

# Discovering and Writing Proofs

© 2009 by Bong Lian. All rights reserved.

This course is an introduction to proofs. Students will learn how to read, to discover for themselves, and ultimately to write proofs, using a generally accepted mathematical language. The latter includes various mathematical notions, vocabulary, notations, conventions, syntax, and logic symbols. Since this is the first exposure to proofs for most students in this course, these various aspects of the language will be learned through many examples and exercises. Peppered throughout the notes, there are many short exercises which form the most essential part of this course. Each of them usually involves just a few lines of work that rely only on one or two important lessons (theorems or notions) students have accumulated up to that point. It is important that students take the initiatives to try those exercises on their own, before seeing them done in class.

# 1. Numbers

## 1.1. Where to begin

Virtually everyone is familiar with counting numbers or the *integers* since grade school. We learned to perform arithmetical operations – addition, subtraction and multiplication – by examples and association like counting with fingers or marbles. We then learned fractions, as a necessary means to break up or divide 1 into smaller units. Thus that 1 foot is broken up into 12 equal units or inches means that 1 inches is  $1/12$  of a foot. Hence our system of numbers has been enlarged to include all fractions or *rational numbers*, and their arithmetic now includes inversion or division of integers.

Well, now that we have grown up, can we give a better definition of the integers without relying on fingers or marbles? It turns out that any reasonable mathematical definition we attempt to give of the integers, though may be dressed up in a more sophisticated mathematical language, would end up being pretty much the same as the operational definition associated with finger counting. For beginners, it is better to start by accepting those numbers on faith and gaining a solid understanding of how the arithmetical operations on those numbers work, before attempting to provide a mathematical foundation for those numbers. That's what we will do.

Should we stop at fractions? Anyone could have easily discovered through a grade school art project that fractions alone are inadequate or fine enough for quantifying measurements. For example, imagine that our art teacher says: let's begin with a square sheet

with an area size of 2 feet square... How long must each side of that sheet be? That naturally leads to the equation

$$x^2 = 2.$$

It has been known, since at least the ancient Greeks, that you can never find a fraction  $x$  that makes this equation hold. Yet, ancient craftsmans already knew how to cut, using only primitive tools, a square of area exactly 2. It is therefore easy to accept that a positive number exists to satisfy that equation, even though it can't be written as a fraction. So, we invent a symbol for this non fractional number:  $\sqrt{2}$ . Thus we need to enlarge our system of numbers again in order to include this and their obvious cousins like  $\pm\sqrt{2}$ ,  $\pm\sqrt{3}$ ,  $\pm\sqrt{5}$ ,... and any number you can get by performing the four aforementioned basic arithmetical operations on these numbers. These are known as *constructible numbers*.

Is this system of constructible numbers enough for quantifying basic measurements? Unfortunately not. For instance, you can't even represent the area size enclosed by a circle of radius 1, or the arclength of this circle, in this system. So, we enlarge our number system yet again, by including a non constructible number  $\pi$  representing the area of the unit circle. But that seems to open up a can of worms: what about  $\sqrt{\pi}$  and many of its obvious cousins? What's the endgame?

Fortunately, by the 19th century, a satisfactory resolution to this conundrum was finally found. It is based on the ancient idea that we can represent numbers in the decimal system by a string, possibly infinitely long, of integers between 0 and 9. Indeed, algorithms for computing the decimal string for numbers like  $\sqrt{2}$ ,  $\sqrt{3}$  or  $\pi$  had been well understood long before any foundation of numbers was created. All four basic arithmetical operations on decimal numbers were also understood. By mid 19th century, the consistency of those computational algorithms were proved as corollaries to the foundation of a system called the system of *real numbers*.

For beginners, without questioning its foundation or its consistency, we can think of this real number system as a basket or a *set* we denote by  $\mathbf{R}$ , whose members are the decimal numbers. It is a set that is endowed with the aforementioned four basic arithmetical operations, each of which can be described algorithmically, as we have learned since grade school. To proceed along this line, we must also learn the basic rules or properties that govern the four basic operations, which we now review. Do we ever go back to learning the

proper foundation of  $\mathbf{R}$ , and if so, when? This is usually the starting point of a standard introduction to analysis. So, be patient.

Let's begin.

## 1.2. Arithmetic

A member of the set  $\mathbf{R}$  is called a real number or a number for short (again, you can think of this as a decimal number.) The symbolic phrase  $a \in \mathbf{R}$  means to say that  $a$  is a member of  $\mathbf{R}$ . That is to say,  $a$  is a number. There are two numbers, 0 or 1 (0.000... or 1.000...,) that play special roles in arithmetic. Furthermore,  $\mathbf{R}$  is equipped with the following four basic arithmetical operations. Two of the operations are *binary* – involving two numbers, and the other two operations are *unary* – involving just one number. For any  $a, b \in \mathbf{R}$ ,

- i. *Addition.* Given any  $a, b \in \mathbf{R}$ , addition produces a third number, which is called the sum of  $a$  and  $b$ , and is denoted by the symbol  $a + b$ .
- ii. *Multiplication.* Given any  $a, b \in \mathbf{R}$ , multiplication produces a third number, which is called the product of  $a$  and  $b$ , and is denoted by symbol  $ab$ .
- iii. *Negation.* Given any  $a \in \mathbf{R}$ , negation produces a second number, which is called the negative of  $a$ , and is denoted by  $-a$ .
- iv. *Inversion.* Given any  $a \in \mathbf{R}$  not equal to 0, inversion produces a second number, which is called the reciprocal of  $a$ , and is denoted by  $1/a$ .

The numbers 0,1 and the four operations are compatible in ways that are governed by the following rules, which we shall refer to as the *Arithmetic Laws*. To state them, let  $a, b, c$  be arbitrary numbers.

*Neutrality Laws.*  $a + 0 = a$ ,  $a \cdot 1 = a$ . Thus we say that 0 is additively neutral and that 1 is multiplicatively neutral.

*Negation Law.*  $a + (-a) = 0$ .

*Inversion Law.*  $a(1/a) = 1$ .

*Commutative Laws.*  $a + b = b + a$  and  $ab = ba$ .

*Associative Laws.*  $(a + b) + c = a + (b + c)$  and  $(ab)c = a(bc)$ .

*Distributive Law.*  $a(b + c) = ab + bc$ .

Each of these laws can be proved as a corollary to the foundation of  $\mathbf{R}$ . But even though we can't prove them without reaching back to the foundation, our experience working with decimal numbers certainly make accepting these laws palatable.

**Exercise.** Using the Arithmetic Laws, prove that for  $a, x \in \mathbf{R}$ ,  $a + x = 0$  if and only if  $x = -a$ . Likewise, prove that  $ax = 1$  if and only if  $x = 1/a$ .

### 1.3. Comparing numbers – ordering

As important as they are, the Arithmetic Laws are still inadequate for deducing some of the commonly used facts about numbers. For instance, there is no way to prove using these laws alone that there is a number  $x$  such that

$$x^2 = 2.$$

To prove this would require reaching back to the foundation of  $\mathbf{R}$ . Again, our goal is to gain a solid understanding of various operational aspects of  $\mathbf{R}$  first before reaching back to its foundation. Indeed, without questioning its validity, it is not hard to give a precise algorithm (whose validity ultimately rests on the foundation of  $\mathbf{R}$ ) to find the decimal number representing  $\sqrt{2}$ . Thus as part of our starting point, we shall state and assume a theorem (without proof) to that effect. In order to formulate the theorem in sufficient generality, we consider an analytical (as opposed to arithmetical) aspect of  $\mathbf{R}$ . This is basically the notion of positivity of a number.

There is a special relation  $>$  between numbers which we call an *ordering*. For any  $a, b \in \mathbf{R}$ , exactly one of the following three statement is true:

- i.  $a > b$ , in which case we say that  $a$  is greater than  $b$  or that  $b$  is less than  $a$ ;
- ii.  $b > a$ ;

iii.  $a = b$ .

For convenience, we sometimes write  $a < b$  to say  $b > a$ . When  $a > 0$  we say that  $a$  is positive. When  $a < 0$ , we say that  $a$  is negative. When either i. or iii. is true, we write  $a \geq b$ , and  $b \leq a$  means the same. When either ii. or iii. is true, we write  $b \geq a$ , and  $a \leq b$  means the same. The relation  $>$  is compatible with the operations of addition and multiplications in ways governed by the following rules, which we shall refer to as the *Ordering Laws*. To state them, let  $a, b, c \in \mathbf{R}$ .

*Translation Law.* If  $a > b$  then  $a + c > b + c$ .

*Transitive Law.* If  $a > b$  and  $b > c$  then  $a > c$ .

*Positive Product Law.* If  $a > 0$  and  $b > 0$  then  $ab > 0$ .

Again, each of these inequalities or laws can be proved as a corollary to the foundation of  $\mathbf{R}$ .

**Exercise.** Using the Arithmetic and the Ordering Laws to prove that for  $a \in \mathbf{R}$ ,  $a > 0$  if and only if  $-a < 0$ .

**Exercise.** Likewise, prove that if  $a > 0$  and  $b < 0$  then  $ab < 0$ . (Hint:  $b < 0$  is equivalent to  $-b > 0$ .)

It is often helpful to first discover the reasons why an assertion is true before writing a formal proof. Discovering those reasons often time involves trial and error, asking a series of questions, sometimes starting from the assertion and working backward. Once the reasons become clear, writing a formal proof amounts to enumerating those reasons step by step in proper English. As an illustration, let's prove that for  $a, b, c \in \mathbf{R}$ , if  $a > b$  and  $c > 0$  then  $ac > bc$ .

$$ac \stackrel{?}{>} bc$$

$$ac - bc \stackrel{?}{>} 0 \quad (\text{Translation})$$

$$(a - b)c \stackrel{?}{>} 0 \quad (\text{Distributive, Commutative})$$

At this point, it should be clear that the last inequality holds under the assumptions that  $a - b > 0$  (or  $a > b$ ) and that  $c > 0$ , by the Positive Product Law. Now comes the writing of a formal proof.

Proof: Assume  $a > b$  and  $c > 0$ . Then  $a - b > b - b = 0$  by the Translation Law. Then  $(a - b)c > 0$  by the Positive Product Law. Hence  $ac - bc > 0$  by the Commutative and Distributive Laws. Finally,  $ab > bc$  by the Translation Law again.  $\square$

We are now ready to return to the question of square roots.

**Theorem 1.1.** (*Square Root*) For any given positive number  $a$ , there is a unique positive number  $x$  such that  $x^2 = a$ .

For  $a > 0$ , the theorem says that the equation  $x^2 = a$  has exactly one positive solution, which we call the square root of  $a$  and denote it by  $\sqrt{a}$ . Again, this theorem can't be proved without reaching back to the foundation of  $\mathbf{R}$ . But we shall explore some of the corollaries of this theorem. For instance, how does the operation of taking square root interact with multiplication and the relation  $>$ ?

**Corollary 1.2.** For any positive numbers  $a, b$ , we have  $\sqrt{ab} = \sqrt{a}\sqrt{b}$  and  $\sqrt{a^2} = a$ .

Proof: By definition,  $\sqrt{ab}$  is the solution to the equation

$$x^2 = ab.$$

So, to prove the asserted equality, it is enough to verify that  $\sqrt{a}\sqrt{b}$  is also a solution. By the Commutative and Associative Laws,

$$(\sqrt{a}\sqrt{b})(\sqrt{a}\sqrt{b}) = (\sqrt{a})^2(\sqrt{b})^2 = ab.$$

This proves our first assertion. Our second assertion follows from the first by setting  $a = b$ .

$\square$

**Corollary 1.3.** For any positive numbers  $a, b$ , if  $a > b$  then  $a^2 > ab > b^2$  and  $\sqrt{a} > \sqrt{b}$ .

There are three assertions to prove. Again, it is easy to discover the reasons behind the first two assertions by working backward with each.

Proof: Assume  $a > b$ . By the Translation Law  $a - b > 0$ . By the Positive Product Laws,  $a(a - b) > 0$ . By the Distributive and the Translation Laws,  $a^2 > ab$ . The assertion

$ab > b^2$  can be proved similarly. Here we make an observation that will be used below: by the Transitive Law, our first two assertions imply that  $a^2 > b^2$  under the assumption that  $a > b > 0$ .

To prove  $\sqrt{a} > \sqrt{b}$ , we use something called a counter positive argument (also called a proof-by-contradiction argument) – proving an assertion by first supposing that the assertion is false, and then try to derive a conclusion that contradicts the initial assumption. Here our initial assumption is  $a > b$ . So, we suppose  $\sqrt{a} > \sqrt{b}$  is false, which means that  $0 < \sqrt{a} \leq \sqrt{b}$ . By the observation we made above,

$$a = (\sqrt{a})^2 \leq (\sqrt{b})^2 = b.$$

This contradicts the initial assumption  $a > b$ , and the proof is complete.  $\square$

We have the following generalization of the Square Root Theorem.

**Theorem 1.4.** (*n*th Root) *Let  $n$  be a positive integer. For any given positive number  $a$ , there is a unique positive number  $x$  such that  $x^n = a$ .*

We call *the* solution to the equation  $x^n = a$  in the theorem the *n*th root of  $a$  and denote it by  $\sqrt[n]{a}$ .

**Corollary 1.5.** *For any positive numbers  $a, b$ , we have  $\sqrt[n]{ab} = \sqrt[n]{a} \sqrt[n]{b}$  and  $\sqrt[n]{a^n} = a$ .*

**Corollary 1.6.** *If  $a > b > 0$  then  $a^n > b^n$  and  $\sqrt[n]{a} > \sqrt[n]{b}$ .*

**Exercise.** By imitating the proofs of the corollaries to the Square Root Theorem, prove the preceding two corollaries.

## 1.4. Proving and solving basic inequalities

The relation  $>$  on  $\mathbf{R}$  allows us to compare any two given numbers by means of an inequality. It is often useful in analysis to be able to compare two different expressions of a bunch of numbers using an inequality. We explore a few of them here.

For  $x \in \mathbf{R}$ , we define the operation

$$|x| = \begin{cases} x & \text{if } x \geq 0 \\ -x & \text{if } x < 0 \end{cases}$$

The number  $|x|$  is called the absolute value of  $x$ . Note that  $|x| \geq x$  and that  $|x|^2 = x^2$ , for any  $x \in \mathbf{R}$ .

**Exercise.** For  $x, y \in \mathbf{R}$ , prove that  $|x| \leq y$  if and only if  $-y \leq x \leq y$ .

How does the absolute value operation interact with arithmetic operations?

**Proposition 1.7.** For  $x, y \in \mathbf{R}$ , we have  $|xy| = |x||y|$ .

Proof:

$$|xy|^2 = (xy)^2 = x^2y^2 = |x|^2|y|^2 = (|x||y|)^2.$$

Taking square root on both sides (Square Root Theorem!), and noting that both numbers  $|xy|$  and  $|x||y|$  are nonnegative yields

$$|xy| = |x||y|. \quad \square$$

**Proposition 1.8.** (*Triangle Inequality*) For  $x, y \in \mathbf{R}$ , we have  $|x + y| \leq |x| + |y|$ .

Let's discover the reasons by working backward from the conclusion.

$$\begin{aligned} |x + y| &\stackrel{?}{\leq} |x| + |y| \\ (x + y)^2 &= |x + y|^2 \stackrel{?}{\leq} (|x| + |y|)^2 \\ x^2 + y^2 + 2xy &\stackrel{?}{\leq} x^2 + y^2 + 2|x||y| \\ xy &\stackrel{?}{\leq} |x||y| = |xy| \end{aligned}$$

The last inequality is true, and observe that each step above is *reversible*. In other words, each line is implying the next line as well as being implied by the next line (check this!) It is now easy to write down a formal proof by following the reversing the steps (exercise.)

The next so-called algebraic-geometric mean inequality compares two different expressions involving two numbers  $x, y$ , one being the average or algebraic mean and the other being the geometric mean.

**Proposition 1.9.** (*AGM<sub>2</sub> Inequality*) For positive numbers  $a, b$ ,

$$\sqrt{ab} \leq \frac{a+b}{2}.$$

Let's work backward again.

$$\begin{aligned} \sqrt{ab} &\leq \frac{a+b}{2} \\ (2\sqrt{ab})^2 &\leq (a+b)^2 \\ 4ab &\leq a^2 + b^2 + 2ab \\ 0 &\leq (a-b)^2. \end{aligned}$$

Again, the last inequality is true, and each step above is reversible. We can now write down a formal proof (exercise.)

You can arbitrarily make up two expressions involving one or more numbers and ask if one expression is *always* less than or equal to the other expression. Of course, the answer is typically no. But it may still be interesting to solve the inequality, i.e. to find out what values of those numbers involved would make the inequality true. Let's solve the inequality

$$\left| \frac{x-2}{x+1} \right| \leq 1.$$

That is, give a simple description of the set of values of  $x$  that make this inequality true. Clearly, the left side only makes sense for  $x \neq -1$ . Otherwise, the inequality is true if and only if

$$\left| \frac{x-2}{x+1} \right|^2 \leq 1.$$

Clearing denominator yields,

$$(x-2)^2 \leq (x+1)^2.$$

Expanding and simplifying this yields

$$\frac{1}{2} \leq x.$$

Since each step above is reversible, we find that the set of solutions consists of all numbers with values at least  $\frac{1}{2}$ .

Next, we prove an analogue of *AGM<sub>2</sub>* involving four numbers.

**Proposition 1.10.** (*AGM<sub>4</sub> Inequality*) For positive numbers  $a, b, c, d$ ,

$$\sqrt[4]{abcd} \leq \frac{a + b + c + d}{4}.$$

Proof: If we assume that the  $AGM_4$  is true, then for any positive numbers  $x, y, z, w$ , by setting  $a = x^4, b = y^4, c = z^4, d = w^4$ , we see that

$$AGM'_4 : 4xyzw \leq x^4 + y^4 + z^4 + w^4$$

is true. Conversely, if we assume the  $AGM'_4$  is true for any positive numbers  $x, y, z, w$ , then the  $AGM_4$  is true (by setting  $x = \sqrt[4]{a}, y = \sqrt[4]{b}, z = \sqrt[4]{c}, w = \sqrt[4]{d}$ .) Thus proving  $AGM_4$  is equivalent to proving  $AGM'_4$ .

We shall prove the latter. First, we have

$$AGM'_2 : 2uv \leq u^2 + v^2$$

is true for any positive numbers  $u, v$ . This implies that

$$4xyzw = 2xy2zw \leq (x^2 + y^2)(z^2 + w^2) = x^2z^2 + x^2w^2 + y^2z^2 + y^2w^2.$$

By  $AGM'_2$ , on the right side the first term  $x^2z^2$  is  $\leq \frac{x^4+z^4}{2}$  (by setting  $u = x^2$  and  $v = z^2$ .)

We can do the same for the remaining three terms. So, we get

$$4xyzw \leq \frac{x^4 + z^4}{2} + \frac{x^4 + w^4}{2} + \frac{y^4 + z^4}{2} + \frac{y^4 + w^4}{2} = x^4 + y^4 + z^4 + w^4.$$

This proves the  $AGM'_4$ .  $\square$

**Proposition 1.11.** (*AGM<sub>3</sub> Inequality*) For positive numbers  $a, b, c$ ,

$$\sqrt[3]{abc} \leq \frac{a + b + c}{3}.$$

Proof: Again, proving this is equivalent to proving

$$AGM'_3 : 3tuv \leq t^3 + u^3 + v^3$$

for any positive numbers  $t, u, v$ . The idea is to use the  $AGM'_4$  above, with  $w = \sqrt[3]{xyz}$ . We get

$$4(xyz)^{4/3} = 4xyz(xyz)^{1/3} \leq x^4 + y^4 + z^4 + (xyz)^{4/3}.$$

(Fractional powers of a number are defined and studied in your next homework assignment.)

It follows that

$$3(xyz)^{4/3} \leq 4xyz(xyz)^{1/3} \leq x^4 + y^4 + z^4$$

is true for any positive numbers  $x, y, z$ . Given positive numbers  $t, u, v$ , by choosing  $x = t^{3/4}, y = u^{3/4}, z = v^{3/4}$ , we see that the  $AGM'_3$  above follow.  $\square$

## 2. Sets

### 2.1. What is a set?

In mathematics, there is a small number of terms or notions which are accepted as atomic or primitive, and which we do not attempt to define. The notion of a *set* is one of them. What do we do with something that is not even defined? We agree on a list of rules that govern exactly how we are allowed to use the term. For us beginners, it is simply too involved to get into all the details of the rules that govern the usage of *set*.

For us, it is better to start with the following simple operational rule, whenever the term *set* is used.

*To specify a set, we specify a membership test.*

Thus, we specify a set  $A$  by spelling out a membership test (also called a *defining condition*.) What is that? Roughly speaking, a membership test can be thought of as a black box  $T$  that will accept one input  $x$ , and that is capable of deciding if the input  $x$  is a member of  $A$ . Think of the box as having a red and a green light. Each time we wish to decide if  $x$  is a member of  $A$ , we can feed  $x$  into  $T$ . If the green light goes off, then we say that  $x$  is a member of  $A$ . If the red light goes off, then we say that  $x$  is not a member of  $A$ . We now discuss two formal ways to specify membership tests.

*Universe.* To specify a membership test, we can specify a function

$$T : U \rightarrow \{true, false\}.$$

Here  $U$  is a *pre-determined* set, we call *the universe*, on which we wish apply the membership test;  $T$  will produce exactly one value, either true or false (but not both) each time a test input  $x$  from  $U$  is given. In other words, this membership test  $T$  is nothing but a rule that separates those test inputs that belong in  $A$  and those that do not.

It may seem circular and disconcerting to define a set  $A$ , by having to *first* provide another set  $U$ . But the point is that  $U$  must be a set that is considered universally accepted, so that the ultimate validity of  $A$  would rest on an object we don't question. That's life in mathematics – a universal language – and as such, we are allowed to define something only *in terms of* things that are already considered universally accepted.

For example, let's specify the set  $S$  of all integers divisible by 3. So, in this example we can use the set of integers  $\mathbf{Z}$  to be our universe, which is considered universally accepted. Our membership test in this case will yield  $T(x) = \text{true}$  if  $x$  is an integer divisible by 3, and  $T(x) = \text{false}$  if  $x$  is an integer not divisible by 3. Symbolically, we can write

$$S = \{x \in \mathbf{Z} | T(x) = \text{true}\}.$$

(Sometimes, a colon  $:$  is used in place of  $|$ , and sometimes  $T(x) = \text{true}$  is written simply as  $T(x)$ .) Or, if we wish to be more direct, we can write

$$S = \{x \in \mathbf{Z} | x \text{ is divisible by } 3\}.$$

Verbally, this reads “let  $S$  be the set of all  $x$  is a member of  $\mathbf{Z}$  such that  $x$  is divisible by 3.” For instance, 6 is a member of  $S$  because  $T(6) = \text{true}$ . But 2 is not a member of  $S$  because  $T(2) = \text{false}$ . Symbolically, we can write  $6 \in S$  and  $2 \notin S$ .

Note that in this example, we are allowed to mention the set  $\mathbf{Z}$  (without actually defining it first) because there is a universal agreement of what  $\mathbf{Z}$  exactly is. This will always be the case when it comes to discussing sets – we always discuss a set in the backdrop of something else that has been pre-determined or universally accepted and agreed upon.

*Enumeration or Indexing.* It is not always convenient or possible to specify a set by *first* giving a universe. For example, suppose we want  $S$  to be the set whose members are 1,2, and 3. It is clearly an overkill (or silly) to write

$$\{x \in \mathbf{Z} | x \text{ is } 1, 2, \text{ or } 3\}$$

when  $\{1, 2, 3\}$  conveys the same meaning much more clearly. In the latter, we specifies  $S$  by enumerating all its members. When no confusion can arise, we can also enumerate members of a set by using a pattern. For example,  $\{1, 2, 3, \dots\}$  is an acceptable specification of the set of natural numbers, since the intended meaning of “...” is clearly understood to be that “the next number in the list is obtained from the preceding one by adding 1 to it.”

Similarly, we can declare members of a set by labeling or indexing them using *another set* – again, a universally accepted one. For example

$$\{2k | k \in \mathbf{Z}\}$$

specifies the set of all numbers of the form  $2k$  where  $k$  is an integer. In this example we can, of course, use evenness as a membership test to specify this set. Later, we will discuss an example of a set that is difficult to specify using a universe, but can be specified using indexing.

## 2.2. Specifying vs. describing a set

In practice, we often encounter a set that is specified by a membership test that is obscure looking at first, despite being perfectly acceptable as a membership test. For example, in a previous exercise we *specify*

$$A = \{x \in \mathbf{R} \mid \left| \frac{x-2}{x+1} \right| \leq 1\}.$$

This is the set of solutions to the inequality  $\left| \frac{x-2}{x+1} \right| \leq 1$ . As a test, this inequality is perfectly acceptable, since for *each* given number  $x$ , it is capable of deciding if  $x$  makes the inequality true. However, without further insight, it is not clear what this set “looks” like. To *describe* it, we showed that that inequality is equivalent to  $x \geq \frac{1}{2}$ . In other words, a number passes the test  $\left| \frac{x-2}{x+1} \right| \leq 1$  if and only if it passes the test  $x \geq \frac{1}{2}$ . So  $A$  is equal to the set

$$\{x \in \mathbf{R} \mid x \geq \frac{1}{2}\}.$$

It is clear that this is a much better description of  $A$  than its original specification.

The kinds of problems most often encountered in mathematics are about giving a good description of a set, initially specified by some complicated tests or conditions.

What does it mean to say that two sets  $A$  and  $B$  are equal? To say that  $A = B$  is to assert that the following statements are both true:

- (1) If  $x$  is a member of  $A$ , then  $x$  is a member of  $B$ .
- (2) If  $x$  is a member of  $B$ , then  $x$  is a member of  $A$ .

Let's denote the membership test for  $A$  by  $T : U \rightarrow \{T, F\}$  and that for  $B$  by  $T' : U' \rightarrow \{T, F\}$ . Then (1) and (2) are equivalent respectively to

- (1)' If  $x \in U$  and if  $T(x)$  is true, then  $x \in U'$  and  $T'(x)$  is true.
- (2)' If  $x \in U'$  and if  $T'(x)$  is true, then  $x \in U$  and  $T(x)$  is true.

So, to prove the assertion  $A = B$ , we can do so by proving both (1) *and* (2). We can also do so by proving both (1)' and (2)'.

**Example.** Let  $A = \{x \in \mathbf{Z} \mid (x+1)^2 \leq 4\}$  and  $B = \{-3, -2, -1, 0, 1\}$ . Let's prove that  $A = B$ . To prove statement (1), assume  $x \in A$ . This says that  $x$  is an integer and  $(x+1)^2 \leq 4$ . Since  $(x+1)^2 = |x+1|^2$ , taking square roots yields

$$|x+1| \leq 2.$$

By a previous exercise, this is equivalent to  $-2 \leq x+1 \leq 2$ . Adding -1, we get  $-3 \leq x \leq 1$ . We have shown that  $A$  is the set of all integers  $x$  passing the test  $-3 \leq x \leq 1$ . We can enumerate the integer solutions to this inequality. They are *exactly* the members of  $B$ . Note that while we set out to prove statement (1), we have ended up proving both (1) and (2) at once.

### 2.3. Making new sets from existing sets

*Cartesian products.* If  $S$  and  $S'$  are given sets, the *Cartesian* product of  $S$  and  $S'$  is the set

$$S \times S' = \{(a, a') \mid a \in S, a' \in S'\}.$$

It is the set of all pairs with first and second entries being members of  $S, S'$  respectively.

For example,  $\mathbf{R} \times \mathbf{R}$ , also denoted by  $\mathbf{R}^2$ , has the usual pictorial interpretation as the Cartesian plane.

More generally, given a positive integer, if  $A$  is a given set, we denote by  $A^k$  the set of all lists  $(a_1, \dots, a_k)$ , also called  $k$ -tuples, whose entries  $a_1, \dots, a_k$  are arbitrary members of  $A$ .

*Union, Intersection, and Difference.* Let  $A$  and  $B$  be sets. Their union, denoted by  $A \cup B$ , is the set consisting of all members in  $A$  or in  $B$  (or in both.) Thus  $x \in A \cup B$  if and only if  $x \in A$  or  $x \in B$ . The intersection of  $A$  and  $B$ , denoted by  $A \cap B$ , is the set consisting of all members which are both in  $A$  and in  $B$ . Thus  $x \in A \cap B$  if and only if  $x \in A$  and  $x \in B$ . We say that  $A$  and  $B$  are *disjoint* if  $A \cap B$  is empty, in which case we write  $A \cap B = \emptyset$ . The difference of  $A$  and  $B$ , denoted by  $A - B$ , is the set consisting of all members which are in  $A$  but not in  $B$ . We say that  $B$  is a subset of  $A$  and we write  $B \subset A$ , if every member of  $A$  is a member of  $B$ . Thus,  $B \subset A$  if and only if  $B - A = \emptyset$ . In this case, we call the set  $A - B$  the complement of  $B$  in  $A$ , and we denote it by  $B^c$ , if the role of  $A$  is understood. If  $B$  is not a subset of  $A$ , we write  $B \not\subset A$ . Note that  $A = B$  if and only if  $A \subset B$  and  $B \subset A$ .

One can use a Venn diagram to depict these various notions we have just introduced.

You should verify the following examples:

- i.  $\{1, 2, 3\} \cup \{3, 4\} = \{1, 2, 3, 4\}$ .
- ii.  $\{1, 2, 3\} \cap \{3, 4\} = \{3\}$ .
- iii.  $\{1, 2, 3\} - \{3, 4\} = \{1, 2\}$ .
- iv.  $\{3, 4\} - \{1, 2, 3\} = \{4\}$ .
- v.  $\{1, 2\} \subset \{1, 2, 3\}$ , but  $\{1, 2, 3\} \not\subset \{1, 2\}$ .
- vi.  $\{x \in U \mid T(x) = \text{true}\} \cup \{x \in U \mid T'(x) = \text{true}\} = \{x \in U \mid T(x) = \text{true} \text{ or } T'(x) = \text{true}\}$ .
- vii.  $\{x \in U \mid T(x) = \text{true}\} \cap \{x \in U \mid T'(x) = \text{true}\} = \{x \in U \mid T(x) = \text{true} \text{ and } T'(x) = \text{true}\}$ .

- viii.  $\{x \in U | T(x) = \text{true}\} - \{x \in U | T'(x) = \text{true}\} = \{x \in U | T(x) = \text{true and } T'(x) = \text{false}\}$ .
- ix.  $\{4k | k \in \mathbf{Z}\} \subset \{2k | k \in \mathbf{Z}\}$ , but  $\{2k | k \in \mathbf{Z}\} \not\subset \{4k | k \in \mathbf{Z}\}$ .

**Exercise.** Let  $\mathbf{N}$  be the set of positive integers and  $A = \{(x, y) \in \mathbf{N}^2 | (2 - x)(2 + y) > 2(y - x)\}$ . Prove that  $A = B$  where  $B = \{(1, 1), (1, 2), (1, 3), (2, 1), (3, 1)\}$ .

*Power set.* This is another important way to make a new set from an existing one. Let  $A$  be a given set. The power set of  $A$  is declared to be the set  $P(A)$  consisting of all *subsets* of  $A$ . Thus,  $x$  is member of  $A$  if and only if  $x \subset A$ . Here is an example.

- x.  $P(\{1, 2\}) = \{\emptyset, \{1\}, \{2\}, \{1, 2\}\}$ , which has 4 members.

**Exercise.** Write down  $P(\{1, 2, 3\})$ , which has 8 members. Can you guess how many members  $P(\{1, 2, \dots, n\})$  has, for a given positive integer  $n$ ?

A word of caution is in order here. The way we have declared the power set  $P(A)$  does not exactly conform to the standard we set out at the beginning of this chapter, because we did not name a pre-determined universe, on which the membership test “ $x \subset A$ ” for  $P(A)$  is supposed to apply. The standard we set out, though simple, turns out to be too restrictive here, because it is difficult to provide such a universe for specifying  $P(A)$ . However, the existence of  $P(A)$  is universally accepted as an axiom in set theory. We shall see later that for each subset of  $A$ , we can define what we call the characteristic function of that subset. We can then use such functions to index members of  $P(A)$ .

## 3. Functions

Roughly speaking, a function or a map is a rule of association from one set to another set. Why do we need functions? Functions are relations between two sets that provide ways for comparing structures of those two sets. For instance, as we shall see in a later lecture, functions are essential for comparing the relative sizes of the two sets. Functions are often used to compare operations we wish to perform on those sets.

### 3.1. What is it?

Here is an operational definition.

*To specify a function, we specify two sets  $A$  and  $B$ , and a rule that assigns exactly one member of  $B$  to each member of  $A$ .*

If we name this function  $f$ , then we refer to  $A$  as the *domain* of  $f$  and  $B$  the *target* of  $f$ ; we write  $f : A \rightarrow B$  and denote by  $f(a)$  the member of  $B$  that  $f$  assigns to  $a \in A$ . We also call the element  $f(a)$  the *image* of  $a$  under  $f$  or the *value* of  $f$  at  $a$ . The symbolic phrase  $a \mapsto f(a)$  has the same meaning. If  $X$  is a subset of  $A$ , then the set  $\{f(a) \in B | a \in X\}$  is called the image of  $X$  under  $f$ , and is denoted by  $f(X)$ . If  $Y$  is subset of  $B$ , then the set  $\{a \in A | f(a) \in Y\}$  is called the *pre-image* of  $Y$  under  $f$ , and is denoted by  $f^{-1}(Y)$ .

For example,  $f : \mathbf{R} \rightarrow \mathbf{R}, x \mapsto x^2$ , specifies a function whose domain and target are both the set of real numbers. It assigns to each number  $x$  its square  $x^2$ . We can

also give a formula for this rule of assignment by writing  $f(x) = x^2$ . You can verify that  $f(\{0, 1\}) = \{0, 1\}$  and  $f^{-1}(\{0, 1\}) = \{-1, 0, 1\}$ .

**Proposition 3.1.** *Let  $f : A \rightarrow B$  be a function. If  $C, D$  are subsets of  $A$  and  $E, F$  are subsets of  $B$ , then the following are true:*

- i.  $f(C \cap D) \subset f(C) \cap f(D)$
- ii.  $f(C \cup D) = f(C) \cup f(D)$
- iii.  $f^{-1}(E \cap F) = f^{-1}(E) \cap f^{-1}(F)$
- iv.  $f^{-1}(E \cup F) = f^{-1}(E) \cup f^{-1}(F)$ .

Proof: We prove i and iii. The remaining ones can be proved similarly.

To prove i, let  $b \in f(C \cap D)$ . Then  $b = f(a)$  for some  $a \in C \cap D$ . Since  $a \in C$ ,  $f(a) \in f(C)$ . Since  $a \in D$ ,  $f(a) \in f(D)$ . It follows that  $f(a) \in f(C) \cap f(D)$ . This proves i.

For iii, we have

$$\begin{aligned} f^{-1}(E \cap F) &= \{x \in A \mid f(x) \in E \cap F\} \\ &= \{x \in A \mid f(x) \in E \text{ and } f(x) \in F\} \\ &= \{x \in A \mid f(x) \in E\} \cap \{x \in A \mid f(x) \in F\} \\ &= f^{-1}(E) \cap f^{-1}(F). \quad \square \end{aligned}$$

**Exercise.** Give an example to show that  $f(C \cap D) \neq f(C) \cap f(D)$  in general.

A function  $f : A \rightarrow B$  may be specified by several different rules, each applied to a particular subset of  $A$ . In this case, we must make sure that the different rules are consistent with one another. The following are examples of some of the common mistakes in specifying a function  $f : \mathbf{R} \rightarrow \mathbf{R}$ .

- i. Let  $f(x) = 2$  if  $x > 0$  and  $f(x) = -2$  if  $x < -1$ . This rule of assignment is inadequate, since it specifies no value for  $f(x)$  when  $-1 \leq x \leq 0$ .

- ii. Let  $f(x) = 2$  if  $x > 0$  and  $f(x) = -2$  if  $x < 1$ . This rule of assignment consists of two different rules that are inconsistent with one another, since they specify two different values for  $f(x)$  when  $0 < x < 1$ . A rule of assignment must assign exactly one value  $f(x)$ , to each  $x \in \mathbf{R}$ .
- iii. Let  $f(x) = x^2$  if  $x \leq 1$  and  $f(x) = x$  if  $x \geq 1$ . This is a good rule of assignment, despite the appearance that  $f(1)$  is specified by two different looking formulas, because those formulas yield the same value at  $x = 1$ .

What does it mean to say that two functions  $f : A \rightarrow B$  and  $g : C \rightarrow D$  are equal?

*That  $f = g$  means exactly that the three conditions are true:  $A = C$ ,  $B = D$ , and  $f(a) = g(a)$  for each  $a \in A = C$ .*

Conversely, to say that  $f \neq g$  is to say that one or more of those three conditions are false.

### 3.2. Real valued functions

These are functions whose target set is  $\mathbf{R}$ . Thus we call a function  $f : A \rightarrow \mathbf{R}$  from a set  $A$  to  $\mathbf{R}$  a real valued function defined on  $A$ . These functions are special in that we can add and multiply any two of them. More precisely, if  $f : A \rightarrow \mathbf{R}$  and  $g : A \rightarrow \mathbf{R}$  are functions, then we define new functions  $f + g : A \rightarrow \mathbf{R}$  and  $fg : A \rightarrow \mathbf{R}$ , respectively called the sum and product of  $f, g$ , by the rules that for all  $a \in A$ ,

$$(f + g)(a) = f(a) + g(a), \quad (fg)(a) = f(a)g(a).$$

These operations allow us to build more complicated functions by using some simple functions as building blocks. For example, the function  $f : \mathbf{R} \rightarrow \mathbf{R}$ ,  $f(x) = x^2 + \pi x + 1$ , can be built out of the elementary function  $g(x) = x$  and the constant functions 1 and  $\pi$ , by repeatedly applying the addition and multiplication operations to them.

More generally, let's start with the following  $n$  functions  $g_1, \dots, g_n$  defined on  $\mathbf{R}^n$ , where  $g_i : \mathbf{R}^n \rightarrow \mathbf{R}$ ,  $g_i(x_1, \dots, x_n) = x_i$  for  $i = 1, 2, \dots, n$ . A *monomial* in  $n$  variables is the

product of some number of these functions with a constant function. For example,  $\frac{1}{2}x_1^2x_2^4$  is a monomial. A sum of some number of monomials is called a *polynomial*. For example,

$$f(x, y, z) = x^2 + y^2 + z^2 + 2xy + 2xz + 2yz$$

defines a polynomial in three variables.

*Boundedness.* There are many qualitative ways to understand real valued functions. For instance, in calculus we can understand maxima and minima (if any) of a function  $f : \mathbf{R} \rightarrow \mathbf{R}$  by plotting the geometric shape of its *graph* in the plane  $\mathbf{R}^2$ . Another qualitative aspect of a real valued function is whether or not it can have arbitrarily large (positive or negative) values. We say that a real valued function  $f : A \rightarrow \mathbf{R}$  is *bounded*, if there is a number  $M$  such that  $|f(a)| \leq M$  for all  $a \in A$ . In this case, we call  $M$  an (absolute) bound of  $f$ . For instance, let's consider a function  $f : \mathbf{R} \rightarrow \mathbf{R}$  defined on  $\mathbf{R}$  that is bounded by  $M$ . Then pictorially, the graph of  $f$  lies in between two horizontal lines in the “ $xy$ -plane”  $\mathbf{R}^2$ , namely, the lines given by the equations  $y = M$  and  $y = -M$ . Here are some examples of bounded and unbounded functions:

- i.  $f : \{x \in \mathbf{R} \mid |x| < 1\} \rightarrow \mathbf{R}$ ,  $f(x) = 2x^2$ , is bounded, since  $2x^2 \leq 3$  for all  $x$  in the domain of  $f$ .
- ii.  $f : \mathbf{R} \rightarrow \mathbf{R}$ ,  $f(x) = 2x^2$ , is *not* bounded, since  $2x^2$  can be made to exceed any given number  $M$  by a suitable choice of  $x$  in the domain of  $f$ . The last two examples illustrate the danger of looking at the formula for  $f$  alone. You must consider the domain of  $f$  as well in deciding boundedness.
- iii.  $f : \mathbf{R} \rightarrow \mathbf{R}$ ,  $f(x) = \frac{1+x}{1+x^2}$ , is bounded, since  $|\frac{1+x}{1+x^2}| \leq \frac{1+|x|}{1+x^2} \leq 2$  for all  $x \in \mathbf{R}$ .
- iv.  $f : \mathbf{R} \rightarrow \mathbf{R}$ ,  $f(x) = \frac{1+x^4}{1+x^2}$ , is not bounded, since  $\frac{1+x^4}{1+x^2} \geq \frac{1}{2}x^2$  for all  $x \in \mathbf{R}$  (prove it!) and  $\frac{1}{2}x^2$  can be made to exceed any given number  $M$  by a suitable choice of  $x \in \mathbf{R}$ . This example illustrates a useful trick: comparing one function with a function whose unboundedness is easier to understand.

### 3.3. More basic vocabulary about functions

*Injection, Surjection, and Bijection.* Consider a function  $f : A \rightarrow B$  from  $A$  to  $B$ . We say that  $f$  is surjective or that  $f$  is onto or that  $f$  is a surjection, if  $f(A) = B$ . This

is to say, each member of  $B$  can be expressed as  $f(a)$ , for some  $a \in A$ . We say that  $f$  is injective or that  $f$  is one-to-one or that  $f$  is an injection, if any two distinct members of  $A$  have distinct images under  $f$ . That is to say, if  $a, b \in A$  and if  $a \neq b$  then  $f(a) \neq f(b)$ . We say that  $f$  is bijective or that  $f$  is one-to-one onto or that  $f$  is a bijection, if  $f$  is both injective and surjective. To illustrate these various notions, we consider the following functions  $f : \mathbf{N} \rightarrow \mathbf{N}$ , from the set of positive integers  $\mathbf{N}$  to itself. You should verify each assertion.

- i. Let  $f(k) = k + 1$ ,  $k \in \mathbf{N}$ .  $f$  is injective but not surjective.
- ii. Let  $f(k)$  be 1 plus the number of distinct prime divisors of  $k \in \mathbf{N}$ .  $f$  is surjective but not injective.
- iii. Let  $f(k) = k - 1$  if  $k$  is even and  $f(k) = k + 1$  if  $k$  is odd.  $f$  is bijective.
- iv. Let  $f(k)$  be the sum of all divisors of  $k \in \mathbf{N}$ .  $f$  is neither injective nor surjective.

A bijection pairs up members of the domain and the target without repeating or missing any members of either set. More precisely, let  $f : A \rightarrow B$  be a bijection. Then it assigns exactly one member of  $A$  to each member of  $B$ . In other words, the pre-image set  $f^{-1}(\{b\}) \subset A$  has just one member, for each  $b \in B$ . We can therefore define a “partner” function from  $B$  to  $A$  as follows: assigns to each  $b \in B$ , the one and only member of the set  $f^{-1}(\{b\})$ . This partner function is called the *inverse* of  $f$  and is denoted by  $f^{-1}$ . (Though this symbol is being used to convey two different meanings, they are clearly compatible.) Restated symbolically, any given bijection  $f$  and its inverse  $f^{-1} : B \rightarrow A$  have the property that for  $a \in A$  and  $b \in B$ ,

$$f^{-1}(f(a)) = a, \quad f(f^{-1}(b)) = b.$$

The next result shows that we can use this property as a way to test for bijectivity of a function.

**Proposition 3.2.** (*Bijection Test*) *For a given function  $g : X \rightarrow Y$  to be bijective, it is necessary and sufficient that we have a function  $h : Y \rightarrow X$ , such that  $h(g(x)) = x$  for all  $x \in X$  and that  $g(h(y)) = y$  for all  $y \in Y$ . In this case,  $h = g^{-1}$  and it is also a bijection.*

**Proof:** Assume that  $g$  is bijective. We saw that its inverse  $g^{-1} : Y \rightarrow X$  has the property we want for the  $h$  we seek. Conversely, assume that we have a function  $h : Y \rightarrow X$  that

has the property stated in the proposition. So, if  $x_1, x_2 \in X$  are different members, then  $h(g(x_1)) = x_1$  and  $h(g(x_2)) = x_2$  are different, implying that  $g(x_1)$  and  $g(x_2)$  are different. This shows that  $g$  is injective. Since  $g(h(y)) = y$  for all  $y \in Y$ , it follows that  $g : X \rightarrow Y$  is surjective.

Finally, since  $g$  is bijective, given any  $y \in Y$  there is a unique  $x \in X$  such that  $g(x) = y$ . So,

$$g^{-1}(y) = g^{-1}(g(x)) = x = h(g(x)) = h(y).$$

It follows that  $g^{-1} = h$ . Note that  $h$  and  $g$  satisfy a condition in which their appearances are symmetric. Since this condition implies that  $g$  is bijective, it must also imply that  $h$  is bijective. This proves our last assertion.  $\square$

Consider example iii above. By definition of  $f$ , we have  $f(f(k)) = k$  for all  $k \in \mathbf{N}$ . Thus, the Bijectivity Test implies that  $f^{-1} = f$ .

In the next example, we make use of the fact that a positive integer  $k$  divided by 3 has exactly one of three possible remainders: 0, 1, or 2.

- v. Let  $f(k) = k + 1$  if  $k/3$  has remainder 1 or 2, and  $f(k) = k - 2$  if  $k/3$  has remainder 0. Then  $f$  is a bijection. Its inverse is given by  $f^{-1}(k) = k - 1$  if  $k/3$  has remainder 0 or 2, and  $f^{-1}(k) = k + 2$  if  $k/3$  has remainder 1.

Proof: Let's give the purported inverse of  $f$  a name. For  $k \in \mathbf{N}$ , let  $h(k) = k - 1$  if  $k/3$  has remainder 0 or 2, and  $h(k) = k + 2$  if  $k/3$  has remainder 1. To show that  $h$  is indeed the inverse of  $f$ , we apply the Bijectivity Test. Let  $k \in \mathbf{N}$ . We shall compute  $h(f(k))$  and  $f(h(k))$  in three separate cases.

If  $k/3$  has remainder 0, then  $f(k)/3 = (k - 2)/3 = (k - 3 + 1)/3$  has remainder 1 and  $h(k)/3 = (k - 1)/3$  has remainder 2, implying that  $h(f(k)) = h(k - 2) = k - 2 + 2 = k$  and  $f(h(k)) = f(k - 1) = k - 1 + 1 = k$ . So,  $h(f(k)) = k = f(h(k))$  in this case. Likewise, the same is true when  $k/3$  has remainder 1 or 2. Thus the Bijectivity Test implies that  $f$  is bijective with inverse  $h$ .  $\square$

**Example.** *Characteristic functions.* Let  $A$  be a given set. A function  $f : A \rightarrow \{0, 1\}$  is called a characteristic or a binary function on  $A$ . For each subset  $S$  of  $A$ , we can define a

characteristic function  $\chi_S : A \rightarrow \{0, 1\}$  as follows. Let  $\chi_S(a) = 1$  if  $a \in S$  and  $\chi_S(a) = 0$  if  $a \notin S$ . We call  $\chi_S$  the characteristic function of  $S$ .

**Proposition 3.3.** *If  $S$  and  $S'$  are distinct subsets of  $A$ , then  $\chi_S \neq \chi_{S'}$ .*

Proof: To prove  $\chi_S \neq \chi_{S'}$  is to prove that those two functions have different domains or that they have different targets or that they have different values at some  $a \in A$ . Since we know they have the same domain  $A$  and the same target  $\{0, 1\}$ , we want to show that they have different values at some point  $a \in A$ . Since  $S \neq S'$ , there is at least one  $a \in S - S'$  or there is at least one  $a \in S' - S$ . In the first case,  $\chi_S(a) = 1$  and  $\chi_{S'}(a) = 0$ . In the second case,  $\chi_{S'}(a) = 1$  and  $\chi_S(a) = 0$ . In each case, the values of  $\chi_S$  and  $\chi_{S'}$  are different at  $a$ .  
 $\square$

Given a characteristic function  $f : A \rightarrow \{0, 1\}$ ,  $f^{-1}(\{1\})$  is the subset of  $A$  on which the value of  $f$  is 1. Thus,  $f = \chi_S$  where  $S = f^{-1}(\{1\})$ . Let  $C(A)$  be the set consisting of all characteristic functions  $f : A \rightarrow \{0, 1\}$ . We have just shown that

$$C(A) = \{\chi_S \mid S \subset A\}.$$

**Corollary 3.4.** *Define the function  $\phi : P(A) \rightarrow C(A)$ ,  $S \mapsto \chi_S$ , from the power set  $P(A)$  to  $C(A)$ . Then  $\phi$  is a bijection.*

Proof: We have just shown that  $\phi$  is surjective. The preceding proposition says that  $\phi$  is injective.  $\square$

This shows that  $P(A) = \{\phi^{-1}(f) \mid f \in C(A)\}$ , allowing us to index the set  $P(A)$  using the set  $C(A)$  of characteristic functions on  $A$ .

There will be a homework problem in which you use this to find the size of  $P(\{1, 2, \dots, n\})$ .