester/Melbourne/Sydney, ISBN 0-521-31802-5.

Cohen, P. J. [1963] *A minimal model for set theory*, American Mathematical Society. Bulletin. New Series, **69**, 537–540.

Gödel, K. [1931] *über formal unentscheidbare sätze der principia mathematica und verwandter systeme*, Monatshefte für Mathematik und Physik, **38**, 173–189.

Moore, G. H. [1982] *Zermelo's Axiom of Choice: Its Origins, Development, and Influence*, Springer-Verlag, New York–Heidelberg–Berlin, ISBN 0-387-90670-3.

Suppes, P. [1960] *Axiomatic Set Theory*, The University Series in Undergraduate Mathematics, Van Nostrand Reinhold Co., London, reprint: [Suppes 1972].

— [1972] *Axiomatic Set Theory*, Dover Publications, Inc., New York, reprint of 1960 edition.

Symbol Index