Lecture Notes

for

Elementary Set Theory

Part II

Textbook By
Richard P. Millspaugh
University of North Dakota

Lecture Notes Prepared By
Bryce A. Christopherson
University of North Dakota

Lecture Notes for Elementary Set Theory

Chapter 3

Set Theory

3.1 What is a Set?

3.1.1 Naive set theory

Early work with sets assumed only that any collection that could be clearly specified (there is a rule for determining whether or not something is in the collection) could be considered a set, and that two sets were the same if they contained the same elements. However, it turns out that there are in fact some restrictions. Perhaps the most important is that sets are not allowed to be elements of themselves.

3.1.2 Russell's Paradox

In 1902, philosopher and mathematician Bertrand Russell published his famous paradox. This paradox works only if we allow the possibility that a set might be an element of itself. To avoid this kind of problem, we will not allow any set to be an element of itself. We will talk more about Russell's version of this paradox later, but for now let's deal with a popularization that doesn't depend so much on the language of sets.

Example 3.1. The Barber's Paradox. *A certain small town has only one barber. He shaves only those men in town who do not shave themselves. Who shaves the barber?*

3.2

Elements and Subsets

3.2.1 Elements

We will think of a set as a collection of objects, called it's *elements*. These objects can be any objects we are interested in talking about. For a set to be well-defined, we must have a way to determine whether or not a given object is in the set. We consider two sets to be equal if they contain exactly the same elements. Note that an object is either an element of a set or not, the elements of a set do not occur in a particular order and the same object cannot be an element of the set more than once.

One way to indicate a set is to simply list all of the elements between set braces; i.e., between the symbols { and }. The set of positive integers less than 3 is $\{1, 2\}$. We will find it useful to have a special notation for the set which has no elements. To this end, we let the symbol $\varnothing$ denote the set $\{\}$ (i.e. $\varnothing = \{\}$).

Example 3.2. *Let*

$$A = \{1, 2, 4, 8\}, \quad B = \{8, 2, 1, 4\}, \quad C = \{1, 2, 1, 4, 8\}, \quad D = \{1, 2, 3, 4, 8\}.$$

Of these sets, $A = B = C$ *because they each contain the same elements.*

It doesn't matter that we listed the elements in a different order when we defined A *than when we defined* B*, or that we listed* 1 *as an element of* C *more than once.*

The set D *is different from the others since it contains the element* 3 *and the other sets do not.*

Example 3.3. *Let $X = \{1, 2, \{2\}\}$. Note that this set contains some elements that are numbers and some elements that are sets of numbers. We consider those different objects.*

For example 1 is an element of X, but $\{1\}$ is not an element of X. On the other hand, both 2 and $\{2\}$ are elements of X.

We indicate that the object x is an element of the set A by writing $x \in A$. We write $x \notin A$ if x is not an element of A.

We indicate that the object x is an element of the set A by writing $x \in A$. We write $x \notin A$ if x is not an element of A. Listing all of the elements of a set works well when the set contains only a few elements, but what about sets with many elements? In this case, we use *set builder* notation to denote the set. An important part of this notation is the use of the vertical bar | (some texts use a colon in place of the vertical bar), read "such that." So the set $A = \{x \mid x^2 - 2 = 0\}$ is read "the set of all elements x such that $x^2 - 2 = 0$." An object will be an element of this set if and only if it satisfies the equation $x^2 - 2 = 0$.

There are also a few sets that are useful enough to deserve their own special notation. In particular, $\mathbb{N}$ denotes the set of natural numbers, $\mathbb{Z}$ the set of integers, $\mathbb{Q}$ the set of rational numbers, $\mathbb{R}$ the set of real numbers, and $\mathbb{C}$ the set of complex numbers.

3.2.2 Subsets

Definition 3.4. Given two sets A and B we say that A is a *subset* of B, or that A is contained in B, if every element of A is also an element of B. In this case we write $A \subset B$ (sometimes $A \subseteq B$).

Since every natural number is also an integer and every integer is a real number, $\mathbb{N} \subset \mathbb{Z} \subset \mathbb{R}$.

Example 3.5. *Define the sets*

$$A = \{1, 2, 3\}, \quad B = \{1, 2, \{3\}, 4\}, \quad C = \{1, 2, 3, \{4\}\}.$$

Then A *is a subet of* C *because each element of* A *is also an element of* C.

Note that A *is not a subset of* B *because* 3 *is an element of* A, *but not an element of* B*–remember that* 3 *and* $\{3\}$ *are different objects.*

Given any set A, it is certainly true that every element of A is an element of A. In other words, every set is a subset of itself. Note also that the empty set is a subset of every set A since there are no elements of the empty set which are not in A.

Definition 3.6. A subset B of A is said to be a *proper subset* of A if $B \neq \emptyset$ and $B \neq A$.

Theorem 3.7. *If* $A \subset B$ *and* $B \subset C$, *then* $A \subset C$.

Proof. Suppose that $a \in A$. Since $A \subset B$ it follows that $a \in B$. Now $B \subset C$, so we may say that $a \in C$ as desired.

3.2.3 Universal sets

In many instances, all of the sets we may be interested in are subsets of some particular set U. In this case we say that U is a *universal set*, or that U is the *universe*. For example, all of the functions we study in single variable calculus have domains and ranges that are subsets of the universal set $\mathbb{R}$.

3.2.4 Equality of sets

We say that two sets are equal if they contain exactly the same elements. In other words, $A = B$ means that every element of A is an element of B and every element of B is an element of A. A good test for this is the following one:

Theorem 3.8. *Given any two sets* A *and* B*,* $A = B$ *if and only if* $A \subset B$ *and* $B \subset A$.

This theorem will become one of our most valuable tools for showing that two sets are equal.

3.3 Operations on Sets

3.3.1 Union and intersection

Definition 3.9. The *union* of two sets A and B is the set:

$$A \cup B = \{x \mid (x \in A) \vee (x \in B)\}$$

Definition 3.10. The *intersection* of A and B is the set:

$$A \cap B = \{x \mid (x \in A) \wedge (x \in B)\}$$

Example 3.11. *Define* $A = \{1, 2, 3, 4, 5\}$ *and* $B = \{2, 4, 6, 8, 10\}$. *Then*

$$A \cup B = \{1, 2, 3, 4, 5, 6, 8, 10\} \text{ and } A \cap B = \{2, 4\}.$$

The following theorem summarizes several useful algebraic properties of unions and intersections.

Theorem 3.12. *Let* A, B, *and* C *be any sets in the universal set* U. *Then:*

(i) (Commutative Laws)
- *(a)* $A \cup B = B \cup A$
- *(b)* $A \cap B = B \cap A$

(ii) (Associative Laws)
- *(a)* $A \cup (B \cup C) = (A \cup B) \cup C$
- *(b)* $A \cap (B \cap C) = (A \cap B) \cap C$

(iii) (Distributive Laws)
- *(a)* $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
- *(b)* $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$

(iv) (Idempotence)
- *(a)* $A \cup A = A$
- *(b)* $A \cap A = A$

(v) (Identities)
- *(a)* $A \cup \varnothing = A$
- *(b)* $A \cap U = A$

Proof. We prove some of these properties here and leave the remainder for the exercises.

Commutative Laws. Note that $P \vee Q$ is equivalent to $Q \vee P$, so it follows immediately from our definition that $A \cup B = B \cup A$. Similarly we have $A \cap B = B \cap A$ since $P \wedge Q$ is equivalent to $Q \wedge P$.

Associative Laws. We prove the Associative Law for unions here. Rather than applying Theorem 3.8, we show that $(x \in A \cup (B \cup C)$ is equivalent to $x \in (A \cup B) \cup C$ using the associativity of disjunction at $(*)$

$$\begin{aligned}
x \in A \cup (B \cup C) &\Leftrightarrow (x \in A) \vee (x \in B \cup C) \\
&\Leftrightarrow (x \in A) \vee \big((x \in B) \vee (x \in C)\big) \\
&\Leftrightarrow \big((x \in A) \vee (x \in B)\big) \vee (x \in C) \qquad (*) \\
&\Leftrightarrow (x \in A \cup B) \vee (x \in C) \\
&\Leftrightarrow x \in (A \cup B) \cup C
\end{aligned}$$

Distributive Laws. We show that union distributes over intersection using the fact that disjunction distributes over conjunction at (**).

$$\begin{aligned} x \in A \cup (B \cap C) &\Leftrightarrow (x \in A) \vee (x \in B \cap C) \\ &\Leftrightarrow (x \in A) \vee \big((x \in B) \wedge (x \in C)\big) \\ &\Leftrightarrow \big((x \in A) \vee (x \in B)\big) \wedge \big((x \in A) \vee (x \in C)\big) \qquad (**) \\ &\Leftrightarrow (x \in A \cup B) \wedge (x \in A \cup C) \\ &\Leftrightarrow x \in (A \cup B) \cap (A \cup C) \end{aligned}$$

Idempotence. We prove that $A \cap A = A$. This time we make use of Theorem 3.8. Assume first that $x \in A \cap A$, then by definition $x \in A$ and $x \in A$. From this we may deduce that $x \in A$. Since this is true for every $x \in A \cap A$, we have shown that $A \cap A \subset A$. Now assume that $y \in A$, then we have $y \in A$ and $y \in A$. It follows that $y \in A \cap A$ for every $y \in A$, so $A \subset A \cap A$. We now apply Theorem 3.8 to see that $A \cap A = A$.

Identities. We prove here that $A \cup \varnothing = A$. Assume first that $x \in A \cup \varnothing$, then $x \in A$ or $x \in \varnothing$. By definition we know that $x \notin \varnothing$, so it follows that $x \in A$. We have shown that $A \cup \varnothing \subset A$. Now suppose that $y \in A$, then we may deduce that $y \in A$ or $y \in \varnothing$. By definition it follows that $y \in A \cup \varnothing$ and we have shown that $A \subset A \cup \varnothing$. We may now apply Theorem 3.8 to see that $A \cup \varnothing = A$ as desired.

3.3.2 Complements

Definition 3.13. The *relative complement* of B in A (also called the set difference) is the set:

$$A \setminus B = \{a \mid a \in A \text{ and } a \notin B\}$$

In a given universe U, we may also define the *complement* of a set A to be the set:

$$A' = U \setminus A$$

Theorem 3.14 (DeMorgan). *Let* A, B, *and* C *be sets. Then:*

(i) $A \setminus (B \cup C) = (A \setminus B) \cap (A \setminus C)$, *and*

(ii) $A \setminus (B \cap C) = (A \setminus B) \cup (A \setminus C)$.

Proof of Theorem 3.14(i). (i)

$$\begin{aligned}
x \in A \setminus (B \cup C) &\Leftrightarrow (x \in A) \wedge (x \notin B \cup C) \\
&\Leftrightarrow (x \in A) \wedge \neg(x \in B \vee x \in C) \\
&\Leftrightarrow (x \in A) \wedge \big(\neg(x \in B) \wedge \neg(x \in C)\big) \\
&\Leftrightarrow (x \in A) \wedge (x \notin B) \wedge (x \notin C) \\
&\Leftrightarrow \big((x \in A) \wedge (x \notin B)\big) \wedge \big((x \in A) \wedge (x \notin C)\big) \\
&\Leftrightarrow (x \in A \setminus B) \wedge (x \in A \setminus C) \\
&\Leftrightarrow x \in (A \setminus B) \cap (A \setminus C)
\end{aligned}$$

(ii)

$$\begin{aligned} x \in A \setminus (B \cap C) &\Leftrightarrow (x \in A) \land (x \notin B \cap C) \\ &\Leftrightarrow (x \in A) \land \neg(x \in B \land x \in C) \\ &\Leftrightarrow (x \in A) \land \big(\neg(x \in B) \lor \neg(x \in C)\big) \\ &\Leftrightarrow (x \in A) \land \big((x \notin B) \lor (x \notin C)\big) \\ &\Leftrightarrow \big((x \in A) \land (x \notin B)\big) \lor \big((x \in A) \land (x \notin C)\big) \\ &\Leftrightarrow (x \in A \setminus B) \lor (x \in A \setminus C) \\ &\Leftrightarrow x \in (A \setminus B) \cup (A \setminus C) \end{aligned}$$

3.3.3 Cartesian products

Definition 3.15. The *(Cartesian) product* of two sets A and B is the set:

$$A \times B = \{(a, b) \mid a \in A \text{ and } b \in B\},$$

where (a, b) denotes an ordered pair, not an interval.

Note that $A \times B$ is simply the set of all ordered pairs with first coordinates in A and second coordinates in B. For example, the Cartesian plane used in Calculus is the set $\mathbb{R} \times \mathbb{R}$.

Theorem 3.16. *For any sets* A, B, *and* C*:*

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

(ii) $(A \cap B) \times C = (A \times C) \cap (B \times C)$

(iii) $(A \setminus B) \times C = (A \times C) \setminus (B \times C)$

Proof. We leave parts (ii) and (iii) as exercises.

(i) First suppose that $(x, y) \in (A \cup B) \times C$. By definition we have $x \in A \cup B$ and $y \in C$. Since $x \in A \cup B$, either $x \in A$ or $x \in B$. If $x \in A$, then we have $x \in A$ and $y \in C$, so $(x, y) \in A \times C$. If $x \in B$, then we have $x \in B$ and $y \in C$, so $(x, y) \in B \times C$. We can now say that $(x, y) \in A \times C$ or $(x, y) \in B \times C$, so $(x, y) \in (A \times C) \cup (B \times C)$. Hence $(A \cup B) \times C \subset (A \times C) \cup (B \times C)$.

To see that the converse is true, suppose that $(x, y) \in (A \times C) \cup (B \times C)$. Either $(x, y) \in A \times C$ or $(x, y) \in B \times C$. If $(x, y) \in A \times C$, then $x \in A \subset (A \cup B)$ and $y \in C$, so $(x, y) \in (A \cup B) \times C$. If $(x, y) \in B \times C$, then $x \in B \subset (A \cup B)$ and $y \in C$, so $(x, y) \in (A \cup B) \times C$. Hence $(A \times C) \cup (B \times C) \subset (A \cup B) \times C$. Therefore, $(A \cup B) \times C = (A \times C) \cup (B \times C)$ as desired.

Since $A \times B$ is a set of ordered pairs, $A \times B \neq B \times A$. The proof of the following theorem involves making obvious changes to the previous proof and checking that all of the details still work.

Theorem 3.17. *For any sets* A, B, *and* C*:*

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

(ii) $A \times (B \cap C) = (A \times B) \cap (A \times C)$

(iii) $A \times (B \setminus C) = (A \times B) \setminus (A \times C)$

It may be tempting to assume that we could combine these two theorems somehow and obtain statements like $(A \times B) \setminus (C \times D) = (A \setminus C) \times (B \setminus D)$. This statement is generally false, however.

Example 3.18. *Let* $A = \{1, 2, 3\}$, $B = \{5, 6\}$, $C = \{1, 2\}$, *and* $D = \{6\}$. *Then:*

$$(A \times B) \setminus (C \times D) = \{(1, 5), (2, 5), (3, 5), (3, 6)\}$$

but

$$(A \setminus C) \times (B \setminus D) = \{(3, 5)\}$$

3.4 Collections of Sets

Some of the structures used in pure mathematics require the use of sets whose elements are other sets. It is customary, though not necessary, to refer to these kinds of sets as collections of sets.

3.4.1 The power set of a set

Definition 3.19. Let A be any set. The *power set* of A is the set $P(A) = \{B \mid B \subset A\}$.

In words, the power set of A is the set whose elements are the subsets of A. Note that for any set A we have $\varnothing \in P(A)$ and $A \in P(A)$, so $P(A)$ is nonempty for every set A. The power set of A is frequently denoted 2^A.

Theorem 3.20. *For any sets A and B, $A \subset B$ if and only if $P(A) \subset P(B)$.*

Proof. First assume that $A \subset B$ and let $X \in P(A)$. By definition $X \subset A$, so Theorem 3.7 implies that $X \subset B$. Hence $X \in P(B)$ and $P(A) \subset P(B)$. Conversely, if we assume that $P(A) \subset P(B)$, then $A \in P(A) \subset P(B)$ and $A \subset B$ by definition.

Theorem 3.21. *For any sets A and B, $P(A \cap B) = P(A) \cap P(B)$.*

Proof. First note that $A \cap B \subset A$ by an exercise in the homework. So, Theorem 3.20 implies that $P(A \cap B) \subset P(A)$. The same reasoning shows that $P(A \cap B) \subset P(B)$, so we have $P(A \cap B) \subset P(A) \cap P(B)$ in the homework.
Now suppose that $X \in P(A) \cap P(B)$, then $X \in P(A)$ and $X \in P(B)$. By definition $X \subset A$ and $X \subset B$, so $X \subset A \cap B$ by a problem from the homework. It follows that $P(A \cap B) \supset P(A) \cap P(B)$, so $P(A \cap B) = P(A) \cap P(B)$ as desired.

Chapter 4

Relations

Introduction

The first mathematical relation most students become aware of is the standard order on the natural numbers, which is later extended to the integers and finally the reals. A relation is a much more general concept, and by now you have probably worked with several others, perhaps without thinking about them as relations. In this chapter we define what we mean by a relation between sets, then define an equivalence relation on a set and use that idea to develop the set of rational numbers.

4.1

Relations

Definition 4.1. Given two sets A and B, a *relation* from A to B is a subset of $A \times B$. The relations we will be most interested in are usually from a set A to itself, in which case we say that the relation is a relation on A. If S is a relation, we frequently use the notation xSy to indicate that the ordered pair (x, y) is in S.

Let's look at some examples of relations.

Example 4.2. *The usual ordering $<$ on $\mathbb{R}$ is a relation on $\mathbb{R}$. We don't usually think of this relation as a set of ordered pairs, but we could. Define a subset L of $\mathbb{R} \times \mathbb{R}$ by $L = \{(a, b) \mid b - a \text{ is positive}\}$. In other words, an ordered pair (a, b) is in the relation if $a < b$.*

Example 4.3. *The set of points on a circle is another example of a relation on $\mathbb{R}$. For example, we could let $C = \{(x, y) \mid x^2 + y^2 = 1\}$.*

Example 4.4. *A relation need not be something you've encountered before or that you necessarily have a use for. Let A denote the set of people in this room, and B the set of possible hair colors. One might define a relation H from A to B by: $H = \{(x, y) \mid y \text{ is } x\text{'s hair color}\}$.*

Definition 4.5. Let A be a nonempty set and let $\sim$ be a relation on A. We say that $\sim$ is:

(i) *reflexive* if $a \sim a$ for every $a \in A$;

(ii) *symmetric* if $a \sim b$ implies $b \sim a$ for every $a, b \in A$;

(iii) *antisymmetric* if $a \sim b$ and $b \sim a$ imply $a = b$ for every $a, b \in A$.

(iv) *transitive* if $a \sim b$ and $b \sim c$ imply $a \sim c$ for every $a, b, c \in A$.

Note that symmetry and antisymmetry are not logical opposites, though the names may lead you to believe otherwise. It is possible for a relation to satisfy both of these properties, or to satisfy neither of them.

Example 4.6. *Let $<$ denote the relation "less than" on $\mathbb{R}$. Determine which of the properties listed above are satisfied by $<$.*

- *$<$ is not reflexive since $1 \not< 1$*
- *$<$ is not symmetric since $1 < 7$ and $7 \not< 1$*
- *$<$ is antisymmetric because the hypotheses $a < b$ and $b < a$ are never both satisfied*
- *$<$ is transitive since $a < b$ and $b < c$ implies that $a < c$*

4.2

Equivalence Relations

Consider the following question: Are $\pi/4$ and $25\pi/4$ measurements of the same angle? It's certainly true that when we place these angles in standard position they have the same terminal side, which means that we can treat them as the same much of the time. On the other hand, if we think of the angle as a rotation, these two angles are certainly different. If you're playing pin the tail on the donkey, being spun through an angle of $25\pi/4$ is going to make you dizzier than being spun through an angle of $\pi/4$. Intuitively, the terminal side of the angle in this setting tells us what direction we end up pointing in. The angle tells us how far we rotated to end up pointing in that direction. When we are only concerned with the terminal side of an angle, how do we tell when two numbers are measures for *equivalent* angles? You probably learned in a trigonometry or precalculus class that two measurements represent equivalent angles when they differ by an integer multiple of 2π. In other words, two angles α and β are said to be *equivalent* when there is an integer n such that $\alpha - \beta = 2\pi n$. In the following example we consider this relation between real numbers.

Example 4.7. *For $x, y \in \mathbb{R}$ we define $x \sim y$ if there is an integer n such that $x - y = 2\pi n$. Show that $\sim$ is reflexive, symmetric, and transitive.*

- *For any $x \in \mathbb{R}$ we have $x - x = 0 = 2\pi(0)$. Since $0 \in \mathbb{Z}$, $x \sim x$ and $\sim$ is relexive.*
- *Suppose that $x \sim y$, then there is an $n \in \mathbb{Z}$ so that $x - y = 2\pi n$. It follows that $y - x = 2\pi(-n)$. Since $-n \in \mathbb{Z}$ we have $y \sim x$ and $\sim$ is symmetric.*
- *Let $x, y, z \in \mathbb{R}$; assume that $x \sim y$ and $y \sim z$. By definition there are integers $n, m \in \mathbb{Z}$ such that $x - y = 2\pi n$ and $y - z = 2\pi m$. It follows that:*

$$x - z = (x - y) - (y - z) = 2\pi n + 2\pi m = 2\pi(n + m)$$

Now we have $x \sim z$, which shows that $\sim$ is transitive.

Definition 4.8. A relation on a set A that is reflexive, symmetric, and transitive is said to be an *equivalence relation* on A.

4.2.1 Equivalence classes

Definition 4.9. Let A be a set and let $\sim$ be an equivalence relation on A. For any $a \in A$ the set $[a] = \{b \in A \mid b \sim a\}$ is called the *equivalence class* of a.

Given a set A and an equivalence relation on A, the set of equivalence classes form a *partition* of the set X. In other words, the collection of equivalence classes is a collection of nonempty subsets of A with the proprties that every element of A is in some equivalence class and no two distinct equivalence classes intersect each other. We formalize this in the following theorem:

Theorem 4.10. *Let* A *be a nonempty set and* $\sim$ *an equivalence relation on* A*. For each* $a \in A$*, let* $[a] = \{b \in A \mid a \sim b\}$*.*

(i) For each $a \in A$*,* $a \in [a]$*.*

(ii) If $[a] \neq [b]$*, then* $[a] \cap [b] = \varnothing$*.*

Proof. We begin with part (i): Since $\sim$ is reflexive, $a \sim a$ for each $a \in A$. By definition this implies that $a \in [a]$ and (i) is satisfied.

Now, for part (ii): Suppose that $[a] \cap [b] \neq \varnothing$ and let $c \in [a] \cap [b]$. Then $c \sim a$ and $c \sim b$.

Since $c \sim a$ and $\sim$ is symmetric, it follows that $a \sim c$. Now we have $a \sim c$ and $c \sim b$, so $a \sim b$ by transitivity. Let $x \in [a]$, then $x \sim a$ by definition. From transitivity it follows that $x \sim b$, so $x \in [b]$. Hence $[a] \subset [b]$.

Since $a \sim b$, it must also be true that $b \sim a$ by symmetry. Let $y \in [b]$, then $y \sim b$ by definition. From transitivity it follows that $y \sim a$, so $y \in [a]$. Hence $[b] \subset [a]$.

Since $[a] \subset [b]$ and $[b] \subset [a]$, $[a] = [b]$ by Theorem 3.8.

Example 4.11. *Consider the equivalence relation $\equiv_5$ on $\mathbb{Z}$ defined by: $a \equiv_5 b$ if there is an integer n so that $b - a = 5n$. There are five equivalence classes associated with this equivalence relation:*

$$[0] = \{\dots, -10, -5, 0, 5, 10, \dots\}$$
$$[1] = \{\dots, -9, -4, 1, 6, 11, \dots\}$$
$$[2] = \{\dots, -8, -3, 2, 7, 12, \dots\}$$
$$[3] = \{\dots, -7, -2, 3, 8, 13, \dots\}$$
$$[4] = \{\dots, -6, -1, 4, 9, 14, \dots\}$$

The equivalence classes are sets of integers that have the same remainder when divided by 5.

In the previous example, note that we had several ways to refer to each equivalence class. For example $[1] = [16]$, so we may as well have used $[16]$ as a name for this equivalence class. In this context the numbers 1 and 16 (or any other member of the class) are called *representatives* of this equivalence class, and any representative can be used to name the equivalence class.

4.3

The rational numbers

In this section we are going to show how to use the ideas in this chapter to construct the set of rational numbers. We assume that the sets $\mathbb{N}$ of natural numbers and $\mathbb{Z}$ of integers are given and have their usual properties. Note that in this text the number 0 is not an element of the set of natural numbers. We will use the characterization of the rationals as the set of numbers that can be expressed as a fraction of integers, so we begin by defining the set $\mathbb{F} = \mathbb{Z} \times \mathbb{N}$. Note that $\mathbb{F}$ is the set of all ordered pairs (a, b) where $a \in \mathbb{Z}$ and $b \in \mathbb{N}$. We want to think of $\mathbb{F}$ as a set of fractions of integers, so we denote an ordered pair $(a, b) \in \mathbb{F}$ by a/b or $\dfrac{a}{b}$.

Note that a fraction is not quite the same thing as a rational number, for example $1/2$ and $3/6$ are different fractions that represent the same rational number. We define an equivalence relation $\cong$ on $\mathbb{F}$ that clarifies when two fractions represent the same rational as follows:

Definition 4.12. Given two fractions a/b and c/d in $\mathbb{F}$, we say that $a/b \cong c/d$ if $ad = bc$.

Theorem 4.13. *The relation $\cong$ is an equivalence relation on $\mathbb{F}$.*

Proof. To see that $\cong$ is reflexive, note that for every $a/b \in \mathbb{F}$ we have $ab = ba$, so $a/b \cong a/b$.

Now suppose that $a/b \cong c/d$, then by definition $ad = bc$. Using the fact that multiplication is commutative we see that $cb = da$, so $c/d \cong a/b$ and $\cong$ is symmetric.

It remains to show that $\cong$ is transitive, which is a little bit more work. Assume that $a/b \cong c/d$ and $c/d \cong e/f$, then by definition we have $ad = bc$ and $cf = de$. Making use of our hypotheses, we perform some algebra, making sure that we do not try to divide by 0:

$$
\begin{aligned}
ad &= bc \\
(ad)f &= (bc)f \\
(ad)f &= b(cf) \\
(ad)f &= b(de) \\
(af)d &= (be)d \\
af &= be
\end{aligned}
$$

Now we have $a/b \cong e/f$ and $\cong$ is transitive as desired.

We now define the set of rationals to be the set of equivalence classes of fractions. In this construction the two fractions $1/2$ and $3/6$ do indeed represent the same rational number since they are representatives of the same equivalence class. This construction might be a bit unsatisfying in the sense that it tells us nothing about how we might add or multiply rational numbers. Even if we know how to add fractions, how would we add two equivalence classes of fractions? One possibility is that we might add two equivalence classes by adding their representatives, so for example we might try:

$$\left[\frac{1}{2}\right] \oplus \left[\frac{2}{3}\right] = \left[\frac{1(3)+2(2)}{2(3)}\right] = \left[\frac{5}{6}\right]$$

There is a potential problem with this, though. There are infinitely many choices of fractions representing each rational number. Are we sure that we arrive at the same result for all of those possible choices? The next theorem says that we do.

Theorem 4.14. *Suppose that $a/b \cong x/y$ and $c/d \cong w/z$, where a/b, c/d, x/y, and w/z are all in $\mathbb{F}$, then:*

$$\frac{ad+bc}{bd} \cong \frac{xz+yw}{yz}.$$

Proof. By hypothesis we have $ay = bx$ and $cz = dw$. We want to show that $(ad+bc)(yz) = (bd)(xz+yw)$. Using our hypotheses we have:

$$\begin{aligned}(ad+bc)(yz) &= (ad)(yz)+(bc)(yz)\\ &= (ay)(dz)+(cz)(by)\\ &= (bx)(dz)+(dw)(by)\\ &= (bd)(xz)+(bd)(yw)\\ &= (bd)(xz+yw),\end{aligned}$$

which completes the proof.

The proof of the following theorem is left as an exercise.

Theorem 4.15. *Suppose that $a/b \cong x/y$ and $c/d \cong w/z$, where a/b, c/d, x/y, and w/z are all in $\mathbb{F}$, then:*

$$\frac{ac}{bd} \cong \frac{xw}{yz}.$$

Theorem 4.14 and Theorem 4.15 allow us to make the following definitions, where we may choose any representative from each equivalence class.

Definition 4.16. Let $[a/b]$ and $[c/d]$ be rational numbers, then we define addition and multiplication by:

$$\left[\frac{a}{b}\right] \oplus \left[\frac{c}{d}\right] = \left[\frac{ad+bc}{bd}\right] \quad \text{and} \quad \left[\frac{a}{b}\right] \otimes \left[\frac{c}{d}\right] = \left[\frac{ac}{bd}\right].$$

Chapter 5

Functions

Introduction

As anybody who has taken a calculus class knows, functions are important in mathematics. In this chapter we will define a function between two sets as a special type of relation between those sets. This definition grew from attempts to resolve various paradoxes that were discovered in the 19th century that challenged the intuitive characterizations of functions that were accepted at the time.

5.1 Definition

You probably think of functions in several ways based on what you have seen in previous courses. Functions can be rules for computing things, some people think of them as machines (plug in one number and get out another), and of course we all know that a graph is only the graph of a function if it passes the vertical line test. Most of these ideas match the way that mathematicians thought of functions (and the way they worked with them) at some time or another. It wasn't until the late 19th century that the concept of a function was defined in terms of sets. We present such a definition here:

Definition 5.1. Let A and B be sets. A *function* f from A to B is a relation from A to B with the additional property that for each $a \in A$ there is exactly one ordered pair (a, b) in f having a as a first coordinate. In this case, the set A is called the *domain* of f and the set B is called the *codomain* of f. If f is a function from A to B, we denote this by writing $f : A \to B$.

Do not confuse the codomain of a function with its range. The codomain of a function is the set that second coordinates must come from. The range is the subset of the codomain containing all of those second coordinates. Sometimes those sets are the same, but sometimes they are not.

Note that our definition requires that every element of the domain be the first coordinate of one, and only one, ordered pair in a function. It does not, however, require that each element of the codomain be a second coordinate, or that it be the second coordinate of only one ordered pair. Satisfying these requirements would make our function surjective or injective, respectively. We discuss these kinds of functions in Section 5.2.

Before going any further, let's consider a couple of examples.

Example 5.2. *Consider the sets* $A = \{1, 2, 3, 4\}$ *and* $B = \{1, 3, 5, 7\}$*. Define the following relations from* A *to* B*:*

(i) $f = \{(1, 3), (2, 1), (3, 5), (4, 7)\}$

(ii) $g = \{(1, 1), (3, 3)\}$

(iii) $h = \{(1, 1), (2, 3), (3, 5), (4, 1)\}$

(iv) $j = \{(1, 1), (2, 3), (3, 5), (4, 7), (2, 7)\}$

(v) $\mathrm{Pat} = \{(1, 1), (2, 3), (3, 5), (4, 7), (1, 1)\}$

The relation f *is certainly a function. Each element of* A *is a first coordinate of one and only one ordered pair.*

The relation g *is not a function because 2 and 4 are elements of* A*, but are not first coordinates of ordered pairs.*

The relation h *is a function. The fact that 1 is a second coordinate of two ordered pairs, or that 7 is not the second coordinate of any, do not violate our definition.*

The relation j *is not a function because 2 is the first coordinate of two distinct ordered pairs.*

Finally, Pat *is also a function. We make two comments here. First, the fact that the ordered pair* $(1, 1)$ *is listed twice does not violate our definition of a function. Each ordered pair is either in the relation or not, it cannot be two distinct elements of the relation. Second, while we will usually use letters like* f *or* g *to denote functions, and adhering to this convention makes life easier for us, there is no real requirement that we do so. We can name a relation, and thus also a function, in some other manner if there is a reason to do so.*

The functions in the previous example are obviously constructed to fit the definition, but may not strike you as the kinds of things you think of as functions. What about the kinds of functions we are used to?

Example 5.3. *Define the following relations from $\mathbb{R}$ to $\mathbb{R}$.*

(i) $f = \{(x, y) \mid x \in \mathbb{R} \text{ and } y = x^2\}$

(ii) $g = \{(x, y) \mid x \in \mathbb{R} \text{ and } x = y^2\}$

The relation f is a function (make sure you understand why). In fact, it is a function that should be familiar to you. The second coordinates are squares of the first coordinates, so this is the usual squaring function on $\mathbb{R}$.

The relation g is not a function from $\mathbb{R}$ to $\mathbb{R}$. Despite appearances, there are elements of $\mathbb{R}$ which are not first coordinates: -1 is one such element, but any negative number will do. This relation also violates our definition in another way. Both $(4, 2)$ and $(4, -2)$ satisfy the definition of g, so 4 is the first coordinate of more than one ordered pair in g. In fact, every positive real number is the first coordinate of two ordered pairs in g.

It probably seems unnatural to you to write the squaring function $f : \mathbb{R} \to \mathbb{R}$ as we did in Example 5.3. Wouldn't it be easier to write this function as $f(x) = x^2$, the same way we did in algebra or calculus classes? Once we know that f is a function, we can use this notation.

Suppose that $f : A \to B$ is a function. If $a \in A$ and $(a, b) \in f$, we say that b is the *value* of the function f at a and denote this by writing $b = f(a)$.

Now the notation $f(x) = x^2$ indicates that for each $x \in \mathbb{R}$, x^2 is the value of the function at x. Please be careful to distinguish between the name of the function f and an arbitrary value of the function $f(x)$.

Example 5.4. *Here are some examples of important types of functions. Assume each of the sets is nonempty.*

(i) *Let* A *be any set. The* identity function $i : A \to A$ *is defined by* $i(a) = a$ *for each* $a \in A$.

(ii) *Let* A *and* B *be sets and let* $b_0 \in B$. *The function* $k : A \to B$ *defined by* $k(a) = b_0$ *for all* $a \in A$ *is called a* constant function.

(iii) *Let* A *and* B *be any sets. The* coordinate projections $\pi_A : A \times B \to A$ *and* $\pi_B : A \times B \to B$ *are defined by* $\pi_A(a, b) = a$ *and* $\pi_B(a, b) = b$ *for each* $(a, b) \in A \times B$.

(iv) *Let* A *be any subset of* $\mathbb{R}$. *The* characteristic function $\chi_A : \mathbb{R} \to \{0, 1\}$ *of* A *is defined by*

$$\chi_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases}$$

Note: try and sketch a graph of the characteristic function of $\mathbb{Q}$.

5.1.1 Binary Operations

Standard addition and multiplication on $\mathbb{R}$ are both examples of functions used to compute a single real number from two given real numbers. This kind of function is important enough to deserve it's own name.

Definition 5.5. A *binary operation* on a set X is a function $* : X \times X \to X$. If $*$ is a binary operation on X, we usually use the notation $x * y$ to denote the value $*(x, y)$.

Example 5.6. *The standard addition and multiplication operations on $\mathbb{N}$, $\mathbb{Z}$, or $\mathbb{R}$ are all examples on binary operations. Note that they are all* different *binary operations.*

Example 5.7. *Subtraction is a binary operation on the sets $\mathbb{Z}$ or $\mathbb{R}$. Subtraction on $\mathbb{N}$ is not a binary operation on $\mathbb{N}$ since, for example, $7 - 12$ is not defined on $\mathbb{N}$.*

Definition 5.8. We say that a binary operation $*$ on a set X is:

- *commutative* if $x * y = y * x$ for all $x, y \in X$.
- *associative* if $(x * y) * z = x * (y * z)$ for all $x, y, z \in X$.

Addition and multiplication on $\mathbb{Z}$ (or on $\mathbb{N}$ or $\mathbb{R}$ for that matter) are commutative and associative, which you have probably known since your first algebra class. Subtraction on $\mathbb{Z}$ is not commutative since, for example, $4 - 2 \neq 2 - 4$. Subtraction on $\mathbb{Z}$ also fails to be associative since, for example, $(5 - 1) - 3 = 1 \neq 7 = 5 - (1 - 3)$.

Example 5.9. *Define the operation $*$ on $\mathbb{Z}$ by $a * b = (ab)^2$. We claim that $*$ is commutative.*

To see this, let a and b be arbitrary integers. Then:

$$\begin{aligned} a * b &= (ab)^2 \\ &= a^2 b^2 \\ &= b^2 a^2 \\ &= (ba)^2 \\ &= b * a \end{aligned}$$

We leave it as an exercise to determine whether or not $*$ is associative.

5.2 Injective, Surjective, and Bijective Functions

Definition 5.10. A function $f : X \to Y$ is said to be:

- *injective* or *one-to-one* if for all $x, y \in X$, if $f(x) = f(y)$, then $x = y$.
- *surjective* or *onto* if for each $y \in Y$, there is an $x \in X$ such that $f(x) = y$.
- *bijective* or *a one-to-one correspondence* if f is both injective and surjective.

Example 5.11. *Let's consider the functions we defined in Example 5.4.*

(i) *For any nonempty set* A*, the identity function* $i : A \to A$ *is bijective. Assume first that* $a, b \in A$ *with* $i(a) = i(b)$*. Since* $i(a) = a$ *and* $i(b) = b$*, this implies that* $a = b$ *and* i *is injective. To see that* i *is surjective, let* c *be any element of* A*. Then* $i(c) = c$*, so* i *is surjective.*

(ii) *In general, constant functions are neither injective nor surjective. Assume for the moment that* A *and* B *each have more than one element and let* $k : A \to B$ *be the constant function* $k(a) = b_0$*. Choose two elements* $a_1 \neq a_2$ *in* A*, then* $k(a_1) = b_0 = k(a_2)$ *and* k *is not injective. If* $b \neq b_0$ *and* $b \in B$*, then* $b \neq k(a)$ *for any* $a \in A$ *and* k *is not surjective.*

(iii) *Next we consider the coordinate projection* $\pi_A : A \times B \to A$*. Once again, we assume that* A *and* B *have more than one element each. The function* π_A *is surjective. To see this suppose that* $a \in A$ *and choose some* $b_0 \in B$*, then* $\pi_A(a, b_0) = a$*. Choose* $a \in A$ *and* $b_1 \neq b_2$ *both in* B*. Then* $(a, b_1) \neq (a, b_2)$ *but* $\pi_A(a, b_1) = a = \pi_A(a, b_2)$*, so* π_A *is not injective. The coordinate projection* $\pi_B : A \times B \to B$ *is also surjective but not injective. The proofs are similar.*

(iv) *Finally, consider the characteristic function* $\chi_A : \mathbb{R} \to \{0, 1\}$ *of some subset* A *of* $\mathbb{R}$*. Once again we assume that* A *has more than one element. For any two elements* $a \neq b$ *of* A *we have* $\chi_A(a) = 1 = \chi_A(b)$*, so* χ_A *is not injective. If* $a \in A$ *and* $c \in \mathbb{R} \setminus A$*, then* $\chi_A(a) = 1$ *and* $\chi_A(c) = 0$*. Since* 0 *and* 1 *are the only elements of the codomain,* χ_A *is surjective. Note however that the proof that* χ_A *is surjective requires that we be able to find points in* A *and points in* $\mathbb{R} \setminus A$*. If* $A = \mathbb{R}$*, then* χ_A *is not surjective because* $\mathbb{R} \setminus A = \varnothing$*.*

5.3

Compositions of Functions

Definition 5.12. Let $f : X \to Y$ and $g : Y \to Z$ be functions. The *composition* is the function from X to Z defined by $g \circ f(x) = g(f(x))$. In other words to find $g \circ f(x)$, first find $f(x)$, then plug the result into the function g. In terms of ordered pairs, the composition is

$$g \circ f = \{(x, z) \mid (x, y) \in f \text{ and } (y, z) \in g \text{ for some } y \in Y\}.$$

Theorem 5.13. *Let $f : X \to Y$ and $g : Y \to Z$ be functions.*

(i) If f and g are injective, then $g \circ f : X \to Z$ is injective.

(ii) If f and g are surjective, then $g \circ f : X \to Z$ is surjective.

Proof. For part (i), assume that $f : X \to Y$ and $g : Y \to Z$ are injective. Let x_1 and x_2 be distinct elements of X. Since f is injective, $f(x_1)$ and $f(x_2)$ are distinct elements of the set Y. Now since g is injective, $g(f(x_1))$ and $g(f(x_2))$ are distinct elements of Z. Therefore $g \circ f$ is injective as desired.

For part (ii), assume that $f : X \to Y$ and $g : Y \to Z$ are surjective. Let z be an arbitrary element of Z. Since g is surjective, there must be some element $y \in Y$ such that $g(y) = z$. Now since f is surjective there must be an element $x \in X$ with $f(x) = y$, so $g(f(x)) = g(y) = z$ and $g \circ f : X \to Z$ is surjective as desired.

Combining the two parts of Theorem 5.13, we have the following:

Corollary 5.14. *If $f : X \to Y$ and $g : Y \to Z$ are bijective functions, then $g \circ f : X \to Z$ is bijective.*

It is natural to ask whether or not the converses of the statements in Theorem 5.13 are true. We consider the converse of Theorem 5.13 *(i)* in the following example and theorem.

Example 5.15. *Define $f : \mathbb{N} \to \mathbb{R}$ by $f(n) = n^2$. Since no two natural numbers (which are all positive) have the same square, f is injective. Define $g : \mathbb{R} \to \mathbb{R}$ by $g(x) = x^2$. Since $g(-2) = 4 = g(2)$, g is not injective. Now we consider the composition $g \circ f : \mathbb{N} \to \mathbb{R}$ of these functions. For each $n \in \mathbb{N}$ we have $g(f(n)) = g(n^2) = (n^2)^2 = n^4$. Let $m, n \in \mathbb{N}$ with $m^4 = n^4$, then*

$$0 = m^4 - n^4 = (m^2 + n^2)(m - n)(m + n),$$

so $m = n$ or $m = -n$. But m and n are both positive, so $m \neq -n$ and it must be true that $m = n$. Hence $g \circ f : \mathbb{N} \to \mathbb{R}$ is injective.

This example shows that it is possible for $g \circ f$ to be injective when g is not. The next result says that if $g \circ f$ is injective, then it does follow that f is injective.

Theorem 5.16. *Let $f : X \to Y$ and $g : Y \to Z$ be functions such that $g \circ f : X \to Z$ is injective. Then $f : X \to Y$ must be injective.*

Proof. We prove that if $f : X \to Y$ is not injective, then $g \circ f : X \to Z$ is not injective, which is the contrapositive of the desired proposition. Suppose that f is not injective, then there must be elements $x_1 \neq x_2$ in X with $f(x_1) = f(x_2)$. Since g is a function, it must be true that $g(f(x_1)) = g(f(x_2))$. Since $x_1 \neq x_2$ but $g \circ f(x_1) = g \circ f(x_2)$, $g \circ f$ is not injective.

We leave the proof of the corresponding result for surjective functions as an exercise.

Theorem 5.17. *Let $f : X \to Y$ and $g : Y \to Z$ be functions such that $g \circ f : X \to Z$ is surjective. Then $g : Y \to Z$ must be surjective.*

5.3.1 Inverses of functions

Definition 5.18. If $f : X \to Y$ is a function, the *inverse* of f is the relation g from Y to X given by:

$$g = \{(y, x) \mid (x, y) \in f\}$$

By this definition every function will have an inverse relation. Note however that the inverse of a function is not generally a function, as we see in the following example.

Example 5.19. *Let $f : \mathbb{R} \to \mathbb{R}$ be the function defined by $f(x) = x^2$, then the inverse of f is the relation $g = \{(x^2, x) \mid x \in \mathbb{R}\}$. Note that g is not a function since both $(4, 2)$ and $(4, -2)$ are in g.*

Definition 5.20. If $f : X \to Y$ is a function and the inverse of f is also a function, we say that f is *invertible* and use $f^{-1} : Y \to X$ to denote the inverse.

Question 5.21. When is a function $f : X \to Y$ invertible?

Thinking back to what you've seen in previous courses for a moment, you are likely to have been taught that a function from $\mathbb{R}$ to $\mathbb{R}$ will have an inverse if it passes the "horizontal line test." Do you remember why? One way to think of this is that the graph of the inverse is the reflection through the line $y = x$ of the graph of the function. Since we want the reflection (the graph of the inverse) to pass the vertical line test, the graph of the original function should pass the horizontal line test. There should be some relationship between this condition and the answer to our question.

For the graph of a function to pass the horizontal line test, i.e. no horizontal line intersects the graph more than once, means that no two points of the graph have the same height. In other words, no two distinct points on the graph have the same y-coordinate. Rephrasing, if (x_1, y) and (x_2, y) are both on the graph, then $x_1 = x_2$. But this is just our definition of what it means for a function to be injective. Perhaps injective functions and invertible functions are the same thing? We can show that every invertible function is injective.

Theorem 5.22. *If $f : X \to Y$ is an invertible function, then f is injective.*

Proof. Suppose that f is not injective, then there are two elements $x_1 \neq x_2$ of X such that $f(x_1) = f(x_2)$. Let $y = f(x_1)$. By definition both (y, x_1) and (y, x_2) must be elements of the inverse of f. Since $x_1 \neq x_2$, this implies that the inverse of f is not a function. Therefore, f is not invertible.

Unfortunately, this is not a complete answer to our question because the converse of Theorem 5.22 is not true. We have found a condition that all invertible functions must satisfy, but not all functions that satisfy this condition are invertible. The following example illustrates the difficulty.

Example 5.23. *Define $f : \mathbb{N} \to \mathbb{N}$ by $f(n) = n + 1$. This function is injective (if $m + 1 = n + 1$, then $m = n$), but not invertible according to our definition. The inverse relation g contains all ordered pairs of the form $(n + 1, n)$ where $n \in \mathbb{N}$. For g to be a function from $\mathbb{N}$ to $\mathbb{N}$, every element of $\mathbb{N}$ must be the first coordinate of exactly one ordered pair in g. The problem here is that 1 is in $\mathbb{N}$, but $1 \neq n + 1$ for any $n \in \mathbb{N}$, so 1 is not the first coordinate of any of the ordered pairs in g. Therefore g is not a function from $\mathbb{N}$ to $\mathbb{N}$.*

For a function $f : X \to Y$ to be invertible, every element of the codomain Y must be a first coordinate of some ordered pair in the inverse. That in turn means that every element of the codomain must be the second coordinate of some ordered pair in f. This is exactly what it means to say that f is surjective. This leads us to believe the following:

Theorem 5.24. *If $f : X \to Y$ is invertible, then f is surjective.*

Proof. Assume that $f : X \to Y$ is not surjective, then there is some element $y \in Y$ so that $y \neq f(x)$ for any $x \in X$. In other words, y is not the second coordinate of an ordered pair in f. By definition then, y will not be the first coordinate of any ordered pair in the inverse of f. Hence the inverse of f is not a function from Y to X and f is not invertible.

We are now ready to give a complete answer to our question in the form of the following theorem.

Theorem 5.25. *A function $f : X \to Y$ is invertible if and only if it is bijective.*

Proof. If f is invertible, then we may use Theorem 5.22 and Theorem 5.24 to establish the fact that f is bijective.

To see that the converse is true, suppose that $f : X \to Y$ is a bijective function. Let $g = \{(y, x) \mid (x, y) \in f\}$ be the inverse of f. We must show that g is a function from Y to X, i.e. that every $y \in Y$ is the first coordinate of exactly one ordered pair in g. Let $y \in Y$. Since f is surjective, $y = f(x)$ for some $x \in X$. By definition $(x, y) \in f$ implies that $(y, x) \in g$, so y is the first coordinate of at least one ordered pair in g. Now suppose that (y, x_1) and (y, x_2) are both in g. By definition this means that $f(x_1) = y$ and $f(x_2) = y$. Since f is injective this implies that $x_1 = x_2$. It follows that y cannot be the first coordinate of more than one ordered pair in g, so g is a function from Y to X and f is invertible.