MATH 147 Supplementary Material

Drawn from *Real analysis with real applications*
by Kenneth R. Davidson and Allan P. Donsig

1. The Language of Mathematics

The language of mathematics has to be precise, because mathematical statements must be interpreted with as little ambiguity as possible. Indeed, the rigour in mathematics is much greater than in law. There should be no doubts, reasonable or otherwise, when a theorem is proved. It is either completely correct, or it is wrong. Consequently, mathematicians have adopted a very precise language so that statements may not be misconstrued.

In complicated situations, it is easy to fool yourself. By being very precise and formal now, we can build up a set of tools that will help prevent mistakes later. The history of mathematics is full of stories in which mathematicians have fooled themselves with incorrect proofs. Clarity in mathematical language, like clarity in all other kinds of writing, is essential to communicating your ideas.

We begin with a brief discussion of the logical usage of certain innocuous words *if, then, only if, and, or* and *not*. Let A, B, C represent statements that may or may not be true in a specific instance. For example, consider the statements

A. It is raining.
B. The sidewalk is wet.

The statement "If A, then B" means that whenever A is true, it follows that B must also be true. We also formulate this as "A **implies** B." This statement does not claim either that the sidewalk is wet or that it is not. It tells you that if you look outside and see that it is raining, then without looking at the sidewalk, you will know that the sidewalk is wet as a result. As in the English language, "if A, then B" is a conditional statement meaning that only when the hypothesis A is verified can you deduce that B is valid. One also writes "Suppose A. Then B" with essentially the same meaning.

On the other hand, A implies B is quite different from B implies A. For example, the sidewalk may be wet because

C. The lawn sprinkler is on.

The statement "if B, then A" is known as the **converse** of "if A, then B." This amounts to reversing the direction of the implication. As you can see from this example, one may be true but not the other.

We can also say "A if B" to mean "if B, then A." The statement "B only if A" means that in order that B be true, it is necessary that A be true. A bit of thought reveals that this is yet another reformulation of "if A, then B." For reasons of clarity, these two expressions are rarely used alone and are generally restricted to the combined statement "A if and only if B." Parsing this sentence, we arrive at two statements "A if B" and "A only if B." The former means "B implies A" and the latter means "A implies B." Together they mean that either both statements are true or both are false. In this case, we say that statements A and B are **equivalent**.

The words *and*, *or*, and *not* are used with a precise mathematical meaning that does not always coincide with English usage. It is easy to be tripped up by these changes in meaning; be careful. "Not A" is the **negation** of the statement A. So "not A" is true if and only if A is false. To say that "A and B" is true, we mean that both A is true and B is true. On the

other hand, "A or B" is true when at least one is true, but both being true is also possible. For example, the statement "if A or C, then B' means that if either A is true or C is true, then B is true.

Consider these statements about an integer n:

D. n is even.
E. n is a multiple of 4.
F. There is an integer k so that $n = 4k + 2$.

The statement "not F" is "there is no integer k so that $n = 4k + 2$." The statement "D and not F" says that "n is even, and there is no integer k so that $n = 4k + 2$." One can easily check that this is equivalent to statement E. Here are some valid statements:

(1) D if and only if (E or F).
(2) If (D and not E), then F.
(3) If F, then D.

In the usual logical system of mathematics, a statement is either true or false, even if one cannot determine which is valid. A statement that is always true is a **tautology**. For example, "A or not A" is a tautology. A more complicated tautology known as **modus ponens** is "If A is true, and A implies B, then B is true." It is more common that a statement may be true or false depending on the situation. For example., statement D may be true or false depending on the value assigned to n.

The words *not* and *and* can be used together, but you must be careful to interpret statements accurately. The statement "not (A and B)" is true if (A and B) is false. If A is false, then (A and B) is false. Likewise if B is false, then (A and B) is false. While if both A and B are true, then (A and B) is true. So "not (A and B)" is true if either A is false or B is false. Equivalently, one of "not A" or "not B" is true. Thus "not (A and B)" means the same thing as "(not A) or (not B)."

This kind of thinking may sound pedantic, but it is an important way of looking at a problem from another angle. The statement "A implies B" means that B is true whenever A is true. Thus if B is false, A cannot be true, and thus A is false. That is, "not B implies not A." For example, if the sidewalk is not wet, then it is not raining. Conversely, if "not B implies not A", then "A implies B." Go through the same reasoning to see this through. You may have to use that "not (not A)" is equivalent to A. The statement "not B implies not A" is called the **contrapositive** of "A implies B." This discussion shows that the two statements are equivalent.

In addition to the converse and contrapositive of the statement "A implies B," there is the negation, "not (A implies B)." For "A implies B" to be false, there must be *some instance* in which A is true and B is false. Such an instance is called a **counterexample** to the claim that "A implies B." So the truth of A has no direct implication on the truth of B. For example, "not (C implies B)" means that it is possible for the lawn sprinkler to be on, yet the sidewalk remains dry. Perhaps the sprinkler is in the backyard, well out of reach of the sidewalk. It does not allow one to deduce any sensible conclusion about the relationship between B and C *except* that there are counterexamples to the statement "C implies B."

G. If 2 divides 3, then 10 is prime.
H. If 2 divides n, then $n^2 + 1$ is prime.

One common point of confusion is the fact that false statements can imply anything. For example, statement G is a tautology because the condition "2 divides 3" is never satisfied, so

one never arrives at the false conclusion. One the other hand, H is sometimes false (e.g., when $n = 8$).

Another important use of precise language in mathematics is the phrases **for every** (or **for all**) and **there exists**, which are known as **quantifiers**. For example,

I. For every integer n, the integer $n^2 - n$ is even.

This statement means that every substitution of an integer for n in $n^2 - n$ yields an even integer. This is correct because $n^2 - n = n(n - 1)$ is the product of the two integers n and $n - 1$, and one of them is even.

On the other hand, look at

J. For every integer $n \geq 0$, the integer $n^2 + n + 41$ is prime.

The first few terms 41, 43, 47, 53, 61, 71, 83, 97, 113, 131 are all prime. But to disprove this statement, it only takes a single instance where the statement fails. Indeed, $40^2+40+41 = 41^2$ is not prime. So this statement is false. We established this by demonstrating instead that

K. There is an integer n so that $n^2 + n + 41$ is not prime.

This is the negation of statement J, and exactly one of them is true.

Things can get tricky when several quantifiers are used together. Consider

L. For every integer m, there is an integer n so that 13 divides $m^2 + n^2$.

To verify this, one needs to take each m and prove that n exists. This can be done by noting that $n = 5m$ does the job since $m^2 + (5m)^2 = 13(2m^2)$. On the other hand, consider

M. For every integer m, there is an integer n so that 7 divides $m^2 + n^2$.

To disprove this, one needs to find just one m for which this statement is false. Take $m = 1$. To show that this statement is false for $m = 1$, it is necessary to check *every* n to make sure that $n^2 + 1$ is not a multiple of 7. This could take a rather long time by brute force. However observe that every integer may be written as $n = 7k \pm j$ where j is 0, 1, 2 or 3. Therefore

$$n^2 + 1 = (7k \pm j)^2 + 1 = 7(7k^2 \pm 2j) + j^2 + 1.$$

Note that $j^2 + 1$ takes the values 1, 2, 5 and 10. None of these is a multiple of 7, and thus all of these possibilities are eliminated.

The order in which quantifiers is critical. Suppose the words in the statement L are reordered as

N. There is an integer n so that for every integer m, 13 divides $m^2 + n^2$.

This has exactly the same words as statement L, but it claims the existence of an integer n that works with *every* choice of m. We can dispose of this by showing that for every possible n, there is at least one value of m for which the statement is false. Let us consider $m = 0$ and $m = 1$. If N is true, then for the number n satisfying the statement, we would have that both n^2+1 and n^2+0 are multiples of 13. But then 13 would divide the difference, which is 1. This contradiction shows that n does not validate statement N. As n was arbitrary, we conclude that N is false.

Exercises for Section 1

A. Which of the following are statements? That is, can they be true or false?
(a) Are all cats black?
(b) All integers are prime.
(c) $x + y$.
(d) $|x|$ is continuous.
(e) Don't divide by zero.

B. Which of the following statements implies which others?
(1) X is a quadrilateral.
(2) X is a square.
(3) X is a parallelogram.
(4) X is a trapezoid.
(5) X is a rhombus.

C. Give the converse and contrapositive statements of the following:
(a) An equilateral triangle is isosceles.
(b) If the wind blows, the cradle will rock.
(c) If Jack Sprat could eat no fat and his wife could eat no lean, then together they can lick the platter clean.
(d) (A and B) implies (C or D).

D. Which of the following statements is true? For those that are false, write down the negation of the statement.
(a) For every $n \in \mathbb{N}$, there is an $m \in \mathbb{N}$ so that $m > n$.
(b) For every $m \in \mathbb{N}$, there is an $n \in \mathbb{N}$ so that $m > n$.
(c) There is an $m \in \mathbb{N}$ so that for every $n \in \mathbb{N}$, $m \geq n$.
(d) There is an $n \in \mathbb{N}$ so that for every $m \in \mathbb{N}$, $m \geq n$.

E. Three young hoodlums accused of stealing CDs make the following statements:
(1) Ed: "Fred did it, and Ted is innocent."
(2) Fred: "If Ed is guilty, then so is Ted."
(3) Ted: "I'm innocent, but at least one of the others is guilty."
(a) If they are all innocent, who is lying?
(b) If all these statements are true, who is guilty?
(c) If the innocent told the truth and the guilty lied, who is guilty?
HINT: Remember that false statements imply anything.

F. Let A, B, C, D, E be statements. Make the following inferences.
(a) Suppose that (A or B) and (A implies B). Prove B.
(b) Suppose that ((not A) implies B) and (B implies (not C)) and C. Prove A.
(c) Suppose that (A or (not D)), ((A and B) implies C), ((not E) implies B), and D. Prove (C or E).

2. The Role of Proofs

Mathematics is all about proofs. Mathematicians are not as much interested in *what* is true as in *why* it is true. For example, you were taught in high school that the roots of the quadratic equation $ax^2 + bx + c = 0$ are $\dfrac{-b \pm \sqrt{b^2 - 4ac}}{2a}$ provided that $a \neq 0$. A serious class would not have been given this as a fact to be memorized. It would have been justified by the technique of *completing the square*. This raises the formula from the realm of magic to the realm of understanding.

There are several important reasons for teaching this argument. The first goes beyond intellectual honesty and addresses the real point, which is that you shouldn't accept mathematics (or science) on faith. The essence of scientific thought is understanding why things work out the way they do.

Second, the formula itself does not help you do anything beyond what it is designed to accomplish. It is no better than a quadratic solver button that could be built into your calculator. The numbers a, b, c go into a black box and two numbers come out or they don't—you might get an error message if $b^2 - 4ac < 0$. At this stage, you have no way of knowing if the calculator gave you a reasonable answer, or why it might give an error. If you know where the formula comes from, you can analyze all of these issues clearly.

Third, knowledge of the proof makes further progress a possibility. The creation of a new proof about something that you don't yet know is much more difficult than understanding the arguments someone else has written down. Moreover, understanding these arguments makes it easier to push further. It is for this reason that we can make progress. As Isaac Newton once said, "If I have seen further than others, it is by standing on the shoulders of giants." The first step toward proving things for yourself is to understand how others have done it before.

Fourth, if you understand that the *idea* behind the quadratic formula is completing the square, then you can always recover the quadratic formula whenever you forget it. This nugget of the proof is a useful method of data compression that saves you the trouble of memorizing a bunch of arcane formulae.

It is our hope that most students reading this book already have had some introduction to proofs in their earlier courses. If this is not the case, the examples in this section will help. This may be sufficient to tackle the basic material in this book. But be warned that some parts of this book require significant sophistication on the part of the reader.

Direct Proofs. We illustrate several proof techniques that occur frequently. The first is **direct proof**. In this technique, one takes a statement, usually one asserting the existence of some mathematical object, and proceeds to verify it. Such an argument may amount to a computation of the answer. On the other hand, it might just show the existence of the object without actually computing it. The crucial distinction for existence proofs is between those that are **constructive proofs**—that is, those that give you a method or algorithm for finding the object—and those that are **nonconstructive proofs**—that is, they don't tell you how to find it. Needless to say, constructive proofs do something more than nonconstructive ones, but they sometimes take more work.

Every real number x has a decimal expansion $x = a_0.a_1a_2a_3\ldots$, where a_i are integers and $0 \leq a_i \leq 9$ for all $i \geq 1$. This expansion is **eventually periodic** if there are integers N and $d > 0$ so that $a_{n+d} = a_n$ for $n \geq N$.

Occasionally a direct proof is just a straightforward calculation or verification.

2.1. Theorem. *If the decimal expansion of a real number x is eventually periodic, then x is rational.*

Proof. Suppose that N and $d > 0$ are given so that $a_{n+d} = a_n$ for $n \geq N$. Compute $10^N x$ and $10^{N+d}x$ and observe that

$$\begin{aligned} 10^{N+d}x &= b.a_{N+1+d}a_{N+2+d}a_{N+3+d}a_{N+4+d}\cdots \\ &= b.a_{N+1}a_{N+2}a_{N+3}a_{N+4}\cdots \\ 10^N x &= c.a_{N+1}a_{N+2}a_{N+3}a_{N+4}\cdots, \end{aligned}$$

where b and c are integers that you can easily compute. Subtracting the second equation from the first yields

$$(10^{N+d} - 10^N)x = b - c.$$

Therefore, $x = \dfrac{b-c}{10^{N+d} - 10^N}$ is a rational number. ■

The converse of this statement is also true. We will prove it by an existential argument that does not actually exhibit the exact answer, although the argument does provide a method for finding the exact answer. The next proof is definitely more sophisticated than a computational proof. It still, like the last proof, has the advantage of being constructive.

We need a simple but very useful fact.

2.2. Pigeonhole Principle.
If $n+1$ items are divided into n categories, then at least two of the items are in the same category.

This is evident after a little thought, and we do not attempt to provide a formal proof. Note that it has variants that may also be useful. If $nd+1$ objects are divided into n categories, then at least one category contains $d+1$ items. Also, if infinitely many items are divided into finitely many categories, then at least one category has infinitely many items.

2.3. Theorem. *If x is rational, then the decimal expansion of x is eventually periodic.*

Proof. Since x is rational, we may write it as $x = \frac{p}{q}$, where p, q are integers and $q > 0$. When an integer is divided by q, we obtain another integer with a remainder in the set $\{0, 1, \ldots, q-1\}$. Consider the remainders r_k when 10^k is divided by q for $0 \leq k \leq q$. There are $q+1$ numbers r_k, but only q possible remainders. By the Pigeonhole Principle, there are two integers $0 \leq k < k+d \leq q$ so that $r_k = r_{k+d}$. Therefore, q divides $10^{k+d} - 10^k$ exactly, say $qm = 10^{k+d} - 10^k$.

Now compute

$$\frac{p}{q} = \frac{pm}{qm} = \frac{pm}{10^{k+d} - 10^k} = 10^{-k}\frac{pm}{10^d - 1}.$$

Divide $10^d - 1$ into pm to obtain quotient a with remainder b, $0 \leq b < 10^d - 1$. So

$$x = \frac{p}{q} = 10^{-k}\big(a + \frac{b}{10^d - 1}\big),$$

where $0 \leq b < 10^d - 1$. Write $b = b_1 b_2 \ldots b_d$ as a decimal number with exactly d digits even if the first few are zero. For example, if $d = 4$ and $b = 13$, we will write $b = 0013$. Then consider

the periodic (or repeating) decimal

$$r = 0.b_1b_2\ldots b_db_1b_2\ldots b_db_1b_2\ldots b_d\ldots.$$

Using the proof of Theorem 2.1, we find that $(10^d-1)r = b$ and thus $r = \frac{b}{10^d-1}$. Observe that $10^k x = a + r = a.b_1b_2\ldots b_d\ldots$ has a repeating decimal expansion. The decimal expansion of $x = 10^{-k}(a+r)$ begins repeating every d terms after the first k. Therefore, this expansion is eventually periodic. ■

Proof by Contradiction. The second common proof technique is generally called **proof by contradiction**. Suppose that we wish to verify statement A. Now either A is true or it is false. We assume that A is false and make a number of logical deductions until we establish as true something that is clearly false. No false statement can be deduced from a logical sequence of deductions based on a valid hypothesis. So our hypothesis that A is false must be incorrect, whence A is true.

Here is a well-known example of this type.

2.4. Theorem. *$\sqrt{3}$ is an irrational number.*

Proof. Suppose to the contrary that $\sqrt{3} = a/b$, where a, b are positive integers with no common factor. (This proviso of no common factor is crucial to setting the stage correctly. Watch for where it gets used.) Manipulating the equation, we obtain

$$a^2 = 3b^2.$$

When the number a is divided by 3, it leaves a remainder $r \in \{0,1,2\}$. Let us write $a = 3k+r$. Then

$$\begin{aligned} a^2 &= (3k+r)^2 \\ &= 3(3k^2+2kr) + r^2 \\ &= \begin{cases} 9k^2 & \text{if } r = 0 \\ 3(3k^2+2k)+1 & \text{if } r = 1 \\ 3(3k^2+4k+1)+1 & \text{if } r = 2 \end{cases} \end{aligned}$$

Observe that a^2 is a multiple of 3 only when a is a multiple of 3. Therefore, we can write $a = 3c$ for some integer c. So $9c^2 = 3b^2$. Dividing by 3 yields $b^2 = 3c^2$.

Repeating exactly the same reasoning, we deduce that $b = 3d$ for some integer d. It follows that a and b do have a common factor 3, contrary to our assumption. The reason for this contradiction was the incorrect assumption that $\sqrt{3}$ was rational. Hence, $\sqrt{3}$ is irrational. ■

The astute reader might question why a fraction may be expressed in lowest terms. This is an easy fact that does not depend on deeper facts such as unique factorization into primes. It is merely the observation that if a and b have a common factor, then after it is factored out, one obtains a new fraction a_1/b_1 with a smaller denominator. This procedure must terminate by the time the denominator is reduced to 1, if not sooner. A very crude estimate of how many times the denominator can be factored is b itself.

The same reasoning is commonly applied to verify "A implies B." It is enough to show that "A and not B" is always false. For then if A is true, it follows that not B is false, whence B is true. This is usually phrased as follows: A is given as true. Assume that B is false. If we

can make a sequence of logical deductions leading to a statement that is evidently false, then given that A is true, our assumption that B was false is itself incorrect. Thus B is true.

Proof by Induction. The Principle of Induction is the mathematical version of the domino effect.

2.5. Principle of Induction. Let $P(n)$, $n \geq 1$, be a sequence of statements. Suppose that we can verify the following two statements:

(1) $P(1)$ is true.
(2) If $n > 1$ and $P(k)$ is true for $1 \leq k < n$, then $P(n)$ is true.

Then $P(n)$ is true for each $n \geq 1$.

We note that there is nothing special about starting at $n = 1$. For example, we can also start at $n = 0$ if the statements are numbered beginning at 0. You may have seen step (2) replaced by

($2'$) If $n > 1$ and $P(n-1)$ is true, then $P(n)$ is true.

This requires a stronger dependence on the previous statements and thus is a somewhat weaker principle. However, it is frequently sufficient.

Most students reading this book will have seen how to verify statements like

$$\sum_{k=1}^{n} k^3 = \big(\sum_{k=1}^{n} k\big)^2$$

by induction. As a quick warmup, we outline the proof that the sum of the first n odd numbers is n^2, that is,

$$\sum_{k=1}^{n} (2k-1) = n^2.$$

If $n = 1$, then both sides are 1 and hence equal. Suppose the statement if true for $n-1$, so that

$$\sum_{k=1}^{n-1} (2k-1) = (n-1)^2.$$

Then

$$\sum_{k=1}^{n} (2k-1) = (2n-1) + \sum_{k=1}^{n-1} (2k-1) = 2n - 1 + (n-1)^2 = n^2.$$

By induction, the statement is true for all integers $n \geq 1$.

Next, we provide an example that requires a bit more work and relies on the stronger version of induction. In fact, this example require two steps to get going, not just one.

2.6. Theorem. *The **Fibonacci sequence** is given recursively by*

$$F(0) = F(1) = 1 \quad \textit{and} \quad F(n) = F(n-1) + F(n-2) \quad \textit{for all} \quad n \geq 2.$$

Let $\tau = \dfrac{1+\sqrt{5}}{2}$. *Then* $F(n) = \dfrac{\tau^{n+1} - (-\tau)^{-n-1}}{\sqrt{5}}$ *for all* $n \geq 0$.

Proof. The statements are $P(n)$: $F(n) = \dfrac{\tau^{n+1} - (-\tau)^{-n-1}}{\sqrt{5}}$. Before we begin, observe that

$$\tau^2 = \Big(\frac{1+\sqrt{5}}{2}\Big)^2 = \frac{6+2\sqrt{5}}{4} = \frac{3+\sqrt{5}}{2} = \tau + 1.$$

Therefore, τ is a root of $x^2 - x - 1 = 0$. Now dividing by τ and rearranging yields

$$-\frac{1}{\tau} = 1 - \tau = \frac{1-\sqrt{5}}{2}.$$

Consider $n = 0$. It is generally better to begin with the complicated side of the equation and simplify it.

$$\frac{\tau^1 - (-\tau)^{-1}}{\sqrt{5}} = \frac{1}{\sqrt{5}}\Big(\frac{1+\sqrt{5}}{2} - \frac{1-\sqrt{5}}{2}\Big) = \frac{2\sqrt{5}}{2\sqrt{5}} = 1 = F(0).$$

This verifies the first step $P(0)$.

Right away we have a snag compared with a standard induction. Each $F(n)$ for $n \geq 2$ is determined by the two previous terms. But $F(1)$ does not fit into this pattern. It must also be verified separately.

$$\frac{\tau^2 - (-\tau)^{-2}}{\sqrt{5}} = \frac{1}{\sqrt{5}}\,\frac{(6+2\sqrt{5}) - (6-2\sqrt{5})}{4} = \frac{4\sqrt{5}}{4\sqrt{5}} = 1 = F(1).$$

This verifies statement $P(1)$.

Now consider the case $P(n)$ for $n \geq 2$, assuming that the statements $P(k)$ are known to be true for $0 \leq k < n$. In particular, they are valid for $k = n-1$ and $k = n-2$. Therefore,

$$\begin{aligned} F(n) &= F(n-1) + F(n-2) \\ &= \frac{\tau^n - (-\tau)^{-n}}{\sqrt{5}} + \frac{\tau^{n-1} - (-\tau)^{1-n}}{\sqrt{5}} \\ &= \frac{\tau^{n-1}(\tau+1) - (-\tau)^{-n}(1-\tau)}{\sqrt{5}} \\ &= \frac{\tau^{n-1}(\tau^2) - (-\tau)^{-n}(-\tau^{-1})}{\sqrt{5}} = \frac{\tau^{n+1} - (-\tau)^{-n-1}}{\sqrt{5}}. \end{aligned}$$

Thus $P(n)$ follows from knowing $P(n-1)$ and $P(n-2)$. The Principle of Induction now establishes that $P(n)$ is valid for each $n \geq 0$. ■

We will several times need a slightly stronger form of induction known as **recursion**. Simply put, the Principle of Recursion states that after an induction argument has been established, one has *all* of the statements $P(n)$. This undoubtedly seems to be what induction says. The difference is a subtle point of logic. Induction guarantees that each statement $P(n)$ is true, one at a time. To take all infinitely many of them at once requires a bit more. In order to deal with this rigorously, one needs to discuss the axioms of set theory, which takes us outside of these notes. However it is intuitively believable, and we will take this as valid.

Exercises for Section 2

A. Let $a \neq 0$. Prove that the quadratic equation $ax^2 + bx + c = 0$ has real solutions if and only if the discriminant $b^2 - 4ac \geq 0$. HINT: Complete the square.

B. Prove that the following numbers are irrational.
(a) $\sqrt[3]{2}$ (b) $\log_{10} 3$ (c) $\sqrt{3} + \sqrt[3]{7}$ (d) $\sqrt{6} - \sqrt{2} - \sqrt{3}$

C. Prove by induction that $\sum_{k=1}^{n} k^3 = \left(\sum_{k=1}^{n} k\right)^2 = \left(\frac{n(n+1)}{2}\right)^2$.

D. The **binomial coefficient** $\binom{n}{k} := \frac{n!}{k!(n-k)!}$.
(a) Prove that $\binom{n+1}{k} = \binom{n}{k} + \binom{n}{k-1}$.
(b) Prove by induction that $\sum_{k=0}^{n} \binom{n}{k} = 2^n$.
(c) Prove that $(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$ for all real numbers x and y.

E. Prove by induction that every integer $n \geq 2$ factors as the product of prime numbers. HINT: You need the statements $P(k)$ for all $2 \leq k < n$ here.

F. (a) Prove directly that if $a, b \geq 0$, then $\frac{a+b}{2} \geq (ab)^{1/2}$.
(b) If $a_1, \ldots, a_{2^n} \geq 0$, show by induction that $\dfrac{a_1 + \cdots + a_{2^n}}{2^n} \geq (a_1 a_2 \ldots a_{2^n})^{1/2^n}$.
(c) If $a_1, \ldots, a_m$ are positive numbers, choose $2^n \geq m$ and set $a_i = \dfrac{a_1 + \cdots + a_m}{m}$ for $m < i \leq 2^n$. Apply part (b) to deduce the **arithmetic mean–geometric mean inequality**, $\dfrac{a_1 + \cdots + a_m}{m} \geq (a_1 a_2 \ldots a_m)^{1/m}$.

G. Fix an integer $N \geq 2$. Consider the remainders $q(n)$ obtained by dividing the Fibonacci number $F(n)$ by N, so that $0 \leq q(n) < N$. Prove that this sequence is periodic with period $d \leq N^2$ as follows:
(a) Show that there are integers $0 \leq i < j \leq N^2$ such that $q(i) = q(j)$ and $q(i+1) = q(j+1)$. HINT: Pigeonhole.
(b) Show that if $q(i+d) = q(i)$ and $q(i+1+d) = q(i+1)$, then $q(n+d) = q(n)$ for all $n \geq i$. HINT: Use the recurrence relation for $F(n)$ and induction.
(c) Show that if $q(i+d) = q(i)$ and $q(i+1+d) = q(i+1)$, then $q(n+d) = q(n)$ for all $n \geq 0$. HINT: Work backward using the recurrence relation.

H. Consider the following "proof" by induction. We will argue that all students receive the same mark in calculus. *Let $P(n)$ be the statement that every set of n students receives the same mark. This is evidently valid for $n = 1$. Now look at larger n. Suppose that $P(n-1)$ is true. Given a group of n people, apply the induction hypothesis to all but the last person in the group. The students in this smaller group all have the same mark. Now repeat this argument with all but the first person. Combining these two facts, we find that all n students have the same mark. By induction, all students have the same mark.*

This is patently absurd, and you are undoubtedly ready to refute this by saying that Paul has a much lower mark than Mary. But you must find the mistake in the induction argument, not just in the conclusion.

3. The Real Numbers and Their Arithmetic

Doing calculus in a rigorous way starts with understanding the properties of the real numbers. Readers will be familiar, in some sense, with the real numbers from studying calculus. A completely rigorous development of the real numbers requires checking many details. We attempt to justify one definition of the real numbers without carrying out the proofs.

Intuitively, we think of the real numbers as the points on a line stretching off to infinity in both directions. However, to make any sense of this, we must label all the points on this line and determine the relationship between them from different points of view. First, the real numbers form an algebraic object known as a field, meaning that one may add, subtract, and multiply real numbers and divide by nonzero real numbers. There is also an order on the real numbers compatible with these algebraic properties, and this leads to the notion of distance between two points.

All of these nice properties are shared by the set of rational numbers:

$$\mathbb{Q} = \left\{ \frac{a}{b} : a, b \in \mathbb{Z}, b \neq 0 \right\}.$$

The ancient Greeks understood how to construct all fractions geometrically and knew that they satisfied all of the properties mentioned above. However, they were also aware that there were other points on the line that could be constructed but were not rational, such as $\sqrt{3}$. While the Greeks were focussed on those numbers that could be obtained by geometric construction, we have since found other reasonable numbers that do not fit this restrictive definition. The most familiar example is perhaps π, the area of a circle of radius one. Like the Greeks, we accept the fact that $\sqrt{3}$ and π are bona fide numbers that must be included on our real line.

We will define the real numbers to be objects with an infinite decimal expansion. A subtle point is that an infinite decimal expansion is used only as a name for a point and does mean the sum of an infinite series. It is crucial that we do not use limits to define the real numbers because we deduce properties of limits from the definition.

This construction of the real numbers appears to be strongly dependent on the choice of 10 as the base. We are left with the nagging possibility that the number line we construct depends on the number of digits on our hands. For this reason, some purists prefer a base-independent method of defining the real numbers, albeit a more abstract one. Our construction does yield the same object, independent of choice of base; but the proof requires considerable work.

We outline the construction of the real numbers using infinite decimal expansions, including the necessary operations. These operations extend the familiar ones on the rational numbers and our definitions of the new operations on the reals will use the existing operations on the rationals. We focus on the order and addition of real numbers, and the additive inverse, i.e., the negative of a real number.

We will define a real number by using an infinite decimal expansion such as

$$\begin{aligned}
\tfrac{1}{3} &= 0.33333333333333333333333333333333333333333333333333\ldots \\
\tfrac{1}{4} &= 0.25000000000000000000000000000000000000000000000000\ldots \\
\sqrt{3} &= 1.73205080756887729352744634150587236694280525381038\ldots \\
\pi &= 3.14159265358979323846264338327950288419716939937510 5\ldots \\
e &= 2.71828182845904523536028747135266249775724709369995 9\ldots
\end{aligned}$$

and, in general,

$$x = a_0.a_1a_2a_3a_4a_5a_6a_7a_8a_9a_{10}a_{11}a_{12}a_{10}a_{11}a_{12}a_{13}a_{14}a_{15}a_{16}a_{17}a_{18}\cdots .$$

That is, an **infinite decimal expansion** is a function from $\{0\} \cup \mathbb{N}$ into $\mathbb{Z}$ so that for all $n \geq 1$, n is mapped into $\{0, 1, \ldots, 9\}$.

It should be emphasized that at this point, the infinite decimal expansion is a *name* for a particular *real number*. Such an expansion has, a priori, no connection to the rational numbers or to any kind of infinite series. There is no notion of limits or convergence involved here.

To relate infinite decimal expansions to our geometric idea of the real line, start with a line and mark two points on it; and call the left-hand one 0 and the right-hand one 1. Then we can easily construct points for every integer $\mathbb{Z}$, equally spaced along the line. Now divide each interval from an integer n to $n + 1$ into 10 equal pieces, marking the cuts as $n.1$, $n.2$, ..., $n.9$. Proceed in this way, cutting each interval of length 10^{-k} into 10 equal intervals of length 10^{-k-1} and mark the endpoints by the corresponding number with $k+1$ decimals. In this way, all finite decimals are placed on the line.

Be warned that, by this construction, the point usually called $-5/4$ will be marked -2.75, for example, because we think of this as $-2 + .75$:

$$-\tfrac{5}{4} = -2.750000000000000000000000000000000000000000000000000\ldots$$

It turns out that some parts of the construction are simpler if we do this. After we have finished the construction, we will revert to the standard notation for negative decimals.

It seems clear that, given a number $a_0.a_1a_2a_3\ldots$, there will be a point on this line, call it x, with the property that for each positive integer k, x lies in the interval between the two rational numbers $y = a_0.a_1\ldots a_k$ and $y + 10^{-k}$. For example,

$$3.141592653589 \leq \pi \leq 3.141592653590. \tag{3.1}$$

The decimal expansion of x up to the kth decimal approximates x to an accuracy of at least 10^{-k}. We will use point of view to define the ordering on real numbers To make this precise, we'll have to define the ordering of real numbers.

Is there any point in distinguishing between the real number x and its infinite decimal expansion? It might seem that every real number should have a unique infinite decimal expansion, but this is not quite true. What about the real numbers determined by the expansions $1.000000000\ldots$ and $0.999999999\ldots$? Call them 1 and z, respectively. Clearly these are *different* infinite decimal expansions. However, for each positive integer k, we expect

$$1 - 10^{-k} = 0.\underbrace{99999999999999}_{k} \leq z \leq 1.$$

Thus the difference between z and 1 is arbitrarily small. It would create quite an unintuitive line if we decided to make z and 1 different real numbers. To fit in with our intuition, we must agree that $z = 1$. That means that some real numbers (precisely all those numbers with a finite decimal expansion) have two different expansions, one ending in an infinite string of zeros, and the other ending with an infinite string of nines. For example,

$$0.12500\ldots = 0.12499999\ldots.$$

This is an equivalence relation (see Exercise **??**). Formally, each real number is an equivalence class of infinite decimal expansions given by the identification defined above. Each equivalence class will have one or two elements, depending on whether or not the number is divisible by some power of 10. The set of all real numbers is denoted by $\mathbb{R}$.

To recognize the rationals as a subset of the reals, we need a function that sends a fraction a/b to an infinite decimal expansion. The rational numbers are distinguished among all real numbers by the fact that their decimal expansions are eventually periodic. See Theorems 2.1 and 2.3 in the Supplementary Material. Using the technique used in the proof of Theorem 2.3, we can define a function $F : \mathbb{Q} \to \mathbb{R}$ by defining $F(a/b)$ to be the infinite decimal expansion of a/b. This is a bit pedantic, and after this section, we will just think of $\mathbb{Q}$ as a subset of $\mathbb{R}$, i.e., identify a/b with the associated real number. The function F is injective, sending different fractions to different infinite decimal expansions (see Exercise **??**).

What we need to do next is to extend the ordering and the addition and multiplication operations on $\mathbb{Q}$ to all of $\mathbb{R}$. The following theorem gives all of the properties that we expect these new operations to have. However, there are many details to check.

3.2. Theorem. *The relation $<$ on $\mathbb{R}$ satisfies:*

(1) *For any $a, b, c \in \mathbb{R}$, if $a < b$ and $b < c$, then $a < c$.*
(2) *For any $a, b \in \mathbb{R}$, exactly one of the following is true:*
   (i) $a = b$,
   (ii) $a < b$,
   (iii) $b < a$.
(4) *If $r, s \in \mathbb{Q}$, then $r < s$ if and only if $F(r) < F(s)$.*

The operation $+$ satisfies:

(1) *For all $a, b \in \mathbb{R}$, $a + b = b + a$.*
(2) *For all $a, b, c \in \mathbb{R}$, $(a + b) + c = a + (b + c)$.*
(3) *There is an element $0_{\mathbb{R}} \in \mathbb{R}$ so that for all $a \in \mathbb{R}$, $a + 0_{\mathbb{R}} = a$.*
(4) *For each $a \in \mathbb{R}$, there is an element, called $-a$, so that $a + (-a) = 0_{\mathbb{R}}$.*
(5) *For all $r, s \in \mathbb{Q}$, $F(r + s) = F(r) + F(s)$.*

The operation $\times$ satisfies:

(6) *For all $a, b \in \mathbb{R}$, $a \times b = b \times a$.*
(7) *For all $a, b, c \in \mathbb{R}$, $(a \times b) \times c = a \times (b \times c)$.*
(8) *There is an element $1_{\mathbb{R}} \in \mathbb{R}$ so that for all $a \in \mathbb{R}$, $a \times 1_{\mathbb{R}} = a$.*
(9) *For each $a \in \mathbb{R} \backslash \{0_{\mathbb{R}}\}$, there is an element, called a^{-1}, in $\mathbb{R}$ so that $a \times a^{-1} = 1_{\mathbb{R}}$.*
(10) *For all $r, s \in \mathbb{Q}$, $F(r \cdot s) = F(r) \times F(s)$.*

The two operations $+$ and $\times$ together satisfy:

(11) *For all $a, b, c \in \mathbb{R}$, $a \times (b + c) = a \times b + a \times c$.*

The arithmetic operations relate to the order:

(12) *For all $a, b, c \in \mathbb{R}$, if $a > b$ then $a + c > b + c$.*
(13) *For all $a, b, c \in \mathbb{R}$, if $a > b$ and $c > 0$ then $ac > bc$.*

The Archimedean property is satisfied:

(14) *For all $a \in \mathbb{R}$, if $a \geq 0$ and for all $n \in \mathbb{N}$, $F(1/n) > a$, then $a = 0$.*

There are many properties that follow easily from these. For example, $0_{\mathbb{R}}$ is unique and $F(0) = 0_{\mathbb{R}}$, so we can identify $0_{\mathbb{R}}$ with 0. Similarly, $1_{\mathbb{R}}$ is unique and can be identified with $1 \in \mathbb{Q}$.

Instead of proving all the parts of this theorem, we explain how the ordering and addition are defined and sketch some of the arguments used to prove the required properties.

First, we have a built-in order on the line given by the placement of the points. This extends the natural order on the finite decimals. Notice that between any two distinct finite decimal numbers, there are (infinitely many) other finite decimal numbers. Now if x and y are distinct real numbers given by infinite decimal expansions, these expansions will differ at some finite point. This enables us to find finite decimals in between them. Because we know how an infinite decimal expansion should compare to its finite decimal approximants [using equations such as (**??**)], we can determine which of x or y is larger. For example, if

$$x = 2.7342118284590452354000064338325028841971693993\ldots \tag{3.3}$$

$$y = 2.7342118284590452353999928747135224977572470936\ldots \tag{3.4}$$

then $y < x$ because

$$y < 2.734211828459045235399993 < 2.734211828459045235400000 < x.$$

In fact, because their decimal expansions come from different real numbers, knowing that, in the first digit where they differ, that digit of y is less than the corresponding digit of x forces $y < x$.

We should also verify the order properties in Theorem **??**.

Second, we should extend the arithmetic properties of the rational numbers to all real numbers—namely addition, multiplication, and their inverse operations—and verify all of the field axioms. This is done by making all the operations consistent with the order. For example, if $x = x_0, x_1x_2\ldots$ and $y = y_0.y_1y_2\ldots$ are real numbers and k is a positive integer, then we have the finite decimal approximants

$$a_k = x_0.x_1\ldots x_k \le x \le a_k + 10^{-k} \quad \text{and} \quad b_k = y_0.y_1\ldots y_k \le y \le b_k + 10^{-k},$$

and so we want to have

$$a_k + b_k \le x + y \le a_k + b_k + 2\cdot 10^{-k}. \tag{3.5}$$

Since the lefthand and righthand sum use only rational numbers, we know what they are, and this determines the sum $x + y$ to an accuracy of $2 \cdot 10^{-k}$, for each k.

However, computing the exact sum of two infinite decimals is more subtle. The first digit of $x + y$ may not be determined exactly after any fixed finite number of steps, even though the sum can be determined to any desired accuracy. To see why this is the case, consider

$$\begin{aligned} x &= 0.\overbrace{999999\ldots999999}^{10^{15}\text{ nines}}\overbrace{0123456789\ldots0123456789}^{10^4\text{ repetitions}}31415\ldots \\ y &= 0.\underbrace{999999\ldots999999}_{10^{15}\text{ nines}}\underbrace{9876543210\ldots9876543210}_{10^4\text{ repetitions}}a9066\ldots. \end{aligned}$$

When we add $x + y$ using the first k decimal digits for any $k \le 10^{15}$, we obtain

$$1.\overbrace{999999\ldots999999}^{k-1\text{ nines}}8 \le x + y \le 2.\overbrace{000000\ldots000000}^{k\text{ zeros}}.$$

Taking $k = 10^{15}$ does not determine if the first digit of $x + y$ is 1 or 2, even though we know the sum to an accuracy of $2 \cdot 10^{-10^{15}} = 2/10^{1{,}000{,}000{,}000{,}000{,}000}$. When we proceed with the

computation using one more digit, we obtain

$$1.\overbrace{999999\ldots999999}^{10^{15}-1 \text{ nines}}89 \leq x+y \leq 1.\overbrace{999999\ldots999999}^{10^{15}-1 \text{ nines}}91.$$

All of a sudden, not only is the first digit determined, but so are the next $10^{15}-1$ digits.

A new period of uncertainty now occurs, again because of the problem that a long string of nines can *roll over* to a string of zeros like the odometer in a car. After using another 10^5 digits, we obtain a different result depending on whether $a \leq 4$, $a = 5$ or 6, or $a \geq 7$. When $a = 4$, we get

$$1.\overbrace{9999\ldots9999}^{10^{15}-1 \text{ nines}}8\overbrace{9999\ldots9999}^{10^4 \text{ nines}}7 \leq x+y \leq 1.\overbrace{9999\ldots9999}^{10^{15}-1 \text{ nines}}8\overbrace{9999\ldots9999}^{10^4 \text{ nines}}9.$$

So the digits are now determined for another 10^4+1 places. When $a = 7$, we obtain

$$1.\overbrace{9999\ldots9999}^{10^{15}-1 \text{ nines}}9\overbrace{000\ldots0000}^{10^4 \text{ zeros}}0 \leq x+y \leq 1.\overbrace{9999\ldots9999}^{10^{15}-1 \text{ nines}}9\overbrace{000\ldots0000}^{10^4 \text{ zeros}}2.$$

Again, the next 10^4+1 digits are now determined. However, when $a = 5$ or $a = 6$, these digits of the sum are still ambiguous. For $a = 5$, we have

$$1.\overbrace{9999\ldots9999}^{10^{15}-1 \text{ nines}}8\overbrace{9999\ldots9999}^{10^4 \text{ nines}}8 \leq x+y \leq 1.\overbrace{9999\ldots9999}^{10^{15}-1 \text{ nines}}9\overbrace{000\ldots0000}^{10^4 \text{ zeros}}0,$$

so the 10^{15}-th decimal digit is still undetermined.

The important thing to recognize is that these difficulties are not a serious impediment to defining the sum of two real numbers using infinite decimals. Suppose that, no matter how large k is, looking at the first k digits of x and y does not tell us if the first digit of $x+y$ is a 1 or a 2. In terms of Equation (**??**), this means that, for each k, the interval from $a_k + b_k$ to $a_k + b_k + 2 \cdot 10^{-k}$ contains 2. Since the length of the intervals goes to zero, it seems intuitively clear that the only real number in all of these intervals is 2.

In general, by considering *all* of the digits of x and y, we can write down a definition of $x+y$ as an infinite decimal. We may not be able to specify an algorithm to compute the sum, but then we cannot represent all of even one infinite decimal expansion in a computer either.

In real life, knowing the sum to, say, within 10^{-15} is much the same as knowing it to 15 decimal places. So we are content, on both theoretical and practical grounds, that we have an acceptable working model of addition.

Because of our non-standard definition of the infinite decimal expansions of negative numbers, constructing the negative of an infinite decimal is not just a matter of flipping the sign. If $x = x_0.x_1x_2\ldots$ represents a real number, we can define

$$-x = (-x_0 - 1).(9 - x_1)(9 - x_2)\ldots.$$

Then one can see that

$$x + (-x) = -1.999\ldots.$$

The right hand side is one of the two infinite decimal expansions for 0, so $-x$ is an additive inverse for x.

The issues are similar for the other arithmetic operations: multiplication and multiplicative inverses. It is crucial that these operations are consistent with order, as this means that they are also continuous (respect limits). Carrying out all the details of this program is tedious but not especially difficult.

The key points of this section are that we can define real numbers as infinite decimal expansions (with some identifications) and that we can define the order and all the field operations in terms of infinite decimals. Moreover, the result fits our intuitive picture of the real line, so we have the order and arithmetic properties that we expect.

Exercises for Section 3

A. If $x \neq y$, explain an algorithm to decide if $x < y$ or $y < x$. Does your method break down if $x = 0.9999\ldots$ and $y = 1.0000\ldots$?

B. If $a < b$ and $x < y$, is $ax < by$? What additional order hypotheses make the conclusion correct?

C. Define the **absolute value function** by $|x| = \max\{x, -x\}$.
(a) Prove that $|xy| = |x|\,|y|$ and $|x^{-1}| = |x|^{-1}$.
(b) Prove the **triangle inequality**: $|x + y| \leq |x| + |y|$. HINT: Consider x and y of the same sign and different signs as separate cases.

D. Prove by induction that $|x_1 + x_2 + \cdots + x_n| \leq |x_1| + |x_2| + \cdots + |x_n|$.

E. Prove the **reverse triangle inequality**: $\big||x| - |y|\big| \leq |x - y|$.

F. (a) Prove that if $x < y$, then there is a rational number r with a finite decimal expansion such that $x < r < y$.
(b) Prove that if $x < y$, then there is an irrational number z such that $x < z < y$.
HINT: Use (a) and add a small multiple of $\sqrt{2}$ to r.

G. Verify Property **??** of Theorem **??**, i.e., show that if $p, q \in \mathbb{Q}$, then $F(p) + F(q) = F(p + q)$.

H. (a) Explain how $x + y$ is worked out for

$$x = 2.1357\overbrace{999999\ldots999999}^{10^7 \text{ nines}}\overbrace{0123456789\ldots0123456789}^{10^{19} \text{ repetitions}}34524\ldots$$
$$y = 3.8642\underbrace{999999\ldots999999}_{10^7 \text{ nines}}\underbrace{9876543210\ldots9876543210}_{10^{19} \text{ repetitions}}39736\ldots.$$

(b) How many digits of x and y must we know to determine the first 6 digits of $x + y$?
(c) How many digits of x and y must we know to determine the first 10^8 digits of $x + y$?

I. Prove that the decimal expansion of a rational number is eventually periodic. Conversely, show that any decimal expansion which is eventually periodic represents a rational number. HINT: Any positive integer q divides $10^n - 10^m$ for some $m < n$ in $\mathbb{N}$.

J. Suppose that $r \neq 0$ is a rational number and that x is irrational. Show that $r + x$ and rx are irrational.

K. If m and n are integers, show that $\left|\sqrt{3} - \dfrac{m}{n}\right| \geq \dfrac{1}{5n^2}$.
HINT: Rationalize the numerator and use the irrationality of $\sqrt{3}$.

L. Define precisely the infinite decimal expansion associated to a fraction $a/b \in \mathbb{Q}$. Show that this function is one-to-one as a map from $\mathbb{Q}$ into the real numbers.

M. Describe an algorithm for adding two infinite decimals. You should work from 'left to right', determining the decimal expansion in order, as much as possible. When are you assured that you know the integer part of the sum? In what circumstance does it remain ambiguous?
HINT: Given infinite decimal expansions a and b, start by defining a carry function $\gamma : \{0\} \cup \mathbb{N} \to \{0, 1\}$ and then define the decimal expansion of $a + b$ in terms of $a(n) + b(n) + \gamma(n)$.

N. Explain why the associative property of addition for real numbers, $x + (y + z) = (x + y) + z$, follows from knowing it for for finite decimals.

4. Sets and Functions

Set theory is a large subject in its own right. We assume without discussion the existence of a sensible theory of sets and leave a full and rigorous development to books devoted to the subject. Our goal here to summarize the "intuitive" parts of set theory that we need.

Sets. A **set** is a collection of elements; for example, $A = \{0, 1, 2, 3\}$ is a set. This set has four elements, 0, 1, 2, and 3. The order in which they are listed is not relevant. A set can have other sets as elements. For example, $B = \{0, \{1, 2\}, 3\}$ has three elements, one of which is the set $\{1, 2\}$. Note that 1 is *not* an element of B, and that A and B are different.

We use $a \in A$ to denote that a is an element of the set A and $a \notin A$ to denote "not $(a \in A)$." The empty set $\varnothing$ is the set with no elements. We use the words *collection* and *family* as synonyms for sets. It is often clearer to talk about "a collection of sets" or "a family of sets" instead of "a set of sets." We say that two sets are equal if they have the same elements.

Given two sets A and B, we say A is a **subset** of B if every element of A is also an element of B. Formally, A is a subset of B if "$a \in A$ implies $a \in B$," or equivalently using quantifiers, $a \in B$ for all $a \in A$. If A is a subset of B, then we write $A \subset B$. This allows the possibility that $A = B$. It also allows the possibility that A has no elements, that is, $A = \varnothing$. We say A is a **proper subset** of B is $A \subset B$ and $A \neq B$. Notice that "$A \subset B$ and $B \subset A$" if and only if "$A = B$." Thus, if we want to prove that two sets, A and B, are equal, it is equivalent to prove the two statements $A \subset B$ and $B \subset A$.

You should recognize that there is a distinction between membership in a set and a subset of a set. For the sets A and B defined at the beginning of this section, observe that $\{1, 2\} \subset A$ and $\{1, 2\} \in B$. The set $\{1, 2\}$ is not a subset of B nor an element of A. However, $\big\{\{1, 2\}\big\} \subset B$.

There are a number of ways to combine sets to obtain new sets. The two most important are **union** and **intersection**. The union of two sets A and B is the set of all elements that are in A or in B, and it is denoted $A \cup B$. Formally, $x \in A \cup B$ if and only if $x \in A$ or $x \in B$. The intersection of two sets A and B is the set of all elements that are both in A and in B, and it is denoted $A \cap B$. Formally, $x \in A \cap B$ if and only if $x \in A$ and $x \in B$. Using our example, we have

$$A \cup B = \{0, 1, 2, \{1, 2\}, 3\} \quad \text{and} \quad A \cap B = \{0, 3\}.$$

Similarly, we may have an infinite family of sets A_γ indexed by another set Γ. What this means is that for every element γ of the set Γ, we have a set A_γ indexed by that element. For example, for n a positive integer, let A_n be the set of positive numbers that divide n, so that $A_{12} = \{1, 2, 3, 4, 6, 12\}$ and $A_{13} = \{1, 13\}$. Then this collection A_n is an infinite family of sets indexed by the positive integers, $\mathbb{N}$.

For infinite families of sets, intersection and union are defined formally in the same way. The union is

$$\bigcup_{\gamma \in \Gamma} A_\gamma = \{x : \text{there is a } \gamma \in \Gamma \text{ such that } x \in A_\gamma\}$$

and the intersection is

$$\bigcap_{\gamma \in \Gamma} A_\gamma = \{x : x \in A_\gamma \text{ for every } \gamma \in \Gamma\}.$$

In a particular situation, we are often working with a given set and subsets of it, such as the set of integers $\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$ and its subsets. We call this set our **universal set**. Once we have a universal set, say U, and a subset, say $A \subset U$, we can define the **complement** of A to be the collection of all elements of U that are not in A. The complement is denoted A'.

Notice that the universal set can change from problem to problem, and that this will change the complement.

Given a universal set U, we can specify a subset of U as all elements of U with a certain property. For example, we may define the set of all the integers that are divisible by two. We write this formally as

$$\{x \in \mathbb{Z} \text{ so that } 2 \text{ divides } x\}.$$

It is traditional to use a vertical bar | or a colon : for "so that," so that we can write the set of even integers as $2\mathbb{Z} = \{x \in \mathbb{Z} \mid 2 \text{ divides } x\}$. Similarly, we can write the complement of A in a universal set U as

$$A' = \{x \in U : x \notin A\}.$$

Given two sets A and B, we define the **relative complement** of B in A, denoted $A \setminus B$, to be

$$A \setminus B = \{x \in A : x \notin B\}.$$

Notice that B need not be a subset of A. Thus, we can talk about the relative complement of $2\mathbb{Z}$ in $\{0, 1, 2, 3\}$, namely

$$\{0, 1, 2, 3\} \setminus 2\mathbb{Z} = \{1, 3\}.$$

In our example, $A \setminus B = \{1, 2\}$. Curiously, $B \setminus A = \{\{1, 2\}\}$, the set consisting of the single element $\{1, 2\}$.

Finally, we need the idea of the **Cartesian product** of two sets, denoted $A \times B$. This is the set of ordered pairs $\{(a, b) : a \in A \text{ and } b \in B\}$. For example,

$$\{0, 1, 2\} \times \{2, 4\} = \{(0, 2), (1, 2), (2, 2), (0, 4), (1, 4), (2, 4)\}.$$

More generally, if $A_1, \ldots, A_n$ is a finite collection of sets, the Cartesian product is written $A_1 \times \cdots \times A_n$ or $\prod_{i=1}^{n} A_i$, and consists of **n-tuples**

$$A_1 \times \cdots \times A_n = \{(a_1, a_2, \ldots, a_n) : a_i \in A_i \text{ for } 1 \leq i \leq n\}.$$

If $A_i = A$ is the same set for each i, then we write A^n for the product of n copies of A. For example, $\mathbb{R}^3$ consists of all triples (x, y, z) with arbitrary real coefficients x, y, z. There is also a notion of the product of an infinite family of sets. We will not have any need of it, but we warn the reader that such infinite products raise subtle questions about the nature of sets.

Functions. A function f from A to B is a rule that assigns an element $f(a) \in B$ to each element $a \in A$. Such a rule may be very complicated with many different cases. In set theory, a very general definition of function is given that does not require the use of undefined terms such as *rule*. This definition specifies a function in terms of its graph,

$$G(f) = \{(a, f(a)) : a \in A\} \subset A \times B.$$

Not every subset of $A \times B$ is the graph of a function.

4.1. Definition. If A and B are nonempty sets, a subset $G(f) \subset A \times B$ is the **graph of a function** $f : A \to B$ if

- for each $a \in A$, there is exactly one $b \in B$ so that $(a, b) \in G(f)$,

If b is the *unique* element of B such that $(a, b) \in G(f)$, then we write $f(a) = b$.

The property of a subset of $A \times B$ that makes it the graph of a function is that $\{b \in B : (a, b) \in G(f)\}$ has precisely one element for each $a \in A$. This is the "vertical line test" for functions.

Sometimes we will use such convenient expressions as "the function x^{-1}." This really means "the function that sends x to x^{-1} for all x such that x^{-1} makes sense."

We call A the **domain** of the function $f : A \to B$ and B is the **codomain**. Far more important than the codomain is the **range** of f, which is

$$\operatorname{ran}(f) := \{b \in B : b = f(a) \text{ for some } a \in A\}.$$

If f is a function from A into B and $C \subset A$, the **image** of C under f is

$$f(C) := \{b \in B : \text{ there is some } c \in C \text{ so that } f(c) = b\}.$$

The range of f is $f(A)$.

Notice that the notation $f(r)$ has two possible meanings, depending on whether r is an element of A or a subset of A. The standard practice of using lowercase letters for elements and uppercase letters for sets makes this notation clear in practice.

The same caveat is applied to the notation f^{-1}. If f maps A into B, the **inverse image** of $C \subset B$ under f is

$$f^{-1}(C) = \{a \in A : f(a) \in C\}.$$

Note that f^{-1} is not used here as a function from B to A. Indeed, the domain of f^{-1} is the set of all subsets of B, and the codomain consists of all subsets of A. Even if $C = \{b\}$ is a single point, $f^{-1}(\{b\})$ may be the empty set or it may be very large. For example, if $f(x) = \sin x$, then

$$f^{-1}(\{0\}) = \{n\pi : n \in \mathbb{Z}\} \quad \text{and} \quad f^{-1}(\{y : |y| > 1\}) = \varnothing.$$

4.2. Definition. A function $f : A \to B$ maps A **onto** B or f is **surjective** if $\operatorname{ran}(f) = B$. In other words, for each $b \in B$, there is *at least one* $a \in A$ such that $f(a) = b$. Similarly, if $D \subset B$, say that f maps A **onto** D if $D \subset \operatorname{ran}(f)$.

A function $f : A \to B$ is **one-to-one** or **injective** if $f(a_1) = f(a_2)$ implies that $a_1 = a_2$ for $a_1, a_2 \in A$. In other words, for each b in the range of f, there is *at most one* $a \in A$ such that $f(a) = b$.

A function from A to B that is both one-to-one and onto is called a **bijection**.

Suppose that $f : A \to B$, $\operatorname{ran}(f) \subset B_0 \subset B$ and $g : B_0 \to C$; then the **composition** of g and f is the function $g \circ f(a) = g(f(a))$ from A into C.

A function is one-to-one if it passes a "horizontal line test." In this context, we can interpret f^{-1} as a function from B to A. This notion has a number of important consequences. In particular, when the ordered pairs in $G(f)$ are interchanged, the new set is the graph of a function known as the **inverse function** of f.

4.3. Lemma. *If $f : A \to B$ is a one-to-one function, then there is a unique one-to-one function $h : f(A) \to A$ so that*

$$h(f(a)) = a \text{ for all } a \in A \quad \text{and} \quad f(h(b)) = b \text{ for all } b \in f(A).$$

We call h the inverse function of f and denote it by f^{-1}.

Proof. Let $H \subset f(A) \times A$ be defined by

$$H = \{(b, a) \in f(A) \times A : (a, b) \in G(f)\}.$$

By definition of $f(A)$, for each $b \in f(A)$, there is an $a \in A$ with $(a, b) \in G(f)$. Since f is one to one, there is at most one $a \in A$ with $(a, b) \in G(f)$. Thus, there is exactly one $a \in A$ such that $(b, a) \in H$. Therefore H is the graph of a function h.

Suppose that $h(b_1) = h(b_2)$. Then there is some $a \in A$ so that (b_1, a) and (b_2, a) are in $G(h) = H$. Thus, (a, b_1) and (a, b_2) are in $G(f)$. But f is a function, so by the vertical line test, $b_1 = b_2$. Hence, h is one-to-one.

Finally, if $a \in A$, and $b = f(a)$, then $(b, a) \in G(h)$, so $h(b) = a$ and thus $h(f(a)) = h(b) = a$. Similarly, $f(h(b)) = b$ for all $b \in f(A)$. ■

We can express the relation between a one-to-one function and its inverse in terms of the identity maps. The **identity map** on a set A is $\mathrm{id}_A(a) = a$ for $a \in A$. When only one set A is involved, we use id instead of id_A.

4.4. Corollary. *If $f : A \to B$ is a bijection, then f^{-1} is a bijection and it is the unique function $h : B \to A$ so that $h \circ f = \mathrm{id}_A$ and $f \circ h = \mathrm{id}_B$. That is, $f^{-1}(f(a)) = a$ for all $a \in A$ and $f(f^{-1}(b)) = b$ for all $b \in B$.*

Exercises for Section 4

A. Which of the following statements is true? Prove or give a counterexample.
(a) $(A \cap B) \subset (B \cup C)$
(b) $(A \cup B') \cap B = A \cap B$
(c) $(A \cap B') \cup B = A \cup B$
(d) $A \setminus B = B \setminus A$
(e) $(A \cup B) \setminus (A \cap B) = (A \setminus B) \cup (B \setminus A)$
(f) $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
(g) $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
(h) If $(A \cap C) \subset (B \cap C)$, then $(A \cup C) \subset (B \cup C)$.

B. How many different sets are there that may be described using two sets A and B and as many intersections, unions, complements and parentheses as desired?
HINT: First show that there are four minimal nonempty sets of this type.

C. The **power set** $P(X)$ of a set X is the set consisting of all subsets of X, including $\varnothing$.
(a) Find a bijection between $P(X)$ and the set of all functions $f : X \to \{0, 1\}$.
(b) How many different subsets of $\{1, 2, 3, \ldots, n\}$ are there?

D. Let f be a function from A into X, and let $Y, Z \subset X$. Prove the following:
(a) $f^{-1}(Y \cap Z) = f^{-1}(Y) \cap f^{-1}(Z)$
(b) $f^{-1}(Y \cup Z) = f^{-1}(Y) \cup f^{-1}(Z)$
(c) $f^{-1}(X) = A$
(d) $f^{-1}(Y') = f^{-1}(Y)'$

E. Let $f : A \to X$, and let $B, C \subset A$. Prove the following statements. One of these statements may be sharpened to an equality. Prove it, and show by example that the others may be proper inclusions.
(a) $f(B \cap C) \subset f(B) \cap f(C)$
(b) $f(B \cup C) \subset f(B) \cup f(C)$
(c) $f(B) \subset X$
(d) If f is one-to-one, then $f(B') \subset f(B)'$.

F. Suppose that f, g, h are functions from $\mathbb{R}$ into $\mathbb{R}$. Prove or give a counterexample to each of the following statements.
(a) $f \circ g = g \circ f$
(b) $f \circ (g + h) = f \circ g + f \circ h$
(c) $(f + g) \circ h = f \circ h + g \circ h$

G. Suppose that $f : A \to B$ and $g : B \to A$ satisfy $g \circ f = \mathrm{id}_A$. Show that f is one-to-one and g is onto.

H. (a) What should a *two-to-one* function be?
(b) Give an example of a two-to-one function from $\mathbb{Z}$ onto $\mathbb{Z}$.

5. Appendix: Equivalence Relations

Equivalence relations occur frequently in mathematics and will appear occasionally later in this book.

5.1. Definition. Let X be a set, and let R be a subset of $X \times X$. Then R is a **relation** on X. Let us write $x \sim y$ if $(x, y) \in R$. We say that R or $\sim$ is an **equivalence relation** if it is

(1) **(reflexive)** $x \sim x$ for all $x \in X$.
(2) **(symmetric)** if $x \sim y$ for any $x, y \in X$, then $y \sim x$.
(3) **(transitive)** if $x \sim y$ and $y \sim z$ for any $x, y, z \in X$, then $x \sim z$.

If $\sim$ is an equivalence relation on X and $x \in X$, then the **equivalence class** $[x]$ is the set $\{y \in X : y \sim x\}$. By $X/\!\sim$ we mean the collection of all equivalence classes.

5.2. Examples.

(1) Equality is an equivalence relation on any set. Verify this.

(2) Consider the integers $\mathbb{Z}$. Say that $m \equiv n \pmod{12}$ if 12 divides $m-n$. Note that 12 divides $n - n = 0$ for any n, and thus $n \equiv n \pmod{12}$. So it is reflexive. Also if 12 divides $m - n$, then it divides $n - m = -(m - n)$. So $m \equiv n \pmod{12}$ implies that $n \equiv m \pmod{12}$ (i.e., symmetry). Finally, if $l \equiv m \pmod{12}$ and $m \equiv n \pmod{12}$, then we may write $l - m = 12a$ and $m - n = 12b$ for certain integers a, b. Thus $l - n = (l - m) + (m - n) = 12(a + b)$ is also a multiple of 12. Therefore, $l \equiv n \pmod{12}$, which is transitivity.

There are twelve equivalence classes $[r]$ for $0 \le r < 12$ determined by the remainder r obtained when n is divided by 12. So $[r] = \{12a + r : a \in \mathbb{Z}\}$.

(3) Consider the set $\mathbb{R}$ with the relation $x \le y$. This relation is reflexive ($x \le x$) and transitive ($x \le y$ and $y \le z$ implies $x \le z$). However, it is **antisymmetric**: $x \le y$ and $y \le x$ both occur if and only if $x = y$. This is not an equivalence relation.

When dealing with functions defined on equivalence classes, we often define the function on an equivalence class in terms of a representative. In order for the function to be well defined, that is, for the definition of the function to make sense, we must check that we get same value regardless of which representative is used.

5.3. Examples.

(1) Consider the set of real numbers $\mathbb{R}$. Say that $x \equiv y \pmod{2\pi}$ if $x - y$ is an integer multiple of 2π. Verify that this is an equivalence relation. Define a function $f([x]) = (\cos x, \sin x)$. We are really defining a function $F(x) = (\cos x, \sin x)$ on $\mathbb{R}$ and asserting that $F(x) = F(y)$ when

$x \equiv y \pmod{2\pi}$. Indeed, we then have $y = x + 2\pi n$ for some $n \in \mathbb{Z}$. Since sin and cos are 2π-periodic, we have

$$F(y) = (\cos y, \sin y) = (\cos(x + 2\pi n), \sin(x + 2\pi n)) = (\cos x, \sin x) = F(x).$$

It follows that the function $f([x]) = F(x)$ yields the same answer for every $y \in [x]$. So f is well defined. One can imagine the function f as wrapping the real line around the circle infinitely often, matching up equivalent points.

(2) Consider $\mathbb{R}$ modulo 2π again, and look at $f([x]) = e^x$. Then $0 \equiv 2\pi \pmod{2\pi}$ but $e^0 = 1 \neq e^{2\pi}$. So f is not well defined on equivalence classes.

(3) Now consider Example **??** (2). We wish to define multiplication modulo 12 by $[n][m] = [nm]$. To check that this is well defined, consider two representatives $n_1, n_2 \in [n]$ and two representatives $m_1, m_2 \in [m]$. Then there are integers a and b such that $n_2 = n_1 + 12a$ and $m_2 = m_1 + 12b$. Then

$$n_2 m_2 = (n_1 + 12a)(m_1 + 12b) = n_1 m_1 + 12(am_1 + n_1 b + 12ab).$$

Therefore, $n_2 m_2 \equiv n_1 m_1 \pmod{12}$, and multiplication modulo 12 is well defined.

Exercises for Section 5

A. Put a relation on $C[0,1]$ by $f \sim g$ if $f(k/10) = g(k/10)$ for k with $0 \leq k \leq 10$.
(a) Verify that this is an equivalence relation.
(b) Describe the equivalence classes.
(c) Show that $[f] + [g] = [f + g]$ is a well-defined operation.
(d) Show that $t[f] = [tf]$ is well defined for all $t \in \mathbb{R}$ and $f \in C[0,1]$.
(e) Show that these operations make $C[0,1]/\sim$ into a vector space of dimension 11.

B. Consider the set of all infinite decimal expansions $x = a_0.a_1a_2a_3\ldots$, where a_0 is any integer and a_i are digits between 0 and 9 for $i \geq 1$. Say that $x \sim y$ if x and y represent the same real number. That is, if $y = b_0.b_1b_2b_3\ldots$, then $x \sim y$ if (1) $x = y$, or (2) there is an integer $m \geq 1$ such that $a_i = b_i$ for $i < m - 1$, $a_{m-1} = b_{m-1} + 1$, $b_i = 9$ for $i \geq m$ and $a_i = 0$ for $i \geq m$, or (3) there is an integer $m \geq 1$ such that $a_i = b_i$ for $i < m - 1$, $a_{m-1} + 1 = b_{m-1}$, $a_i = 9$ for $i \geq m$ and $b_i = 0$ for $i \geq m$. Prove that this is an equivalence relation.

C. Define a relation on the set $PC[0,1]$ of all piecewise continuous functions on $[0,1]$ by $f \approx g$ if $\{x \in [0,1] : f(x) \neq g(x)\}$ is finite.
(a) Prove that this is an equivalence relation.
(b) Decide which of the following functions are well defined.

(i) $\varphi([f]) = f(0)$ (ii) $\psi([f]) = \int_0^1 f(t)\,dt$ (iii) $\gamma([f]) = \lim_{x \to 1^-} f(x)$

D. Let $d \geq 2$ be an integer. Define a relation on $\mathbb{Z}$ by $m \equiv n \pmod{d}$ if d divides $m - n$.
(a) Verify that this is an equivalence relation, and describe the equivalence classes.
(b) Show that $[m] + [n] = [m + n]$ is a well-defined addition.
(c) Show that $[m][n] = [mn]$ is a well-defined multiplication.
(d) Let $\mathbb{Z}_d$ denote the equivalence classes modulo d. Prove the distributive law:
$[k]([m] + [n]) = [k][m] + [k][n]$.

E. Say that two real vector spaces V and W are **isomorphic** if there is an invertible linear map T of V onto W.
(a) Prove that this is an equivalence relation on the collection of all vector spaces.
(b) When are two finite-dimensional vector spaces isomorphic?