Chapter I
The Basics of Set Theory

1. Introduction

Every mathematician needs a working knowledge of set theory. The purpose of this chapter is to provide some of the basic information. Some additional set theory will be discussed in Chapter VIII.

Sets are a useful vocabulary in many areas of mathematics. They provide a language for stating interesting results. For example, in analysis: "a monotone function from $\mathbb{R}$ to $\mathbb{R}$ is continuous except, at most, on a countable set of points." In fact, set theory had its origins in analysis, with work done in the late 19th century by Georg Cantor (1845-1918) on Fourier series. This work played an important role in the development of topology, and all the basics of the subject are cast in the language of set theory. However sets are not just a tool; like many other mathematical ideas, "set theory" has grown into a fruitful research area of its own.

Moreover, on the philosophical side, most mathematicians accept set theory as a foundation for mathematics – that is, the notions of "set" and "membership in a set" can be taken as the most primitive notions of mathematics, in terms of which all (or nearly all) others can be defined. From this point of view, "everything in mathematics is a set." To put it another way, most mathematicians believe that "mathematics can be embedded in set theory."

So, you ask, what is a set? There are several different ways to try to answer. Intuitively – and this is good enough for most of our work in this course – a set is a collection of objects, called its *elements* or *members*. For example, we may speak of "the set of United States citizens" or "the set of all real numbers." The idea seems clear enough. However, we have not really given a satisfactory definition of a set: to say "a set is a collection of objects..." seems almost circular. After all, what is a "collection"?

In the early days of the subject, writers tried to give definitions of "set," just as Euclid attempted to give definitions for such things as "point" and "straight line" ("a line which lies evenly with the points on itself"). And, as in Euclid's case, these attempts did not really clarify things very much. For example, according to Cantor

> *Unter einer Menge verstehen wir jede Zusammenfassung M von bestimmten wohlunterschiedenen Objekten in unserer Anschauung oder unseres Denkens (welche die Elemente von M genannt werden) zu einem Ganzen* [By a set we are to understand any collection into a whole *M* of definite and separate objects (called the elements of M) of our perception or thought.] (*German seems to be a good language for this kind of talk.*)
>
> More compactly, Felix Hausdorff, around 1914, stated that a set is "a plurality thought of as a unit."

At this point, there are several ways we could proceed. One possibility is simply to take our intuitive, informal notion of a set and go on from there, ignoring any more subtle issues – just as we would not

worry about having definitions for "point" and "line" in beginning to study geometry. Another option might be to try to make a formal definition of "set" in terms of some other mathematical objects (assuming, implicitly, that these objects are "more fundamental" and intuitively understood). As a third approach, we could take the notions of "set" and "set membership" as primitive undefined terms and simply write down a collection of formal axioms that prescribe how "sets" behave.

The first approach is sometimes called naive set theory. ("Naive" refers only to the starting point – naive set theory gets quite complicated.) Historically, this is the way set theory began. The third option would take us into the subject of axiomatic set theory. Although an enormous amount of interesting and useful naive set theory exists, almost all research work in set theory nowadays requires the axiomatic approach (as well as some understanding of mathematical logic).

We are going to take the naive approach. For one thing, the axiomatic approach is not worth doing if it isn't done carefully, and that is a whole course in itself. In addition, axiomatic set theory isn't much fun unless one has learned enough naive set theory to appreciate why some sort of axiomatization would be important. It's more interesting to try to make things absolutely precise after you have a good overview. However as we go along, we will add some tangential comments about the axiomatic approach to help keep things in a more modern perspective.

2. Preliminaries and Notation

Informal Definition 2.1 A set is a collection of objects called its elements (or members). If A is a set and x is an element of A, we write $x \in A$. Otherwise, we write $x \notin A$.

As the informal definition implies, we may also use the word "collection" (or other similar words such as "family") in place of "set." Strictly speaking, these words will be viewed as synonymous. However, we sometimes interchange these words just for variety. Sometime we switch these words for emphasis: for example, we might refer to a set whose elements are also sets as a "collection of sets" or a "family of sets," rather than a "set of sets" – even though these all mean the same thing.

We can describe sets in at least two different ways:

By listing the elements, most useful when the set is a small finite set or an infinite set whose elements can be referred to using an ellipsis "..."

For example: $A = \{1, 2\}$
$\mathbb{N} = \{1, 2, 3, \ldots\}$, the set of natural numbers
$\mathbb{Z} = \{0, \pm 1, \pm 2, \ldots\}$, the set of integers

By abstraction, that is, by specifying a property that describes exactly what elements are in the set. We do this by writing something like $\{x : x$ has a certain property $\}$.

For example: $\mathbb{R} = \{x : x$ is a real number$\}$, the set of all real numbers
$\mathbb{Q} = \{ \frac{p}{q} : p \in \mathbb{Z}, q \in \mathbb{Z}$ and $q \neq 0\}$, the set of rational numbers
$\mathbb{P} = \{x : x \in \mathbb{R}$ and $x \notin \mathbb{Q}\}$, the set of irrational numbers

Suppose we write $\{x : x \in \mathbb{R}$ and $x^2 = -1\}$ or $\{x : x \in \mathbb{R}$ and $x \neq x\}$. No real number is actually a member of either set – both sets are empty. The empty set is usually denoted by the symbol $\emptyset$; it is

occasionally also denoted by { }. Sometimes the empty set is also called the "null set," although nowadays that term is more often used as a technical term for a certain kind of set in measure theory. (*By the way, $\emptyset$ is a Danish letter, not a Greek phi* $= \Phi$ *or* ϕ).

It may seem odd to talk about an empty set and even to give it a special symbol; but otherwise we would need to say to say that $\{x : x \in \mathbb{R} \text{ and } x^2 = -1\}$, which looks perfectly well-formed, is not sets at all. Even worse: consider $S = \{x : x \in \mathbb{Q} \text{ and } x = \alpha^\beta$, where α and β are irrational $\}$. Do you know whether or not there are any such rational numbers x? If you're not sure and if we did not allow an empty set, then you would not be able to decide whether or not S is a set! It's much more convenient simply to agree that $\{x : x \in \mathbb{Q} \text{ and } x = \alpha^\beta$, where α and β are irrational$\}$ is a set and allow the possibility that it might be empty.

Members of sets could be any kind of objects. In mathematics, however, we are not likely to be interested in a set whose members are aardvarks. We will only use sets that contain various mathematical objects. For example, a set of functions

$$C[a, b] = \{f : f \text{ is a continuous real-valued function with domain } [a, b]\}$$

or a set of sets such as

$$\{\{1\}, \{1, 2\}\} \text{ or } \qquad \{\emptyset\} \text{ or } \qquad \{\emptyset, \{\emptyset\}\}.$$

Of course, if "everything in mathematics is a set," then all sets in mathematics can only have other sets as members (because nothing else is available).

> Once we start thinking that everything is a set, then an interesting thought comes up. If x is a set, then either $x = \emptyset$ or there is an element $x_1 \in x$. Since x_1 is a set, either $x_1 = \emptyset$ or there is a set $x_2 \in x_1$, and so on. Is it possible to find a set for which there is an "infinite descending chain" of members
>
> $$... \in x_n \in x_{n-1} \in ... \in x_1 \in x \ ?$$

We say that two sets are equal, $A = B$, if A and B have precisely the same elements, that is, if $x \in A \Leftrightarrow x \in B$. For example, $\{x, x\} = \{x\}$ and $\{x, y\} = \{y, x\}$. Two sets whose descriptions look very different on the surface may turn out, on closer examination, to have exactly the same elements and therefore be equal. For example, $\{x : x \in \mathbb{R} \text{ and } x^2 = -1\} = \{x : x \in \mathbb{R} \text{ and } x \neq x\}$, and a scrupulous reader might verify that

$$\{x : x \in \mathbb{R} \text{ and } x^5 + 5x^4 - 29x^3 - 109x^2 - 8x + 140 = 0\} = \{-7, -2, 1, 5\}.$$

We say that *A* is a subset of *B* , written $A \subseteq B$, provided each element of *A* is also a member of B, that is, for all x, $x \in A \Rightarrow x \in B$. If $A \subseteq B$ but $A \neq B$ we say A is a proper subset of B. Clearly, $A = B$ if and only if $A \subseteq B$ and $B \subseteq A$ are both true. (*Notation: "if and only if" is often abbreviated to "iff".*)

Note that $A \subseteq B$ and $B \subseteq C$ implies that $A \subseteq C$.

You should look carefully at each of the following true statements to be sure the notation is clear:

$x \in A$ iff $\{x\} \subseteq A$

$\mathbb{N} \subseteq \mathbb{Q} \subseteq \mathbb{R} \subseteq \mathbb{C}$ (the set of complex numbers)

$\emptyset \neq \{\emptyset\}$ $\quad$ $\emptyset \subseteq \{\emptyset\}$ $\quad$ $\emptyset \in \{\emptyset\}$

$\emptyset \subseteq \emptyset$ $\quad$ $\emptyset \notin \emptyset$ $\quad$ $\emptyset \subseteq A$ for any set *A*

$\emptyset \in \{\emptyset\} \in \{\{\emptyset\}\}$, but $\emptyset \notin \{\{\emptyset\}\}$ (so $A \in B \in C$ doesn't imply $A \in C$)

Note that $\emptyset \neq \{\emptyset\}$. The set on the left is empty, while the set on the right has one member, namely the set $\emptyset$. This might be clearer with the alternate notation: $\{\,\} \neq \{\,\{\,\}\,\}$. The set on the left is analogous to an empty paper bag, while the set on the right is analogous to a bag with an empty bag inside.

We define the power set of a set A, written as $\mathcal{P}(A)$, to be the set of all subsets of *A*. In symbols, $\mathcal{P}(A) = \{B : B \subseteq A\}$.

Since $\emptyset$ is a subset of every set, we have $\emptyset \in \mathcal{P}(A)$ for every set A.

Since $A \subseteq A$ for any set A, we also have $A \in \mathcal{P}(A)$ for every set A.

$\mathcal{P}(\{1,2\}) = \{\emptyset\,, \{1\}, \{2\}, \{1,2\}\}$

$\mathcal{P}(\{1\}) = \{\emptyset\,, \{1\}\}$

$\mathcal{P}(\emptyset) = \{\emptyset\,\}$

The last three examples suggest that a set A with n elements has 2^n subsets (*to define a subset* $B \subseteq A$: *for each element* x *in* A, *there are two choices* $-$ *"yes" or "no"* $-$ *about whether* x *to put into* B. *For each* x, *the choice is independent of the other choices. So there are* 2^n *ways to pick the elements to form a subset* B.)

Exercises

E1. Use induction to prove that if a set X has n elements, then $\mathcal{P}(X)$ has 2^n elements.

E2. Use Exercise 1 to explain the meaning of the identity

$$\binom{n}{0} + \binom{n}{1} + \binom{n}{2} + \ldots + \binom{n}{n-1} + \binom{n}{n} = 2^n \qquad (n = 0, 1, 2 \ldots)$$

3. Paradoxes

The naive approach to sets seems to work fine until someone really starts trying to cause trouble. The first person to do this was Bertrand Russell who, around 1902, created Russell's Paradox:

> It makes sense to ask whether a set is one of its own members – that is, for a given set A, to ask whether $A \in A$ is true or false. The statement $A \in A$ is false for the sets you immediately think of: for example $\{1,2\} \notin \{1,2\}$. However, is the infinite set
>
> $$A = \{\emptyset,\ \{\emptyset\}, \{\emptyset,\ \{\emptyset\}\}, \{\emptyset,\ \{\emptyset\}, \{\emptyset,\ \{\emptyset\}\}\}, \{\emptyset,\ \{\emptyset\}, \{\emptyset,\ \{\emptyset\}\}, \{\emptyset,\ \{\emptyset\}, \{\emptyset,\ \{\emptyset\}\}\}\}, \ldots\}$$
>
> a member of itself? How about an even more complicated infinite set? Could it happen that $A \in A$? Whatever the answer, it makes sense to ask.
>
> According to our naive approach for defining sets, we can define a set $\mathfrak{A}$ by writing $\mathfrak{A} = \{A : A \notin A\}$, so that $\mathfrak{A}$ is the set of all sets which are not members of themselves. Then we can ask, for this new set $\mathfrak{A}$, whether $\mathfrak{A} \in \mathfrak{A}$ is true or false. If $\mathfrak{A} \in \mathfrak{A}$, then $\mathfrak{A}$ must satisfy the requirement for being a member of $\mathfrak{A}$, that is, $\mathfrak{A} \notin \mathfrak{A}$. On the other hand, if $\mathfrak{A} \notin \mathfrak{A}$, then $\mathfrak{A}$ does meet the membership requirement for $\mathfrak{A}$, so $\mathfrak{A} \in \mathfrak{A}$. Thus, each of the only two possibilities about the set $\mathfrak{A}$ (that $\mathfrak{A} \in \mathfrak{A}$ or $\mathfrak{A} \notin \mathfrak{A}$) leads to a contradiction!

Russell's Paradox illustrates that dilemmas can arise from using the method of abstraction to define sets too casually. A way out is to refuse to call $\mathfrak{A}$ a set. To do that, in practice, we will insist that whenever we define a set by abstraction, we only form subsets of sets that already exist. That is, in defining a set by abstraction, we should always write $\{x : x \in X$ and ... $\}$ or, for short,
$\{x \in X :$ ... $\}$ where X is some set that we already have. The result is a subset of *X*. Since the preceding definition of $\mathfrak{A}$ doesn't follow this form, $\mathfrak{A}$ is not guaranteed to be a set.

This is the route taken in axiomatic set theory, and it eliminates Russell's paradox. If X is any set and we consider the set $\mathfrak{A} = \{A \in X\colon A \notin A\}$, the dilemma vanishes. For now, $\mathfrak{A} \notin \mathfrak{A}$ means either that $\mathfrak{A} \in \mathfrak{A}$ (a contradiction) or that $\mathfrak{A} \notin X -$ an alternative conclusion that we can live with.

Russell's Paradox has the same flavor as many "self-referential" paradoxes in logic. For example, some books in the library mention themselves – in the preface the author might say, "In this book, I will discuss ... ". Other books make no mention of themselves. Suppose Library X wishes to make a book listing all books that do not mention themselves. Should the new book list itself? That is Russell's Paradox: if it doesn't mention itself, then it should; and if it does mention itself, then it shouldn't. The resolution of the paradox is that the new book really is intended to list "all books in X which do not mention themselves"— that is, in forming the new book, one is restricted to examining only those books already in the collection X. With this additional qualification, the paradox disappears. Think about the same paradox in the dedication at the front of these notes: are the notes dedicated to Freiwald?

In everyday mathematics, we usually don't have to worry about this kind of paradox. Almost always, when we form a new set, we have (at least in the back of our minds) a larger set X of which it is a subset. Therefore, indulging in a bit of sloppiness, we may sometimes write $\{x :$ ... $\}$ rather than the more correct $\{x \in X :$ ... $\}$ simply because the set X could be supplied on demand, and the notation is simpler.

There is another kind of difficulty we can get into when defining sets by abstraction. It arises from the nature of the description " ... " when we write $\{x \in X : \text{ ... }\}$. This is illustrated by Richard's Paradox:

> Consider $A = \{x \in \mathbb{N} : \ x$ is definable in English using less than 10000 characters$\}$. There are only finitely many English character strings with length < 10000 (a very large number, but finite). Most of these character strings are gibberish, but some of them define a positive integer. So the set A is finite. Therefore there must be natural numbers not in A, and we can pick the smallest such natural number : call it m.
>
> But if A is a well-defined set, then the preceding paragraph has precisely defined m, using fewer than 10000 characters (count them!), so m is in A !

To resolve this paradox carefully involves developing a little more formal machinery than we want to bother with here, but the idea is easy enough. The idea involves requiring that only a certain precise kind of description can be used for the property " ... " when defining a set $\{x \in X : \text{ ... }\}$.

Roughly, these are the descriptions we can form using existing sets, $\in$, the logical quantifiers $\forall$ and $\exists$, and the familiar logical connectives $=$, $\wedge$, $\vee$, $\Rightarrow$, $\Leftrightarrow$ and $\neg$. When all this is made precise (using first order predicate calculus), the "set" A above is not allowed as a set – because the description " ... " used above A is not a "legal" description. Fortunately, the sets we want to write down in mathematics can be described by "legal" expressions. For example, we could define the set of positive real numbers by

$$\mathbb{R}^+ = \{x \in \mathbb{R} : \neg(x = 0) \wedge \exists y\, (y \in \mathbb{R} \wedge x = y^2)\}$$

To summarize: there are dangers in a completely naive, casual formation of "sets." One of the reasons for doing axiomatic set theory is to avoid these dangers by giving a list of axioms that lay out our precise initial assumptions about sets and how they are formed. But fortunately, with some common sense and a little feeling acquired in practice, such dangerous situations rarely arise in the everyday practice of mathematics.

4. Elementary Operations on Sets

We want to have operations that we can use to combine old sets into new ones. The simplest operations are union and intersection.

Informally, the union of two sets is the set consisting of all elements in A or in B. (*Note: when a mathematician says "either p or q", this means "either p or q or both." This is called the "inclusive" use of the word "or."*)

The intersection of two sets is the set of all elements belonging to both A and B. In symbols,

the union of A and B is $\quad A \cup B = \{x : x \in A \text{ or } x \in B\}$, and
the intersection of A and B is $\quad A \cap B = \{x : x \in A \text{ and } x \in B\}$.

Note: You might expect, after the discussion of paradoxes, that the definition of union should read: $A \cup B = \{x \in X : \ x \in A \text{ or } x \in B\}$ and then ask "given A and B, what is X ?"

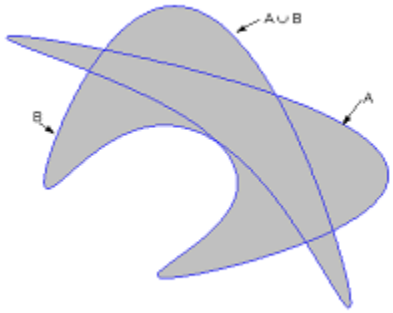
In practice, the sets A and B that we combine are always subsets of some larger set X. Then there is no need to worry because $A \cup B$, we understand, could be written more properly as $\{x \in X : x \in A \text{ or } x \in B\}$. But to cover all possible situations, axiomatic set theory adds a separate axiom (see A5 on p. 12) to guarantee that unions always exist.

This issue doesn't come up for intersections. For example, given sets A and B, we can always write $A \cap B = \{x \in A : \ x \in B\}$.

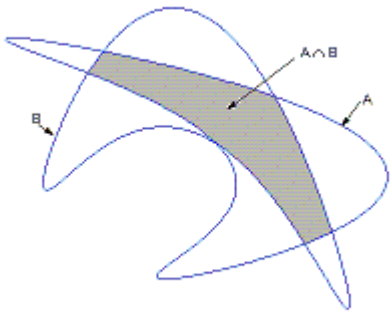
Examples 4.1 $\{1,2\} \cup \{2,3\} = \{1,2,3\}$ $\qquad$ $\{1,2\} \cap \{2,3\} = \{2\}$

$\mathbb{P} \cup \mathbb{Q} = \mathbb{R}$ $\qquad$ $\mathbb{P} \cap \mathbb{Q} = \emptyset$

The union and intersection of two sets can be pictured with "Venn diagrams" :



$A \cup B$



$A \cap B$

We want to be able to combine more than just two sets, perhaps even infinitely many. To express this, we use the idea of an index set.

Definition 4.2 Suppose that for each λ in some set Λ, a set A_λ is given. We then say that the collection $\mathfrak{A} = \{A_\lambda : \lambda \in \Lambda\}$ is indexed by Λ. We might write $\mathfrak{A}$ more informally as $\{A_\lambda\}_{\lambda \in \Lambda}$ or even merely as $\{A_\lambda\}$ if the index set Λ is clearly understood.

Examples 4.3 1) $\mathfrak{A} = \{A_1, A_2\}$ is indexed by the set $\Lambda = \{1, 2\}$

2) Let I_x denote the interval $[0, x]$ of real numbers. Then $\mathfrak{A} = \{I_x : x \in \mathbb{R},\ x \geq 0\}$ is indexed by $\Lambda =$ the set of nonnegative real numbers.

3) $\mathfrak{A} = \{A_1, A_2, ..., A_n, ...\}$ is indexed by $\Lambda = \mathbb{N}$

4) $\mathfrak{A} = \emptyset$ iff $\mathfrak{A}$ can be indexed by $\Lambda = \emptyset$

Definition 4.4 Suppose $\mathfrak{A} = \{A_\lambda : \lambda \in \Lambda\}$. The union of the family $\mathfrak{A}$ is $\bigcup\{A_\lambda : \lambda \in \Lambda\} = \{x :$ for some $\lambda_0 \in \Lambda$, $x \in A_{\lambda_0}\}$. The union is also written more simply, as just $\bigcup \mathfrak{A}$. (So $x \in \bigcup \mathfrak{A}$ iff x is an element of an element of $\mathfrak{A}$.)

The intersection of the family, denoted $\bigcap\{A_\lambda : \lambda \in \Lambda\}$ or, more simply $\bigcap \mathfrak{A}$, is $\{x : x \in A_\lambda$ for all $\lambda \in \Lambda\}$.

When the specific set Λ is understood or irrelevant, we may ignore the "$\lambda \in \Lambda$" and just write $\bigcup A_\lambda$ or $\bigcap A_\lambda$. If $\Lambda = \mathbb{N}$, we might also write $\bigcup \mathfrak{A}$ and $\bigcap \mathfrak{A}$ as $\bigcup_{n=1}^{\infty} A_n$ and $\bigcap_{n=1}^{\infty} A_n$.

Examples 4.5 1) Suppose $\mathfrak{A} = \{A_1, A_2\}$. We can write the union of this family of sets in several different ways: $\bigcup \mathfrak{A} = \bigcup_{i=1}^{2} A_i = A_1 \cup A_2 = \bigcup\{A_i : i = 1, 2\}$.

2) If $I_x = [0, x) \subseteq \mathbb{R}$, then $\bigcup\{I_x : x \geq 0\} = [0, \infty)$ and $\bigcap\{I_x : x \geq 0\} = \emptyset$. $\bigcap\{I_x : x > 0\} = \{0\}$.

3) If $A_n = \left[-\frac{1}{n}, 1 + \frac{1}{n}\right] \subseteq \mathbb{R}$, then $\bigcup\{A_n : n \in \mathbb{N}\} = [-1, 2]$ and $\bigcap\{A_n : n \in \mathbb{N}\} = [0, 1]$

4) If $B_n = [n, \infty) \subseteq \mathbb{R}$, then $\bigcup_{n=1}^{\infty} B_n = [1, \infty)$ and $\bigcap_{n=1}^{\infty} B_n = \emptyset$

5) Suppose each set $A_\lambda \subseteq X$. If $\Lambda = \emptyset$, then $\bigcup\{A_\lambda : \lambda \in \Lambda\} = \emptyset$ and $\bigcap\{A_\lambda : \lambda \in \Lambda\} = X$. *For the intersection: if $x \in X$, then x is in every A_λ (can you name an A_λ that doesn't contain x ?) so $x \in \bigcap\{A_\lambda : \lambda \in \Lambda\}$.*

If this argument bothers you, then you can consider the statement $\bigcap\{A_\lambda : \lambda \in \Lambda\} = X$ to be a just a convention motivated by the idea that "the fewer sets in the family, the larger the intersection should be, so that when $\Lambda = \emptyset$, the intersection should be as large as possible."

Theorem 4.6 1) $A \cup B = B \cup A$, and $A \cap B = B \cap A$.

2) $A \cup (B \cup C) = (A \cup B) \cup C$, and $A \cap (B \cap C) = (A \cap B) \cap C$. More generally,

$$\bigcup_{\lambda\in\Lambda} A_\lambda \cup \left(\bigcup_{\mu\in M} A_\mu \cup \bigcup_{\nu\in N} A_\nu\right) = \left(\bigcup_{\lambda\in\Lambda} A_\lambda \cup \bigcup_{\mu\in M} A_\mu\right) \cup \bigcup_{\nu\in N} A_\nu,$$

which we could also write in an alternate form

$$\bigcup_{\lambda\in\Lambda} A_\lambda \cup \left(\bigcup_{\mu\in M} A_\mu \cup \bigcup_{\nu\in N} A_\nu\right) = \bigcup_{\alpha\in\Lambda\cup M\cup N} A_\alpha$$

The same equations hold with " $\cap$ " replacing " $\cup$ " everywhere.

3) $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$, and $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$. More generally,

$$\left(\bigcup_{\lambda\in\Lambda} A_\lambda\right) \cap \left(\bigcup_{\mu\in M} B_\mu\right) = \bigcup_{\lambda\in\Lambda,\, \mu\in M} (A_\lambda \cap B_\mu) \quad \text{and}$$

$$\left(\bigcap_{\lambda\in\Lambda} A_\lambda\right) \cup \left(\bigcap_{\mu\in M} B_\mu\right) = \bigcap_{\lambda\in\Lambda,\, \mu\in M} (A_\lambda \cup B_\mu)$$

Proof To prove two sets are equal we show that they have the same elements; the most basic way to do this is to show that if x is in the set on the left hand side (LHS) of the proposed equation, then x must also be in the set on the right hand side (RHS) (*thereby proving LHS* $\subseteq$ *RHS*) and vice-versa. All parts of the theorem are easy to prove; we illustrate by proving the last equality:

$$(\bigcap_{\lambda\in\Lambda} A_\lambda) \cup (\bigcap_{\mu\in M} B_\mu) = \bigcap_{\lambda\in\Lambda,\, \mu\in M} (A_\lambda \cup B_\mu)$$

If $x \in$ LHS, then $x \in \bigcap_{\lambda\in\Lambda} A_\lambda$ or $x \in \bigcap_{\mu\in M} B_\mu$.

> If $x \in \bigcap_{\lambda\in\Lambda} A_\lambda$, then $x \in A_\lambda$ for every $\lambda \in \Lambda$, so $x \in A_\lambda \cup B_\mu$ for every $\lambda \in \Lambda$ and every $\mu \in M$, so $x \in$ RHS.
> If $x \in \bigcap_{\mu\in M} B_\mu$, then $x \in B_\mu$ for every $\mu \in M$, so $x \in A_\lambda \cup B_\mu$ for every $\lambda \in \Lambda$and every $\mu \in M$, so $x \in$ RHS.

Therefore LHS $\subseteq$ RHS.

Conversely, if $x \notin$ LHS, then $x \notin \bigcap_{\lambda\in\Lambda} A_\lambda$ and $x \notin \bigcap_{\mu\in M} B_\mu$, so there exist indices λ_0 and μ_0 such that $x \notin A_{\lambda_0}$ and $x \notin B_{\mu_0}$. Then $x \notin A_{\lambda_0} \cup B_{\mu_0}$, so $x \notin$ RHS.

Therefore RHS $\subseteq$ LHS, so RHS $=$ LHS. •

Remarks Part 1) of the theorem states the commutative laws for union and intersection.
Part 2) states the associative laws.
Part 3) states the distributive laws.

Exercise Try to write a generalization of the commutative laws for infinite families.

Definition 4.7 If A and B are sets, then $A - B = \{x \in A : x \notin B\}$ is called the complement of B in A . If the set A is clearly understood, we might simply refer to $A - B$ as the complement of B, sometimes written as B^c or $-B$ or $\widetilde{B}$.

Theorem 4.8 (DeMorgan's Laws) For sets A and B_λ,

1) $A - \bigcup\{B_\lambda : \lambda \in \Lambda\} = \bigcap\{A - B_\lambda : \lambda \in \Lambda\}$, and
2) $A - \bigcap\{B_\lambda : \lambda \in \Lambda\} = \bigcup\{A - B_\lambda : \lambda \in \Lambda\}$

Proof Exercise (*DeMorgan's Laws are very simple but important tools for manipulating sets.*)

Definition 4.9 The set $A \times B = \{(a, b) : a \in A \text{ and } b \in B\}$ is called the product of A and B. This definition can be extended in an obvious way to any finite product of sets: for example $A \times B \times C = \{(a, b, c) : a \in A,\, b \in B,\, c \in C\}$.

If $A = B$ then we sometimes write A^2 for $A \times A$. For example, $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$. Note that $A \times B$ and $B \times A$ are usually not the same. (*In fact,* $A \times B = B \times A$ *iff ... ?*)

$A \times B$ is a set of ordered pairs, and (if everything in mathematics is a set) an ordered pair should itself be defined as a set. The characteristic behavior of ordered pairs is that $(a, b) = (c, d)$ iff $a = c$ and $b = d$. So we want to define (a, b) to be a set having this property. Any set that behaves in this way is as good as any other to agree on as an official definition for an ordered pair. The most commonly used definition is due to the Polish topologist Kazimierz Kuratowski (1896-1980): we define

$$(a, b) = \{\{a\}, \{a, b\}\}$$

If we use this definition then we can prove that $(a, b) = (c, d)$ iff $a = c$ and $b = d$. (*Try it – and remember, in your argument, that a, b, c, d may not be distinct!*)

Exercise: There are other ways one could define an ordered pair. For example, a possible alternate definition is due to the American mathematician Norbert Wiener (1894-1964): we could define

$$(a, b) = \{\{\{a\}, \emptyset\}, \{\{b\}\}\}.$$

Using this definition, prove that $(a, b) = (c, d)$ iff $a = c$ and $b = d$.

> ***Note*** *In the discussion of paradoxes, we stated that sets should only be defined as subsets of other sets. Therefore, a careful definition of $A \times B$ should read:*
>
> $$A \times B = \{(a, b) \in X : a \in A \text{ and } b \in B\}.$$
>
> *So, if we were doing axiomatic set theory, we would have to provide, as part of the definition, a set X which we know exists and define $A \times B$ to be a subset of X. Given A and B, how in general would we supply X? To give the flavor of how axiomatic set theory proceeds, we'll elaborate on this point.*
>
> > *Axiomatic set theory begins with some axioms: the variables $x, y, \ldots$ refer to sets and $\in$ refers to the membership relation among them. In axiomatic set theory, everything is a set, so members of sets are other sets. Therefore we see notation like $x \in y$ in the axioms: "the set x is a member of the set y."*
> >
> > *A partial list of these axioms includes:*
> >
> > *A1)* ***two sets are equal iff they have exactly the same elements****:*
> > *more formally,*
> > $$\forall x \, \forall y \, (x = y \Leftrightarrow \forall z \, (z \in x \Leftrightarrow z \in y))$$
> >
> > *A2)* ***there is an empty set****:*
> > *more formally,*
> > $$\exists x \, (\forall y \, \neg(y \in x))$$
> > (*Notice, by A1, this set x is unique, so we can give this set a name: $\emptyset$*)
> >
> > *A3)* ***any two sets x and y can be "paired" into a new set $\{x, y\}$*** *:*
> > *more formally,*
> > $$\forall x \, \forall y \, \exists z \forall w \, (w \in z \Leftrightarrow (w = x \vee w = y))$$
> >
> > *A4)* ***every set x has a power set y****:*
> > *more formally,*
> > $$\forall x \, \exists y \, \forall z \, (z \in y \Leftrightarrow (\forall w \, (w \in z \Rightarrow w \in x))$$

A5) ***the union u of any set x exists****:*

more formally,

$$\forall x\, \exists u\, \forall z\ (z \in u \Leftrightarrow (\exists y\, (\, z \in y\ \wedge\ y \in x\,)))$$

The last axiom to be mentioned here, A6, is a little harder to write precisely, so we will simply state an informal version of it:

A6) ***for any set x, there is a subset y of x whose members are all the elements in x that meet an appropriate "legal" description "..."****: Roughly,*

$$\forall x\ \exists y\ \forall z\ \ (\, z \in y \Leftrightarrow (z \in x\ \wedge\ \text{“...”})\,)$$

A6) is a fancier (but still somewhat vague) version of saying that we are allowed to define a set y by writing:

$$y = \{z \in x :\ \text{“...”}\ \}$$

These axioms and three others A7) – A9) – which we will not state here – are called the Zermelo-Fraenkel Axioms, or ZF for short. (One of the additional axioms, for example, implies that no set is an element of itself.) The resulting system (the axioms and all the theorems that can be proved from the axioms) is called ZF set theory.

In ZF, by using these axioms, we can make a complete definition of $A \times B$. Suppose $a \in A$ and $b \in B$. By axiom A3), the sets $\{A, B\}$, $\{a, b\}$ and $\{a, a\} = \{a\}$ exist so, by axiom A3) again, the set $\{\{a\}, \{a, b\}\} = (a, b)$ also exists.

By axiom A5) $A \cup B =$ the union of the set $\{A, B\}$ exists, and by axiom 4), the sets $\mathcal{P}(A \cup B)$ and $\mathcal{P}(\mathcal{P}(A \cup B))$ exist. Using these sets, we then notice that

$a \in A \cup B$ and $b \in A \cup B$,	so
$\{a\} \subseteq A \cup B$ and $\{a, b\} \subseteq A \cup B$,	so
$\{a\} \in \mathcal{P}(A \cup B)$ and $\{a, b\} \in \mathcal{P}(A \cup B)$,	so
$\{\{a\}, \{a, b\}\} \subseteq \mathcal{P}(A \cup B)$,	so
$\{\{a\}, \{a, b\}\} = (a, b) \in \mathcal{PP}(A \cup B)$	

Therefore each pair $(\, a,\ b\,)$ which we want to put in the set to be called $A \times B$ happens to be a member of the set $X = \mathcal{PP}(A \cup B)$, whose existence follows from the axioms. Of course, not every element of $\mathcal{PP}(A \cup B)$ is an ordered pair. Putting all this together, we could then make the definition

$$A \times B = \{z \in \mathcal{PP}(A \cup B):\ (\exists a \in A)(\exists b \in B)\ z = (a, b)\}.$$

In the rest of these notes we will usually not refer to the ZF axioms. However, we will occasionally make comments about the axioms when something interesting is going on.

Exercises

E3. Which of the following are true when " $\in$ " is inserted in the blank space? Which are true when " $\subseteq$ " is inserted?

a) $\{\emptyset\}$ ___ $\{\emptyset, \{\emptyset\}\}$
b) $\{\emptyset\}$ ___ $\{\emptyset, \{\{\emptyset\}\}\}$
c) $\{\{\emptyset\}\}$ ___ $\{\emptyset, \{\emptyset\}\}$
d) $\{\{\emptyset\}\}$ ___ $\{\emptyset, \{\{\emptyset\}\}\}$
e) $\{\{\emptyset\}\}$ ___ $\{\emptyset, \{\emptyset,\{\emptyset\}\}\}$

E4. a) Suppose A and $\mathcal{B}$ are sets and $A \in \mathcal{B}$. Prove, or give a counterexample, to each of the following statements:

i) $\mathcal{P}(A) \in \mathcal{P}(\mathcal{B})$
ii) $\mathcal{P}(A) \in \mathcal{P}(\bigcup \mathcal{B})$
iii) $\mathcal{P}(A) \subseteq \mathcal{P}(\bigcup \mathcal{B})$

b) Let $A = \mathcal{P}(\emptyset)$, $B = \mathcal{P}(A)$ and $C = \mathcal{P}(B)$. What is $A \cap B \cap C$?

c) Give an example of a set A which has more than one element and such that whenever $x \in A$, then $x \subseteq A$. (Such a set is called transitive.)

E5. Explain why the following statement is or is not true:

If $A = \{\{\emptyset\}, \{\{\emptyset\}\}\}$, then $\bigcup \{B : B \in \mathcal{P}(A)\} = \{\{\{\emptyset\}\}, \{\{\{\emptyset\}\}\}\}$

E6. a) Prove that $\mathcal{P}(A - B) \cup \mathcal{P}(B) = \mathcal{P}(B - A) \cup \mathcal{P}(A)$ if and only if $(A = B$ or $A \cap B = \emptyset)$.

b) State and prove a theorem of the form:

$\mathcal{P}(A - B) - \{\emptyset\} = \mathcal{P}(A) - \mathcal{P}(B)$ if and only if ...

c) For any sets A and B it is true that $\mathcal{P}(A) \cap \mathcal{P}(B) = \mathcal{P}(A \cap B)$ and that $\mathcal{P}(A) \cup \mathcal{P}(B) \subseteq \mathcal{P}(A \cup B)$. State and prove a theorem of the form:

$\mathcal{P}(A) \cup \mathcal{P}(B) = \mathcal{P}(A \cup B)$ if and only if ...

E7. Suppose $A, B, C,$ are sets with $A \neq \emptyset$, $B \neq \emptyset$ and $(A \times B) \cup (B \times A) = C \times C$. Prove that $A = B = C$.

E8. Suppose $A, B, C,$ and D are sets, with $A \neq \emptyset$ and $B \neq \emptyset$. Show that if

$$(A \times B) \cup (B \times A) = (C \times D) \cup (D \times C),$$

then either $(A = C$ and $B = D)$ or $(A = D$ and $B = C)$.

E9. Let A_1, A_2,..., A_n,... be subsets of a set A. Define $\underline{\lim}\, A_n = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k$.

$\underline{\lim}$ is read "lim inf". In expanded form, the notation means that

$$\underline{\lim}\, A_n = (A_1 \cap A_2 \cap A_3 \cap ...) \cup (A_2 \cap A_3 \cap A_4 \cap ...) \cup (A_3 \cap A_4 \cap A_5 \cap ...) \cup ... .$$

Similarly, define $\overline{\lim}\, A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k$. ($\overline{\lim}$ is read "lim sup")

a) Prove that $\underline{\lim}\, A_n = \{x : x$ is in all but at most finitely many $A_n\}$ and that $\overline{\lim}\, A_n = \{x : x$ is in infinitely many of the $A_n\}$.

b) Prove that $\bigcap_{n=1}^{\infty} A_n \subseteq \underline{\lim}\, A_n \subseteq \overline{\lim}\, A_n \subseteq \bigcup_{n=1}^{\infty} A_n$

c) Assume that all the A_n's are subsets of a set X. Prove that

$$\underline{\lim}\, (X - A_n) = X - \overline{\lim}\, A_n$$

d) Prove that $\underline{\lim}\, A_n = \overline{\lim}\, A_n$ if either $A_1 \subseteq A_2 \subseteq A_3 \subseteq$ ... or $A_1 \supseteq A_2 \supseteq A_3 \supseteq$ ... .

E10. Suppose $A, B, C,$ and D are nonempty sets satisfying

$$A \neq B,\ C \neq D, \text{ and } \{A, B\} \neq \{C, D\}.$$

Prove that

$$(A \times (B - A)) \cup (B \times (A - B)) \neq (C \times (D - C)) \cup (D \times (C - D))$$

5. Functions

Suppose X and Y are sets. Informally, a function (or mapping, or map) f from X into Y is a rule that assigns to each element x in X a unique element y in Y. We call y the image of x and call x a preimage of y. We write $y = f(x)$, and we denote the function by $f : X \to Y$. This informal definition is usually good enough for our purposes.

X is called the domain of f = dom(f). The range of f = ran(f) is the set

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

We can think of giving an "input" x to f and the corresponding "output" is $y = f(x)$. Then dom(f) is the set of "all allowed inputs" and ran(f) is the "set of all outputs" corresponding to those inputs.

The set Y is sometimes referred to as the codomain of f. Of course, ran(f) $\subseteq Y$, but ran(f) may not be the whole codomain Y. When ran(f) = the codomain Y, we say that f is a function from X onto Y, or simply that f is onto. In some books, an onto function is also called a surjection.

If different inputs always produce different outputs, we say that f is a one-to-one (or $1-1$) function. More formally: f is one-to-one if, whenever $a, b \in$ dom(f) and $a \neq b$, then $f(a) \neq f(b)$. In some books, a one-to-one function is also called an injection.

If f is both one-to-one and onto, we call f a bijection between X and Y. A bijection sets up a perfect one-to-one correspondence between the elements of the two sets X and Y. Intuitively, a bijection can exist iff X and Y have the same number of elements.

Examples 5.1 (verify the details where necessary)

1) Suppose $X = Y$ and $f : X \to X$ is given by $f(x) = x$. This bijection is called the identity map on X, sometimes denoted by i_X.

2) If $f : X \to Y$ and $A \subseteq X$, then we can define a new function $g : A \to Y$ by $g(x) = f(x)$ for $x \in A$; g is called the restriction of f to A, denoted $g = f|A$. If f is one-to-one then so is g; but g may not be onto even if f is onto. Note that f and g are different functions when $A \neq X$ – because dom(f) $\neq$ dom(g). For example, $g = \sin|[-\frac{\pi}{2}, \frac{\pi}{2}]$ is a bijection between $[-\frac{\pi}{2}, \frac{\pi}{2}]$ and $[-1, 1]$.

3) Let $f : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ by the rule $f(m, n) = 2^m 3^n$. Then f is one-to-one by the Fundamental Theorem of Arithmetic (which states that each natural number has a unique prime factorization). But f is not onto because, for example, $35 \notin$ range(f).

4) Let $X =$ the interval $(1, \infty) \subseteq \mathbb{R}$, $Y = \mathbb{R}$ and define $f(x) = \int_1^\infty \frac{1}{t^x}\, dt$. (*This definition makes sense since the improper integral converges for* $x > 1$.) For example, $f(2) = \int_1^\infty \frac{1}{t^2}\, dt = 1$. Then $f : X \to Y$ is one-to-one but not onto (why?).

5) Let $f : \{2, 3, 4, ...\} \to \mathbb{N}$ by the rule $f(x) =$ "the least integer ≥ 2 which divides x." For example, $f(2) = 2$, $f(3) = 3$, $f(4) = 2$, $f(5) = 5$, $f(6) = 2$, and $f(9) = 3$. The function f is not one-to-one because $f(2) = f(4)$, and f is not onto because ran(f) is the set of prime numbers.

6) Define $f : \mathbb{N} \to \mathbb{N}$ by $f(n) =$ the n^{th} prime number. For example $f(1) = 2$, $f(2) = 3$, $f(3) = 5$, and $f(4) = 7$. The function f is clearly one-to-one but not onto.

7) Let $\mathbb{S}$ be the set of prime numbers and define $f : \mathbb{N} \to \mathbb{S}$ using the same rule as in Example 6. Then f is a bijection between $\mathbb{N}$ and $\mathbb{S}$. This illustrates that whether $f : X \to Y$ is onto depends on the codomain Y. Although ran(f) is determined in Examples 6) and 7) by the domain and the "rule" f, the codomain is not automatically determined. Without knowing Y, we cannot say whether a function f is onto.

8) Define $\pi : \mathbb{R} \to \{z \in \mathbb{Z} : z \geq 0\}$ by $\pi(x) =$ "the number of primes $\leq x$". Thus, $\pi(1) = 0$, $\pi(2) = 1$, $\pi(2.5) = 1$, and $\pi(5.6) = 3$. The function π is onto (why?) but not one-to-one.

This function appears as part of the famous Prime Number Theorem which states that

$$\lim_{x \to \infty} \frac{\pi(x) \ln(x)}{x} = 1.$$

A proof of the Prime Number Theorem is quite difficult. However, a much simpler fact is that $\lim_{x \to \infty} \pi(x) = \infty$. *Why is this simpler fact true? And does it also follow <u>from</u> the Prime Number Theorem?*

9) Define $g : \mathbb{R} \to \mathbb{R}$ by $g(x) = \sum_{n=0}^{\infty} \frac{x^n}{n!}$. This function is one-to-one but not onto. (Why?)

10) Let A be a set and define $f : A \to \mathcal{P}(\mathcal{P}(A))$ by

$$f(a) = \text{"the set of all subsets of } A \text{ containing } a\text{"} = \{B \in \mathcal{P}(A) : a \in B\}.$$

For example, if $A = \{1, 2, 3\}$, then

$$f(2) = \{\{2\}, \{1, 2\}, \{2, 3\}, \{1, 2, 3\}\}.$$

This function is one-to-one. To see this, suppose that $a \neq b$; then $\{a\} \in f(a)$ and $\{a\} \notin f(b)$, so $f(a) \neq f(b)$. Is the function onto?

11) Define $\pi : X \times Y \to Y$ by $\pi(x, y) = y$, called the <u>projection of $X \times Y$ to Y</u>." When is this projection function $1 - 1$? Must π be onto?

12) Let $X = \{C : C$ is a rectifiable curve in the plane$\}$. (*A rectifiable curve is one for which an arc-length is defined.*) Let $f : X \to [0, \infty)$ by $f(C) =$ "the length of the curve *C*." Is f one-to-one? onto?

Our informal definition of function is good enough for most purposes. However in the spirit of "everything in mathematics is a set," we should be able to give a more precise definition of a "function f from X into Y" as a set. We begin by using sets to define a relation.

Definition 5.2 A relation R between X and Y is a subset of $X \times Y$. If $(x, y) \in R$, we write xRy, meaning that "x is related to y" (by this relation R).

Example 5.3 Consider $\{(x, y) \in \mathbb{R} \times \mathbb{R} : \text{for some } b \in \mathbb{R},\ y = x + b^2\}$. This set of ordered pairs is a relation from $\mathbb{R}$ to $\mathbb{R}$. We could call the relation R, but it actually has a more familiar name, $\leq$:

$$\leq \ = \{(x, y) \in \mathbb{R} \times \mathbb{R} : \text{for some } b \in \mathbb{R},\ y = x + b^2\}$$

The relation $\leq$ is a set. The statement "$x \leq y$" means that the pair $(x, y) \in \leq$. Notice that for each x there are many y's for which $(x, y) \in \leq$: for example $(1, 2) \in \leq$, $(1,\ 4.5) \in \leq$, ...

> **Exercise:** The relation $\leq$ is a set. What sets are the familiar relations $>$, $<$, and $\geq$ on $\mathbb{R}$?

A function f is a special kind of relation between X and Y: one in which every x is related by f to only one y in Y.

Definition 5.4 A function f from X into Y, denoted $f : X \to Y$ is a set $f \subseteq X \times Y$ (*so f is a relation from X to Y*) with the additional properties that:

a) for every $x \in X$, there is a $y \in Y$ such that $(x, y) \in f$, and

b) if $(x, y) \in f$ and $(x, z) \in f$, then $y = z$.

X is called the domain of f and the range of f is the set

$$\{y \in Y : \text{ there exists an } x \in X \text{ for which } (x, y) \in f\}$$

If f is a function from X to Y and $(x, y) \in f$, then we use the standard functional notation and write $y = f(x)$ in preference over the relation notation xfy.

Condition a) states that a function f from X into Y is defined at each point of X; condition b) states that f is single-valued – f can't have more than one value for a particular x. For functions $f : \mathbb{R} \to \mathbb{R}$, condition b) just states the familiar condition from precalculus that a vertical line cannot intersect the graph of a function f in more than one point.

Example 5.5 Let $f : \mathbb{R} \to \mathbb{R}$ be given by $f(x) = x^2$. More formally, this function is the set of ordered pairs $f = \{(x, x^2) : x \in \mathbb{R}\} \subseteq \mathbb{R}^2$. A drawing of this set of points in $\mathbb{R}^2$ is the parabola $y = x^2$. In precalculus, this parabola is called the "graph" of the function. But in our formal definition, the parabola is f : that is, f is defined to be the set of pairs which, in precalculus, is called its graph.

Example 5.6 For sets X and Y, we write Y^X to represent the set of all functions from X to Y. For example, $\mathbb{R}^{\mathbb{R}}$ is the set of all real-valued functions of a real variable.

If $X = \emptyset$, then there is exactly one function in Y^X : the empty function $\emptyset$.
(Check: $\emptyset \subseteq \emptyset \times Y$ and $\emptyset$ satisfies the conditions a) and b) in Definition 5.4.)
So $Y^X = \{\emptyset\}$.

If $X \neq \emptyset$ and $Y = \emptyset$, then there are no functions in Y^X so $Y^X = \emptyset$.
(Since $X \times \emptyset = \emptyset$, the only relation from X to Y is $\emptyset$, so $\emptyset$ is the only "candidate" for a function from X to Y. But since $X \neq \emptyset$, $\emptyset$ fails to satisfy part a) in Definition 5.4.)

If X has n elements and Y has m elements ($n, m \in \mathbb{N}$), then there are m^n functions in Y^X. (Why? For each $x \in X$, how many choices are there for $f(x)$?)

It is usually not necessary to think of a function as a set of ordered pairs; the informal view of a function as a "rule" usually is good enough. The formal definition is included partly to reiterate the point of view that "everything in mathematics is a set." However, it can also be very useful notationally. For example, if f and g are functions then, since f and g are sets, it makes sense to form the sets $f \cup g$ and $f \cap g$ and sometimes these sets are new functions.

1) Suppose $f : (-\infty, 0] \to \mathbb{R}$ and $g : [0, \infty) \to \mathbb{R}$ are given by $f(x) = -x$ and $g(x) = x$. As sets, $f = \{(x, -x) : x \leq 0\}$ and $g = \{(x, x) : x > 0\}$. Then the set $h = f \cup g$ is a function with domain $\mathbb{R}$. For $x \in \mathbb{R}$, $h(x) =$?.
Is the set $k = f \cap g$ a function? If so, what is its domain and what is a formula for the function?

2) In general, if f and g are functions, must $f \cup g$ and $f \cap g$ always be functions? If so, what is the domain of each? If not, then what conditions must f and g satisfy so that $f \cup g$ and $f \cap g$ are functions?

Example 5.7 As another illustration of how "everything in mathematics is a set," consider the real number system – it consists of the set $\mathbb{R}$, two operations called addition ($+$) and multiplication ($\cdot$), and an order relation $\leq$. (*Subtraction and division are defined in terms of addition and multiplication.*) Notice that the operations $+$ and $\cdot$ are functions $+ : \mathbb{R}^2 \to \mathbb{R}$ and $\cdot : \mathbb{R}^2 \to \mathbb{R}$.

Therefore, $+ \subseteq \mathbb{R}^2 \times \mathbb{R}$ and, for example, $((2, 3), 5) \in +$. In function notation this would be written $+(2, 3) = 5$, but the more standard way, in this case, is to write $2 + 3 = 5$.

The functions $+$ and $\cdot$ are required to obey certain axioms such as $(x, y) \in +$ iff $(y, x) \in +$, that is, $x + y = y + x$. The order relation $\leq$ is also just a certain subset of $\mathbb{R}^2$ (see Example 5.3).

Therefore we can think of the real number system, – its elements and all its ordering and arithmetic – as being "captured" in the 4 sets: $\mathbb{R}$, $+$, $\cdot$, and $\leq$, and we can gather all this into a single object: the 4-tuple of sets: $(\mathbb{R}, +, \cdot, \leq)$.

But this 4-tuple can be viewed as just a complicated ordered pair: $(((\mathbb{R}, +), \cdot), \leq)$. Since each ordered paired here is just a set, the whole real number system, with all its ordering and operations, can be thought of as one single (complicated) set.

6. More About Functions

Suppose $A \subseteq X$, $B \subseteq Y$ and that $f : X \to Y$. The image of the set A, denoted $f[A]$, is the set of all images of the elements of A. More precisely, $f[A] = \{y \in Y : \ y = f(a) \text{ for some } a \in A\}$. A little less formally we could also write $f[A] = \{f(a) : a \in A\}$.

The inverse image set of B, denoted $f^{-1}[B]$, is the subset of X consisting of all preimages of all the elements from B. More precisely, $f^{-1}[B] = \{x \in X : f(x) \in B\}$.

Example 6.1 Suppose $f : \mathbb{R} \to \mathbb{R}$ is the function given by $f(x) = x^2$. Then

$f[\,[0,2]\,] = [0,4]$ $\qquad$ $f^{-1}[\,[0,9]\,] = [\,-3,3]$

$f[\,\{2\}] = \{4\}$ $\qquad$ $f^{-1}[\,\{4\}] = \{\,-2,2\}$

$f[\,[\,-3,\,-2]\,] = [4,9]$ $\qquad$ $f^{-1}[\,[\,-5,\,-4]\,] = \emptyset$

The function f is not onto since $f[\mathbb{R}] \neq \mathbb{R}$.

When B is a one point set such as $\{4\}$, we will often write $f^{-1}[\,4]$ or even $f^{-1}(4)$ rather than the more formal $f^{-1}[\{4\}]$.

The next theorem gives some frequently used properties of image and inverse image sets.

Theorem 6.2 Suppose $f\colon X \to Y$, that $A_\lambda \subseteq X$ and $B_\lambda \subseteq Y$ $\quad (\lambda \in \Lambda)$

1) $f[\emptyset] = \emptyset$ and $f[X] \subseteq Y$ $\qquad$ 1′) $f^{-1}[\emptyset] = \emptyset$ and $f^{-1}[Y] = X$

2) if $A_1 \subseteq A_2$, then $f[A_1] \subseteq f[A_2]$ $\qquad$ 2′) if $B_1 \subseteq B_2$, then $f^{-1}[B_1] \subseteq f^{-1}[B_2]$

3) $f[\bigcup A_\lambda] = \bigcup f[A_\lambda]$ $\qquad$ 3′) $f^{-1}[\bigcup B_\lambda] = \bigcup f^{-1}[B_\lambda]$

4) $f[\bigcap A_\lambda] \subseteq \bigcap f[A_\lambda]$ $\qquad$ 4′) $f^{-1}[\bigcap B_\lambda] = \bigcap f^{-1}[B_\lambda]$

Proof The proof of each part is extremely simple. As an example, we prove 4′).

Let LHS and RHS denote the left and right sides of 4′). Then $x \in$ LHS iff $f(x) \in \bigcap B_\lambda$ iff $f(x) \in B_\lambda$ for every $\lambda \in \Lambda$ iff $x \in f^{-1}[B_\lambda]$ for every $\lambda \in \Lambda$ iff $x \in$ RHS. Thus LHS and RHS have the same members and are equal. •

Example 6.3 In part 4), " $\subseteq$ " cannot be replaced by " $=$ ". For example, suppose $A_1 = \{x \in \mathbb{R} : x < 1\}$, $A_2 = \{x \in \mathbb{R} : x > 1\}$ and let f be the constant function $f(x) = 1$. Then $f[A_1] = f[A_2] = \{1\}$, so $f[A_1] \cap f[A_2] = \{1\}$, but $f[A_1 \cap A_2] = f[\emptyset] = \emptyset$.

Definition 6.4 Suppose $f : X \to Y$ and $g : Y \to Z$ are functions. Their composition $g \circ f$ is a function from X to Z formed by first appplying f, then g. More precisely, the composition $(g \circ f) : X \to Z$ is defined by $(g \circ f)(x) = g(f(x))$.

$$X \xrightarrow{f} Y \xrightarrow{g} Z$$
$$\underbrace{\qquad\qquad}_{g \circ f}$$

If we have another function $h: Z \to W$, we can form the "triple compositions" $h \circ (g \circ f)$ and $(h \circ g) \circ f$. It is easy to check that these functions $X \to W$ are the same. In other words, the associative law holds for composition of functions and we may write $h \circ g \circ f$ without worrying about parentheses.

$$X \xrightarrow{f} Y \xrightarrow{g} Z \xrightarrow{h} W$$
$$\underbrace{\qquad\qquad\qquad}_{h \circ g \circ f}$$

Here is useful observation about compositions and inverse image sets: if $A \subseteq Z$, then

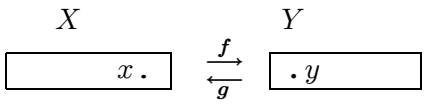
$$(g \circ f)^{-1}[A] = f^{-1}[g^{-1}[A]]$$

To check this, note that $x \in$ RHS iff $f(x) \in g^{-1}[A]$ iff $g(f(x)) \in A$ iff $(g \circ f)(x) \in A$ iff $x \in$ LHS.

If $f : X \to Y$ is a bijection, then f gives a perfect one-to-one correspondence between the members of X and Y. In that case, we can define a function $g : Y \to X$ as follows.

> Suppose $y \in Y$. Then $y = f(x)$ for some $x \in X$ (*because f is onto*) and there is only one such x (*because f is one-to-one*). Define $g(y) = x$. Then $g : Y \to X$ and $g(y) = x$ if and only if $f(x) = y$.

More formally, $g = \{(y, x) \in Y \times X : x \in f^{-1}(y)\}$. Because f is a bijection, the inverse image set $f^{-1}(y)$ contains exactly one member, x.



It follows that $f(g(y)) = y$ and $g(f(x)) = x$, that is, $f \circ g = i_Y : Y \to Y$ and $g \circ f = i_X : X \to X$. The function *g* defined above is clearly the only possible function with these two properties.

Of course, g is also a bijection and the whole discussion could be carried out starting with g and using it to define a function f with the property that $f \circ g = i_Y : Y \to Y$.

Definition 6.5 Suppose $f : X \to Y$ and $g : Y \to X$ and that $g \circ f = i_X$ and $f \circ g = i_Y$. Then we say f and g are inverse functions to each other. We write $g = f^{-1}$ and $f = g^{-1} = (f^{-1})^{-1}$.
(*The idea is that inverse functions "undo" each other.*)

The preceding discussion proves the first part of the following theorem.

Theorem 6.6 Suppose $f: X \to Y$. Then

1) If f is a bijection, there is a unique bijection $g : Y \to X$ for which $g \circ f = i_X$ and $f \circ g = i_Y$.

2) If f has an inverse function g, then f is a bijection.

Proof To prove part 2), we note that if $f(a) = f(b)$, then $g(f(a)) = g(f(b))$. Since $g \circ f = i_X$, this implies that $a = b$, so f is one-to-one. Also, for any $y \in Y$ we have $g(y) \in X$, and since $f \circ g = i_Y$, we get that $f(g(y)) = y$. Therefore y is the image of the point $x = g(y)$ in X, so f is onto. •

Comment on notation: For <u>any</u> function $f : X \to Y$ and $B \subseteq Y$, we can write the <u>inverse image set</u> $f^{-1}[B]$, whether or not f is bijective and has an inverse function. In particular, for <u>any</u> function f we might write $f^{-1}(y)$ as a shorthand for $f^{-1}[\{y\}]$ = the inverse image <u>set</u> of the one-point set $\{y\}$.

<u>If</u> f <u>happens to be bijective,</u> then the notation $f^{-1}(y)$ is ambiguous: " f^{-1} " might mean the inverse image operation for which $f^{-1}(y) = \{x\}$, or " f^{-1} " might mean the inverse function, in which case $f^{-1}(y) = x$.

This double use of the symbol f^{-1} to denote the inverse image set operation (defined for <u>any</u> function) and also to denote the inverse function (<u>when</u> f <u>is bijective</u>) usually causes no confusion in practice – the intended meaning is clear from the context.

In a situation where the ambiguity might cause confusion, we simply avoid the shorthand and use the more awkward notation $f^{-1}[\{y\}]$ for the inverse image set..

As a simple exercise in notation, convince yourself that f is bijective iff $f^{-1}[\{y\}]$ is a one-element subset of X for every $y \in Y$.

Examples 6.7 (In each case, verify the details.)

1) Let $\ln : (0, \infty) \to \mathbb{R}$ be defined by $\ln x = \int_1^x \frac{1}{t}\, dt \quad (x > 0)$. This function is bijective (why?) and therefore has an inverse function $\ln^{-1} : \mathbb{R} \to (0, \infty)$. The function $\ln^{-1}$ is often called exp, so $\exp \circ \ln : (0, \infty) \to (0, \infty)$ and $\exp(\ln x) = x$ for every $x > 0$. Similarly, $\ln \circ \exp : \mathbb{R} \to \mathbb{R}$ and $\ln(\exp x) = x$ for each $x \in \mathbb{R}$. (*It turns out, of course, that $\exp x = e^x$, although that equation requires proof.*)

2) The function $f : \mathbb{N} \to \{2, 4, 6....\} = \mathbb{E}$ given by $f(n) = 2n$ is bijective and has the inverse function $g : \mathbb{E} \to \mathbb{N}$ given by $g(e) = \frac{e}{2}$.

3) A <u>linear</u> function $f : \mathbb{R}^n \to \mathbb{R}^n$ can be expressed as multiplication by some $n \times n$ matrix M: $f(x) = M \cdot x$. Then f is bijective iff $\det(M) \neq 0$ and, in that case, $f^{-1} : \mathbb{R}^n \to \mathbb{R}^n$ is multiplication by the inverse matrix M^{-1} and

$$f^{-1}(f(x)) = M^{-1} \cdot M \cdot x = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ & & \vdots & \\ 0 & 0 & \dots & 1 \end{bmatrix} \cdot x = x$$

These are facts you should know from linear algebra.

> **Exercise** (*if you have had an analysis course*) Prove that if $f : \mathbb{R} \to \mathbb{R}$ is continuous and onto, then f has an inverse iff f is strictly monotone. (*Hint: Remember the Intermediate Value Theorem. If f is not strictly monotone, then there must exist three points $a < c < b$ such that $f(c)$ is either $\leq$ or $\geq$ both of $f(a)$ and $f(b)$.*)

If we look carefully, we see that part 1) of Theorem 6.6 can be broken into two pieces.

Theorem 6.8 1) $f\colon X \to Y$ is one-to-one iff there exists a function $g : Y \to X$ such that $g \circ f = i_X$ (unless $X = \emptyset$, $Y \neq \emptyset$, and $f = \emptyset$; why?)

2) $f\colon X \to Y$ is onto iff there exists a function $g : Y \to X$ such that $f \circ g = i_Y$.

Proof The ideas are quite similar to the argument establishing the existence of an inverse for a bijection. The proof of 1) is left as an exercise. We will illustrate by proving part 2), which is of special interest because of a subtle point that comes up.

If such a g exists, then for each $y \in Y$ we have $f(g(y)) = y$, so y is the image of the point $x = g(y) \in X$. Therefore f is onto.

Conversely suppose f is onto so that, for each $y \in Y$, we know that $f^{-1}(y) \neq \emptyset$. Choose an arbitrary element $x \in f^{-1}(y)$ and define $g(y) = x$. Clearly, $f \circ g = i_Y$. (Note that if $Y = \emptyset$, then we must have $f = \emptyset$ and $X = \emptyset$. The description above then defines $g = \emptyset$, and $g \circ f = i_Y = \emptyset$ is still true.) •

> *The subtle point of the proof lies in the definition of g which, in set notation, reads:*
>
> $$g = \{(y, x) \in Y \times X : \ x \text{ is an element } \underline{\text{arbitrarily}} \ \underline{\text{chosen}} \ \text{ from } f^{-1}(y)\}$$
>
> *In discussing paradoxes, we stated that a set g could only be formed by abstraction as $g = \{(y, x) \in Y \times X$: ... "some legal description"... $\}$. If one is doing axiomatic set theory, the phrase "an arbitrarily chosen element" is not legal; it is "too vague." More precisely, it cannot be properly written as a description in the first order predicate calculus and therefore the definition of the function g would be invalid.*
>
> *The axioms for set theory (ZF) tell us that certain sets exist ($\emptyset$, for example) and give methods to create new sets from old ones. Roughly, these methods are of two types:*
>
> *i) "from the top down" – forming a subset of a given set*
> *ii) "from the bottom up" – somehow piecing together a new set from old ones (for example, by union, or by pairing). It is understood that the "piecing together" must be done in a finite number of steps – for example, we cannot say "apply the Pairing Axiom infinitely many times to get the set z ..."*
>
> *<u>In the present situation</u>,*

i) we tried to define g ("from the top down") by describing g as a certain subset of $Y \times X$; the problem is that we can't say precisely <u>how</u> to pick each x and therefore we can't use the Subset Axiom A6.

ii) if we try to define g "from the bottom up," we can begin by using the fact that f is onto: this means that

$$\forall y \in Y \ \exists x \ x \in f^{-1}(y)$$

Therefore for any particular $y \in Y$, we can form an ordered pair (y, x) with $y \in f^{-1}(x)$. <u>But if there are infinitely many pairs</u> (y, x), we have no axiom that allows us to "gather" these pairs into a single set g

Also, none of the ZF axioms A7) – A9) (which we didn't state) is any help here.

In spite of all this, our informal description of g seems intuitively sound, so another axiom called the <u>Axiom of Choice</u> (AC, for short) is usually added to the set theory axioms ZF so that our intuitive argument can be justified. In one of its many equivalent forms, the axiom of choice states:

[*AC*] *If $\mathcal{A} = \{A_y : y \in Y\}$ is a family of nonempty sets, then there exists a function $h : Y \to \bigcup\{A_y : y \in Y\}$ such that, for each y, $h(y) \in A_y$.*

AC guarantees the existence of a function (set) h that "chooses" an element $h(y)$ from each set A_y and, here, that's just what we need. If we use the sets $A_y = f^{-1}(y)$, then AC gives us the function h and we can then use h to define

$$g = \{(y, x) : y \in Y \wedge x = h(y)\}$$

Defining the function g is <u>not always</u> such a delicate matter. In some <u>special cases</u>, we can avoid the whole problem and don't need AC. For example:

i) if $X = \mathbb{N}$, we could quite specifically define g by saying "let x be the smallest element in $g^{-1}(y)$." In other words, we could write a perfectly precise definition of g "from the top down" using the language of first order predicate calculus:

$$g = \{(y, x) \in Y \times X : \ x \in f^{-1}(y) \ \wedge \ \forall z \, (z \in f^{-1}(y) \Rightarrow z \geq x) \ \}.$$

ii) if Y were finite, we could index its members so that $Y = \{y_1, y_2, ..., y_n\}$ for some $n \in \mathbb{N}$. Then we could proceed "from the bottom up" to write the <u>finite</u> list of statements

$$\begin{array}{l} \exists x_1 \in f^{-1}(y_1) \\ \exists x_2 \in f^{-1}(y_2) \\ \qquad \vdots \\ \exists x_n \in f^{-1}(y_n) \end{array}$$

Using these $x_1, ..., x_n$ we can write a definition of g :

$$g = \{(y, x) \in Y \times X : (y = y_1 \wedge x = x_1) \vee ... \vee (y = y_n \wedge x = x_n)\}\}$$

(The description is "legal" since it involves only finitely many terms.)

Generally speaking, AC is required when it is necessary to choose an element from each nonempty set in an infinite family and there is no way to describe precisely which element to choose from each set. When the choice can be explicitly stated (as when $X = \mathbb{N}$ above), AC is not necessary.

To borrow a non-mathematical example from the philosopher Bertrand Russell: if you have an infinite collection of pairs of socks, you need AC to create a set consisting of one sock from each pair, but if you have an infinite collection of pairs of shoes, you don't need AC to create a set containing one shoe form each pair— because you can precisely describe each choice: "from each pair, pick the left shoe."

The Axiom of Choice, when added to the other axioms of set theory, makes it possible to prove some very nice results. For example, in real analysis, AC can be used to show the existence of a "nonmeasurable" set of real numbers. AC can also be used to show that every vector space (even one that's not finite dimensional) has a basis. AC is equivalent to a mathematical statement called Zorn's Lemma (see Chapter VIII) which you may have met in another course.

The cost of adding AC to the axioms ZF is that it also make it possible to prove some very counter-intuitive results about infinite sets. Here is a famous example:

> *The Banach-Tarski paradox (1924) states that it is possible to divide a solid ball into six pieces which can be reassembled by rigid motions to form two balls of the same size as the original. The number of pieces was subsequently reduced to five by R. M. Robinson in 1944, although the pieces are extremely complicated. (Actually, it can be done with just four pieces if the single point at the center of the ball is ignored..)*

Most mathematicians are content to include AC with the other along with the other axioms and simply to be amused by some of the strange results it can produce. This is because AC seems intuitively very plausible and because many important mathematical results rely on it. We will adopt this attitude and use AC freely when it's needed (and perhaps even when it isn't!), usually without calling attention to the fact.

The system ZF together with the Axiom of Choice is referred to as ZFC set theory for short.

Definition 6.9 A sequence in a set X is a function $f : \mathbb{N} \to X$. The terms of the sequence are $f(1) = x_1$, $f(2) = x_2$, ... , $f(n) = x_n$, ... We often denote a sequence informally by the notation (x_n).

For example, the function $f(n) = 2n + 1$ defines the sequence whose terms are $3, 5, 7, 9, ...$ The n^{th} term of the sequence is $x_n = 2n + 1$, and we might refer to the sequence as (x_n) or $(2n + 1)$.

In the spirit of "everything in mathematics is a set": since a sequence in X is a function from $\mathbb{N}$ to X, a sequence (formally) is just a set $-$ in this case, a special subset of $\mathbb{N} \times X$.

Exercises

E11. a) Show that if $(x, y) \in A$, then $x \in \bigcup\bigcup A$ and $y \in \bigcup\bigcup A$

b) Show that if f is a function, then dom f and ran $f \subseteq \bigcup\bigcup f$

c) Show that if $f : A \to B$, then $f \subseteq \mathcal{PP}(A \cup B)$

E12. Let $\mathbb{R}$ denote the real numbers and $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ be given by $f(x) = 3x^2 - 2x + 1$ and $g(x) = |2x - 1|$. Find the range of $f \circ f$, $f \circ g$, $g \circ f$, and $g \circ g$.

E13. a) How many bijections exist from the set $X = \{1, 3, 5,..., 99\}$ to the set $Y = \{2, 4, 6,..., 100\}$?

b) How many 1-1 maps are there from $X = \{1, 3, 5..., 99\}$ into $Y = \{2, 4, 6,...., 100\}$?

c) How many 1-1 maps are there from $X = \{1, 3, 5..., 99\}$ into $Y = \{1, 2, 3, ..., 100\}$?

E14. Define $f : \mathcal{P}(X) \times \mathcal{P}(Y) \to \mathcal{P}(X \times Y)$ by $f((A, B)) = A \times B$. Prove that f is onto if and only if one of the sets X or Y contains no more than one point.

E15. Let $C[a, b]$ be the set of all continuous functions $f : [a, b] \to \mathbb{R}$. Define a function $\Gamma : C[a, b] \to \mathbb{R}$ by $\Gamma(f) = f(\frac{a+b}{2})$. Is Γ $1 - 1$? onto?

E16. Suppose $p(t) = (r(t), s(t))$ is a one-to-one map from $\mathbb{R}$ into $\mathbb{R}^2$. Define a new map $q : \mathbb{R} \to \mathbb{R}^3$ by $q(t) = (r(t), r(s(t)), s(s(t)))$. Prove that q is one-to-one.

E17. Let $\mathcal{L}$ be the set of straight lines in $\mathbb{R}^2$ which do not pass through the origin $(0, 0)$. Describe geometrically a bijection $f : \mathcal{L} \to \mathbb{R}^2 - \{(0, 0)\}$.

E18. Let $f : X \to X$ and let $f^n : X \to X$ denote the result of composing f with itself n times. Suppose that for every $x \in X$, there exists an $n \in \mathbb{N}$ such that $f^n(x) = x$ (note that n may depend on x). Prove that f is a bijection.

E19. Let $C(\mathbb{R})$ denote the set of all continuous real-valued functions with domain $\mathbb{R}$, that is, $C(\mathbb{R}) = \{f : f \in \mathbb{R}^{\mathbb{R}}$ and f is continuous$\}$. Define a map $I : C(\mathbb{R}) \to C\mathbb{R})$ as follows:

for $f \in C(\mathbb{R})$, $I(f)$ is the function given by $I(f)(x) = \int_0^x f(t)\, dt$.

Is *I* one-to-one? onto? (*Hint: The Fundamental Theorem of Calculus is useful here.)*

E20. Let f be the bijection between the set of nonnegative integers and the set $\mathbb{Z}$ of all integers defined by

$$f(n) = \begin{cases} \frac{n}{2}, & \text{if } n \text{ is even} \\ -\frac{(n+1)}{2}, & \text{if } n \text{ is odd} \end{cases}$$

Now define a mapping $g : \mathbb{Z} \to \mathbb{Q}$ as follows:

> For any integer $m > 1$ we can factor m in a unique way into a product of primes $m = \prod_{i=1}^{k} p_i^{n_i}$, where $p_i < p_{i+1}$ and each n_i is a positive integer. For each $m > 1$, let
>
> $$g(m) = \prod_{i=1}^{k} p_i^{f(n_i)}$$
>
> Define $g(1) = 1$, $g(0) = 0$ and, for negative integers k, define $g(k) = -g(-k)$.

Prove that g is a bijection between $\mathbb{Z}$ and $\mathbb{Q}$.

E21. Let $a_1 = 0$ and for $n = 2, 3, ...$ define $a_n = \sum_{i=1}^{n-1} i = \frac{n(n-1)}{2}$. Show that the function $f : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ given by $f(m, n) = a_{m+n-1} + n$ is a bijection.

E22. Let $\mathcal{B}$ infinite subsets of $\mathbb{N}$. Define $f : \mathcal{B} \to (0, 1]$ as follows:
for $B \in \mathcal{B}$, $f(B)$ = the " binary decimal" $0.x_1x_2\,x_3...\,x_n...$ where $x_n = 1$ if $n \in B$ and $x_n = 0$ if $n \notin B$. *For example, if* $\mathbb{E} = \{2, 4, 6, ...\} \in \mathcal{B}$, *then* $f(\mathbb{E}) = 0.010101...$*base 2*

Prove or disprove that f is onto.

E23. "Measurable sets" in $\mathbb{R}^n$ are defined in elementary measure theory as a way to introduce a notion of integration that is more general than the Riemann integral. It is true, but not trivial to prove, that not every subset of $\mathbb{R}^n$ is measurable – that is, there exist nonmeasurable sets. Each measurable subset X is assigned a number $\mu(X)$ that "measures" its size. For example, $\mu([a, b]) = b - a$.

The following argument claims to show in a very simple way that a nonmeasurable subset of $[0, 1]$ exists. Find the error in the argument. (*Hint: You only need to know that if* $X \subseteq [0, 1]$ *and* X *is measurable, then* $0 \le \mu(X) \le 1$. *Of course, the argument should sound very suspicious because it seems that it doesn't matter whether or not you actually know the definition of measurable set.*)

> Assume each $X \subseteq [0, 1]$ is measurable and has Lebesgue measure $\mu(X)$. Then $\mu(X) \in [0, 1]$ and the number $\mu(X)$ may or may not be in the set X.
> Let $B = \{\mu(X) : X \subseteq [0, 1] \text{ and } \mu(X) \notin X\}$. Since $B \subseteq [0, 1]$, B is measurable by assumption.
>
> But $\mu(B) \in B$ iff $\mu(B) \notin B$ which is impossible. Therefore not every $X \subseteq [0, 1]$ can be measurable. •

7. Infinite Sets

We can classify sets as either finite or infinite and, as we will see later, infinite sets can be further classified into different "sizes." We have already used the words "finite" and "infinite" informally, but now we want to make a more careful definition.

Definition 7.1 The sets A and B are equivalent, written $A \sim B$, if there exists a bijection $f: A \to B$.

Clearly, $A \sim A$; $A \sim B$ implies $B \sim A$; and if $A \sim B$ and $B \sim C$, then $A \sim C$.

Definition 7.2 The set A is called infinite if there is a one-to-one map $f : \mathbb{N} \to A$. A is called finite if A is not infinite.

If A is infinite, then $f[\mathbb{N}]$ is a subset of A equivalent to $\mathbb{N}$, so we could say that A is infinite iff A contains a "copy" of $\mathbb{N}$.

Since we have defined "finite" as "not infinite," we should prove some things about finite sets using that definition. For example:

A is finite iff $A = \emptyset$ or A is equivalent to $\{1, 2, ..., n\}$ for some $n \in \mathbb{N}$.
A subset of a finite set is finite.
A union of a finite number of finite sets is finite.
The image of a finite set under a map f is finite.
The power set of a finite set is finite.

These statements are easy to believe and the proofs are not really difficult but they are tedious induction arguments and will be omitted.

Examples 7.3

1) $A \times B$ is equivalent to $B \times A$. To see this, consider the function given by $f(a, b) = (b, a)$.

2) The mapping $f : \mathbb{N} \to \mathbb{E} = \{2, 4, 6, ... \}$given by $f(n) = 2n$ is a bijection, so $\mathbb{N} \sim \mathbb{E}$. Thus, an infinite set may be equivalent to a proper subset of itself. (*This fact was noticed by Galileo in the 17th century.)* In fact the following theorem shows that this property actually characterizes infinite sets.

Theorem 7.4 A is infinite if and only if A is equivalent to a proper subset of itself.

Proof Suppose A is infinite. Then there is a one-to-one map $f : \mathbb{N} \to A$. Let $f(n) = a_n$. Break A into two pieces: $A = f[\mathbb{N}] \cup (A - f[\mathbb{N}])$. Define $g : A \to (A - \{a_1\})$ by

$$g(x) = \begin{cases} a_{n+1} & \text{if } x = a_n \in f[\mathbb{N}] \\ x & \text{if } x \in A - f[\mathbb{N}] \end{cases}$$

Then g is a bijection between A and A $- \{a_1\}$, so A is equivalent to a proper subset of itself.

Conversely, suppose there is a bijection $f : A \to B$, where $B \subseteq A$ but $B \neq A$. Then $A - B \neq \emptyset$ so we can pick a point $a_1 \in A - B$. Starting with a_1, apply f repeatedly to get a sequence $a_1,\ f(a_1),\ f(f(a_1)) = f^2(a_1), ..., f^n(a_1), ...$.

All the terms $f^n(a_1)$ must be different. To see this, suppose that two of these points are equal, say $f^n(a_1) = f^{n+j}(a_1)$ for some n, j. Then (because f is one-to-one) we have $f^{n-1}(a_1) = f^{n+j-1}(a_1)$, $f^{n-2}(a_1) = f^{n+j-2}(a_1)$, ... , and so on until we get to $a_1 = f^j(a_1)$. But this is impossible because $f^j(a_1)$ is in B and a_1 is not. Therefore the map $g : \mathbb{N} \to A$ given by $g(n) = f^n(a_1)$ is one-to-one, so A is infinite. •

Definition 7.5 The set A is called countable if there exists a one-to-one map $f: A \to \mathbb{N}$.

Thus, a countable set is one which is equivalent to a subset of $\mathbb{N}$ (namely, the range of f). A countable set may be finite or infinite. (*In some books, the word "countable" is defined to mean "countable and infinite."*)

Examples 7.6

1) If $A \subseteq \mathbb{N}$, then A is countable (because $A \sim A \subseteq \mathbb{N}$).

2) Any countable set A is either finite or equivalent to $\mathbb{N}$.

Proof: Suppose $A \sim J \subseteq \mathbb{N}$, where J is infinite. Let j_1 be the least element of J, let j_2 be the least element in $J - \{j_1\}$, ... , and in general, let $j_n =$ the least element in the set $J - \{j_1,\ \ldots, j_{n-1}\}$. Then the map $f : \mathbb{N} \to J$ given by $f(n) = j_n$ is a bijection, so $J \sim \mathbb{N}$. Therefore $A \sim \mathbb{N}$.

By (2), A is countable and infinite if and only if there is a bijection $g : \mathbb{N} \to A$. The function g is a one-to-one sequence whose range is A. Therefore a countable infinite set is one whose elements can be listed as a sequence: $a_1 = g(1),\ a_2 = g(2),\ \ldots,\ a_n = g(n),\ \ldots$.

3) The set $\mathbb{Q}^+ = \{x \in \mathbb{Q} : x > 0\}$ is countable. This is the first really surprising example. It means that we can list all the positive rationals in a sequence even though it is not clear, at first, how to do this. In the usual order on $\mathbb{R}$, there is a third rational between any two rationals so we can't simply list the rationals, for example, in order of increasing size. But the definition of "countable" doesn't require that our list have any connection to size.

One way to create the list is to use a "diagonal argument" invented by Cantor. Begin by imagining all the positive rationals arranged into the following "infinite matrix:"

$$
\begin{array}{cccccccc}
\frac{1}{1} \to & \frac{2}{1} & \frac{3}{1} \to & \frac{4}{1} & \frac{5}{1} & \frac{6}{1} & \frac{7}{1} & \frac{8}{1} \ \ldots\ldots \\
\swarrow & \nearrow & \swarrow & & & & & \\
\frac{1}{2} & \frac{2}{2} & \frac{3}{2} & \frac{4}{2} & \frac{5}{2} & \frac{6}{2} & \frac{7}{2} & \frac{8}{2} \ \ldots\ldots \\
\downarrow \nearrow & & \swarrow & & & & & \\
\frac{1}{3} & \frac{2}{3} & \frac{3}{3} & \frac{4}{3} & \frac{5}{3} & \frac{6}{3} & \frac{7}{3} & \frac{7}{3} \ \ldots\ldots \\
\swarrow & & & & & & & \\
\frac{1}{4} & \frac{2}{4} & \frac{3}{4} & \frac{4}{4} & \frac{5}{4} & \frac{6}{4} & \frac{7}{4} & \frac{8}{4} \ \ldots\ldots \\
. & . & . & . & . & . & . & . \\
. & . & . & . & . & . & . & . \\
. & . & . & . & . & . & . & .
\end{array}
$$

Then create a bijection $g : \mathbb{N} \to \mathbb{Q}^+$ by moving back and forth along the diagonals (skipping over a rational from the matrix if it has been listed previously): the sequence begins $g(1) = \frac{1}{1}$, $g(2) = \frac{2}{1}$, $g(3) = \frac{1}{2}$, $g(4) = \frac{1}{3}$, $g(5) = \frac{3}{1}$, $g(6) = \frac{4}{1}$, $g(7) = \frac{3}{2}$, $g(8) = \frac{2}{3}$, $g(9) = \frac{1}{4}$, $g(10) = \frac{1}{5}$, ...

Of course, we can't picture the whole "infinite matrix" and we don't have a formula $g(n) = ...$. However we have described a definite computational procedure: you would have no trouble finding $g(15)$, and you could find $g(9999)$ with enough time and patience. The function g is clearly one-to-one and onto (for example, with some effort you could find the n for which $g(n) = \frac{379}{211}$).

> *Those who prefer formulas can consider the following alternate approach. Each natural number has a unique representation in base 11, using the numerals 0,1,2,3,...,9,/ (with the symbol / representing 10). For a positive rational p/q reduced to lowest terms, we may reinterpret the symbol string "p/q" as a natural number by thinking of it as a base 11 numeral. For example, 23/31 would be interpreted $in\ base\ 11$ as 2(11)^4 + 3(11)^3 + 10(11)^2 + 3(11)^1 + 1(11)^0 = 34519 (in base 10). In this way, we define a one-to-one function f from $\mathbb{Q}^+$ into $\mathbb{N}$. For example, f(23/31) = 34519. So the set $\mathbb{Q}^+$ is countable.*

We can show that the set of negative rationals, $\mathbb{Q}^-$, is countable using a similar diagonal argument. Or, we can notice that the function $f : \mathbb{Q}^+ \to \mathbb{Q}^-$ given by $f(x) = -x$ is a bijection; since $\mathbb{Q}^-$ is equivalent to a countable set, $\mathbb{Q}^-$ is also countable.

Of course, introducing the term "countable" would be a waste of words if all sets were countable. Cantor also proved that uncountable infinite sets exist. This means that not all infinite sets are equivalent. Some are "bigger" than others!

Theorem 7.7 The interval of real numbers $(0, 1)$ is uncountable.

The heart of the proof is another "diagonal" argument. A technical detail in the proof depends on the following fact about decimal representations of real numbers:

> *Sometimes two different decimal expansions can represent the same real number. For example, 0.10000... = 0.09999... (why?). However, two different decimal expansions can represent the same real number only if one of the expansions ends in an infinite string of 0's and the other ends in an infinite string of 9's.*

Proof Consider any function $f : \mathbb{N} \to (0, 1)$. We will show f cannot be onto and therefore no bijection exists between $\mathbb{N}$ and $(0, 1)$. To begin, write decimal expansions of all the numbers in ran(f):

$$\begin{aligned}
r_1 &= f(1) = 0.x_{11}x_{12}x_{13}\ldots x_{1n}\ldots \\
r_2 &= f(2) = 0.x_{21}x_{22}x_{23}\ldots x_{2n}\ldots \\
r_3 &= f(3) = 0.x_{31}x_{32}x_{33}\ldots x_{3n}\ldots \\
&\vdots \\
r_n &= f(n) = 0.x_{n1}x_{n2}x_{n3}\ldots x_{nn}\ldots \\
&\vdots
\end{aligned}$$

where each x_{ij} is one of the digits $0, 1, ..., 9$. Now define a real number $y = 0.y_1y_2y_3 \ldots y_n \ldots$. by choosing

$$\begin{cases} y_n = 1 & \text{if } x_{nn} \neq 1 \\ y_n = 2 & \text{if } x_{nn} = 1 \end{cases}$$

Then $y \in (0,1)$ and y is not equal to any of the numbers $r_n = f(n)$. To see that $y \neq r_n$, notice that i) by construction, the decimal expression for y differs from r_n in the n^{th} decimal place; and ii) y is not an alternate decimal representation of r_n because y does not end in an infinite string of 0's or 9's. Therefore $y \notin \text{ran}(f)$, so f is not onto. •

Of course, we could start over, adding y to the original list. But then the same construction could be repeated. The list can never be complete, that is, $\text{ran}(f) = (0,1)$ is impossible.

> *A different way to try to dodge the technical difficulty about non-uniqueness of representations might be: whenever a number $f(n)$ has two different decimal representations, include both in the list and then define y. Is there any problem with that approach?*

The next theorem tells us some important properties of countable sets.

Theorem 7.8 1) Any subset of a countable set is countable.

2) If A_n is countable for each $n \in \mathbb{N}$, then $\bigcup_{n=1}^{\infty} A_n$ is countable.

3) If $A_1, A_2, ..., A_n$ are countable, then $A_1 \times A_2 \times ... \times A_n$ is countable.

Proof To show that a set is countable, we need to produce a one-to-one map g of the set into $\mathbb{N}$.

1) If A is countable, we have, by definition, a one-to-one map $f : A \to \mathbb{N}$. If $B \subseteq A$, then $g = f \mid B : B \to \mathbb{N}$ is also one-to-one, so B is also countable.

2) If the A_n's are not disjoint, then consider the sets $B_1 = A_1$, $B_2 = A_2 - A_1$, ... , $B_n = A_n - (A_1 \cup ... \cup A_{n-1})$, ... The B_n's are countable and for any m, n we have that $B_n \cap B_m = \emptyset$. In addition, $\bigcup_{n=1}^{\infty} B_n = \bigcup_{n=1}^{\infty} A_n$ so it's sufficient to prove that $\bigcup_{n=1}^{\infty} B_n$ is countable. Thus we will not lose any generality if we assume at the beginning of the proof that the A_n's are disjoint from each other.

For each n, there is a one-to-one function $f_n : A_n \to \mathbb{N}$. Let p_n denote the n^{th} prime number and define $g : \bigcup_{n=1}^{\infty} A_n \to \mathbb{N}$ as follows:

if $x \in \bigcup_{n=1}^{\infty} A_n$, then x is in exactly one set A_n;
for $x \in A_n$, let $g(x) = p_n^{f_n(x)}$.

Then g is well-defined because the A_n's are disjoint, and g is one-to-one (*why?*).

3) Let p_1, ... , p_n be the first n primes and, as before, let f_n be a one-to-one map from A_n into $\mathbb{N}$. Define $g : A_1 \times A_2 \times \ ... \ \times A_n \to \mathbb{N}$ by

$$g(a_1, ..., a_n) = p_1^{f_1(a_1)} \cdot p_2^{f_2(a_2)} \cdot \ \ldots \ \cdot p_n^{f_n(a_n)}$$

The function g function is $1-1$ (*why?*). •

Example 7.9

1) A union of a finite number of countable sets A_1, ... , A_k is countable: we can let $A_{k+1} = A_{k+2} = ... = \emptyset$ and then use part 2) of Theorem 7.8 to conclude that $\bigcup_{i=1}^{k} A_i = \bigcup_{i=1}^{\infty} A_i$ is countable.

We can combine this result with Theorem 7.8.2 to state: if Λ is a countable index set and A_λ is a countable set (for each $\lambda \in \Lambda$), then $\bigcup_{\lambda \in \Lambda} A_\lambda$ is countable.. Stated more informally: a union of countably many countable sets is countable.

2) The set $\mathbb{Q} = \mathbb{Q}^+ \cup \{0\} \cup \mathbb{Q}^-$ is countable because it is the union of three countable sets.

3) If $B \subseteq A$ and B is uncountable, then A is uncountable – because if A were countable, then its subset B would be countable by Theorem 7.8.1.

For example, the set $\mathbb{R}$ of real numbers is uncountable because its subset $(0, 1)$ is uncountable.. This means that you cannot index the real numbers using $\mathbb{N}$: you cannot write "Let $\mathbb{R} = \{r_1, r_2, r_3, ... \}$."

4) $\mathbb{R} = \mathbb{P} \cup \mathbb{Q}$, where $\mathbb{P}$ is the set of irrational numbers. Since $\mathbb{R}$ is uncountable and $\mathbb{Q}$ is countable, the set $\mathbb{P}$ must be uncountable. Thus there are "more" irrational numbers than rationals.

5) $\mathbb{R}^n$ $(n \geq 1)$ is uncountable, because it contains the "x_1-axis" $\mathbb{R} \times \{0\} \times ... \times \{0\}$, a subset which is clearly equivalent to $\mathbb{R}$.

6) Suppose $A = \{a_1, a_2, \ ... \ , a_n, \ ... \} \subseteq \mathbb{R}$ and that $\epsilon > 0$. Let I_n be the open interval centered at a_n with length $\frac{\epsilon}{2^{n+1}}$, that is, $I_n = (a_n - \frac{\epsilon}{2^{n+2}}, \ a_n + \frac{\epsilon}{2^{n+2}})$. Then $A \subseteq \bigcup_{n=1}^{\infty} I_n$, and the total length of all the intervals I_n is $\sum_{n=1}^{\infty} \frac{\epsilon}{2^{n+1}} = \frac{\epsilon}{2} < \epsilon$. This shows that any countable subset of $\mathbb{R}$ can be "covered" by a sequence of open intervals whose total length is arbitrarily small. (If this does not seem surprising, suppose $A = \mathbb{Q}$; if you think that $\bigcup_{n=1}^{\infty} I_n$ must be $\mathbb{R}$, try to prove that $\sqrt{2}$ must be in $\bigcup_{n=1}^{\infty} I_n$.)

More informally: take a piece of string with length $\frac{\epsilon}{2}$. Cut it in half and lay one half on $\mathbb{R}$ with its center at a_1. Cut the remaining piece in half again and lay one half on $\mathbb{R}$ with center at a_2. Continue in this way, always cutting the remaining piece in half and using one piece to cover the next element from A. (Of course, some of the pieces of string, when laid down, may overlap with earlier pieces.) These pieces from the original length of string cover all the rationals.

Exercises

E24. Prove or give a counterexample:

a) if $A \sim B$ are equivalent and $C \sim D$ are equivalent, then $A \cap C \sim B \cap D$.
b) if $A \sim B$ are equivalent and $C \sim D$ are equivalent, then $A \cup C \sim B \cup D$.
c) if $A \sim B$, then $A - B \sim B - A$.
d) if $A - B \sim B - A$, then $A \sim B$.
e) if $A, B,$ and C are nonempty sets and $A \times B \sim A \times C$, then $B \sim C$.
f) if A is infinite and B is countable, then $A \cup B \sim B$.
g) if A is infinite and B is countable, then $A \cup B \sim A$.

E25. If A is uncountable and B is countable, prove that $A \sim A - B$.

E26. a) Give an explicit formula for a bijection between the intervals $(-\infty, 7)$ and $(0, \infty)$.
b) Give an explicit formula for a bijection between the sets $[0, 1]$ and $(0, 1) \cup \{2\}$.

E27. Prove that the set of all real numbers in the interval $(0, 1)$ that have a decimal expansion using only even digits is uncountable.

E28. Is the following statement true? If not, what additional assumptions about A and B will make it true?

$A \sim B$ iff there is a set $f = \{(a, b) : a \in A, b \in B\}$ such that each element of A and each element of B occur in exactly one pair (a, b).

E29. A subset B of X is called *cocountable* if $X - B$ is countable and *cofinite* if $X - B$ is finite. (*The name is an abbreviation: cocountable = complement is countable.*)

a) Prove that if B and C are cocountable subsets of X, then $B \cup C$ and $B \cap C$ are also cocountable. Is the analogous result true for cofinite sets?
b) Show how to write the set of irrational numbers, $\mathbb{P}$, as an intersection of countably many cofinite subsets of $\mathbb{R}$.

E30. A collection $\mathcal{A}$ of sets is called pairwise disjoint if whenever $A, B \in \mathcal{A}$ and $A \neq B$, then $A \cap B = \emptyset$. For each statement, provide a proof or a counterexample:

a) If $\mathcal{A}$ is a collection of pairwise disjoint circles in the plane, then $\mathcal{A}$ is countable.
b) If $\mathcal{A}$ is a collection of pairwise disjoint circular disks in the plane, then $\mathcal{A}$ is countable.

E31. Prove that there cannot exist an uncountable collection of pairwise disjoint open intervals in $\mathbb{R}$.

E32. Let ℓ be a straight line in the plane. Prove that the set $\ell \cap (\mathbb{Q} \times \mathbb{Q})$ is either empty, contains exactly one point, or is countably infinite. In each case, give an equation for a specific line ℓ to illustrate that case.

E33. Suppose that $f : [0, 1] \to \mathbb{R}$ has the following property:

there is a fixed constant, *M,* such that for every finite set $\{x_1, x_2, \ldots, x_n\} \subseteq [0, 1]$, the following inequality is true:

$$| f(x_1) + \ldots + f(x_n) | < M$$

a) Give an example of such a function f where ran (f) is infinite.
b) Prove that for such a function *f,* $\{x \in [0, 1] : f(x) \neq 0\}$ is countable.

E34. What is wrong with the following argument?

For each irrational number $p \in \mathbb{P}$, pick an open interval (a_p, b_p) with rational endpoints and centered at p. There are only countably many possible pairs (a_p, b_p) because $\mathbb{Q} \times \mathbb{Q}$ is countable. Since the interval (a_p, b_p) is centered at p, the function $\Phi : \mathbb{P} \to \mathbb{Q} \times \mathbb{Q}$ given by $\Phi(p) = (a_p, b_p)$ is one-to one, . Therefore $\mathbb{P}$ is equivalent to a subset of $\mathbb{Q} \times \mathbb{Q}$, so $\mathbb{P}$ is countable.

E35. a) A sequence s in $\mathbb{N}$ is called an arithmetic progression if $\exists d \in \mathbb{N}$ such that $s_{n+1} = s_n + d$ for every $n \in \mathbb{N}$. Prove that the set of arithmetic progressions in $\mathbb{N}$ is countable.

b) A sequence s in $\mathbb{N}$ is called eventually constant if $\exists k, l \in \mathbb{N}$ such that $s_n = l$ for all $n \geq k$. Prove that the set of eventually constant sequences in $\mathbb{N}$ is countable.

c) A sequence s in $\mathbb{N}$ is called eventually periodic if $\exists k, p \in \mathbb{N}$ such that $s_n = s_{n+p}$ for all $n \geq k$. Prove that the set of all eventually periodic sequences in $\mathbb{N}$ is countable.

E36. For a set A, let $\mathcal{W}(A) = \{f \in A^{\{0,1,\ldots,n\}} : n \in \mathbb{N}\}$. (*We can think of A as an "alphabet" and $\mathcal{W}(A)$ as the set of all "finite sequences" or "words" formed from this alphabet.*)

Prove that if A is countable, then $\mathcal{W}(A)$ is countable.

E37. Let A be an uncountable subset of $\mathbb{R}$. Prove that there is a subset of distinct elements $\{a_n : n = 1, 2, \ldots\} \subseteq A$ such that $\sum_{n=1}^{\infty} a_n$ diverges.

E38. Let Y be a set and, for each $n \in \mathbb{N}$, suppose $f_n : \mathbb{R} \to Y$. Let g be a function $g : \mathbb{R} \to Y$ such that, for every $n \in \mathbb{N}$, the set $\{x \in \mathbb{R} : g(x) = f_n(x)\}$ is countable. Prove that there is a point $x_0 \in \mathbb{R}$ such that for all n, $g(x_0) \neq f_n(x_0)$.

What property of $\mathbb{R}$ makes the proof work?

E39. For a set $S \subseteq \mathbb{R}$ and $t \in \mathbb{R}$, let $S + t$ denote the "translated set" $\{s + t : s \in S\}$.

a) Show that if S is countable, then $\exists t \in \mathbb{R}$ such that $S + t \subseteq \mathbb{P}$.

b) For sequences *s* and *t* in $\mathbb{R}$ (that is, for $s, t \in \mathbb{R}^{\mathbb{N}}$), let $s + t \in \mathbb{R}^{\mathbb{N}}$ be the sequence defined by $(s + t)(n) = s(n) + t(n)$. Show that if $\mathcal{S}$ is a countable subset of $\mathbb{R}^{\mathbb{N}}$, then $\exists\ t \in \mathbb{R}^{\mathbb{N}}$ such that $s + t \in \mathbb{P}^{\mathbb{N}}$ for every $s \in \mathcal{S}$.

E40. Give a proof or a counterexample for the following statement:

> If $\mathcal{O} = \{O_\lambda : \lambda \in \Lambda\}$ is a collection of open intervals in $\mathbb{R}$ with the property that $\mathbb{Q} \subseteq \bigcup \mathcal{O}$, then $\bigcup \mathcal{O} = \mathbb{R}$.

E41. a) Show how to write $\mathbb{N}$ as the union of infinitely many pairwise disjoint infinite subsets.

b) Show how to write $\mathbb{N}$ as the union of uncountably many sets with the property that, given any two of them, one is a subset of the other.

c) Show how to write $\mathbb{N}$ as the union of uncountably many sets with the property that any two of them have finite intersection. (*Such sets are called almost disjoint.*)

(*Hint: These statements that actually are true for any infinite countable set, not just for* $\mathbb{N}$. *For parts b) and c), you may find it easier to solve the problem for* $\mathbb{Q}$ *instead, and then use a bijection to "convert" your solution for* $\mathbb{Q}$ *into a solution for the set* $\mathbb{N}$.)

E42. a) Imagine an infinite rubber stamp which, when applied to the plane, inks over all concentric circles with irrational radii around its center. What is the minimum number of stampings necessary to ink over the whole plane?

b) What if the stamp, instead, inks over all concentric circles with rational radii around its center?

E43. Suppose $f : \mathbb{R} \to \mathbb{R}$. Let $A = \{a \in \mathbb{R} : \lim_{x \to a} f(x)$ exists but is not equal to $f(a)\}$. Prove that *A* is countable.

[*Hint: One way to start is to define, for* $r \in \mathbb{Q}$, $A_r^- = \{a \in A : f(a) < r < \lim_{x \to a} f(x)\}$ *and* $A_r^+ = \{a \in A : \lim_{x \to a} f(x) < r < f(a)\}$. *Then* $A = \bigcup_{r \in \mathbb{Q}} (A_r^- \cup A_r^+)$ *(why?). Prove that* A_r^- *and* A_r^+ *are countable.*]

E44. Let *D* be a countable set of points in the plane, $\mathbb{R}^2$. Prove there exist sets A and B such that $D = A \cup B$, where the set A has finite intersection with every horizontal line in the plane and B has finite intersection with every vertical line in the plane.

Notes: 1) This problem is fairly hard. You might get an idea by starting the easy special case of $D = \mathbb{N} \times \mathbb{N}$ *2) The statement that "*$\mathbb{R}^2$ *can be written as the union of two sets A and B where A has countable intersection with every horizontal line and B has countable intersection with every vertical line" is, in fact, equivalent to the continuum hypothesis* (*see p. 40 and Exercise VIII.E.26*).

E45. We say that a function $f : \mathbb{R} \to \mathbb{R}$ has a local maximum at a point x if there exists an open interval (a, b) containing x such that $f(x) \geq f(y)$ for all $y \in (a, b)$ – in other words, $f(x)$ is the (absolute) maximum value of f on the interval (a, b).

a) Give an example of a nonconstant f which has a local maximum at every point. Then modify your example, if necessary, to get an example where ran(f) is infinite.

b) Suppose f has a local maximum at every point $x \in \mathbb{R}$. Prove that ran(f) is countable.
(*Hint: if $f(x) = y \in$ ran(f), pick an interval (a, b) with rational endpoints containing x and such that y is the maximum of f on (a, b).*)

8. Two Mathematical Applications

The relatively simple facts that we know about countable and uncountable sets are enough to prove an interesting fact about the real numbers.

Definition 8.1 A real number r is called algebraic if r is a root to a nonconstant polynomial equation

$$a_n x^n + a_{n-1}x^{n-1} + \ ... \ + a_1 x + a_0 = 0 \quad (*)$$

where the coefficients $a_0, \ ... \ , a_n$ are all integers. (*It is clearly equivalent to require that the coefficients be rational numbers – because we could multiply both sides by some integer to "clear the fractions" and get a polynomial equation with integer coefficients and the same roots.)*

A real number is rational if and only if it is the root of a first degree polynomial equation with integer coefficients: the rational $\frac{p}{q}$ is a root of the equation $qx - p = 0$. Therefore rational numbers are algebraic. The algebraic numbers are a natural generalization to include certain irrationals: for example, $\sqrt{2}$ is algebraic, since it is a root of the quadratic equation $x^2 - 1 = 0$; and $-1 + 2\sqrt{2}$ is algebraic because it is a root of $x^3 + x^2 - 9x + 7 = 0$. A reasonable question would be: are there any nonalgebraic real numbers?

(*Complex roots of (*) are called complex algebraic numbers.)*

Theorem 8.2 The set $\mathbb{A}$ of all algebraic real numbers is countable.

Proof For each $n \geq 1$, let $\mathcal{P}_n$ be the set of all polynomials with integer coefficients and degree n. Each polynomial $P \in \mathcal{P}_n$ is precisely described by the $(n+1)$-tuple containing its coefficients: $(a_n, a_{n-1}, \ ... \ , a_0)$. Since P has degree n, we know that $a_n \neq 0$ and so $(a_n, a_{n-1}, \ ... \ , a_0)$ is in the (countable) set $(\mathbb{Z} - \{0\}) \times \mathbb{Z}^n$. The function $f : \mathcal{P}_n \to (\mathbb{Z} - \{0\}) \times \mathbb{Z}^n$ given by $f(P) = (a_n, a_{n-1}, \ ... \ , a_0)$ is a bijection, so $\mathcal{P}_n$ is countable.

For each polynomial P, let $Z(P)$ be the set of real roots of the equation $P(x) = 0$. $Z(P)$ is a finite set. Then $R_n = \bigcup\{Z(P) : P \in \mathcal{P}_n\}$ is a union of a countable collection of finite sets, so R_n is countable. (*For a particular value $n_0 \in \mathbb{N}$, R_{n_0} is the set of all algebraic numbers that are roots of an equation (*) that has degree n_0; for example, $R_1 = \mathbb{Q}$.*)

By Theorem 7.8(2), $\mathbb{A} = \bigcup_{n=1}^{\infty} R_n$ is countable. •

Definition 8.3 A real number which is not algebraic is called transcendental.

> (*Euler called these numbers "transcendental" because they "transcend the power of algebraic methods." To be more politically correct, we might call these numbers "polynomially challenged."*)

Corollary 8.4 Transcendental numbers exist.

Proof Let $\mathbb{T}$ be the set of transcendental numbers. Since $\mathbb{R} = \mathbb{A} \cup \mathbb{T}$ and $\mathbb{A}$ is countable, $\mathbb{T}$ cannot be empty. •

In fact, this short proof shows much more: not only is $\mathbb{T}$ nonempty, but $\mathbb{T}$ must be uncountable! There are "more" transcendental numbers than algebraic numbers. This is an example of a "pure existence

proof" – meaning that it does not tell us any parrticular transcendental numbers, nor does it give us a way to construct one. To do that is harder. Transcendental numbers were first shown to exist (using different methods) by Liouville in 1844. The numbers e (Hermite, 1873) and π (Lindemann, 1882) are transcendental. One method for producing transcendental numbers is contained in a theorem of Gelfand (1934) from from algebraic number theory; it states that

> ff α is an algebraic number, $\alpha \neq 0, 1$, and β is an algebraic irrational, then α^β is transcendental.

For example, the theorem implies that $\sqrt{2}^{\sqrt{2}}$ is transcendental. The number e^π is also transcendental. This follows from Gelfand's Theorem (which allows <u>complex</u> algebraic numbers) because:

> $e^\pi = e^{-i^2\pi} = (e^{i\pi})^{-i}$ and $e^{i\pi} = \cos\pi + i\sin\pi = -1$. So $e^\pi = (-1)^{-i}$, which is transcendental by Gelfand's Theorem.

As a second application to a different part of mathematics, we will prove a simple theorem from analysis.

Theorem 8.5 A monotone function $f : [a, b] \to \mathbb{R}$ has at most countably many points of discontinuity. (S*ince* $[a, b]$ *is uncountable, this impl9ies that a monotone function on* $[a, b]$ *must be continuous "at most points.*")

Proof Assume $x \leq y$ implies $f(x) \leq f(y)$. (*If* f *is decreasing, simply apply the following argument to the increasing function* $-f$.) Therefore, at each point $c \in (a, b)$ we have

$$\lim_{x \to c^-} f(x) \leq f(c) \leq \lim_{x \to c^+} f(x) \quad (\textit{why must these one-sided limits exist}\ ?)$$

Let $j(c)$ denote the "jump of f at c" $= \lim_{x \to c^+} f(x) - \lim_{x \to c^-} f(x)$. Since f is increasing, $j(c) \geq 0$, and f is discontinuous at c if and only if $j(c) > 0$.

Let $A_n = \{c \in (a, b) : j(c) > \frac{1}{n}\}$. The set A_n is finite because the sum of any set of jumps cannot be more than $f(b) - f(a)$. Furthermore, if $j(c) > 0$, then $j(c) > \frac{1}{n}$ for some sufficiently large n, so $c \in A_n$ for some n. Therefore every point c of discontinuity of f in (a, b) must be in the countable set $\bigcup_{n=1}^{\infty} A_n$. (The function might also be discontinuous at an endpoint a or b, but the set of discontinuities would still be countable.) •

Corollary 8.6 A monotone function $f : \mathbb{R} \to \mathbb{R}$ has at most countably many points of discontinuity.

Proof For every $k \in \mathbb{Z}$, the set D_k of discontinuities of f in the interval $[k, k + 1]$ is countable, by Theorem 8.5. So $\bigcup_{k \in \mathbb{Z}} D_k$, the set of all discontinuities, is countable. •

As a sort of "converse," it is possible to prove that if A is <u>any</u> countable subset of $\mathbb{R}$, then there exists a monotone function $f : \mathbb{R} \to \mathbb{R}$ for which A is precisely the set of points of discontinuity. (You can find an argument in *A Primer of Real Functions* (Boas, p. 129), or see Exercise E66 below.)

9. More About Equivalent Sets

So far, we have seen two kinds of infinite set, countable and uncountable. In this section, we explore the idea of uncountable sets in more detail. Notice that we have no right to assume two sets are equivalent simply because both are uncountable.

There are many equivalent uncountable subsets of $\mathbb{R}$. The following theorem gives some examples. The proof could be made much easier by using the Cantor-Schroeder-Bernstein Theorem, which we will discuss later (p. 42). However, at this stage, it is instructive to give a direct proof.

Theorem 9.1 Suppose $a < b$. The following intervals in $\mathbb{R}$ are equivalent:

$$\mathbb{R} = (-\infty, \infty) \sim (a, b) \sim (0, 1) \sim (a, b] \sim (0, 1] \sim [a, b) \sim [0, 1) \sim [a, b]$$

$$\sim [0, 1] \sim [0, \infty) \sim [a, \infty) \sim (0, \infty) \sim (a,\infty) \sim (-\infty, 0] \sim (-\infty, a]$$

$$\sim (-\infty, 0) \sim (-\infty, a).$$

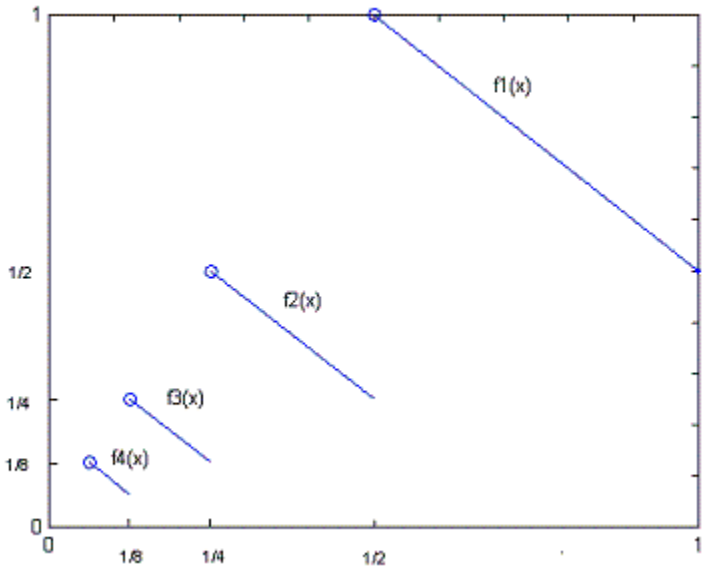
(*Note: for technical reasons that we will discuss later, it is convenient to consider* $\emptyset$ *and all one-point sets* $\{a\}$ *as intervals. These, it turns out, are the only finite intervals. Theorem 9.1 says that all* <u>*infinite*</u> *intervals in* $\mathbb{R}$ *are equivalent.*)

Proof The linear map $f(x) = (b - a)x + a$ can be used to show that $(0, 1) \sim (a, b)$, $(0, 1] \sim (a, b]$, $[0, 1) \sim [a, b)$, and that $[0, 1] \sim [a, b]$.

The bijection $\tan : \left(-\frac{\pi}{2}, \frac{\pi}{2}\right) \to \mathbb{R}$ proves that $\mathbb{R} \sim \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$.

To show that $(0, 1] \sim (0, 1)$, define functions:

$$f_n(x) = \frac{3}{2^n} - x \text{ for } \frac{1}{2^n} < x \leq \frac{1}{2^{n-1}} \text{ for each } n \in \mathbb{N}$$



Each f_n is a bijection from $\left(\frac{1}{2^n}, \frac{1}{2^{n-1}}\right]$ to $\left[\frac{1}{2^n}, \frac{1}{2^{n-1}}\right)$, as illustrated above. Let $f = \bigcup_{n=1}^{\infty} f_n$.

(*The graph of* f *consists of all the separate pieces, shown above, taken together.*)

Then (*check!*) f is a function with $\text{dom}(f) = \bigcup_{n=1}^{\infty} \text{dom}(f_n) = (0, 1]$ and $\text{ran}(f) = \bigcup_{n=1}^{\infty} \text{ran}(f_n)$

$= (0, 1)$. It is easy to check that f is a bijection. (*Of course, f is not continuous, but that is not required in the definition of set equivalence.*)

We can extend the definition of f to prove that $[0, 1] \sim [0, 1)$. Specifically, use f to define a function $g : [0, 1] \to [0, 1)$ by

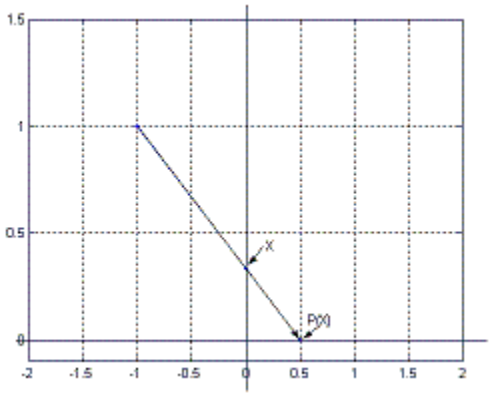
$$g(x) = \begin{cases} f(x) & \text{if } x \in (0, 1] \\ 0 & \text{if } x = 0 \end{cases}$$

Since f is a bijection, so is g.

It is very simple to show that $[0, 1) \sim (0, 1]$, that $(-\infty, 0] \sim [0, \infty)$, and that $(-\infty, 0) \sim (0, \infty)$: just use the maps $h(x) = 1 - x$ and $k(x) = -x$.

The function ln : $(0, \infty) \to \mathbb{R}$ proves that $(0, \infty) \sim \mathbb{R}$.

We can use a "projection" to show that $[0, 1)$ is equivalent to $[0, \infty)$. Imagine a ray of light that emanates from $(-1, 1)$, passes through the point x on the y-axis ($0 \leq x < 1$) and then hits the x-axis at a point that we call $p(x)$. (*See the following illustration*). Then $p : [0, 1) \to [0, \infty)$ is a bijection.



Those who prefer formulas can check that a formula for $p : [0, 1) \to [0, \infty)$ is $p(x) = \frac{x}{1-x}$.

The few remaining equivalences such as $(a, \infty) \sim (0, \infty)$ are left for the reader to prove. •

At this point, we might ask "Are all uncountable sets are equivalent?"

Example 9.2 The set $\mathbb{R}^{\mathbb{R}}$ is not equivalent to $\mathbb{R}$. To see this, we use an argument whose flavor is similar to the "diagonal" argument Cantor used to prove that $(0, 1)$ is uncountable.

Consider any function $\phi : \mathbb{R} \to \mathbb{R}^{\mathbb{R}}$. We will show that ϕ cannot be onto, and therefore no bijection can exist between $\mathbb{R}$ and $\mathbb{R}^{\mathbb{R}}$. If $x \in \mathbb{R}$, then $\phi(x) \in \mathbb{R}^{\mathbb{R}}$. Note that $\phi(x)$ is not a number: $\phi(x)$ is a function from $\mathbb{R} \to \mathbb{R}$; to emphasize this, we will temporarily write $\phi(x) = \phi_x$. Then for $y \in \mathbb{R}$, it makes sense to evaluate ϕ_x at a point $y \in \mathbb{R}$ to get a number $\phi_x(y) \in \mathbb{R}$.

Now look at the function $f \in \mathbb{R}^{\mathbb{R}}$ defined by

$$f(x) = \begin{cases} 1 & \text{if } \phi_x(x) = 0 \\ 0 & \text{if } \phi_x(x) \neq 0 \end{cases}$$

We claim that for each $x \in \mathbb{R}$, $\phi(x) = \phi_x \neq f$. This is because the two functions ϕ_x and f have different values at the point $x : \phi_x(x) \neq f(x)$ because of the way f was defined. Therefore $f \notin \text{ran}(\phi)$, so ϕ is not onto.

Therefore no subset of $\mathbb{R}$ can be equivalent to $\mathbb{R}^{\mathbb{R}}$. To see this, suppose $A \subseteq \mathbb{R}$ and that $\psi : A \to \mathbb{R}^{\mathbb{R}}$. If ψ were onto, we could use ψ to create an onto map ϕ from $\mathbb{R}$ to $\mathbb{R}^{\mathbb{R}}$ (which is impossible!) as follows: pick any function $f \in \mathbb{R}^{\mathbb{R}}$ and define

$$\phi(x) = \begin{cases} \psi(x) & \text{if } x \in A \\ f & \text{if } x \in \mathbb{R} - A. \end{cases}$$

Therefore $\mathbb{R}^{\mathbb{R}}$ is an uncountable set which is not equivalent to any subset of $\mathbb{R}$. In particular, $\mathbb{R}^{\mathbb{R}}$ is not equivalent to $\mathbb{N}$. Intuitively, $\mathbb{R}^{\mathbb{R}}$ is "bigger" than any subset of $\mathbb{R}$.

The fact that not all uncountable sets are equivalent leads to an interesting question. We have already seen (*see Example 7.6(2)*) that every subset of $\mathbb{N}$ is either finite or equivalent to the whole set $\mathbb{N}$— there are no "intermediate" sizes for subsets of $\mathbb{N}$. How about for $\mathbb{R}$? Could there be a subset of $\mathbb{R}$ which is uncountable but not equivalent to $\mathbb{R}$? Intuitively, such an infinite set would be "bigger than $\mathbb{N}$ but smaller than $\mathbb{R}$." It turns out that this question is "undecidable" ! What does that mean?

A conjecture of Cantor called the continuum hypothesis (or CH, for short), says that the answer to the question is "no"— that is, CH states that every subset of $\mathbb{R}$ is either countable or equivalent to $\mathbb{R}$. (*"Continuum" is an old word for the real number line.*) Cantor was unable to prove his conjecture.
In an address to the International Congress of Mathematicians in 1900, David Hilbert presented a list of research problems he considered most important for mathematicians to solve in the new century; the first problem on the list was CH.

Now if one believes (philosophically) that the set of real numbers actually exists "somewhere out there" (in some Platonic sense, or, say, in the mind of God) then any proposition about these real numbers, such as CH, must be either true or false – we simply have to figure out which. But to prove such a proposition mathematically requires an argument based on some set of assumptions (axioms) and theorems in terms of which we have made our mathematical definition of $\mathbb{R}$. Since $\mathbb{R}$ is defined mathematically in terms of set theory, a proof of CH (or of its negation) needs to be a proof derived from the axioms of set theory.

The usual collection of axioms for set theory – we have seen some of these axioms – is called the Zermelo-Fraenkel system (ZF). If the Axiom of Choice is added for good measure, the axiom system is referred to as ZFC. The fact is that the standard axioms for set theory (ZFC) are not sufficient to prove either CH or its negation. Since CH cannot be proven from ZFC, we say CH is independent of ZFC. But, since the denial of CH also cannot be proven from these axioms, we say that CH is also consistent with ZFC. Taken together, these statements say that CH is undecidable in ZFC, so one can add either CH or $\sim$ CH as an additional axiom to ZFC without fear of introducing a contradiction. These facts were established by Kurt Gödel (1906-1978) who showed in 1939 that ZFC could not

prove the denial of CH, and Paul J. Cohen (1934-2007) who showed in 1963 that ZFC could not prove CH.

> *"In the early 1960's, a brash, young and extremely brilliant Fourier analyst named Paul J. Cohen (people who knew him in high school assure me he was always brash and brilliant) chatted with a group of colleagues at Stanford about whether he would become more famous by solving a certain Hilbert problem or by proving that CH is independent of ZFC. This (informal) committee decided that the latter problem was the ticket. (To be fair, Cohen had been interested in logic and recursive functions for several years; he may have conducted this seance just for fun.) Cohen went off and learned the necessary logic and, in less than a year, had proved the independence. This is certainly one of the most amazing intellectual achievements of the twentieth century, and Cohen was awarded the Fields Medal for the work. But there is more.*
>
> *Proof in hand, Cohen flew off to Princeton to the Institute for Advanced Study to have his result checked by Kurt Gödel. Gödel was naturally skeptical, as Cohen was not the first person to claim to have solved the problem; and Cohen was not even a logician! Gödel was also at this time beginning his phobic period. (Toward the end of his life, Gödel became convinced that he was being poisoned, and he ended up starving himself to death.) When Cohen went to Gödel's home and knocked on the door, it was opened six inches and a hoary hand snatched the manuscript and slammed the door. Perplexed, Cohen departed. However, two days later, Cohen received an invitation for tea at Gödel's home. His proof was correct: the master had certified it."*
>
> *Mathematical Anecdotes*,
> Steven G. Krantz, Mathematical Intelligencer, v. 12, No. 4,
> 1990, pp. 35-36)

By a curious coincidence, Cohen was an analyst interested in Fourier series, and it was Cantor's work on Fourier series that led to his creation of set theory in the first place.

Therefore CH has a status, with respect to ZFC, like that of the parallel postulate in Euclidean geometry: the other axioms are not sufficient to prove or disprove it. To the other axioms for Euclidean geometry, you can either add the parallel postulate (to get Euclidean geometry) or, instead, add an axiom which denies the parallel postulate (to get some kind of non-Euclidean geometry). Likewise, to ZFC, one may consistently add CH as additional axiom, or get a "different set theory" by adding an axiom which denies CH.

Unlike the situation with AC, most mathematicians prefer to avoid assuming CH or its negation in a proof whenever possible. The reason is that CH (in contrast to AC) does not seem to command intuitive belief and, moreover, no central mathematical results depend on CH. When the use of CH use seems necessary, it is customary to call attention to the fact that it is being used.

To reiterate one last point: if you believe that, in some sense, the real numbers actually exist "out there" beyond ourselves, then you believe that the set $\mathbb{R}$, as defined mathematically within ZFC, is just a model of the "real" real number system. This model may be an inaccurate or incomplete fit to reality. You may therefore continue to believe that for the "real" real numbers, CH is either true or false as a matter of fact. People with this point of view would say that the undecidability of CH within ZFC simply reflects the inadequacy of the axiom system ZFC. Gödel himself seemed to feel that ZFC is inadequate, although for perhaps different reasons:

> *"I believe that ... one has good reason for suspecting that the role of the continuum problem in set theory will be to lead to the discovery of new axioms which will make it possible to disprove Cantor's conjecture."*
>
> Kurt Gödel, "What is Cantor's Continuum Problem?", *Philosophy of Mathematics*, ed. Benacerraf & Putnam, Prentice-Hall, 1964, p. 268

10. The Cantor-Schroeder-Bernstein Theorem

The Cantor-Schroeder-Bernstein Theorem (CSB for short) gives a way to prove that two sets are equivalent without actually constructing a bijection between them. It states that two sets are equivalent if each is equivalent to a subset of the other.

Theorem 10.2 (CSB) Suppose there exist one-to-one functions $f : A \to B$ and $g : B \to A$. Then $A \sim B$.

Proof We divide the set A into three subsets in the following way. For a point $x \in A$, we say "x has ≥ 0 ancestors" if $g^{-1}(x) \subseteq B$ (*this, of course, is always true, so every x has ≥ 0 ancestors*).
We say "x has $\geq$ 1 ancestor" if $g^{-1}(x) \neq \emptyset$, that "x has ≥ 2 ancestors" if $f^{-1}(g^{-1}(x)) \neq \emptyset$, and so on. In general, "x has $\geq n$ ancestors" if the inverse image set resulting from the alternating application of first g^{-1}, then f^{-1}, then g^{-1}, ..., n times produces a nonempty set. We say "x has exactly n ancestors" if x has $\geq n$ ancestors but x does not have $\geq n+1$ ancestors. We say "x has infinitely many ancestors" if x has $\geq n$ ancestors for every $n \in \mathbb{N}$. Let

$$A_E = \{x \in A : x \text{ has an even number of ancestors}\},$$
$$A_O = \{x \in A : x \text{ has an odd number of ancestors}\}, \quad \text{and}$$
$$A_I = \{x \in A : x \text{ has an infinite number of ancestors}\}$$

Clearly, these sets are disjoint and $A = A_E \cup A_O \cup A_I$. Define ancestors for a point $y \in B$ in a similar way, beginning the definition with an application of f^{-1} rather than g^{-1} and divide B into the sets B_E, B_O, and B_I. The maps $f \mid A_E : A_E \to B_O$ and $f \mid A_I : A_I \to B_I$ are bijections (why?).

The function $f \mid A_O$ maps A_O into B_E, but it may not be onto (why?). However, the map $g \mid B_E : B_E \to A_O$ is a bijection, so it has an inverse bijection $(g \mid B_E)^{-1} : A_O \to B_E$. We can now define a bijection h from A to B by piecing these maps together:

$$h = (f \mid A_E) \cup (f \mid A_I) \cup (g \mid B_E)^{-1}$$

More explicitly,

$$h(x) = \begin{cases} f(x) & \text{if } x \in A_E \cup A_I \\ (g|B_E)^{-1}(x) & \text{if } x \in A_O \end{cases} \qquad \bullet$$

Our previous examples of equivalent sets were simply examples: they were not used in the proof of the CSB Theorem. We can therefore, without circular reasoning, use the CSB Theorem to give easier proofs of many of those equivalences, as the first two examples below illustrate.

Examples 10.3

1) If $A \subseteq \mathbb{N}$, then A is finite or equivalent to $\mathbb{N}$. For if $A \subseteq \mathbb{N}$, we have $i : A \to \mathbb{N}$ by $i(x) = x$, and, if A is also infinite we have, by definition, a one-to-one map $f : \mathbb{N} \to A$. Therefore, $A \sim \mathbb{N}$.

2) $(0, 1) \sim [0, 1]$ since each set is equivalent to a subset of the other:

$(0, 1) \sim (0, 1) \subseteq [0, 1]$ and [0,1] is easily seen to be equivalent (using a linear map) to $\left[\frac{1}{4}, \frac{1}{2}\right] \subseteq (0, 1)$.

3) $(0, 1) \sim (0, 1)^2$. We have the one-to-one map $f(x) = (x, \frac{1}{2})$ mapping the interval into the square, and we can easily define a one-to-one map in the opposite direction as follows:

for $(x, y) \in (0, 1)^2$, write the binary expansions of x and y

$$x = 0.x_1x_2x_3...x_n...{}_{two} \quad \text{and} \quad y = 0.y_1y_2y_3...y_n...{}_{two}$$

choosing, in both cases, a binary expansion that does not end with an infinite string of 1's. Then define $g : (0, 1)^2 \to (0, 1)$ by "interlacing" the digits to create a base 10 decimal:

$$g(x, y) = 0.x_1y_1x_2y_2x_3y_3...x_ny_n...$$

The function g is one-to-one: suppose $(x, y) \neq (x', y')$. Then $g(x, y)$ and $g(x', y')$ have different decimal expansions. Neither decimal expansion ends in an infinite string of 9's, so $g(x, y)$ and $g(x', y')$ are different real numbers.

4) A simple exercise, left to the reader, is to show that if $A \sim B$ and $C \sim D$, then $A \times C \sim B \times D$. Then, because $(0, 1) \sim [0, 1] \sim [0, 1) \sim \mathbb{R}$, we can immediately write down such equivalences as $\mathbb{R} \sim (0, 1) \sim (0, 1)^2 \sim [0, 1]^2 \sim (0, 1) \times [0, 1) \sim \mathbb{R}^2$. (We intuitively think of $(0, 1)$ as 1-dimensional and $(0, 1)^2$ as 2-dimensional, so we see that an equivalence between sets doesn't necessarily preserve our intuitive idea of dimension.).

Since $\mathbb{R} \sim \mathbb{R}^2$, it follows that $\mathbb{R} \sim \mathbb{R} \times \mathbb{R} \sim \mathbb{R} \times \mathbb{R}^2 = \mathbb{R}^3$. Similarly, $\mathbb{R}^n \sim \mathbb{R}^m$ for any $m, n \in \mathbb{N}$.

Example 10.4 (A tangential comment) Although $\mathbb{R}^2 \sim \mathbb{R}$, there is no continuous bijection between them. In fact, if $f : \mathbb{R}^2 \to \mathbb{R}$ is continuous, then f cannot even be one-to-one. The proof is an exercise in the use of the Intermediate Value Theorem from calculus.

> Define a function ϕ of one variable by holding one variable in f constant: $\phi(x) = f(x, 0)$. Let $f(0, 0) = \phi(0) = a$ and $f(1, 0) = \phi(1) = b$. We can assume $a < b$ (if $a = b$, f is not one-to-one and we're done). Since f is continuous, so is ϕ. By the Intermediate Value Theorem, ϕ assumes all values between a and b; in particular, for some $c \in (0, 1)$, $\phi(c) = \frac{a+b}{2}$.
>
> For that c, define $\psi(y) = f(c, y)$. Because ψ is continuous at 0, $\psi(y)$ is close to $\psi(0) = \phi(c) = \frac{a+b}{2}$ for y close to 0. Specifically, we can force $a < \psi(y) < b$ for all y in some sufficiently short interval $(-\delta, \delta)$. In particular, then $\psi(\frac{\delta}{2}) \in (a, b)$.
>
> But we know that ϕ assumes all values between a and b, so $\psi(\frac{\delta}{2}) = \phi(z)$ for some z in (0,1). Therefore $f(c, \frac{\delta}{2}) = \psi(\frac{\delta}{2}) = \phi(z) = f(z, 0)$. Since $(c, \frac{\delta}{2}) \neq (z, 0)$, f is not one-to-one. •
>
> (*Note: we did not even use the full strength of the assumption that f is continuous. We needed only to know that ϕ and ψ were continuous – a weaker statement, as you should know from advanced calculus. This means there cannot even be a 1 – 1 map of the plane to the line which is continuous in each variable separately— a condition which is weaker than assuming f is continuous.*)

A similar argument shows that there cannot be a one-to-one continuous map $f: [0, 1]^2 \to [0, 1]$. Therefore, in particular, there cannot be a continuous bijection from $[0, 1]^2$ to $[0, 1]$.

However, in the "opposite" direction, it is possible to construct a continuous map $f: [0, 1] \to [0, 1]^2$ which is onto. Such a map is called a space-filling curve, and such a map is often constructed in an advanced calculus course.

Can there be a continuous onto map from $\mathbb{R}^2 \to \mathbb{R}$?

11. More About Subsets

We have already seen that there is more than one "size" of infinite set: some are countable, others are uncountable. Moreover, we have seen uncountable sets of two different sizes: $\mathbb{R}$ and $\mathbb{R}^{\mathbb{R}}$. In fact, there are infinitely many different sizes. We will see this using the power set operation $\mathcal{P}$. But first we will prove a result about subsets of countable sets that is frequently useful.

Theorem 11.1 The set $\mathcal{F}(A)$ of all finite subsets of a countable set A is countable.

Proof If A is finite, it has only finitely many subsets. So assume A is countable and infinite. For convenience, we assume A is the set of all prime numbers (*Be sure you understand why this is no loss of generality!*) Then each finite $B \subseteq A$ is a set of primes and we can define $f : \mathcal{F}(A) \to \mathbb{N}$ by

$$f(B) = \begin{cases} 1 & \text{if } B = \emptyset \\ \text{the product of the primes in } B & \text{if } B \neq \emptyset \end{cases}$$

By the Fundamental Theorem of Arithmetic, f is one-to-one; so $\mathcal{F}(A)$ is countable. •

Recall that Y^A denotes the set of all functions from A into Y. When $Y = \{0, 1\}$, we will also use the notation 2^A for the set $\{0, 1\}^A$. Thus, 2^A is the set of all "binary functions" with domain A. The following theorem gives us two important facts.

Theorem 11.2 For any set A,

1) $2^A \sim \mathcal{P}(A)$
2) A is not equivalent to $\mathcal{P}(A)$ (so $A \not\sim 2^A$)

Proof 1) For each $B \subseteq A$, define $\chi_B : A \to \{0, 1\}$ by $\chi_B(x) = \begin{cases} 1 & \text{if } x \in B \\ 0 & \text{if } x \notin B \end{cases}$.

The function χ_B is called the characteristic function of B.

Define $\phi : \mathcal{P}(A) \to 2^A$ by $\phi(B) = \chi_B$. The function ϕ pairs each subset of A with its characteristic function. The function ϕ is clearly one-to-one because different subsets have different characteristic functions. The function ϕ is also onto since every function $f : A \to \{0, 1\}$ is the characteristic function of a subset of A: $f = \chi_B = \phi(B)$, where $B = \{x \in A : f(x) = 1\}$. So $\mathcal{P}(A) \sim 2^A$.

2) Suppose $\psi : A \to \mathcal{P}(A)$. We will show that ψ cannot be onto. For each $x \in A$, $\psi(x)$ is a subset of A, so it makes sense to ask whether or not $x \in \psi(x)$. Let $B = \{x \in A : x \notin \psi(x)\}$.
Of course, $B \in \mathcal{P}(A)$.

Suppose $x \in A$. If $x \in \psi(x)$, then $x \notin B$, so $B \neq \psi(x)$. But if $x \notin \psi(x)$, then $x \in B$ so, again, $B \neq \psi(x)$. Either way, $\psi(x) \neq B$, so $B \notin \text{ran}(\psi)$ •

We use the notation $\mathcal{P}^2(A)$ for $\mathcal{P}(\mathcal{P}(A))$ and, more generally, $\mathcal{P}^n(A)$ for $\mathcal{P}(\mathcal{P}^{n-1}(A))$. The preceding theorem says that no set is equivalent to its power set, so $\mathcal{P}^{n-1}(A)$ is not equivalent to $\mathcal{P}^n(A)$. We can prove even a bit more.

Corollary 11.3 No two of the sets in the sequence A, $\mathcal{P}(A)$, $\mathcal{P}^2(A)$, ... , $\mathcal{P}^n(A)$, ... are equivalent.

Proof For any j, there is a one-to-one function $i_j : \mathcal{P}^j(A) \to \mathcal{P}^{j+1}(A)$ given by $i_j(a) = \{a\}$. Now suppose there is a bijection $f : \mathcal{P}^j(A) \to \mathcal{P}^{j+k}(A)$ for some j, k. Then we would have the following maps:

$$\mathcal{P}^j(A) \xrightarrow{i_j} \mathcal{P}^{j+1}(A) \xrightarrow{i_{j+1}} \text{.....} \xrightarrow{i_{j+k-2}} \mathcal{P}^{j+k-1}(A) \xrightarrow{i_{j+k-1}} \mathcal{P}^{j+k}(A)$$
$$\big|\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\uparrow$$
$$f$$

Then $i_{j+k-1} : \mathcal{P}^{j+k-1}(A) \to \mathcal{P}^{j+k}(A)$ and $i_{j+k-2} \circ ... \circ i_j \circ f^{-1} : \mathcal{P}^{j+k}(A) \to \mathcal{P}^{j+k-1}(A)$ are both one-to-one. But then the CSB Theorem would tell us that $\mathcal{P}^{j+k-1}(A) \sim \mathcal{P}^{j+k}(A)$, which is impossible by Theorem 11.2(2). •

Therefore, by repeated applications of the power set operation, we can generate an infinite sequence of sets no two of which are equivalent – for example, $\mathbb{N}$, $P(\mathbb{N})$, $\mathcal{P}^2(\mathbb{N})$, ... , $\mathcal{P}^n(\mathbb{N})$, ... In the "exponential" notation of Theorem 11.2(1), we can also write this sequence as $\mathbb{N}$, $2^{\mathbb{N}}$, $2^{2^{\mathbb{N}}}$, ... Thus we have infinitely many different "sizes" of infinite set.

Examples 11.4

1) $2^{\mathbb{N}} = \{0,1\}^{\mathbb{N}} \sim \mathcal{P}(\mathbb{N}) \not\sim \mathbb{N}$, so we conclude that the set of all binary sequences is uncountable.

2) Because $\mathcal{P}(\mathbb{N})$ is uncountable and because $\mathbb{N}$ has only countably many finite subsets (Theorem 11.1), we conclude that $\mathbb{N}$ has uncountably many infinite subsets.

Therefore every infinite set has uncountably many infinite subsets (*explain the details!*).

3) Using the CSB Theorem, we can state another paradox similar to Russell's Paradox.

> Let $B = \{\{b\} : b \text{ is a set}\}$, so B is the "set of all one-element sets." Then the maps $\mathcal{P}(B) \to B$ and $B \to \mathcal{P}(B)$ given (in both directions!) by $v \to \{v\}$ are one-to-one. Then the CSB Theorem gives that $B \sim \mathcal{P}(B)$.

What's wrong? (*The paradox is dealt with in the same way as Russell's Paradox.*)

Since $2^{\mathbb{N}} \not\sim \mathbb{N}$, we can ask whether $2^{\mathbb{N}}$ is equivalent to some other familiar set. The next theorem answers this question.

Theorem 11.5 $2^{\mathbb{N}} \sim \mathbb{R}$

Proof We use the CSB Theorem. If $f \in 2^{\mathbb{N}}$, then $f : \mathbb{N} \to \{0, 1\}$. Define $\phi : 2^{\mathbb{N}} \to \mathbb{R}$ by $\phi(f) = \sum_{n=1}^{\infty} \frac{f(n)}{10^n} = \sum_{n=1}^{\infty} \frac{a_n}{10^n}$ where $a_n = f(n)$. In other words, $\phi(f)$ is the real number whose decimal expansion is $0.a_1a_2$ ... a_n ... . If $f \neq g \in 2^{\mathbb{N}}$, then $\phi(f)$ and $\phi(g)$ have different decimal expansions, and neither expansion ends in a string of 9's (*each* a_n *is either* 0 *or* 1). Therefore $\phi(f) \neq \phi(g)$, so ϕ is one-to-one.

To get a one-to-one function from $\mathbb{R}$ into $2^{\mathbb{N}}$, we begin by defining a function $\psi : \mathbb{R} \to \mathcal{P}(\mathbb{N})$. List the rational numbers in a sequence $\mathbb{Q} = \{q_1, q_2, q_3, ..., q_n, ... \}$ and, for each $x \in \mathbb{R}$, let $\psi(x) = \{n \in \mathbb{N}: \ q_n < x\}$. Then $\psi(x) \in \mathcal{P}(\mathbb{N})$.

Since two different real numbers x and y always have a rational between them, the set of rationals $< x$ is different from the set of rationals $< y$ – in other words, $\psi(x) \neq \psi(y)$, so ψ is one-to-one. Then we can compose ψ with a bijection $g : \mathcal{P}(\mathbb{N}) \to 2^{\mathbb{N}}$ to get a one-to-one function $g \circ \psi : \mathbb{R} \to 2^{\mathbb{N}}$.

By the CSB Theorem, $2^{\mathbb{N}} \sim \mathbb{R}$. •

In fact, it is fairly easy to generalize Theorem 11.5 replacing $2^{\mathbb{N}}$ *with* $k^{\mathbb{N}}$*, where* $k^{\mathbb{N}}$ *is short for* $\{0, 1, \ ... \ , k-1\}^{\mathbb{N}}$.

12. Cardinal Numbers

To each finite set A we can assign, in principle, a number called the cardinal number or cardinality of the set). It answers the question: "How many elements does the set have?" The symbol $|A|$ represents the cardinal number of A. For example, $|\emptyset| = 0$, $|\{0\}| = 1$, and $|\{0, 1\}| = 2$. In practice, of course, this might be difficult. For example, if $A = \{p \in \mathbb{N} : \ p$ is prime and $p < 10^{100}\}$, then $|A| = ?$
Of course, there must be a correct value for $|A|$, but actually finding it would be hard. (*You could make a rough estimate using the Prime Number Theorem – see Example 5.1(8), p. 16:* $\pi(10^{100}) \approx \frac{10^{100}}{\ln(10^{100})} \approx 4.34 \times 10^{97}$.)

We have already stated informally that nonequivalent infinite sets have "different sizes". To make this more precise, we assume that to each set A there is associated an "object" denoted $|A\ |$ and called the cardinal number of A (or cardinal of A for short), and that this is done in such a way that $|A| = |B|$ if and only if $A \sim B$.

> *Note: How to justify this assumption precisely doesn't matter right now. That question belongs in a more advanced set theory course. It turns out that* $|A|$ *can be precisely defined in ZFC. Like everything in mathematics,* $|\ A\ |$ *is itself a certain set and, of course,* $|A|$ *is defined in terms of the given set* A*. For these notes, it is enough to know that two sets have the same cardinal number if and only if the sets are equivalent.*

There are standard symbols for the cardinal numbers associated to certain everyday sets:

$|\,\emptyset\,| \quad = 0$

$|\,\{0\}\,| \quad = 1$ (Of course, $|\,\{a\}\,| = 1$ for any a, because $\{0\} \sim \{a\}$)

$|\{0,1\}| \; = 2$ (Of course, $|\,\{a,b\}\,| = 2$ for any $a \neq b$)

$\vdots$

$|\,\mathbb{N}\,| \quad = \aleph_0$ (So $\aleph_0$ is the cardinal number for every countable infinite set; for example, $|\mathbb{N} \times \mathbb{N} \times \mathbb{N}| = |\mathbb{Q}| = ... = \aleph_0$)

$|\,\mathbb{R}\,| \quad = c$

The symbols $\aleph$ and c were chosen by Cantor in his original work. The symbol $\aleph$ is "aleph," the first letter of the Hebrew alphabet (but Cantor was Lutheran). We read $\aleph_0$ as "aleph-zero" (or the more British "aleph-nought" or "aleph-null.")

Cantor chose "*c*" for $|\mathbb{R}|$ since c is the first letter of the word "continuum" (*an old word for the real line*). Because we know that the following sets are equivalent, we can write, for example, that $|[0,1]| = |(0,1)| = |(0,1)^2| = |\mathbb{R}| = |\mathbb{R}^2| = |\mathbb{R}^3| = \; ... \; = c$.

What is $|\,\mathbb{P}\,|$? Why? (*You <u>can</u> give a definite answer for this, but be careful: you can't assume that an uncountable subset of* $\mathbb{R}$ *must be equivalent to* $\mathbb{R}$. *See the discussion of CH on pp. 40-42.*)

13. Ordering the Cardinals

We have talked informally about some infinite sets being "bigger" than others. We can make this precise by defining an order " $\leq$ " between cardinal numbers. Throughout this section, M, N, P are sets with $|\,M\,| = m$, $|\,N\,| = n$, and $|\,P\,| = p$ (N is not necessarily the set natural numbers, $\mathbb{N}$). We say that the sets $M, N,$ and P <u>represent</u> the cardinal numbers $m, n,$ and p.

Definition 13.1 1) $m \leq n$ means that M is equivalent to a subset of N. (*We also write* $m \leq n$ *as* $n \geq m$.)

2) $m < n$ (or $n > m$) means that $m \leq n$ but $m \neq n$. (*We also write* $m < n$ *as* $n > m$.)

Thus $m < n$ means that M is equivalent to a subset of N but M is not equivalent to N. According to the CSB Theorem, this is the same as saying that M is equivalent to a subset of N but N is not equivalent to a subset of M.

There is a detail to check. The relation $m \leq n$ is defined <u>using</u> the sets M and N. Another person might choose different sets, say $m = |M'|$ and $n = |N'|$ to represent the cardinal numbers. Would that person necessarily come to the same conclusion that $m \leq n$? We need to check that our definition is <u>independent of which representing sets are chosen</u> or, in other words, that $\leq$ has been <u>well-defined</u>. This is easy to do.

> Suppose M' and N' also represent m and n. If M is equivalent to a subset of N, then there are bijections f and g and a one-to-one map h as pictured below:

$$\begin{array}{ccc} & \overset{f}{} & \\ M & \rightarrow & M' \\ & & \\ h \downarrow & & \downarrow k \\ & \overset{g}{} & \\ N & \rightarrow & N' \end{array}$$

Then $k = g \circ h \circ f^{-1}$ is a one-to-one map $k : M' \rightarrow N'$, so M' is equivalent to a subset of N'. Therefore the question "Is $m \leq n$?" does not depend on which representing sets we use – that is, the relation $\leq$ is well-defined.

Theorem 13.2 For any cardinal numbers $m, n,$ and p :

1) $m \leq m$
2) $m \leq n$ and $n \leq p$ implies $m \leq p$
3) $m \leq n$ and $n \leq m$ implies $m = n$
4) $m \leq n$ and $n < p$ implies $m < p$
5) at most one of the relations $m < n$, $m = n$, and $m > n$ holds
6*) at least one of the relations $m < n$, $m = n$, and $m > n$ holds.

Proof The proofs of 1) and 2) are obvious.

For 3), notice that we are given that each of M and N is equivalent to a subset of the other. Then $M \sim N$ (by the CSB Theorem), so $m = n$.

4) If $n < p$, then $n \leq p$ so, by part 2), $m \leq p$. If $p = m$, then $p \leq n$ and therefore $n = p$ by part 3). But this contradicts the hypothesis that $n < p$. Therefore $p \neq m$, so $m < p$.

5) By definition of $<$, $m = n$ excludes $m < n$ and $n < m$. And if $m < n$ and $n < m$ were both true, then $m \leq n$ and $n \leq m$, so $m = n$, which is impossible.

6*) The proof is postponed. •

It seems like the proof of part 6) of the theorem shouldn't be difficult. Informally, you simply pair an element of M with an element of N, and keep repeating this process until either a bijection is created between M and N or until one of the sets has no remaining elements. If one of the sets is used up before the other, then it has the smaller cardinal. In fact, this works for finite sets, but for infinite sets it is hard to make precise what "keep repeating this process until ..." means. To make the argument precise, the Axiom of Choice has to be used in some form. We could develop the machinery to complete the proof here, but it would digress too much from the main ideas. So for use in our, examples, we'll simply assume part 6*) for now.*

From parts 5) and 6*) of the theorem, we immediately get the following corollary.

Corollary 13.3 For any two cardinals m and n, exactly one of the relations $m < n$, $m = n$, or $m > n$ is true.

Examples 13.4

1) If $k = |K|$ is a finite cardinal, then K is equivalent to a subset of $\mathbb{N}$ but K is not equivalent to $\mathbb{N}$. And if m is an infinite cardinal number, there is a one-to-one function $f : \mathbb{N} \to M$. Therefore $k < \aleph_0 \leq m$, so $\aleph_0$ is the <u>smallest</u> infinite cardinal number.

2) $\aleph_0 < c < \mid \mathbb{R}^{\mathbb{R}} \mid$ (*Explain*)

14. The Arithmetic of Cardinal Numbers

We want to define exponentiation, addition and multiplication for cardinal numbers. Of course, for finite cardinals, these operations will coincide with the ordinary arithmetic operations in $\mathbb{N}$.

We begin with exponentiation. As in the previous section, $m, n,$ and p will denote cardinals represented by sets $M, N,$ and P.

Definition 14.1 $n^m = \mid N^M \mid$

Thus, n^m is the number of functions from the set M into the set N. As with the definition of the order relation $\leq$, we must check that exponentiation is well-defined. (That is, if one person calculates $c^{\aleph_0}$ using $|\mathbb{R}^{\mathbb{N}}|$ and another person uses $|\,(0,1)^{\mathbb{Q}}|$, will they get the same answer?)

If $m = |M| = |M'|$ and $n = |N| = |N'|$, we must show that $N^M \sim (N')^{M'}$, that is, we must produce a bijection $\phi : N^M \to (N')^{M'}$.

By hypothesis, we have bijections $f : M \to M'$ and $g : N \to N'$. For $h \in N^M$, define a function $\phi(h) = g \circ h \circ f^{-1} \in (N')^{M'}$.

<u>ϕ is one-to-one</u>: Suppose $h \neq k \in N^M$. Then for some $m \in M$, $h(m) \neq k(m)$. Let $m' = f(m)$. Then $\phi(h)(m') = g(h(f^{-1}(m'))) = g(h(m))$ and
$\phi(k)(m') = g(k(f^{-1}(m'))) = g(k(m))$.
But $h(m) \neq k(m)$ and g is one-to-one so $\phi(h)(m') = g(h(m)) \neq g(k(m)) = \phi(k)(m')$. Therefore the functions $\phi(h)$ and $\phi(k)$ have different values at m, so $\phi(h) \neq \phi(k)$. Therefore ϕ is one-to-one.

<u>ϕ is onto</u>: Exercise.

Examples 14.2

1) $\mid [0,1]^{(0,1)} \mid = \mid \mathbb{R}^{\mathbb{R}} \mid = c^c$

2) $2^{\aleph_0} = |2^{\mathbb{N}}| = c$, since $2^{\mathbb{N}} \sim \mathbb{R}$. Thus, there are c different binary sequences. As remarked earlier, it is not hard to generalize this to show that $k^{\aleph_0} = c$ is true for any integer $k > 1$.

Students sometimes confuse the <u>fact</u> that $2^{\aleph_0} = c$ with the continuum hypothesis. The continuum hypothesis states that there is no cardinal m such that $\aleph_0 < m < 2^{\aleph_0} = c$.

In other words, CH states that $2^{\aleph_0}$ is the immediate successor (in terms of size) of the cardinal number $\aleph_0$. Here is still another way of putting it: let us <u>define</u> $\aleph_1$ to be the immediate

successor of $\aleph_0$ (*assuming for now that an immediate successor must exist*). *Then CH simply states that* $c = \aleph_1$. *To go a little further, the generalized continuum hypothesis* (*GCH for short*) *states that for any infinite cardinal* m, *the "immediate successor" is* 2^m.

(*Since ZFC cannot prove CH, certainly ZFC cannot prove the stronger statement GCH. In fact, GCH is undecidable in ZFC.*)

3) For any set M, M is equivalent to $\{\{m\} : m \in M\} \subseteq \mathcal{P}(M)$. Therefore $m \leq 2^m$. But $m \neq 2^m$ because M is not equivalent to 2^M. Therefore $m < 2^m$ for all cardinals m. It follows that $m < 2^m < 2^{2^m} < 2^{2^{2^m}} <$ ... $<$ ... is an infinite increasing sequence of infinite cardinals. In particular, for $m = \aleph_0$, we have $\aleph_0 < 2^{\aleph_0} = c < 2^{2^{\aleph_0}} < \ldots$

We now define multiplication and addition of cardinal numbers.

Definition 14.3 Suppose $m = |M|$and $n = |N|$,

1) $m \cdot n$ (or simply mn) means $|M \times N|$
2) $m + n$ means $|M \cup N|$, provided that M and N are disjoint.

Part 2) of the definition requires that to add m *and* n, *we choose disjoint representing sets* M *and* N. *This is always possible because if* $M \cap N \neq \emptyset$, *we can replace* M *and* N *by the equivalent disjoint sets* $M \times \{0\}$ *and* $N \times \{1\}$.

You should check that multiplication and addition are well-defined – that is, the operations are independent of the sets chosen to represent the cardinals m and n.

Theorem 14.4 For cardinal numbers $m, n, p,$ and q :

1) $m + n = n + m$ 1′) $mn = nm$
2) $(m + n) + p = m + (n + p)$ 2′) $m(np) = (mn)p$
3) $p(m + n) = pm + pn$

If $m \leq n$ and $p \leq q$, then

4) $m + p \leq n + q$ 4′) $mp \leq nq$

Proof The proofs are all simple. For example, we prove 4′).

Suppose m, n, p, q are represented by the sets M, N, P, Q. Since $m \leq n$ and $p \leq q$, there are one-to-one functions $f : M \to N$ and $g : P \to Q$. Define $h : M \times P \to N \times Q$ by $h(m, p) = (f(m), g(p))$. Then h is one-to-one, so $mp \leq nq$. •

Addition Examples 14.5

1) For any m, $m + 0 = m$ because $M \cup \emptyset \sim M$.

2) For finite n, $n + \aleph_0 = \aleph_0 + \aleph_0 = \aleph_0$ because the union of two countable sets is countable.

3) For finite n, $n + c = \aleph_0 + c = c + c = c$.

$c + c = c$ is true because $(-\infty, 0) \cup [0, \infty) = \mathbb{R}$. Then write the inequalities

$$0 \leq n \leq \aleph_0 \leq c \quad \text{and}$$
$$c \leq c \leq c \leq c$$

Add the inequalities and apply Theorem 14.4(4) to get

$$c \leq n + c \leq \aleph_0 + c \leq c + c.$$

But $c + c = c$, so we conclude that $c = n + c = \aleph_0 + c = c + c$.

4) If m is infinite, then $m + \aleph_0 = m$.

To see this, pick M so that $m = |M|$ and $M \cap \mathbb{N} = \emptyset$. M is infinite so there is a one-to-one map $f : \mathbb{N} \to M$ and we can write $M = f[\mathbb{N}] \cup (M - f[\mathbb{N}])$. Because $f[\mathbb{N}]$ is countable, so is $f[\mathbb{N}] \cup \mathbb{N}$. Therefore we have a bijection $g : f[\mathbb{N}] \cup \mathbb{N} \to f[\mathbb{N}]$. We can then define a bijection $h : M \cup \mathbb{N} \to M$ by

$$h(x) = \begin{cases} x & \text{if } x \in M - f[\mathbb{N}] \\ g(x) & \text{if } x \in \mathbb{N} \cup f[\mathbb{N}] \end{cases}$$

The following two facts about addition are also true, but their proofs require more complicated arguments that involve AC. We omit the proofs and, for now, simply assume 5) and 6*).*

5*) If m or n is infinite, then $m + n = \max\{m, n\}$: that is, a sum involving an infinite cardinal number equals the larger of the two cardinals.

6*) If $m < n$ and $p < q$, then $m + p < n + q$.

This result deserves a word of caution. For infinite cardinals, it is <u>not</u> true in general that if $m < n$ and $p \leq q$, then $m + p < n + q$! For example, $\aleph_0 < c$ and $c \leq c$ but $\aleph_0 + c = c + c \ (= c)$.

7) The World's Longest Song: "$\aleph_0$ Bottles of Beer on the Wall" (*It fits the music better with the British pronunciation "Aleph-nought."*)

Multiplication Examples 14.6

1) For any m, $m \cdot 0 = 0$ and $m \cdot 1 = m$. These equations are true because $M \times \emptyset \sim \emptyset$

and $M \times \{a\} \sim M$.

2) $\aleph_0^2 = \aleph_0$.
If we view $\aleph_0^2$ as shorthand for $\aleph_0 \cdot \aleph_0$, the equation is true because $\mathbb{N} \times \mathbb{N} \sim \mathbb{N}$. But an alert reader might point out that we could also interpret $\aleph_0^2$ as an exponentiation, that is, as $|\,\mathbb{N}^{\{0,1\}}|$. But this gives to the same result because $\mathbb{N}^{\{0,1\}} \sim \mathbb{N} \times \mathbb{N}$. (*We can establish a bijection by pairing each* $f \in \mathbb{N}^{\{0,1\}}$ *with the pair* $(f(0), f(1)) \in \mathbb{N} \times \mathbb{N}$.)

3) If n is finite and $n > 0$, then $n \cdot \aleph_0 = \aleph_0$.
To see this, begin with the inequalities

$$\begin{array}{l} 1 \ \ \leq n \ \ \leq \aleph_0 \\ \aleph_0 \leq \aleph_0 \leq \aleph_0 \end{array}$$

Multiplying and applying Theorem 14.4$(4')$ gives

$$\aleph_0 \leq n \cdot \aleph_0 \leq \aleph_0^2$$

Since $\aleph_0 = \aleph_0^2$, we get that $\aleph_0 = n \cdot \aleph_0 = \aleph_0^2$.

4) $c^2 = c$ because $(0,1)^2 \sim (0,1)$,

5) If n is finite and $n > 0$, then $nc = \aleph_0 \cdot c = c^2 = c$.
Begin with the inequalities

$$\begin{array}{l} 1 \leq \ n \ \leq \aleph_0 \ \leq c \\ c \leq \ \ c \ \ \leq \ c \ \ \leq c \end{array} \quad \text{and multiply to get}$$

$$c \leq nc \leq \aleph_0 \cdot c \leq c^2$$

Since $c = c^2$, we conclude that $c = nc = \aleph_0 \cdot c = c^2$

One additional fact about multiplication of cardinals will be assumed, for now, without proof:

6*) If m is infinite and $n \neq 0$, then $mn = \max\{m, n\}$. In particular, for an infinite cardinal m, we have $m^2 = m$.

In algebraic systems (such as $\mathbb{Z}$, $\mathbb{Q}$ or $\mathbb{R}$) where subtraction is defined, the definition of subtraction is always given in terms of addition: $a - b = c$ is <u>defined</u> to mean $a = b + c$. This shows why <u>subtraction cannot be sensibly defined for infinite cardinal numbers</u>: should we say $\aleph_0 - \aleph_0 = 0$ because $\aleph_0 = \aleph_0 + 0$? or $\aleph_0 - \aleph_0 = 1$ because $\aleph_0 = \aleph_0 + 1$? or $\aleph_0 - \aleph_0 = \aleph_0$ because $\aleph_0 = \aleph_0 + \aleph_0$? Similarly, division is usually defined in terms of multiplication: $\frac{a}{b} = c$ means $a = bc$. Think about why <u>there is also no sensible definition for division involving infinite cardinal numbers</u>.

The following theorem is an excellent check on whether you understand "function notation."

Theorem 14.7 For cardinals $m, n,$ and p : $(m^n)^p = m^{np}$

Proof We want to define a bijection $\phi : (M^N)^P \to M^{(N\times P)}$.

If $f \in (M^N)^P$, then f is a function $P \to M^N$. If $p \in P$, then $f(p) \in M^N$, so $f(p)$ is a function $N \to M$. Therefore, for $n \in N$, $f(p)(n)$ makes sense: it is an element of M. So we can define a function $\phi(f) \in M^{N\times P}$ by the rule $\phi(f)(n,p) = f(p)(n)$.

ϕ is <u>onto</u>: if $g \in M^{N\times P}$, let $f : P \to M^N$ be the function defined by $f(p)(n) = g(n,p)$. Then $f \in (M^N)^P$ and $\phi(f) = g$ because for any pair (n,p), we have $\phi(f)(n,p) = f(p)(n)$
$= g(n,p)$.

ϕ is <u>one-to-one</u>: if $f, g \in (M^N)^P$ and $f \neq g$, then for some $p \in P$, $f(p) \neq g(p)$. Because $f(p)$ and $g(p)$ are different functions $N \to M$, there must be some $n \in N$ for which $f(p)(n) \neq g(p)(n)$. But this says that $\phi(f)(n,p) \neq \phi(g)(n,p)$, so $\phi(f) \neq \phi(g)$. •

Examples 14.8

1) $c^{\aleph_0} = (2^{\aleph_0})^{\aleph_0} = 2^{\aleph_0\cdot\aleph_0} = 2^{\aleph_0} = c$

2) $c = 2^{\aleph_0} \leq \aleph_0^{\aleph_0} \leq c^{\aleph_0} = c$, so $\aleph_0^{\aleph_0} = c$

3) $|\mathbb{R}^{\mathbb{R}}| = c^c = (2^{\aleph_0})^c = 2^{\aleph_0\cdot c} = 2^c$

<u>Caution</u>: We now know that $\aleph_0 < \aleph_0^{\aleph_0} = c$, but that $c^{\aleph_0} = c$. So we might be tempted to conjecture that if $m > \aleph_0$, then $m^{\aleph_0} = m$. This is <u>false</u>.

In fact, for every cardinal number k, it is possible to find an $m > k$ for which $m^{\aleph_0} > m$ and also to find some other cardinal $m > k$ for which $m^{\aleph_0} = m$. The proof of this is a little too complicated to look at now.

Exercises

E46. Prove or disprove: Let $\mathbb{A}$ be the set of algebraic numbers. Then every open interval in $\mathbb{R}$ contains a point of $\mathbb{R} - \mathbb{A}$.

E47. a) Suppose S is a countable subset of $\mathbb{R}$. Prove that there exists a fixed real number, c such that $s + c$ is transcendental for every $s \in$ S.

b) For any set $B \subseteq \mathbb{R}$ and $\alpha \in \mathbb{R}$, we write $B + \alpha$ for the set $\{b + \alpha : b \in B\}$. Find a set $A \subseteq \mathbb{R}$ for which $|A| = c$ and $(\mathbb{Q} + \alpha) \cap (\mathbb{Q} + \beta) = \emptyset$ for all $\alpha \neq \beta \in A$.

E48. You and I play the following infinite game. We take turns (you go first) picking "0" or "1" and use our choices as the consecutive digits of a binary decimal which, when completed, represents a real number in the interval $[0, 1]$. You win if this number is transcendental; I win if it is algebraic. Explain how to make your choices so that you are guaranteed to win, no matter what choices I make. (*Hint: Consider the binary expansions of the algebraic numbers in* $[0, 1]$. *Look at the proof that* $(0, 1)$ *is uncountable.*)

E49. Find the cardinal number of each of the following sets:

a) the set of all convergent sequences of real numbers

b) the set of all straight lines ℓ in the plane for which $|\ell \cap (\mathbb{Q} \times \mathbb{Q})| \geq 2$

c) the set of all sequences $f : \mathbb{N} \to \mathbb{N}$ that are eventually constant (*Note: f is <u>eventually constant</u> if there are natural numbers l and $m \in \mathbb{N}$ such that $f(n) = l$ for all $n \geq m$.*)

d) the set of all differentiable functions $f : \mathbb{R} \to \mathbb{R}$

e) the set of all geometric progressions in $\mathbb{R}$ (*A sequence (x_n) in $\mathbb{R}$ is a <u>geometric progression</u> if there exists a real number $r \neq 0$ such that $a_{n+1} = ra_n$ for every n.*)

f) the set of all strictly increasing sequences $f : \mathbb{N} \to \mathbb{N}$ (*Note: f is <u>strictly</u> <u>increasing</u> if $l, m \in \mathbb{N}$ and $l < m \Rightarrow f(l) < f(m)$.*)

g) the set of all countable subsets of $\mathbb{R}$. (*Hint: part f) could be used.*)

Note: Sometimes one proves $|A| = m$ by giving two separate arguments, one to show $|A| \leq m$ and the other to show $|A| \geq m$. One of these two inequalities often is easy and all the harder job is to show that the other inequality holds.

E50. Find a subset of $\mathbb{Q}$ equivalent to the set of all binary sequences (a_n), or explain why no such subset exists .

E51. Explain why the following statement is true:

The set of all real numbers x which have a decimal expansion of the form

$$x = 0.x_1x_2x_3...x_n0101\overline{01}... \quad (n \text{ may depend on } x)$$

is countable.

E52. Prove that for any collection of sets $\{A_\lambda : \lambda \in \Lambda\}$, there must exist a set Y such that $|Y| > |A_\lambda|$ for every $\lambda \in \Lambda$. (*Hint: use the fact that* $X \not\sim \mathcal{P}(X)$.)

E53. Prove or give a counterexample for the following statement:

> If $\mathcal{C}$ is an uncountable collection of uncountable subsets of $\mathbb{R}$, then at least two sets in $\mathcal{C}$ must have an uncountable intersection.

E54. Prove that if m, n, and p are cardinals, then

a) $m^{n+p} = m^n \cdot m^p$
b) $(m \cdot n)^p = m^p \cdot n^p$

E55. Prove or disprove:

a) $\aleph_0^c = 2^c$
b) $2^{(2^{\aleph_0})} = (2^2)^{\aleph_0}$
c) if $2 \leq m < 2^{\aleph_0}$, then $m^{\aleph_0} = c$
d) if m is infinite and $2 \leq n \leq 2^m$, then $n^m = 2^m$

E56. a) Prove that $\mathbb{R}$ has c countable subsets.
b) Prove or disprove: If A and B have the same number of countable subsets, then $A \sim B$.

E57. Prove or disprove: there are exactly c sequences of the form $(A_1, A_2, \ldots, A_n, \ldots)$ where each A_n is a subset of $\mathbb{Q}$.

E58. Find all unjustified steps in the following "proof" of the continuum hypothesis:

If CH is false, then $\aleph_0 < m < c$ *for some cardinal* m. *Since* $c = \aleph_0^{\aleph_0} \leq m^{\aleph_0} \leq c^{\aleph_0} = c$, *we have* $m^{\aleph_0} = c = 2^{\aleph_0} < 2^m$, *so* $m^{\aleph_0} < 2^m$. *Therefore* $(m^{\aleph_0})^c < (2^m)^c$, *so* $m^c < 2^c$, *which is impossible because* $m > 2$.. *Therefore no such* m *can exist, so CH is true.*

E59. Find all unjustified steps in the following "disproof" of the continuum hypothesis:

We know $c = 2^{\aleph_0} = 2^{\aleph_0^2} = (2^{\aleph_0})^{\aleph_0}$. *However,* $(2^{\aleph_0})^{\aleph_0} > \aleph_0^{\aleph_0}$ (*because* $2^{\aleph_0} > \aleph_0$). *Since* $\aleph_0 > 1$, *we have* $\aleph_0^{\aleph_0} > \aleph_0^1 = \aleph_0$. *Therefore* $c > \aleph_0^{\aleph_0} > \aleph_0$, *so CH is false.*

E60. Since $\mathbb{Q}$ is countable and $|\mathbb{R}| = c$, we know that $\mathbb{P}$ is uncountable. But that does not automatically mean that $|\mathbb{P}| = c$. (*Unless you assume CH, it could happen that* $\aleph_0 < |\mathbb{P}| < c$.)

a) Without using the properties 5*), 6*) of cardinal addition or multiplication and without using CH, prove that:

if $|B| = c$ and A is a countable subset of B, then $|B - A| = c$.

(*Hints: Obviously* $\aleph_0 < |B - A| \leq c$. *One way to show* $|B - A| = c$: *without loss of generality, you can assume* $B = \mathbb{R}^2$. *Then consider vertical lines in* B. *Of course, there are other approaches.*)

b) Using a), deduce that $|\mathbb{P}| = c$.

E61. Prove that for any infinite set E, there is an infinite sequence of disjoint subsets E_1, E_2, E_3,... such that $E = \bigcup_{n=1}^{\infty} E_n$ and $|E_n| = |E|$ for all *n*. (*Hint: The multiplication rule in Example 14.6(6*) that implies that* $m \cdot \aleph_0 = m$ *for any infinite cardinal* m. *You can assume that rule.*)

E62. Call a function $f : X \to Y$ "double-rooted" if $|f^{-1}(y)| = 2$ for every $y \in Y$. Find the number of double-rooted functions $f : \mathbb{Q} \to \mathbb{Q}$.

E63. Assume the Generalized Continuum Hypothesis (see p. 51). Then

True or false (explain): $\mathcal{P}(A) \sim \mathcal{P}(B)$ implies $A \sim B$.

E64. Say that a pair of sets (A, B) has property (*) if all three of the following conditions are true:

i) $A \cup B = \mathbb{N} \times \mathbb{N}$
ii) every horizontal line intersects A in only finitely many points
iii) every vertical line intersects B in only finitely many points.

We saw in Exercise E44 that such pairs (A, B) *exist.*

Prove or disprove: there are exactly *c* different pairs (A, B) with property (*).

E65. Let $D = \{a_1, a_2, ..., a_n, ...\}$ be a countable subset of $\mathbb{R}$ and choose positive numbers ϵ_n for which $\sum_{n=1}^{\infty} \epsilon_n < \infty$. Define $f : \mathbb{R} \to \mathbb{R}$ by $f(x) = \sum_{a_n \leq x} \epsilon_n$. Clearly $f(x) \leq f(y)$ if $x \leq y$.

Prove that f is discontinuous at each point in D and continuous at each point in $\mathbb{R} - D$.

In Theorem 8.5, we saw *that a monotone function* $f : \mathbb{R} \to \mathbb{R}$ *has a countable set of discontinuities; this result is a "sort of" converse. Note that this* f *is continuous from the right at every point.*)

15. A Final Digression

Let S^2 denote the sphere $\{(x, y, z) \in \mathbb{R}^3\text{: } x^2 + y^2 + z^2 = 1\}$. We will prove the following surprising (*or is it not so surprising*?) result: stated informally

> If a countable set D is removed from S^2, it is possible to write the remainder $S^2 - D$ as the union of two subsets A and B which, when rotated, give the whole sphere S^2 back again**.**

For a point $v = (x, y, z) \in S^2$, we use vector notation and write $-v = (-x, -y, -z)$.

Since S^2 is uncountable (*why?*), we can choose a point $v \in S^2$ for which $v \notin D \cup (-D)$ $= \{\pm d : d \in D\}$. Then neither v nor $-v$ is in D. Change the coordinate axes so that the z-axis goes through v and $-v$ (so the "north and south poles" of S^2 are not in D).

For any $C \subseteq S^2$, write $C(\beta)$ to represent the set obtained by rotating C on the surface S^2 around the z-axis through angle β . In other words, a point in $C(\beta)$ comes from taking a point in C and adding β to its "longitude" on S^2. With this notation, the precise statement of what we want to prove is:

> Suppose D is a countable subset of S^2. Then there are subsets A and B of S^2, and there are real numbers α and γ such that $S^2 - D = A \cup B$ and $S^2 = A(\alpha) \cup B(\gamma)$.

First, we claim that we can choose a $\beta \in [0, 2\pi)$ that makes the sets D, $D(\beta)$, $D(2\beta)$, ... , $D(n\beta)$, ... all pairwise disjoint. For any point v (other than the north and south poles), let $\arg(v) \in [0, 2\pi)$ be the longitude of v measured from the great circle through $(1, 0, 0)$ (= the "Greenwich meridian"). $D(k\beta)$ and D($n\beta$) can intersect only if there are points $a, b \in D$ such that

$$\arg(a) + k\beta = \arg(b) + n\beta + 2j\pi \quad (k, n \in \mathbb{N},\, j \in \mathbb{Z}) \qquad (**)$$

This means $D(n\beta)$ and $D(k\beta)$ can intersect <u>only if</u> β satisfies the equation $\beta = \frac{\arg(a) - \arg(b) - 2j\pi}{n-k}$.

But there are only countably to pick a "5-tuple" of values (a, b, j, n, k) to plug into the right side of the equation – because $\aleph_0^5 = \aleph_0$. So there are only countably many values β for which a pair of the sets $D, D(\beta), D(2\beta),$ ... $, D(n\beta),$ ... could intersect. <u>Choose</u> any $\beta \in [0, 2\pi)$ different from these countably many values; then the sets will be pairwise disjoint.

Now, define $T = D \cup D(\beta) \cup D(2\beta) \cup ... \cup D(n\beta) \cup ... = \bigcup_{n=0}^{\infty} D(n\beta)$
and $B = T - D = D(\beta) \cup D(2\beta) \cup ... \cup D(n\beta) \cup ... = \bigcup_{n=1}^{\infty} D(n\beta)$

A rotation through $(-\beta)$ moves each set $D((k+1)\beta)$ onto the set $D(k\beta)$, so $B(-\beta) = T$.

Let $A = S^2 - T$. Then $A \cup B = (S^2 - T) \cup (T - D) = S^2 - D$.

Since $A(0) = A$ and $B(-\beta) = T$, we have $A(0) \cup B(-\beta) = (S^2 - T) \cup T = S^2$. •

Chapter I Review

Explain why each of the following statements is true or provide a counterexample.

1. $\mathbb{R}^{\mathbb{R}} \sim \mathcal{P}(\mathbb{R})$

2. The continuum hypothesis (CH), which states that $2^{\aleph_0} = c$, is independent of the other usual axioms of the set theory.

3. If $f : \mathbb{R} \to \mathbb{R}$ is strictly decreasing, then there are at least 7 points at which f is continuous.

4. There exists a straight line ℓ in the plane such that ℓ contains exactly three points (x, y) where both x and y are rational.

5. In $\mathbb{R}^3$, it is possible to find uncountably many "solid balls" of the form

$$\{(x, y, z) : (x - a)^2 + (y - b)^2 + (z - c)^2 < \epsilon\}$$

such that any two of them are disjoint.

6. The continuum hypothesis is true iff the set of all sequences of 0's and 1's has cardinality c.

7. There are c different infinite sets of prime numbers.

8. Let A be the set of all sequences (a_n) where $a_n \in \{0, 1, 2, 3\}$ and such that $\{n : a_n = k\}$ is infinite for each $k = 0, 1, 2, 3$. Then A is countable.

9. If $\mathcal{C}$ is an uncountable collection of uncountable subsets of $\mathbb{R}$, then at least two sets in $\mathcal{C}$ must have uncountable intersection. (*Hint: recall that* $c^2 = c$)

10. The set of real numbers which are not transcendental is uncountable.

11. $\mathbb{N}^{\mathbb{R}} \sim \mathcal{P}(\mathbb{R})$

12. Let S^2 denote the unit sphere $\{(x, y, z) \in \mathbb{R}^3 : \ x^2 + y^2 + z^2 = 1\}$ and $P = (0, 0, 1)$. There exists a continuous bijection $f : S^2 - \{P\} \to \mathbb{R}$. (*The values of f are in $\mathbb{R}$, not $\mathbb{R}^2$!*)

13. Assume that for an infinite cardinal m, there is never a cardinal strictly between m and 2^m (the Generalized Continuum Hypothesis). Then $\mathcal{P}(A) \sim \mathcal{P}(B)$ implies $A \sim B$.

14. There are exactly c sequences of the form $(A_1, A_2, \ ... \ , A_n, \ ...)$ where each A_n is a subset of $\mathbb{Q}$.

15. There is an algebraic number between any two real numbers.

16. Let $\mathcal{S} \ = \{f \subseteq \mathbb{R}^2 : \$ every horizontal line and every vertical line intersects f in exactly one point$\}$. Then $|S| = 2^c$.

17. There are 2^c subsets of $\mathbb{R}$ none of which contains an interval of positive length.

18. Let $\Gamma : C[a, b] \to \mathbb{R}$ by $\Gamma(f) = f(\frac{a+b}{2})$. Then Γ is one-to-one.

19. Suppose a nonempty set X can be "factored" as $X = Y \times Z$. Then Y and Z are unique.

20. Suppose A, B and C are infinite sets and that $A \sim B \cup C$. Then $A \sim B$ or $A \sim C$.