Foundation of Mathematics

Min Yan

August 26, 2022

Contents

Chapter 1

Logic

1.1 Statement

The following are some *statements*:

- 10 is divisible by 2.
- 2 is divisible by 10.
- n is divisible by 10.
- 2 loves candy.

Based on our common knowledge, some statements can be determined as *true* or *false*. For example, the first statement above is true, and the second is false. However, not every statement has to be true or false by itself. For example, the third statement is true or false depending on the value of n. In other words, the truth of the statement is *conditional*. Moreover, the fourth statement above is simply *meaningless*, and it is also meaningless to say it is true or false.

Exercise 1.1. Which statements are true or false?

1. 5 is bigger than 2.
2. 5 is smaller than 2.
3. 2 is smaller than 5.
4. 5 is not bigger than 2.
5. $5 - 2$ is positive.
6. 5 is bigger than $2n$.
7. $5a$ is bigger than $2a$.

Equivalent statements are different ways of saying the same thing. The following statements are equivalent:

- Bill is older than Bob.
- Bob is younger than Bill.

The following are equivalent mathematical statements:

- 10 is divisible by 2.
- There is an integer n such that $10 = 2n$.
- 10 is an even number.

The following statements are also equivalent:

- The cube of the largest of three consecutive integers cannot be the sum of the cubes of the other two.
- There is no integer n, such that $n^3 + (n+1)^3 = (n+2)^3$.
- If $n^3 + (n+1)^3 = (n+2)^3$, then n cannot be an integer.

We often need to rephrase a mathematical statement into an equivalent one in order to determine whether the statement is true or not.

The following statements are *opposite* (or *negation*) of each other:

- 10 is an even number.
- 10 is an odd number.

The opposite of a statement A is simply "not A". For example, the above pair are opposite because "odd" means "not even" (and "even" means "not odd"). In practice, the statement "not A" is often further rephrased in a more familiar form, such as the following pair of opposite statements:

- The sum of two odd numbers is even.
- There are odd numbers m and n, such that $m + n$ is still odd.

The following is another pair of opposite statements:

- All dogs have four legs.
- Some dogs do not have four legs.

Here "some" means "at least one".

Exercise 1.2. Which statements in Exercise 1.1 are equivalent? Which are opposite?

Statements can be combined by using *and*. The new statement is the *conjunction*. For example, combining the statements "n is divisible by 2" and "n is divisible by 3", we get a new statement

- n is divisible by 2 and 3.

The statement is actually equivalent to

- n is divisible by 6.

For another example, we combine the statements "A dog has four legs" and "A dog has a tail" to get

- A dog has four legs and a tail.

Statements can also be combined by using *or*. The new statement is the *disjunction*. From the statements "n is divisible by 2" and "n is divisible by 3", we get a new statement

- n is divisible by either 2 or 3.

For example, $n = 2, 3, 4, 6, 8, 9, \dots$ fit into the statement, while $n = 1, 5, 7, 11, 13, \dots$ do not fit. The new statement is also equivalent to (any one of) the following two statements:

- If n is not divisible by 2, then n is divisible by 3.
- If n is not divisible by 3, then n is divisible by 2.

For another example, the following are equivalent:

- A dog is male <u>or</u> a dog is female.
- A dog is either male or female.

Exercise 1.3. Which statements are equivalent to "n is divisible by 6"? Which are opposite?

1. n is divisible by 2.
2. There is an integer k such that $n = 3k$.
3. $n = 6k$ for some integer k.
4. $n = 2k$ and $n = 3k'$ for some integers k and k'.
5. n is divisible by 2 or 3.
6. n is not divisible by 3.

7. n is not divisible by 2 and not by 3.

8. n is not divisible by 2 or not by 3.

9. If n is not divisible by 2, then n is not divisible by 3.

Exercise 1.4. Rewrite the following statements into equivalent ones and opposite ones using mathematical symbol expressions.

1. 5 is an odd number.

2. $\sqrt{2}$ is an irrational number.

3. The square of an even number must be even.

4. 10 is not the sum of squares of two integers.

5. The polynomial $t^4 + 2t^3 + 3t + 10$ is not a product of quadratic polynomials with integer coefficients.

Now we discuss how the various concepts about the statements interact with each other. We use capital letters $A, B, C, \dots$ to denote statements.

First, we note that if A and B are equivalent, and B and C are equivalent, then A and C are also equivalent. For an example, since the following are equivalent (by the definition of "divisible by 2"),

- 10 is divisible by 2.
- There is an integer n such that $10 = 2n$.

and the following are equivalent (by the definition of "even number"),

- 10 is divisible by 2.
- 10 is an even number.

we conclude that the following are also equivalent

- There is an integer n such that $10 = 2n$.
- 10 is an even number.

Second, if A and B are opposite, and B and C are equivalent, then A and C are also opposite. For example, the following are opposite,

- $\sqrt{2}$ is an irrational number.
- $\sqrt{2}$ is a rational number.

and the by the definition of rational number, the following are equivalent.

- $\sqrt{2}$ is a rational number.
- $\sqrt{2}$ is the quotient of two integers.

Then we know the following are opposite.

- $\sqrt{2}$ is an irrational number.
- $\sqrt{2}$ is the quotient of two integers.

The opposite of "A *and* B" is "(not A) *or* (not B)", instead of "(not A) and (not B)". For example, the following are opposite statements:

- (*Both*) m and n are even numbers.
- *At least one of* m and n is an odd number.

and the second one is the same as the following.

- *Either* m is an odd number *or* n is an odd number.

Similarly, the opposite of "A or B" is "(not A) *and* (not B)" ("neither A nor B" in more proper English). For example, the following are opposite statements:

- A dog is either male or female.
- A dog is neither male nor female.

The opposite of combinations of statements is similar to the following example of opposite statements:

- A dog has four legs.
- Some dog does not have four legs.

We emphasis that the opposite of "all do (are)" is "some do not (are not)"[1], and the opposite of "some do (are)" is "all do not (are not)".

By common sense, we get more ways of how equivalence and opposite can interact. See Exercise 1.6.

Exercise 1.5. Write down the opposites of the following statements in the most natural way.

1. 10 is divisible by 2 and 5.
2. 10 is not divisible by 3 and 7.

[1] Which is the same as "not all do (are)".

3. There are no integers m and n, such that $10 = m^2 + n^2$.
4. $m > n$ or $m < n$ (m and n are integers).
5. A non-negative number has either one or two square roots.
6. An integer is either positive, or 0, or negative.
7. All students in the class love mathematics.
8. The students in the class are 19 years old and born in Hong Kong.
9. Some students in the class are from Beijing or Shanghai.
10. If I do not feel well, then I will not play badminton.

Exercise 1.6. True or false. The capital letters denote statements.

1. A is equivalent to A
2. If A is equivalent to B, then B is equivalent to A
3. A is opposite to A
4. If A is opposite to B, then B is opposite to A
5. If A is equivalent to B and B is equivalent to C, then A is equivalent to C
6. If A is opposite to B and B is opposite to C, then A is opposite to C
7. If A is opposite to B and B is opposite to C, then A is equivalent to C
8. If A is equivalent to B and C is equivalent to D, then "A and C" is equivalent to "B and D"
9. If A is equivalent to B and C is opposite to D, then "A and C" is equivalent to "B or D"
10. If A is opposite to B and C is opposite to D, then "A and C" is opposite to "B and D"
11. If A is opposite to B and C is opposite to D, then "A or C" is opposite to "D and C"
12. "A and (not B)" is opposite to "(not A) and B"
13. "(not A) and (not B)" is opposite to "not (A and B)"
14. "(not A) and B" is opposite to "A or (not B)"
15. "not (A or B)" is opposite to "(not A) or (not B)"

1.2 Implication

Two statements may be related by *implication*. For example, if the following statement is true,

- n is divisible by 10.

then the statement

- n is an even integer.

also holds. We say the first statement *implies* the second statement. For another example, the statement

- Clifford is a dog, and all dogs have four legs.

implies the statement

- Clifford has four legs.

There are many different ways of presenting an implication. The following are some common expressions:

- A implies B.
- Given A, B is true.
- If A, then B.
- If not B, then not A.
- A, if B.

For example, we can say the following.

- Being divisible by 10 implies being even.
- Given a multiple of 10, the number must be even.
- If n is divisible by 10, then n is an even integer.
- If a number is odd, then it is not divisible by 10.
- A number is even if it is divisible by 10.

Here are more examples:

- Bigger investment implies bigger return.
- Given the heavy rain, we have to cancel the outdoor program.

- If we are healthy, then we will not take medicine.
- We take medicine if we get ill.
- Bad grade is a consequence of not studying hard[2].

If A implies B, then we call A the *sufficient condition* and call B the *necessary condition*. The sufficient condition is the cause and the reason. The necessary condition is the result, the consequence, and the conclusion. In the examples above, "bigger investment", "heavy rain", "healthy", "get ill", "not studying hard" are the sufficient conditions, and "bigger return", "cancel the outdoor program", "not take medicine", "take medicine", "bad grade" are the necessary conditions.

We often use the notation $A \implies B$ or $B \impliedby A$ to indicate an implication. For example, we may write

- (n is) divisible by 10 $\implies$ (n is) even.
- We take medicine $\impliedby$ We get ill.

Naturally, if $A \implies B$ and $B \implies C$, then we may conclude $A \implies C$. For example, by combining the following implications

- divisible by 30 $\implies$ divisible by 10.
- divisible by 10 $\implies$ even.

we get

- divisible by 30 $\implies$ even.

Exercise 1.7. Determine which one implies which one.

1. n is divisible by 2.
2. There is an intger k such that $n = 3k$.
3. $n = 6k$ for some integer k.
4. $n = 2k$ and $n = 3k'$ for some integers k and k'.
5. n is divisible by 2 or 3.
6. n is not divisible by 3.
7. n is not divisible by 2 and not by 3.
8. n is not divisible by 2 or not by 3.

[2] This is yet another way of expressing implication.

9. If n is not divisible by 2, then n is not divisible by 3.

By common sense, the equivalence between A and B means $A \implies B$ and $B \implies A$. We write $A \iff B$ to indicate equivalent statements. Therefore the following sentences mean the same:

- A is equivalent to B.
- B is the necessary and sufficient condition for A.
- A if and only if (often abbreviated as iff) B.

Here are some typical examples:

- An even number is equivalent to multiple of 2.
- $m > n$ is the necessary and sufficient condition for $-m < -n$.
- An integer is a multiple of 6 iff it is a multiple of 2 and a multiple of 3.
- Scoring above 60 is necessary and sufficient for you to pass the course.
- Bill is older than Bob if and only if Bob is younger than Bill.

Common sense also tells us that $A \implies B$ is the same as (not B) $\implies$ (not A). We have seen the following equivalent statements before:

- If n is divisible by 10, then n is an even integer.
- If a number is odd, then it is not divisible by 10.

Here is another example:

- If we are healthy, then we will not take medicine.
- We take medicine if we get ill.

Again by common sense, we know that "$A \implies$ (B or C)" is the same as "(A and not B) $\implies C$". For example, the following are equivalent statements:

- If n is a nonzero number, then n is positive or negative.
- If n is a nonzero number and is not positive, then n is negative.

The following are also equivalent:

- Given that we cannot fly and we want to travel to Shanghai, we will ride a train.
- If we want to travel to Shanghai, then we will either fly an airplane or ride a train.

Next, we discuss the opposite of implication. Specifically, the opposite of the statement "$A \implies B$" means it is possible for A to be true, but B is still not true. In other words, the opposite of "$A \implies B$" is "A and (not B)". For example, the opposites of

- If you work hard, then you get good grade.
- If it rains, then we cancel the game.

are

- You may work hard but still get bad grade.
- We still have the game in case of rain.

The opposites of

- If n is a multiple of 2, then n is a multiple of 3.
- If $n > 10$, then x_n is even.
- If $m > n$, then $x_m > x_n$.

are

- There is an n that is a multiple of 2 but not a multiple of 3.
- x_n is odd for some $n > 10$.
- There are some $m_0 > n_0$, such that $x_{m_0} \le x_{n_0}$.

Exercise 1.8. Determine which ones are equivalent and which ones are opposite.

1. If he wears black shirt and carries an umbrella, then he is either Bill or Bob.
2. If he wears black shirt or carries an umbrella, then he is either Bill or Bob.
3. If he wears white shirt and carries an umbrella, then he is neither Bill nor Bob.
4. If he wears white shirt or does not carry an umbrella, then he is neither Bill nor Bob.
5. If he wears black shirt and is not Bill, then he does not carry an umbrella or he is Bob.
6. If he wears white shirt and is not Bill, then he does not carry an umbrella or he is Bob.
7. If he carries an umbrella and he is not Bob, then he does not ware white shirt or he is Bill.

8. If he wears black shirt or carries an umbrella, and he is not Bob, then he is either Bill.

9. If he is neither Bill nor Bob, then he either does not wears black shirt or does not carry an umbrella.

10. If he is either Bill or Bob, and he wears white shirt, then he does not carry an umbrella.

Exercise 1.9. Write down the opposite.

1. If n is a multiple of 2, then n is not a multiple of 3.

2. If n is a not multiple of 2, then n is a multiple of 3.

3. If $n < 10$, then x_n is even.

4. If $m > n$, then $x_m \leq x_n$.

5. If $m < n$, then $x_m > x_n$.

Exercise 1.10. True or false.

1. "$A \iff B$" is the same as "$(A \implies B)$ and $((\text{not } A) \implies (\text{not } B))$".

2. "$(A \implies B)$ and $(A \implies C)$" is the same as "$A \implies (B \text{ and } C)$".

3. "$(A \implies C)$ and $(B \implies C)$" is the same as "$(A \text{ and } B) \implies C$".

4. "$(A \text{ and } B) \implies C$" implies "$A \implies C$".

5. "$(A \text{ or } B) \implies C$" implies "$A \implies C$".

6. "$A \implies (B \text{ and } C)$" implies "$A \implies B$".

7. "$A \implies (B \text{ or } C)$" implies "$A \implies B$".

8. "$(A \text{ and } B) \implies C$" is the same as "$(A \text{ and } (\text{not } C)) \implies (\text{not } B)$".

9. "$(A \text{ or } B) \implies C$" is the same as "$(A \text{ or } (\text{not } C)) \implies (\text{not } B)$".

10. "$(A \text{ and } B \text{ and } C) \implies D$" is the same as "$(A \text{ and } B) \implies ((\text{not } C) \text{ or } D)$".

1.3 Quantifier

Pay attention to the italic words in the following statements:

- *All* numbers are even or odd.
- *Some* numbers are even.
- *There are* numbers divisible by 2 and 3.
- *All* numbers divisible by 6 are also divisible by 3.
- *All* men are equal.
- *Something* is wrong.
- *Everything* is wrong.
- *For every* number, *there is* a bigger number.
- *Among all* the candidates, *at least one* will be qualified.

Note the two kinds of *quantifiers* used in the statements: Universal quatifier (all, every) and existential quantifier (some, there is, at least one). It is also possible for both kinds of quantifiers to appear in the same statement.

In mathematics, we often need to deal with the statements of the form "for all n, $A(n)$ happens" or "for some n, $A(n)$ happens", where $A(n)$ is a statement depending on a parameter n. The following are some concrete examples:

- For all number n, n^2 is non-negative.
- For all even number n, n^2 is even.
- For every rational number r, there are integers m and n, such that $a = \dfrac{m}{n}$.
- For all n, $x_n > y_n$.
- $x_n \leq y_n$ for some n.
- There is a number n, such that 2 divides n and 3 does not divide $n + 1$.
- $10 = 2n$ for some n.

In the third statement above, the double quantifier "there are m and n" are used. The following are more examples of double quantifiers:

- For any m and n, either $m > n$, or $m = n$, or $m < n$.
- For all $m, n > 0$, we have $mn > 0$.

- For any rational numbers $r > s$, there is a rational number p, such that $r > p > s$.

- For any $\epsilon > 0$, there is n, such that for any $m > n$, we have $|x_m - a| < \epsilon$.

Note that in the last example has three quantifiers.

Exercise 1.11. Determine which statement is true.

1. For any integer n, there an integer k, such that either $n = 2k$ or $n = 2k + 1$.
2. There is an intger k such that $10 = 3k + 1$.
3. $20 = 3k + 1$ for some integer k.
4. For every rational number r, $2r$ is an integer.
5. For some rational number r, $2r$ is an integer.
6. $(m + n)^2 = m^2 + 2mn + n^2$ for any numbers m and n.
7. $(m + n)^2 = m^2 + n^2$ for some nonzero numbers m and n.
8. $(m + n)^3 = m^3 + n^3$ for some nonzero numbers m and n.
9. For any number $\epsilon > 0$, there is a number n, such that for every $m > n$, we have $m\epsilon < 1$.
10. There is a number $\epsilon > 0$, such that for any number n, we can find a number m, such that $1 < m - n < \epsilon$.

Next we discuss the opposite of quantified statements. For the statement

- For all n, $x_n > y_n$.

to be wrong, it means that $x_n > y_n$ does not hold for all n. In other words, $x_n \leq y_n$ (which is the same as "not $x_n > y_n$") holds for *some* choices of n. Thus the opposite statement is

- There is some n_0, such that $x_{n_0} \leq y_{n_0}$.

In general, the opposite of "for *all* n, $A(n)$ happens" is "for *some* n, $A(n)$ *does not* happen", or "for *some* n, (*not* $A(n)$) happens". We emphasis that the universal quantifier is changed to the existential quantifier, and the statement $A(n)$ is changed to its opposite. Along this line, we have the following similar example from everyday life:

- All students in the class are at least 19 years old.

- Some students in the class are younger than 19 years.

Similarly, the opposite of "for *some* n, $A(n)$ happens" is "for *all* n, $A(n)$ *does not* happen", or "for all n, (*not* $A(n)$) happens". For example, the following are opposite:

- There is some n_0, such that $x_n > y_n$.
- For all n, $x_n \le y_n$.

The following are also opposite:

- Some students in the class are named Li.
- No students in the class are named Li.

When quantifiers are used several times, we find the opposite by making the conversions "all $\leftrightarrow$ some" and "$A(n) \leftrightarrow$ (not $A(n)$)" one by one. The opposites of the following statements

- For any m and n, either $m > n$, or $m = n$, or $m < n$.
- $m^2 = n^2$ for some $m \ne n$.
- For any rational numbers $r > s$, there is a rational number p, such that $r > p > s$.
- For any $\epsilon > 0$, there is n, such that for any $m > n$, we have $|x_m - a| < \epsilon$.
- Every student must work hard in every class.

are

- There are m and n, such that $m \le n$, and $m \ne n$, and $m \ge n$.
- For all $m \ne n$, we must have $m^2 = n^2$.
- For some rational numbers $r > s$, we cannot find another rational number p, such that $r > p > s$.
- There is $\epsilon > 0$, such that for any n, we can find $m > n$, such that $|x_m - a| \ge \epsilon$.
- Some students can relax in some classes.

The bottom line for finding the opposite of given statements is to use your common sense. Moreover, the standard phrases such as "for *all* n, $A(n)$ happens" and "If A, then (not B)" may turn out to be awkward for specific examples. So it is always a good idea to write statements in the most easily understood way.

Exercise 1.12. Write down the opposite.

1. For any $n > 0$, there is m such that m divides n.

2. $m^3 + 2mn + n^2 = 0$ for some m and n.

3. For any number x and $\epsilon > 0$, there are integers m and n, such that $\left|x - \frac{m}{n}\right| < \epsilon$.

4. Any people make some mistake sometime.

5. No matter how high the price is, some people will buy it.

6. For every lecture, there are always some students not showing up.

Exercise 1.13. Write down the opposite.

1. For any m, there is n, such that $A(m,n)$ happens.

2. For any m, there is $n > m$, such that $A(m,n)$ does not happen.

3. For some m, $A(m,n)$ happens for any n satisfying $m > n$.

4. For any $\epsilon > 0$, there is n, such that for any m satisfying $2m \geq n + 1$, we have $A(m,n,\epsilon)$.

5. For some $\epsilon > 0$ and any n, we have $A(m,n)$ for some m.

6. For some a and b, there is n, such that $A(a,b,n)$ happens.

Finally, quantifiers may be implicit in some statements. For example, the following statements

- If $n > 2$, then $n^2 > 4$.
- $4m + 6n$ is divisible by 2.

are equivalent to the following, which make explicit use of quantifiers.

- For any n bigger than 2, n^2 is bigger than 4.
- For any m and n, $4m + 6n$ is divisible by 2.

1.4 Proof

Given a mathematical statement, we would like to find out whether the statement is true or false. For the true statement, we need to provide *proof*, which is the process of logically justifying the statement. For the false statement, we need to provide *counterexample*.

The following is a straightforward proof.

Example 1.4.1. For any positive integer n, we prove

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

Since the order of terms does not affect the sum, we have

$$\begin{aligned} 2\sum_{k=1}^{n} k &= [1 + 2 + \cdots + (n-1) + n] + [n + (n-1) + \cdots + 2 + 1] \\ &= [1+n] + [2 + (n-1)] + \cdots + [(n-1) + 2] + [n+1] \\ &= (n+1) + (n+1) + \cdots + (n+1) + (n+1) \\ &= n(n+1). \end{aligned}$$

Here the last equality is due to the fact that the sum consists of n terms. Dividing the whole equality by 2, we get the equality in the theorem.

A legendary story in mathematics is that Gauss[3] was asked to sum the numbers from 1 to 100. The teacher was expecting to take a long rest while the students were working with addition. But Gauss produced the answer in seconds using the method shown in the proof above. The sum

$$1 + 2 + \cdots + 99 + 100 = \frac{100 \times 101}{2} = 5050$$

is an *example* of the theorem. Usually for any given statement, it always helps to try some simple examples to get a feeling. For the formula in the theorem, you are advised to try $n = 2, 3, 4$. Enough example should give you confidence on the truthfulness of the statement and may even provide some clue on how to prove the theorem.

The statement in Example 1.4.1 is of then form "For any n, $A(n)$ happens", with infinitely many possible values of n. For these kinds of theorems, examples cannot be substitute for proof, no matter how many examples you have verified. On the other hand, if you want to show such statement is *wrong*, all it takes is a single *counterexample*. For example, the statement

- For any n, $n(2n-3) \geq 0$.

is true for all $n > 1$. However, the statement does not hold for $n = 1$. the single counterexample for the case $n = 1$ shows the statement is wrong.

Sometimes it is convenient to rephrase a statement into an equivalent form and then carry out the proof. For example, to study statements of the form "if A, then B", we may study "if (not B), then (not A)". In other words, we may try to show that if the conclusion is wrong, then the assumption cannot be true. A proof along is line is a *proof by contraposition*.

[3] Gauss (April 30, 1777 - February 23, 1855), German mathematician, astronomer and physicist, one of the leading mathematicians of all time.

Example 1.4.2. We prove the following statement: If n^2 is even, then n is even.

Assume n is not even. Then n is odd, which means that $n = 2k + 1$ for some integer k. Consequently, $n^2 = 4k^2 + 4k + 1 = 4(k^2 + k) + 1$ is not even.

We just proved that if n is not even, then n^2 is not even. This is equivalent to the statement in the theorem.

Some statements may be of more sophisticated form and we may have several ways of rephrasing them. For example, if a statement is of the form "if A, then B or C", then we may try to prove any of the following equivalent statements:

- If A and (not B), then C.
- If A and (not C), then B.
- If (not B), then (not A) or C.
- If (not B) and (not C), then (not A).

You may find one statement to be much easier to prove than the others.

The proof by contraposition is a special case of the following strategy. We start the proof of a statement A by assuming A is false. Then based on (not A), we make logical deductions. If we get a *contradiction*, then it means that the original assumption (not A) is wrong. Therefore A is true, and the theorem is proved. This method is *proof by contradiction*, or called *reductio ad absurdum* in Latin.

Example 1.4.3. We prove $\sqrt{2}$ is an irrational number[4].

Assume $\sqrt{2}$ is not an irrational number. Then $\sqrt{2}$ is an rational number, and we have $\sqrt{2} = \dfrac{m}{n}$ for some integers m and n. If both m and n are even, then we may cancel the factor 2 from both m, n and still write $\sqrt{2}$ as quotient of smaller numbers. By canceling the the 2-factors repeatedly, we get an expression $\sqrt{2} = \dfrac{m}{n}$ in which either m or n is odd[5].

From $\sqrt{2} = \dfrac{m}{n}$, we get $2n^2 = m^2$. By Example 1.4.2, since m^2 is even, we conclude that m is even, so that $m = 2k$ for some integer k. Then $2n^2 = m^2 = 4k^2$ implies $n^2 = 2k^2$. Thus n^2 is even, which by Example 1.4.2 implies that n is even. Now we conclude that both m and n are even, which contradicts the (additional) assumption that either m or n is odd. The contradiction shows that $\sqrt{2}$ is irrational.

There is one more technique for constructing proofs, the induction. The method will be discussed in the next section.

[4] The fact is usually attributed to Greek mathematician and philosopher Pythegoras (582 BC - 496 BC) or one of his followers, who gave a geometrical proof.

[5] This is the opposite of "both m and n are even".

Proofs are important because mathematics is based on logic reasoning. Even if you do not choose mathematics as your profession, I still urge you to try to understand proofs and write your own proofs. Due to the logic and thinking involved in constructing a proof, practicing proofs will greatly enhance your logical and thinking skill. Such skills are important for whatever you will do in the future. Moreover, writing your own proofs, especially in a clean, fluent, and easy to understand way, is also a good training in communication skill.

In constructing your own proofs, please pay special attention to the following key words: if, then, since, by, therefore, consequently, such that, etc. These words are the key for indicating logical relations between the statements appearing in the proof. You are urged to write down the proofs with all the key words included.

On the other hand, you should avoid using $\therefore, \because, \forall, \exists, \nexists$.

Exercise 1.14. Prove the following statements. Please write down your proof very carefully, with all the key words included.

1. For any integer n, n^2 is even if and only if n is even.

2. For any integers m and n, mn is odd if and only if both m and n are odd.

3. $\sqrt{2}+\sqrt{6}<\sqrt{15}$.

4. For any integer n, n^3-n is a multiple of 6.

5. $\sqrt{\dfrac{3}{2}}$ is an irrational number.

6. $\sqrt[3]{2}$ is an irrational number.

7. The square of any integer is either of the form $4k$ or the form $4k+1$, where k is an integer.

8. The sum of the square of three consecutive integers cannot be of the form $12k-1$ for some integer k.

9. The polynomial $t^4-2t^2+2t+10$ is not a product of two quadratic polynomials with integer coefficients.

Exercise 1.15. Show that the following statements are false.

1. Every positive integer is a sum of squares of three integers.

2. For every positive integers n and k, n^k-n is divisible by k.

3. $\sqrt{2}+\sqrt{3}$ is a rational number.

1.5 Induction

Let $A(n)$ be a statement involving a positive integer n. The truefulness of $A(n)$ may be established by the method of *induction*, which consists of the following steps:

1. Prove $A(1)$ is true.

2. Prove that if $A(n-1)$ is true, then $A(n)$ is also true.

The idea behind the method is quite easy to understand. In the first step, we know

- $A(1)$ is true.

Taking $n = 2$ in the second step and using the truthfulness of $A(1)$, we get

- $A(2)$ is true.

Further taking $n = 3$ in the second step and using the (just obtained) truthfulness of $A(2)$, we get

- $A(3)$ is true.

The pattern goes on and eventually all positive integers are covered. At the end, the two steps together implies that $A(n)$ is true for all n.

Example 1.5.1. We prove the equality in Example 1.4.1 by induction.

Since

$$\sum_{k=1}^{1} k = 1 = \frac{1(1+1)}{2},$$

the equality has been verified for $n = 1$.

Next, assume the theorem holds for $n-1$. In other words, assume we already have (the equality below is called the *inductive assumption*)

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

Then (the inductive assumption is used in the second equality)

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

The induction does not have to start with $n = 1$. The reason for the induction to work also applies to the other starting points (called the *base of induction*).

Example 1.5.2. We prove $2^n > n^2$ for $n \geq 5$.

Since $2^5 = 32$ and $5^2 = 25$, the inequality is true for $n = 5$.

Next, we make the inductive assumption $2^{n-1} > (n-1)^2$. Then

$$2^n = 2^{n-1}2 > 2(n-1)^2 = n^2 + (n^2 - 4n + 2) > n^2,$$

where the inductive assumption is used in the second step and $n^2 - 4n > 0$ for $n \geq 5$ is used in the last step. The inequality is then proved for n.

Example 1.5.3. For $x > y > 0$, we prove $x^n - y^n \leq n(x-y)x^{n-1}$.

For $n = 1$, we have

$$x^1 - y^1 = x - y = 1(x-y)x^0.$$

Suppose $x^n - y^n \leq n(x-y)x^{n-1}$. Then

$$\begin{aligned} x^{n+1} - y^{n+1} &= x(x^n - y^n) + (x-y)y^n \\ &\leq xn(x-y)x^{n-1} + (x-y)y^n \\ &< n(x-y)x^n + (x-y)x^n \\ &= (n+1)(x-y)x^n. \end{aligned}$$

This completes the inductive proof of the inequality. IN fact, the proof shows $x^n - y^n < n(x-y)x^{n-1}$ for $n \geq 2$.

Another variation of the induction is the following two steps:

1. Prove $A(1)$ is true.
2. Prove that if $A(k)$ is true for all $k < n$, then $A(n)$ is also true.

To see why the two steps imply $A(n)$ for all n, we start with the fact established in the first step

- $A(1)$ is true.

Taking $n = 2$ in the second step, the only $k < 2$ is $k = 1$. Since $A(1)$ is already established, we get

- $A(2)$ is true.

Further taking $n = 3$ in the second step, the only $k < 3$ are $k = 1$ and $k = 2$. Since $A(1)$ and $A(2)$ have just been established, we get

- $A(3)$ is true.

The pattern goes on and eventually we find $A(n)$ to be true for all n.

Example 1.5.4. For any natural number n, we prove $n = 2^k m$ for some integer $k \geq 0$ and odd number m.

Since $1 = 2^0 1$ (with $k = 0$ and $m = 1$), the theorem is verified for $n = 1$. Next assume the theorem holds for all numbers $< n$. Now consider two possibilities for n.

If n is odd, then $n = 2^0 n$ for $k = 0$ and $m = n$. The theorem is verified.

If n is even, then $n = 2n'$ for some integer n'. Since $n' < n$, the inductive assumption may be applied to n' and gives us $n' = 2^{k'} m$ for some integer $k' \geq 0$ and odd number m. Then $n = 2 \times 2^{k'} m = 2^{k'+1} m$. Therefore the theorem is also verified.

A statement involving several numbers may be proved by more sophisticated versions of mathematical induction.

In the next example, we define $n! = 1 \cdot 2 \cdots (n-1) \cdot n$, called the *$n$-th factorial*. For example,

$$\begin{aligned}
1! &= 1, \\
2! &= 1 \cdot 2 = 2, \\
3! &= 1 \cdot 2 \cdot 3 = 6, \\
4! &= 1 \cdot 2 \cdot 3 \cdot 4 = 24, \\
5! &= 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 = 120.
\end{aligned}$$

It is also customary to define $0! = 1$.

Example 1.5.5. We prove that the product of n consecutive natural numbers is divisible by $n!$. Here is the more precise statement: For any natural numbers m and n, the number

$$N(m, n) = m(m+1) \cdots (m+n-1)$$

is divisible by $n!$.

The statement involves two positive integers m and n, and we will carry out a double induction.

First, for $n = 1$ (and any m), we have $N(m, 1) = m$, which is indeed divisible by $1! = 1$.

Second, assume $N(m, n-1)$ is divisible by $(n-1)!$. By induction on n, we need to prove that the following statement (the inductive assumption is included in the parenthesis)

- (If $N(m, n-1)$ is divisible by $(n-1)!$, then) $N(m, n)$ is divisible by $n!$.

We will prove the statement by inducting on m.

For $m = 1$, we have $N(1, n) = n!$, which is divisible by $n!$.

Now assume $N(m-1, n)$ is divisible by $n!$. Also keep in mind the assumption that $N(m, n-1)$ is divisible by $(n-1)!$. Then the following equality

$$N(m, n) = N(n-1, m) + N(m, n-1)n$$

and the assumptions imply that $N(m, n)$ is also divisible by $n!$.

The induction on m in complete, so that the second step for the induction on m is also complete. This completes the whole proof.

Let us recap the double induction process in the proof above:

1. Prove $A(m, 1)$.

2. The second step for induction on n is the problem "If $A(m, n-1)$, then $A(m, n)$".

3. Prove $A(1, n)$. (for the problem in the step 2, start induction on m)

4. Prove "If $A(m, n-1)$ and $A(m-1, n)$, then $A(m, n)$". (second step of induction on m)

There are various other ways of doing double inductions.

The proof in Example 1.5.3 seems rather complicated. How do we come up with the idea for such a proof?

First, the problem depends on positive integers m and n, which suggests the possibility of using induction. If we try to induct on m, then we need to assume $N(m-1, n)$ is divisible by $n!$ and try to show $N(m, n)$ is also divisible by $n!$. naturally we would like to compare

$$\begin{aligned} N(m-1, n) &= (m-1)m(m+1)\cdots(m+n-1) && \text{(known to be divisible by } n!) \\ N(m, n) &= m(m+1)\cdots(m+n-2)(m+n-1) && \text{(want to be divisible by } n!) \end{aligned}$$

Observing that most of the two products are the same, except the first term for $N(m-1, n)$ and the last term for $N(m, n)$. This suggests us to consider their difference

$$\begin{aligned} N(m, n) - N(m-1, n) &= m(m+1)\cdots(m+n-1)[(m+n-1)-(m-1)] \\ &= m(m+1)\cdots(m+n-2)(m+n-1)n \\ &= N(m, n-1)n. \end{aligned}$$

If by knowing $N(m-1, n)$ is divisible $n!$, we wish to conclude that $N(m, n)$ is also divisible by $n!$, then the equality above suggests that we need to *further assume* that $N(m, n-1)$ is divisible by $(n-1)!$. Since $n-1$ is one less than n, we realize that the further assumption can fit into another induction scheme, this time on n. This leads to the double induction adopted in the proof.

Before writing a proof, it is very important to analyze your problem. In the analysis, you always must have a clear mind about what you have (the assumptions) and what you want (the conclusions). I would recommend you to write down your analysis on a piece of paper. The actual proof is often the reverse of your analysis.

Exercise 1.16. Prove the statement.

1. For any integer $n \geq 1$, we have $2^n > n$.

2. For any integer $n \geq 1$, the sum of first n odd numbers is n^2.

3. For any positive integer n and any number p, we have $(1+p)^n \geq 1 + np$.

4. The sum of internal angles in an n-sided polygon is $(n-2)\pi$.

Exercise 1.17. Let a sequence a_n be defined by $a_0 = 1$, $a_1 = 2$, and $a_{n+1} = 3a_n - 2a_{n-1}$.

1. Compute a_3, a_4, a_5, a_6 and make a guess on the general formula for a_n.

2. Prove the general formula.

Exercise 1.18. Find a small (as small as possilbe, as a matter of fact) positive integer N, such that $2^N > N^3$. Then prove that $2^n > n^3$ for any integer $n \geq N$.

Exercise 1.19. Prove the statement in Example 1.5.3 by inducting on $m+n$, starting with $m+n=2$.

Exercise 1.20. Analyse the statement in Example 1.5.3 by attempting to induct on n. Can you come up with another proof?

Exercise 1.21. What is wrong with the following "proof" that all horses are the same color? Let $A(n)$ be the statement that, in any group of n horses, all are the same color. This is clearly true when $n = 1$ as any horse is the same color as itself. Next, take any group of n horses and exclude one. the remaining $n-1$ are the same color by the inductive assumption. Now exclude a different horse, so that the remaining $n-1$ (including the one originally excluded) are the same color, by the inductive assumption again. So all n are the same color.

Chapter 2

Set and Map

2.1 Set and Element

Set and element are the most basic concepts of mathematics. Given an *element* x and a *set* X, either x belongs to X (denoted $x \in X$), or x does not belong to X (denoted $x \notin X$). Sometimes we also call an element a *member* or (figuratively) a *point*.

The following sets are presented by listing all the elements:

- $\{1, 2, 3, \dots, n\}$ is the set of all integers between 1 and n.
- $\{3, -2\}$ is the solution set of the equation $x^2 - x - 6 = 0$.
- $\{a, b, c, \dots, x, y, z\}$ is the set of all latin alphabets.
- {red, green, blue} is the set of basic colours.
- {red, yellow} is the set of colours in the Chinese national flag.
- The set of all registered students in this class is the list of names provided by the registration office.

The following sets are presented by describing the properties satisfied by the elements:

- Natural numbers $\mathbb{N} = \{n\colon n \text{ is obtained by repeatedly adding 1 to itself}\}$.
- Prime numbers $\mathbb{P}$ is the set of natural numbers that are not products of two strictly smaller natural numbers.
- Rational numbers $\mathbb{Q} = \{\frac{m}{n}\colon m \text{ and } n \text{ are integers}\}$.
- Open interval $(a, b) = \{x\colon a < x < b\}$.
- Closed interval $[a, b] = \{x\colon a \le x \le b\}$.

- Real polynomials $\mathbb{R}[t] = \{a_0 + a_1 t + a_1 t^2 + \cdots + a_n t^n : a_i \in \mathbb{R}\}$.
- Unit sphere $S^2 = \{(x_1, x_2, x_3) : x_1^2 + x_2^2 + x_3^2 = 1\}$ in $\mathbb{R}^3$.
- $\{1, 2, 3, \dots, n\} = \{x : x \in \mathbb{N}, x \leq n\}$.
- $\{3, -2\} = \{x \in \mathbb{R} : x^2 - x - 6 = 0\}$.

Exercise 2.1. Present the set.

1. Integers $\mathbb{Z}$ (by using $\mathbb{N}$, for example).
2. Unit sphere S^n in $\mathbb{R}^{n+1}$.
3. Unit ball D^3 in $\mathbb{R}^3$.
4. Set of songs you listened in the last week.
5. Set of latin alphabets in your name.

Exercise 2.2. Provide suitable names for the set.

1. $\{(x, y) : x = 0\}$.
2. $\{(x, y) : x = y\}$.
3. $\{(x, y) : x^2 + y^2 = 4\}$.
4. $\{(x, y) : x^2 + y^2 > 4\}$.
5. $\{(x, y) : y < 0\}$.
6. $\{(x, y) : x^2 + 4y^2 = 4\}$.
7. $\{(x, y) : |x| + |y| < 1\}$.
8. $\{(x, y) : |x| < 1, |y| < 1\}$.

Exercise 2.3. Prove that the set of numbers x satisfying $x^2 = 6x - 8$ is the same as the set of even integers between 1 and 5.

The *empty set* $\emptyset$ is the set with no element.

A set X is a *subset* of another set Y if $x \in X$ implies $x \in Y$. In this case, we denote $X \subset Y$ (X is *contained in* Y) or $Y \supset X$ (Y *contains* X). We have the following properties:

1. $\emptyset \subset X$ for any set X.
2. Transitivity: $X \subset Y$ and $Y \subset Z \implies X \subset Z$.

3. $X = Y \iff X \subset Y$ and $Y \subset X$.

The inclusion $\subset$ is also related to the implication in logic. For example, the following statements are rephrased as inclusions

- If $n < 2$, then $n < 3$: $\{n\colon n < 2\} \subset \{n\colon n < 3\}$.
- n^2 is even implies n is even: $\{n\colon n^2 \text{ is even}\} \subset \{n\colon n \text{ is even}\}$.

Moreover, the third property above corresponds to the fact that statement A is equivalent to statement B if and only if A implies B and B implies A.

Example 2.1.1. We try to find suitable $\delta > 0$, such that

$$\{x\colon |x-2| < \delta\} \subset \{x\colon |x^2-4| < 1\}.$$

This is the same as

$$|x-2| < \delta \implies |x^2-4| < 1.$$

Intuitively, this means that, if x is sufficiently close to 2, then x^2 can be within distance 1 from 4.

We analyse the problem

$$|x-2| < \delta \implies 2-\delta < x < 2+\delta \implies (2-\delta)^2 < x^2 < (2+\delta)^2 \overset{?}{\implies} 3 < x^2 < 5.$$

We remark the following:

1. The second "$\implies$" requires $2-\delta > 0$.
2. $|x^2-4| < 1$ is the same as $3 < x^2 < 5$. This means that, if "$\overset{?}{\implies}$" holds, then we get the wanted inclusion.

The condition for "$\overset{?}{\implies}$" to hold is $(2-\delta)^2 \ge 3$ and $(2+\delta)^2 \le 5$. Combined with $2-\delta > 0$ and $\delta > 0$, the precise condition is $\delta \le 2-\sqrt{3}$ and $\delta \le \sqrt{5}-2$. Since $2-\sqrt{3} > \sqrt{5}-2$, we may choose $\delta = \sqrt{5}-2$.

In fact, the answer $\delta \le \sqrt{5}-2$ we found above is the precise (i.e., necessary and sufficient) condition for the inclusion. However, the original question only asks for *one* δ that makes the inclusion valid. A smart person should not overdo a problem.

Here is the more intelligent analysis of the problem

$$|x-2| < \delta \implies |x^2-4| = |x+2||x-2| \le |x+2|\delta.$$

To get the inclusion, it is sufficient to have $|x+2|\delta < 1$. Intuitively, we know x close to 2 implies $x+2$ close to 4, and should be < 5. More precisely, we have $|x-2| < 1 \implies |x+2| < 5$. Therefore

$$|x-2| < \delta \overset{?}{\le} 1 \implies |x^2-4| = |x+2||x-2| \le |x+2|\delta < 5\delta \overset{?}{\le} 1.$$

We get the wanted inclusion if both $\overset{?}{\le}$ are valid. This suggests us to take $\delta = \frac{1}{5}$.

Now we formally present the proof (the following is what you should write in homework and exam).

Take $\delta = \frac{1}{5}$. Then

$$\begin{aligned} |x-2| < \delta &\implies |x-2| < \frac{1}{5} \text{ and } |x+2| < 5 \\ &\implies |x^2-4| = |x+2||x-2| \le 5\frac{1}{5} = 1. \end{aligned}$$

This means the wanted inclusion.

Example 2.1.2. We try to find suitable $\epsilon > 0$, such that

$$\{x\colon |x-2| < 1\} \subset \{x\colon |x^2-4| < \epsilon\}.$$

We analyse the problem

$$|x-2| < 1 \implies |x^2-4| = |x+2||x-2| < 5 \cdot 1 = 5.$$

Therefore it is sufficient to take $\epsilon = 5$.

Now the formal proof: Take $\epsilon = 5$. Then

$$\begin{aligned} |x-2| < 1 &\implies |x-2| < 1 \text{ and } |x+2| < 5 \\ &\implies |x^2-4| = |x+2||x-2| \le 5 \cdot 1 = \epsilon. \end{aligned}$$

This means the wanted inclusion.

Exercise 2.4. Find suitable $\delta > 0$ or $\epsilon > 0$, such that the inclusion hold.

1. $\{x\colon |x-1| < \delta\} \subset \{x\colon |x^2-1| < 0.1\}$.
2. $\{x\colon |x-1| < \delta\} \subset \{x\colon |x^2-1| < 1\}$.
3. $\{x\colon |x+1| < 0.2\} \subset \{x\colon |x^2-1| < \epsilon\}$.
4. $\{x\colon |x+1| < 0.1\} \subset \{x\colon |x^2-1| < \epsilon\}$.

Exercise 2.5. Given $\epsilon > 0$, find $\delta > 0$, such that

$$\{x\colon |x-1| < \delta\} \subset \{x\colon |x^2-1| < \epsilon\}.$$

The inclusion means that, if x is close to 1, then x^2 is also close to 1.

Exercise 2.6. Given $\delta > 0$, find $\epsilon > 0$, such that

$$\{x\colon |x-1| < \delta\} \subset \{x\colon |x^2-1| < \epsilon\}.$$

Exercise 2.7. Find a number $n \in \mathbb{N}$, such that

$$\{m\colon m > n\} \subset \{m\colon \left|\frac{m}{m^2+1}\right| < 0.0001\}.$$

Exercise 2.8. For $m, n \in \mathbb{N}$, let $S_{m,n} = \{k \in \mathbb{N}\colon m \le k \le n\}$. What is the necessary and sufficient condition for $S_{m,n} \subset S_{m',n'}$?

Exercise 2.9. For $r > 0$, let

$$\begin{aligned} F_r &= \{(x,y) \in \mathbb{R}^2\colon x^2 + y^2 \le r^2\}, \\ G_r &= \{(x,y) \in \mathbb{R}^2\colon |x| \le r, |y| \le r\}. \end{aligned}$$

What is the necessary and sufficient condition for $F_r \subset G_{r'}$? What is the necessary and sufficient condition for $G_{r'} \subset F_r$?

Exercise 2.10. Given two functions $f(x,y)$ and $g(x,y)$, we get two subsets of $\mathbb{R}^2$

$$F_r = \{(x,y)\colon f(x,y) < r\}, \quad G_r = \{(x,y)\colon g(x,y) < r\}.$$

If $f(x,y) \le g(x,y)$, what can you say about the inclusion between F_r, G_r? Then take f and g to be $\sqrt{x^2+y^2}$ and $\min\{|x|, |y|\}$ and redo Exercise 2.9.

Exercise 2.11. We say X is a *proper subset* of Y if $X \subset Y$ and $X \ne Y$. Prove that if X is a proper subset of Y and Y is a subset of Z, then X is a proper subset of Z.

Can you make another similar statement?

The *power set* $\mathcal{P}(X)$ (also denoted as 2^X) of a set X is the collection of all subsets of X. For example, the following are all the subsets of the set $\{1,2,3\}$:

$$\emptyset,\ \{1\},\ \{2\},\ \{3\},\ \{1,2\},\ \{1,3\},\ \{2,3\},\ \{1,2,3\}.$$

Therefore the power set of $\{1,2,3\}$ is

$$\mathcal{P}\{1,2,3\} = \big\{\emptyset,\ \{1\},\ \{2\},\ \{3\},\ \{1,2\},\ \{1,3\},\ \{2,3\},\ \{1,2,3\}\big\}.$$

The power set shows that sets themselves can become elements of some other set (which we usually call *collection of sets*). For example, the set

$$\big\{\emptyset,\ \{1,2\},\ \{1,3\},\ \{2,3\}\big\}$$

is the collection of subsets of $\{1,2,3\}$ with even number of elements.

Exercise 2.12. List all elements in $\mathcal{P}\{1,2,3,4\}$ that do not include 2.

Exercise 2.13. List all elements in $\mathcal{P}(\mathcal{P}\{1,2\})$, the power set of the power set of $\{1,2\}$.

Exercise 2.14. How many elements are in the power set of $\{1,2,\dots,n\}$?[1] How many of these contain even number of elements?

2.2 Set Operation

The *union* of two sets X and Y is

$$X \cup Y = \{x\colon x \in X \text{ or } x \in Y\}.$$

Note that in the union, we allow the possibility that x is in both X and Y.

The *intersection* of two sets is

$$X \cap Y = \{x\colon x \in X \text{ and } x \in Y\}.$$

If $X \cap Y = \emptyset$, then there is no common elements shared by X and Y, and we say X and Y are *disjoint*. For disjoint X and Y, we call the union $X \cup Y$ a *disjoint union*. We may also emphasise the disjoint property by writing the union as $X \sqcup Y$.

The *difference* of two sets is

$$X - Y = \{x\colon x \in X \text{ and } x \notin Y\}.$$

In case Y is a subset of X, we also call $X - Y$ the *complement* of Y in X.

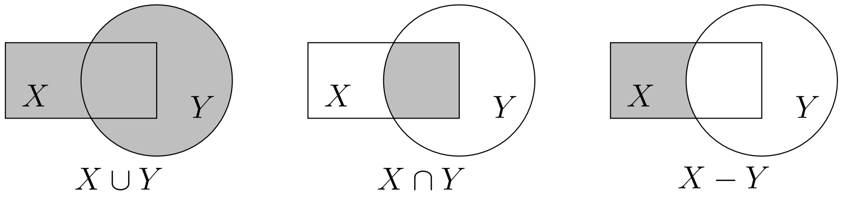

Figure 2.1: Union, intersection, and difference.

The union and intersection has clear connection with the use of “and” and “or” in logic. For example, the following statements are rephrased as unions or intersections

- 42 is divisible by 2 annd 3: $42 \in \{n\colon n \text{ is divisible by } 2\} \cap \{n\colon n \text{ is divisible by } 3\}$.
- Any integer is either odd or even: $\mathbb{Z} \subset \{n\colon n \text{ is even}\} \cup \{n\colon n \text{ is odd}\}$.
- The students in the class are the first year math students: $\{\text{students in the class}\} \subset \{\text{first year students}\} \cap \{\text{math students}\}$.

[1] The answer suggests the reason for the notation 2^X.

- Bob is either playing basketball or badminton: Bob $\in$ {basketball players} $\cup$ {badminton players}.

Moreover, we note that $X \cap Y = \emptyset$ means "if $x \in X$, then $x \notin Y$". For example,

- If an integer is even, then it is not odd: $\{n\colon n \text{ is even}\} \cup \{n\colon n \text{ is odd}\} = \emptyset$.
- If a person is a male, then the person is not female: $\{\text{male}\} \cap \{\text{female}\} = \emptyset$.

The set operations have the following properties:

1. $X \cup \emptyset = X$, $X \cap \emptyset = \emptyset$.
2. $X \cap Y \subset X \subset X \cup Y$.
3. $X - Y = \emptyset \iff X \subset Y$.
4. $X - Y = X \iff X \cap Y = \emptyset$.
5. Commutativity: $Y \cup X = X \cup Y$, $Y \cap X = X \cap Y$.
6. Associativity: $(X \cup Y) \cup Z = X \cup (Y \cup Z)$, $(X \cap Y) \cap Z = X \cap (Y \cap Z)$.
7. Distibutivity: $(X \cup Y) \cap Z = (X \cap Z) \cup (Y \cap Z)$, $(X \cap Y) \cup Z = (X \cup Z) \cap (Y \cup Z)$.
8. Double Complement: If $Y \subset X$, then $X - (X - Y) = Y$.
9. deMorgan's Law: $X - (Y \cup Z) = (X - Y) \cap (X - Z)$, $X - (Y \cap Z) = (X - Y) \cup (X - Z)$.

The proof of the properties is an exercise on the interactions of and, or, not for statements. For example, the proof of the associativity of the union involves the verification of $(X \cup Y) \cup Z \subset X \cup (Y \cup Z)$ and $(X \cup Y) \cup Z \supset X \cup (Y \cup Z)$. The first inclusion is verified as follows

$$\begin{aligned} x \in (X \cup Y) \cup Z &\iff x \in X \cup Y \text{ or } x \in Z \\ &\iff x \in X \text{ or } x \in Y \text{ or } x \in Z \\ &\iff x \in X \text{ or } x \in Y \cup Z \\ &\iff x \in X \cup (Y \cup Z). \end{aligned}$$

The second inclusion can be verified similarly.

We note that deMorgan's law says that the complement operation converts the union to the intersection and vice versa. Figure 2.2 is a pictorial proof of $X - (Y \cup Z) = (X - Y) \cap (X - Z)$.

The (cartesian) *product* of two sets X and Y is

$$X \times Y = \{(x, y)\colon x \in X, y \in Y\}.$$

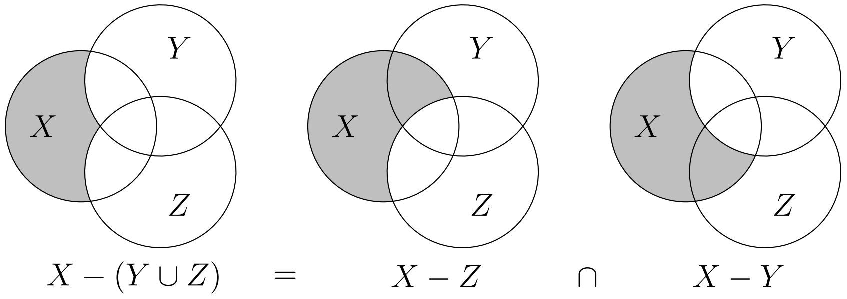

Figure 2.2: deMorgan's law.

We also use X^n to denote the product of n copies of X. For example, $\mathbb{R}^n$ is the Euclidean space of dimension n, consisting of ordered sequence of n real numbers. Moreover, we have $\mathbb{R}^{m+n} = \mathbb{R}^m \times \mathbb{R}^n$.

When several operations are mixed, usually the product $\times$ is taken first, then the union $\cup$ or the intersection $\cap$ is taken, and the difference $-$ is taken last. For example, $x \in (X - Y \times Z) \cap W - U$ means $x \in X$, $x \notin Y \times Z$, $x \in W$, and $x \notin U$. The convention is similar to first taking multiplication and division, and then taking summation and subtraction in numerical computations.

Exercise 2.15. What are the following sets?

1. $\{n\colon n \text{ is even}\} \cap \{n\colon n \text{ is divisible by } 5\}$.

2. $\{n\colon n \text{ is a positive integer}\} \cup \{n\colon n \text{ is a negative integer}\} \cup \{0\}$.

3. $\{n\colon n \text{ is even}\} \cup \{n\colon |n| < 10 \text{ and } n \text{ is an integer}\} - \{n\colon n \neq 2\} \cap \{n\colon n^2 \neq 6n - 8\}$.

4. $\{x\colon 1 < x \leq 10\} - \{x\colon 2x \text{ is an integer}\} \cap \{x\colon 6x^3 + 5x^2 - 33x + 18 = 0\}$.

5. $\{x\colon x > 0\} \times \{y\colon |y| < 1\} - \{(x, y)\colon x + y < 1\} - \{(x, y)\colon x > y\}$.

6. $\{(x, y)\colon x^2 + y^2 \leq 1\} - \{x\colon x \geq 0\}^2$.

Exercise 2.16. Convince yourself the following properties by drawing pictures. Then prove the properties.

1. $(X \cap Y) \cap Z = X \cap (Y \cap Z)$.

2. $(X \cap Y) \cup Z = (X \cup Y) \cap (X \cup Z)$.

3. $X - (Y \cap Z) = (X - Y) \cup (X - Z)$.

4. $(X - Y) \cup (Y - X) = X \cup Y - X \cap Y$.

5. $(X - Y) \cap (Y - X) = \emptyset$.

6. $X \times (Y \cap Z) = (X \cap Y) \times (X \cap Z)$.

7. $X \times Y - X \times Z = X \times (Y - Z)$.

Exercise 2.17. Find all the unions and intersections among $A = \{\emptyset\}$, $B = \{\emptyset, A\}$, $C = \{\emptyset, A, B\}$.

Exercise 2.18. Express the following using sets X, Y, Z and operations $\cup$, $\cap$, $-$:

1. $\{x \colon x \in X \text{ and } (x \in Y \text{ or } x \in Z)\}$.

2. $\{x \colon (x \in X \text{ and } x \in Y) \text{ or } x \in Z\}$.

3. $\{x \colon x \in X,\ x \notin Y, \text{ and } x \in Z\}$.

Exercise 2.19. Let A and B be subsets of X, prove that

$$A \subset B \iff X - A \supset X - B \iff A \cap (X - B) = \emptyset.$$

Exercise 2.20. Given two sets X and Y, the set $X \cup Y$ is a disjoint union of subsets $X - Y$, $Y - X$ and $X \cap Y$.

1. Express the unions of some of the three subsets. How many such union subsets are there?

2. If we start with n sets, how many union subsets can be get?

Exercise 2.21. Which statements are true?

1. $X \subset Z$ and $Y \subset Z \implies X \cup Y \subset Z$.

2. $X \subset Z$ and $Y \subset Z \implies X \cap Y \subset Z$.

3. $X \subset Z$ or $Y \subset Z \implies X \cup Y \subset Z$.

4. $Z \subset X$ and $Z \subset Y \implies Z \subset X \cap Y$.

5. $Z \subset X$ and $Z \subset Y \implies Z \subset X \cup Y$.

6. $Z \subset X \cap Y \implies Z \subset X$ and $Z \subset Y$.

7. $Z \subset X \cup Y \implies Z \subset X$ and $Z \subset Y$.

8. $X - (Y - Z) = (X - Y) \cup Z$.

9. $(X - Y) - Z = X - Y \cup Z$.

10. $X - (X - Z) = Z$.

11. $X \cap (Y - Z) = X \cap Y - X \cap Z$.
12. $X \cup (Y - Z) = X \cup Y - X \cup Z$.
13. $(X \cap Y) \cup (X - Y) = X$.
14. $X \subset U$ and $Y \subset V \iff X \times Y \subset U \times V$.
15. $X \times (U \cup V) = X \times U \cup X \times V$.
16. $X \times (U - V) = X \times U - X \times V$.
17. $(X - Y) \times (U - V) = X \times U - Y \times V$.

2.3 Map

A *map* (also called *transformation*) from a set X (called *domain*) to a set Y (called *range*) is a rule f that assigns, for each $x \in X$, a unique $y = f(x) \in Y$, called the *image* or *value* of x. The rule should be *well-defined* in the following sense:

1. Applicability: The rule applies to any input $x \in X$ and always produces some output $f(x)$.
2. Unambiguity: For any input $x \in X$, the output $f(x)$ is unique (f is *single-valued*).

In case Y is a set of numbers, f is also called a *function*.

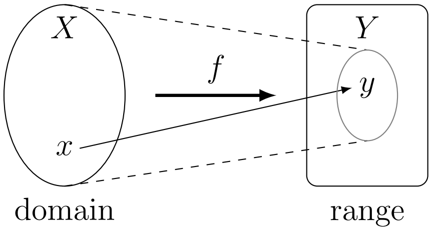

Figure 2.3: Domain, range, and image.

A map may be denoted by the notation

$$f\colon X \to Y,\ x \mapsto f(x),$$

or the notation

$$f(x) = y\colon X \to Y,$$

in which all the ingredients are indicated. Some parts of the notation may be omitted if the part is clear from the context.

Example 2.3.1. By the map (equivalently, $f\colon \mathbb{R} \to \mathbb{R}, x \mapsto 2x^2 - 1$)

$$f(x) = 2x^2 - 1\colon \mathbb{R} \to \mathbb{R},$$

we mean the following process: For any $x \in \mathbb{R}$, we first multiply x to itself, then multiply 2 to the result x^2, and then subtracted the result $2x^2$ by 1. The end result is $2x^2 - 1$. Since each step always works and gives unique outcome, the process is a map.

Example 2.3.2. The square root function $f(x) = \sqrt{x}\colon [0,\infty) \to \mathbb{R}$ is defined by the following process: For any non-negative number x (i.e., $x \in [0,\infty)$), we find a *non-negative* number y, such that multiplying y to itself yields x. Since y always exists and is unique, the process is a map.

Next, we modify the domain $[0,\infty)$ to $\mathbb{R}$ and consider the square root function $f(x) = \sqrt{x}\colon \mathbb{R} \to \mathbb{R}$. The process described above does not work for negative numbers, such as $y = -1$. Therefore the first condition for the process to be a map is violated, and $f(x) = \sqrt{x}\colon \mathbb{R} \to \mathbb{R}$ is not a function.

Finally, suppose we still consider $f(x) = \sqrt{x}\colon [0,\infty) \to \mathbb{R}$, but modifying the process by no longer requiring y to be non-negative. Then for any $x \in [0,\infty)$, the process always works, except two results will be produced (one positive, one negative) in general. Therefore the second condition is violated, and the process is also not a map.

Example 2.3.3. For any set X, the *identity map* is

$$id_X(x) = x\colon X \to X.$$

The map

$$\Delta_X(x) = (x,x)\colon X \to X^2$$

is the *diagonal map.*

For any sets X and Y, and fixed element $b \in Y$, the map

$$c(x) = b\colon X \to Y,$$

is a *constant map.* Moreover, we have two *projection maps*

$$\pi_X(x,y) = x\colon X \times Y \to X, \quad \pi_Y(x,y) = y\colon X \times Y \to Y.$$

Example 2.3.4. The flip of $\mathbb{R}^2$ with respect to the x-axis is a map, and is given by the formula

$$F(x,y) = (x,-y)\colon \mathbb{R}^2 \to \mathbb{R}^2.$$

The rotation of $\mathbb{R}^2$ by angle θ is also a map, and is given by the formula

$$R_\theta(x,y) = (x\cos\theta - y\sin\theta, x\sin\theta + y\cos\theta)\colon \mathbb{R}^2 \to \mathbb{R}^2.$$

See Figure 2.4, in which $v = (x, y)$.

Both are maps because the processs can be applied to all points on the plane and produces unique results. It is in fact easier to understand the map by their pictures than by the formulae.

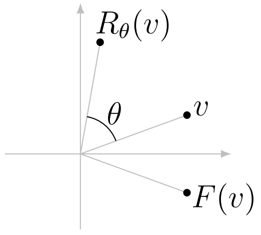

Figure 2.4: Flip and rotation of $\mathbb{R}^2$.

Example 2.3.5. The sign of numbers

$$\text{Sign}: \mathbb{R} \to \{+, 0, -\},\ x \mapsto \begin{cases} +, & \text{if } x > 0 \\ -, & \text{if } x < 0 \\ 0, & \text{if } x = 0 \end{cases}$$

is a map. The *Dirichlet function*

$$f(x) = \begin{cases} 1, & \text{if } x \text{ is rational} \\ 0, & \text{if } x \text{ is irrational} \end{cases} : \mathbb{R} \to \mathbb{R}$$

is also a map.

Both maps are given by descriptions instead of formulae. Even when we have formulae, sometimes it is easier to understand the maps through the descriptions than formulae.

Example 2.3.6. The map

$$\text{Age}\colon \text{People} \to \mathbb{N}$$

is the process of subtracting the birth year from the current year.

Let $X = \{\text{red, green, blue}\}$, and let Y be the set of all latin alphabets. Then the first alphabet map $F\colon X \to Y$ is given by

$$F(\text{red}) = r,\ F(\text{green}) = g,\ F(\text{blue}) = b.$$

Moreover, the all alphabet map $A\colon X \to \mathcal{P}(Y)$ is given by

$$A(\text{red}) = \{d, e, r\},\ A(\text{green}) = \{e, g, n, r\},\ A(\text{blue}) = \{b, e, l, u\}.$$

With self-evident definitions, the following are more maps from everyday life:

- Height: People → Number.
- ID_s: Student → Number.
- ID_p: Professor → Number.
- Instructor: Course → Professor.
- Maker: Product → Manufacturer.
- Capital City: Country → City.
- Population: City → Number.

Exercise 2.22. The following are some attempts to define a "square root" map. Which ones are maps?

1. For $x \in \mathbb{R}$, find $y \in \mathbb{R}$, such that $y^2 = x$. Then $f(x) = y$.
2. For $x \in [0, \infty)$, find $y \in \mathbb{R}$, such that $y^2 = x$. Then $f(x) = y$.
3. For $x \in [0, \infty)$, find $y \in [0, \infty)$, such that $y^2 = x$. Then $f(x) = y$.
4. For $x \in [1, \infty)$, find $y \in [1, \infty)$, such that $y^2 = x$. Then $f(x) = y$.
5. For $x \in [1, \infty)$, find $y \in (-\infty, -1]$, such that $y^2 = x$. Then $f(x) = y$.
6. For $x \in [0, 1)$, find $y \in [0, \infty)$, such that $y^2 = x$. Then $f(x) = y$.
7. For $x \in [1, \infty)$, find $y \in (-\infty, -4] \cup [1, 2)$, such that $y^2 = x$. Then $f(x) = y$.

Exercise 2.23. Describe the processes that define the maps.

1. $2^n : \mathbb{Z} \to \mathbb{R}$.
2. Angle: Two rays emanating from the origin of $\mathbb{R}^2$ → $[0, 2\pi)$.
3. Area_r: Rectangle → $[0, \infty)$.
4. Area_t: Triangle → $[0, \infty)$.
5. Absolute Value: $\mathbb{R} \to \mathbb{R}$.

Given two maps $f\colon X \to Y$ and $g\colon Y \to Z$, such that the range of f and the domain of g are the same set Y, the *composition* $g \circ f$ (or simply denoted gf) is

$$(g \circ f)(x) = g(f(x))\colon X \to Z.$$

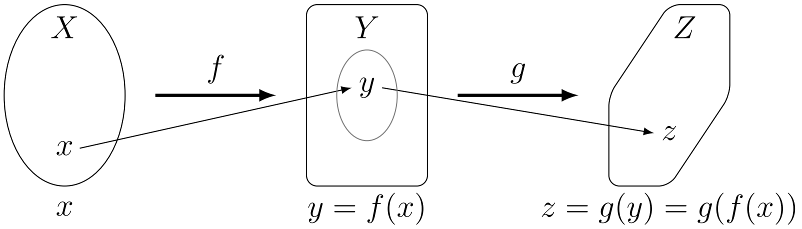

Figure 2.5: Composition of maps.

Example 2.3.7. If maps are given by formulae, then the composition may be computed by *substitution*. For example, for $f(x) = 2x^2 - 1$ and $g(y) = (y+1)^2$, we have $(g \circ f)(x) = [(2x^2-1)+1]^2 = 4x^4$. Such exercises usually takes the following form: If $y = 2x^2 - 1$ and $z = (y+1)^2$, then $z = 4x^4$.

Example 2.3.8. Let $A \subset X$ be a subset. Then we have the natural *inclusion map* $i(a) = a\colon A \to X$. For any map $f\colon X \to Y$, the composition $fi\colon A \to Y$ is the *restriction*[2] of f on A, and is often denoted by $f|_A$.

Example 2.3.9. The composition

$$\text{Population} \circ \text{Capital City}\colon \text{Country} \to \text{Number}$$

is the map "Population of the Capital City". The composition

$$\text{ID}_p \circ \text{Instructor}\colon \text{Course} \to \text{Number}\ .$$

is the map "ID number of the Instructor of the Course".

Example 2.3.10. If we flip $v \in \mathbb{R}^2$ with respect to the x-axis twice, then we get back v. Therefore the composition $F^2 = F \circ F = id$ is the identity map.

The composition $R_{\theta_2}R_{\theta_1}$ means a rotation by angle θ_1 followed by another rotation by angle θ_2. Clearly the effect is the same as a rotation by angle $\theta_1 + \theta_2$. Therefore we have $R_{\theta_2}R_{\theta_1} = R_{\theta_1+\theta_2}$.

A map $f\colon X \to Y$ induces maps between the subsets of X and Y. The *image* of a subset $A \subset X$ is

$$f(A) = \{f(a)\colon a \in A\} = \{y\colon y = f(a) \text{ for some } a \in A\}.$$

This can be considered as a map between the power sets

$$\text{Image}\colon \mathcal{P}(X) \to \mathcal{P}(Y),\ A \mapsto f(A).$$

[2] Because $fi(a) = f(a)$, where a in $fi(a)$ is considered as an element of A, and a in $f(a)$ is considered as an element of X.

In the other direction, the *preimage* of a subset $B \subset Y$ is

$$f^{-1}(B) = \{x \colon f(x) \in B\}.$$

In case B is a single point, we get the preimage

$$f^{-1}(y) = \{x \colon f(x) = y\}$$

of a point $y \in Y$. Preimage is also a map between the power sets

$$\text{Preimage} \colon \mathcal{P}(Y) \to \mathcal{P}(X),\ B \mapsto f^{-1}(B).$$

Example 2.3.11. Consider $f(x) = 2x^2 - 1 \colon \mathbb{R} \to \mathbb{R}$. The image of the whole domain $\mathbb{R}$ is $f(\mathbb{R}) = [-1, \infty)$. We also have

$$f[0,\infty) = f(-\infty, 0] = [-1,\infty), \quad f[2,\infty) = f(-\infty,-2] = [7,\infty).$$

In general, we have

$$f[a,\infty) = f(-\infty,-a] = \begin{cases} [2a^2-1,\infty), & \text{if } a > 0 \\ [-1,\infty), & \text{if } a \le 0 \end{cases}.$$

For the preimage, we have

$$f^{-1}[0,\infty) = (-\infty, -\tfrac{1}{\sqrt{2}}] \cup [\tfrac{1}{\sqrt{2}}, \infty), \quad f^{-1}(1) = \{1, -1\}.$$

Moreover, for any $a \le -1$, we have $f(\mathbb{R}) \subset [a, \infty)$. Then $f^{-1}[a,\infty) = \mathbb{R}$.

Example 2.3.12. For the rotation map R_θ. The image (and the preimage) of any circle centered at the origin is the circle itself. If the circle is not centered at the origin, then the image (and the preimage) is still a circle, but at a different location.

Example 2.3.13. For the first alphabet map F in Example 2.3.6, we have the images

$$F(X) = \{b, g, r\}, \quad F(\{\text{blue}, \text{red}\}) = \{b, r\},$$

and preimages

$$F^{-1}\{a,b,c\} = F^{-1}\{a,b,c,d,e\} = \{\text{blue}\}, \quad F^{-1}\{u,v,w,x,y,z\} = \emptyset.$$

In terms of properties, preimage behaves better than image. For example, we have

$$f^{-1}(A \cup B) = f^{-1}(A) \cup f^{-1}(B), \quad f^{-1}(A \cap B) = f^{-1}(A) \cap f^{-1}(B),$$

and

$$f(A \cup B) = f(A) \cup f(B), \quad f(A \cap B) \subset f(A) \cap f(B).$$

We note that $\subset$ above may not be equality.

The image and preimage behave quite nicely with respect to the composition:

$$(gf)(A) = g(f(A)), \quad (gf)^{-1}(B) = f^{-1}(g^{-1}(B)).$$

Exercise 2.24. Describe the image.

1. $f(x) = (x, 2x)\colon \mathbb{R} \to \mathbb{R}^2$, $A_1 = [0, 1]$, $A_2 = \mathbb{Z}$.
2. $f(\theta) = (\cos\theta, \sin\theta)\colon \mathbb{R} \to \mathbb{R}^2$, $A_1 = [0, \pi]$, $A_2 = \{0, \pi, 2\pi\}$.
3. $f(x, y) = x+y\colon \mathbb{R}^2 \to \mathbb{R}$, $A_1 = \{(x, y)\colon |x|+|y| < 1\}$, $A_2 = \{(x, y)\colon |x|+|y| > 1\}$.
4. $f =$Dirichlet function, $A_1 = \mathbb{Z}$, $A_2 = [0, 1]$, $A_2 = \{\sqrt{2}, \sqrt{3}\}$.

Exercise 2.25. Describe the preimage.

1. $f(x) = (x, 2x)\colon \mathbb{R} \to \mathbb{R}^2$, $B_1 = \{(x, y)\colon x^2 + y^2 \le 1\}$, $B_2 = \{(x, y)\colon |x| + |y| < 1\}$.
2. $f(\theta) = (\cos\theta, \sin\theta)\colon \mathbb{R} \to \mathbb{R}^2$, $B_1 = \{(x, y)\colon x^2 + y^2 \le 1\}$, $B_2 = \{(x, y)\colon x > 0, y < 0\}$.
3. $f(x, y) = x + y\colon \mathbb{R}^2 \to \mathbb{R}$, $B_1 = [0, 1]$, $B_2 = \mathbb{R}$.
4. $f =$Dirichlet function, $B_1 = \{0\}$, $B_2 = \{1\}$, $B_3 = \{2\}$.

Exercise 2.26. What is the image (and the preimage) of a straight line in $\mathbb{R}^2$ under the rotation R_θ? When is the image (or the preimage) the same as the original line?

Exercise 2.27. For a subset $A \subset X$, define the *characteristic function*

$$\chi_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases} : X \to \mathbb{R}.$$

Let $B \subset X$ be another subset of X, and let $Y \subset \mathbb{R}$.

1. Prove $\chi_{A\cap B} = \chi_A\chi_B$ and $\chi_{X-A} + \chi_A = 1$.
2. Express $\chi_{A\cup B}$ in terms of χ_A and χ_B.
3. Describe $\chi_A(B)$.
4. Describe $\chi_A^{-1}(Y)$.

Exercise 2.28. Explain why for any map $f\colon X \to Y$, we have $f^{-1}(Y) = X$.

Exercise 2.29. Let $A, B \subset X$ be subsets. What is the preimage $\Delta^{-1}(A \times B)$ under the diagonal map $\Delta\colon X \to X^2$?

Exercise 2.30. Let $A \subset X$ be a subset, and let $i \colon A \to X$ be the inclusion map. For a subset $B \subset X$, what is the preimage $i^{-1}(B)$?

Exercise 2.31. Suppose we want to combine two maps $f \colon X \to Z$ and $g \colon Y \to Z$ to get a new map $h \colon X \cup Y \to Z$ as follows

$$h(x) = \begin{cases} f(x) & \text{if } x \in X \\ g(x) & \text{if } x \in Y \end{cases}.$$

What is the condition for h to be a map? How are the images and preimages of f, g, h are related?

Exercise 2.32. Which statements are true? If not, whether at least some directions or inclusions are true?

1. $A \subset B \iff f(A) \subset f(B)$.
2. $A \subset B \iff f^{-1}(A) \subset f^{-1}(B)$.
3. $A \cap B = \emptyset \iff f(A) \cap f(B) = \emptyset$.
4. $A \cap B = \emptyset \iff f^{-1}(A) \cap f^{-1}(B) = \emptyset$.
5. $f(A \cup B) = f(A) \cup f(B)$.
6. $f(A \cap B) = f(A) \cap f(B)$.
7. $f(A - B) = f(A) - f(B)$.
8. $f^{-1}(A \cup B) = f^{-1}(A) \cup f^{-1}(B)$.
9. $f^{-1}(A \cap B) = f^{-1}(A) \cap f^{-1}(B)$.
10. $f^{-1}(A - B) = f^{-1}(A) - f^{-1}(B)$.
11. $f(f^{-1}(A)) = A$.
12. $f^{-1}(f(A)) = A$.

Exercise 2.33. Let $f \colon X \to Y$ be a map. Show that the "image map" $F \colon \mathcal{P}(X) \to \mathcal{P}(Y), A \mapsto f(A)$ is indeed a map. Is the similar "preimage map" also a map?

2.4 Onto, One-to-one, and Invertibility

A map $f\colon X \to Y$ is *onto* (or *surjective*) if every element of Y is an image:

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

This means $f(X) = Y$.

The map is *one-to-one* (or *injective*) if different elements have different images:

$$x_1 \neq x_2 \implies f(x_1) \neq f(x_2).$$

This is equivalent to that elements with the same image must be the same

$$f(x_1) = f(x_2) \implies x_1 = x_2.$$

The map is a *one-to-one correspondence* (or *bijective*) if it is one-to-one and onto.

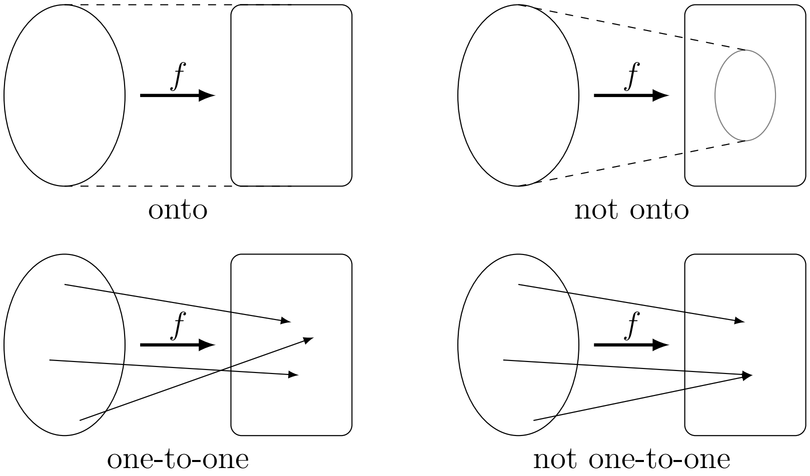

Figure 2.6: Onto and one-to-one.

The following reinterprets the concepts from the viewpoint of solving the equation $f(x) = y$, in which the right side $y \in Y$ is chosen and $x \in X$ is the variable.

Theorem 2.4.1. *Let $f\colon X \to Y$ be a map.*

1. *f is onto $\iff$ For any $y \in Y$, $f(x) = y$ has solutions.*
2. *f is one-to-one $\iff$ If $f(x) = y$ can be solved for x, then the solution is unique.*
3. *f is a one-to-one correspondence $\iff$ For any $y \in Y$, $f(x) = y$ has unique solution.*

For any $y \in Y$, there are three possibilities for the solution of $f(x) = y$:

- $0 \to 1$: No solution, i.e., no x satisfying $f(x) = y$.
- $1 \to 1$: Unique solution, i.e., exactly one x satisfying $f(x) = y$.
- $m \to 1$ (m for multiple): Non-unique solution, i.e., more than one x satisfying $f(x) = y$.

Onto means $0 \to 1$ does not happen. One-to-one means $m \to 1$ does not happen. One-to-one correspondence means only $1 \to 1$ happens.

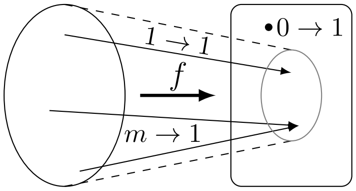

Figure 2.7: Possibilities for the solution of $f(x) = y$.

Example 2.4.1. Consider the function $f(x) = 2x^2 - 1\colon \mathbb{R} \to \mathbb{R}$. By $2x^2 - 1 \ge -1$, there is no $x \in \mathbb{R}$ satisfying $2x^2 - 1 = -2$. Therefore the map is not onto. On the other hand, both $x = 1, -1$ satisfy $2x^2 - 1 = 1$. Therefore the map is not one-to-one. In fact, the map is mostly two-to-one, in the sense that $2x^2 - 1 = y$ has exactly two solutions for $y > -1$.

Next, we improve the function to achieve onto or one-to-one. Since $2x^2 - 1 = y$ has solution if and only if $y \ge -1$, we may reduce the range from $\mathbb{R}$ to $[-1, \infty)$, and get an onto map $f_1(x) = 2x^2 - 1\colon \mathbb{R} \to [-1, \infty)$. We note that, although f, f_1 have the same formula, we regard them as different maps, due to different range set.

The reason for the function to be two-to-one is that, if $2x^2 - 1 = y$, then we also have $2(-x)^2 - 1 = y$. Therefore if we restrict x to be non-negative, there the function $f_2(x) = 2x^2 - 1\colon [0, \infty) \to \mathbb{R}$ becomes one-to-one. Again, we regard f, f_1, f_2 as distinct functions.

If we combine the two modifications, then we get a one-to-one correspondence $f_3(x) = 2x^2 - 1\colon [0, \infty) \to [-1, \infty)$. We note the following functions are also one-to-one correspondences

$$2x^2 - 1\colon \begin{cases} (-\infty, 0] \to [-1, \infty), \\ (-1, 0] \cup [1, \infty) \to [-1, \infty), \\ [1, \infty) \to [1, \infty), \\ (a, \infty) \to (2a^2 - 1, \infty) \text{ for } a > 0. \end{cases}$$

Example 2.4.2. The map Polar: $[0, \infty) \times \mathbb{R} \to \mathbb{R}^2$, $(r, \theta) \mapsto (x, y) = (r\cos\theta, r\sin\theta)$ is onto because any point in the (range) $\mathbb{R}^2$ can be assigned a norm r (distance to the origin) and an angle θ. The map is not one-to-one because the same point in

$\mathbb{R}^2$ can have many different angles, all differ by a multiple of 2π. If we modify the map to $[0,\infty)\times[0,2\pi)\to\mathbb{R}^2$, then the map is almost a one-to-one correspondence, with the origin as the only exception. To get a real one-to-one correspondence, we may modify the map to

$$\{(0,0)\}\cup(0,\infty)\times[0,2\pi)\to\mathbb{R}^2.$$

This means we only assign $r=0$ and $\theta=0$ to the origin.

Example 2.4.3. Consider the map Instructor: Course → Professor. Onto means all professors teach. One-to-one means each professor teaches at most one course. One-to-one correspondence means each professor teaches exactly one course.

Example 2.4.4. Since there is nobody of the age 200, the Age map in Example 2.3.6 is not onto. Since there are millions of people of 20 years old, the map is not one-to-one.

Since there are cities such as Hong Kong that are not capital cities of any country, the Capital City map is not onto. Since no two countries share the same capital city, the map is one-to-one.

Exercise 2.34. Determine onto and one-to-one.

1. $f(x)=(x,2x)\colon\mathbb{R}\to\mathbb{R}^2$.
2. $f(\theta)=(\cos\theta,\sin\theta)\colon\mathbb{R}\to\mathbb{R}^2$.
3. $f(\theta)=(\cos\theta,\sin\theta)\colon[0,2\pi)\to S^1=\{(x,y)\colon x^2+y^2=1\}$.
4. $f(x,y)=(2x-y-1,3x+2y+1)\colon\mathbb{R}^2\to\mathbb{R}^2$.
5. $f(x,y)=(2x-y-1,3x+2y+1)\colon S^1\to\mathbb{R}^2$.
6. $f(x,y)=(2x-y-1,3x+2y+1,x+y)\colon\mathbb{R}^2\to\mathbb{R}^3$.
7. $f(x)=x^3+x\colon\mathbb{R}\to\mathbb{R}$.
8. Sign: $\mathbb{R}\to\{+,0,-\}$.
9. ID_s: Student → Natural Number.
10. Population: City → Natural Number.
11. Price: Book → Positive Number.

Exercise 2.35. Add one more colour to the domain, so that the first alphabet map $F\colon\{\text{red},\text{green},\text{blue}\}\to\{a,b,c,f,g,m,o,p,r\}$ is no longer one-to-one. Moreover, add enough colours to the domain, so that F becomes a one-to-one correspondence.

Exercise 2.36. After the domain is made smaller or the range is made larger, how are the onto and one-to-one properties changed?

Exercise 2.37. If gf is onto, prove that g is onto. If gf is one-to-one, prove that f is one-to-one.

Exercise 2.38. If f and g are onto, prove that gf is onto. If f and g are one-to-one, prove that gf is one-to-one.

Exercise 2.39. Rephrase the possible solutions of $f(x) = y$ in terms of the preimage $f^{-1}(y)$. Then interpret onto and one-to-one in terms of $f^{-1}(y)$.

Exercise 2.40. Given a map $f\colon X \to Y$, we may modify the range from Y to $f(X)$ and think of f as a new map $\hat{f}\colon X \to f(X)$. Explain that $\hat{f}$ is onto. Moreover, find the condition for $\hat{f}$ to be a one-to-one correspondence.

Exercise 2.41. When is the "image map" in Exercise 2.33 onto or one-to-one? What about the "preimage map"?

A map $f\colon X \to Y$ is *invertible* if there is another map $g\colon Y \to X$ in the opposite direction, such that

$$gf = id_X, \quad fg = id_Y.$$

In other words, we have

$$g(f(x)) = x, \quad f(g(y)) = y$$

for all $x \in X$ and $y \in Y$.

The map g is the *inverse* of f, and is denoted $g = f^{-1}$. Since the definition is symmetric in f, g, we also know f is the inverse of g, and $(f^{-1})^{-1} = f$.

We remark that the notation $f^{-1}(B)$ for the preimage does not imply f is invertible. Of course, if f happens to be invertible, then the preimage $f^{-1}(B) = g(B)$ is the image of the inverse map g.

Suppose f is invertible, with inverse g. The equality $f(g(y)) = y$ means $g(y)$ is a solution of the equation $f(x) = y$. Therefore $f(x) = y$ has solution for all y, and f is onto.

On the other hand, if $f(x_1) = y = f(x_2)$, then the equality $g(f(x)) = x$ implies $x_1 = gf(x_1) = g(y) = gf(x_2) = x_2$. This means the solution is unique, and f is one-to-one.

We have proved that invertible implies one-to-one, which is the only if part of the following result.

Theorem 2.4.2. *A map $f\colon X \to Y$ is invertible if and only if it is a one-to-one correspondence.*

Proof. We only need to prove the "if" part. This means that we assume f is onto and one-to-one. Then we need to find the inverse map g.

We define a map $g\colon Y \to X$ by the following process: For any $y \in Y$, find $x \in X$ satisfying $f(x) = y$. Then define $g(y) = x$.

Since f is onto, there is $x \in X$ satisfying $f(x) = y$ for any y. In other words, the process is "applicable" to any $y \in Y$.

Since f is one-to-one, if both x_1 and x_2 satisfy $f(x) = y$ for the same y, then

$$f(x_1) = y = f(x_2) \implies x_1 = x_2.$$

In other words, for any y, the element x produced by the process is "unambiguous".

Therefore g is indeed a map. It remains to verify $fg(y) = y$ and $gf(x) = x$.

First, for $y \in Y$, $y' = fg(y)$ is obtained by first finding $g(y) = x \in X$ satisfying $f(x) = y$, and then computing $y' = f(x)$. We clearly have $y' = f(x) = y$. This proves $fg(y) = y$.

Second, for $x \in X$, $x' = gf(x)$ is obtained by first applying f to get $y = f(x)$, and then finding x' satisfying $f(x') = y$. Since we already have $y = f(x)$, we may simply choose $x' = x$ in the second step. This proves $gf(x) = x$. □

Example 2.4.5. To find the inverse of the map $f(x_1, x_2) = (x_1+2x_2+1, 2x_1+3x_2-1) : \mathbb{R}^2 \to \mathbb{R}^2$, we need to solve $f(x_1, x_2) = (y_1, y_2)$, which is the system of equations

$$\begin{aligned} x_1 + 2x_2 + 1 &= y_1, \\ 2x_1 + 3x_2 - 1 &= y_2. \end{aligned}$$

The system has unique solution $x_1 = -3y_1 + 2y_2 + 5, x_2 = 2y_1 - y_2 - 3$. Therefore f is invertible with $f(x_1, x_2) = (-3y_1 + 2y_2 + 5, 2y_1 - y_2 - 3)$.

If the system does not always have unique solution (i.e., either no solution for some (y_1, y_2), or the solution is not unique for some (y_1, y_2)), then the map is not invertible.

Example 2.4.6. Finding the inverse of the function $f(x) = 2x^2 - 1\colon \mathbb{R} \to \mathbb{R}$ is the same as solving the equation $2x^2 - 1 = y$, and the solution should be unique.

We note that the solution does not always exist. For example, $2x^2 - 1 = -2$ has no solution. We also note that the solution may not be unique. For example, $2x^2 - 1 = 7$ has two solutions $x = 2, -2$. Either is the reason for f to be *not* invertible.

In Example 2.4.1, we modify the domain and range to get a one-to-one correspondence $f_3(x) = 2x^2 - 1\colon [0, \infty) \to [-1, \infty)$. For $y \in [-1, \infty)$, the inverse $x = f_3^{-1}(y)$ is obtained by solving $2x^2 - 1 = y$ for $x \in [0, \infty)$. In other words, we find the non-negative solution of the equation, which is

$$f_3^{-1}(y) = \sqrt{\tfrac{y+1}{2}}\colon [-1, \infty) \to [0, \infty).$$

In Example 2.4.1, we also give another one-to-one correspondence modification $f_4(x) = 2x^2 - 1\colon (-1, 0] \cup (1, \infty) \to [-1, \infty)$. Finding $x = f_4^{-1}(y)$ means solving $2x^2 - 1 = y$ for $x \in (-1, 0] \cup [1, \infty)$. By $f_4(-1, 0] = [0, 1)$ and $f_4[1, \infty) = [1, \infty)$, we need to divide into two cases $0 \le y < 1$ and $y \ge 1$. The solution is

$$f_4^{-1}(y) = \begin{cases} -\sqrt{\frac{y+1}{2}}, & \text{if } 0 \le y < 1 \\ \sqrt{\frac{y+1}{2}}, & \text{if } y \ge 1 \end{cases} \colon [-1, \infty) \to (-1, 0] \cup (1, \infty).$$

Example 2.4.7. The flip F with respect to the x-axis in Example 2.3.10 satisfies $FF = id$. Therefore the flip is invertible, and $F^{-1} = F$.

The rotation R_θ by angle θ can be reversed by the rotation $R_{-\theta}$ by angle $-\theta$. Therefore $R_\theta^{-1} = R_{-\theta}$. The formula for the inverse is

$$\begin{aligned} R_\theta^{-1}(y_1, y_2) &= R_{-\theta}(y_1, y_2) \\ &= (y_1 \cos(-\theta) - y_2 \sin(-\theta), y_2 \sin(-\theta) + y_1 \cos(-\theta)) \\ &= (y_1 \cos\theta + y_2 \sin\theta, -y_2 \sin\theta + y_1 \cos\theta). \end{aligned}$$

Example 2.4.8. The first alphabet map $F\colon \{\text{red, green, blue}\} \to \{r, g, b\}$ is invertible, with $F^{-1}(r) = \text{red}, F^{-1}(g) = \text{green}, F^{-1}(b) = \text{blue}$. The enlarged first alphabet map $F_1\colon \{\text{red}, \text{green}, \text{blue}, \text{yellow}\} \to \{b, g, r, w, y\}$ is one-to-one but not onto , and is therefore not invertible. The further enlarged first alphabet map $F_2\colon \{\text{red}, \text{green}, \text{blue}, \text{yellow}, \text{black}, \text{white}\} \to \{b, g, r, w, y\}$ is onto but not one-to-one, and is therefore also not invertible.

Exercise 2.42. Find the inverse map.

1. $f(x, y) = (2x - y - 1, 3x + 2y + 1)\colon \mathbb{R}^2 \to \mathbb{R}^2$.
2. $f(x, y, z) = (x, x + y, x + y + z)\colon \mathbb{R}^3 \to \mathbb{R}^3$.
3. $f(x, y) = (2x + 1, 3x + x^2 - y)\colon \mathbb{R}^2 \to \mathbb{R}^2$.
4. $f(x) = x^4 + 4x^2 + 4\colon (-\infty, 0] \to [4, \infty)$.

Exercise 2.43. We have several one-to-one correspondences at the end of Example 2.4.1. Find their inverses.

Exercise 2.44. In Example 2.4.2, we improved the polar map to become a one-to-one correspondence $\{(0, 0)\} \cup (0, \infty) \times [0, 2\pi) \to \mathbb{R}^2$. Find the inverse of this improved polar map.

Exercise 2.45. Prove that, if two of f, g, gf are invertible, then the third is invertible. Moreover, we have $(gf)^{-1} = f^{-1}g^{-1}$.

Exercise 2.46. A *right inverse* of $f\colon X \to Y$ is a map g, such that $fg = id$. Prove that f has right inverse if and only if f is onto.

Exercise 2.47. A *left inverse* of $f\colon X \to Y$ is a map g, such that $gf = id$. Prove that f has left inverse if and only if f is one-to-one.

2.5 Equivalence Relation

The goal of this section is to discuss the following three equivalent concepts: equivalence relation, partition, and quotient.

An *equivalence relation* on a set X is a collection of pairs, denoted $x \sim y$ for $x, y \in X$, such that the following properties are satisfied:

1. Reflexivity: $x \in X \implies x \sim x$.
2. Symmetry: $x \sim y \implies y \sim x$.
3. Transitivity: $x \sim y$ and $y \sim z \implies x \sim z$.

Example 2.5.1. For integers $x, y \in \mathbb{Z}$, define $x \sim y$ if $x - y$ is even. This is an equivalence relation: First, by $x - x = 0$ being even, we get the reflexivity. Second, if $x - y$ is even, then $y - x = -(x - y)$ is also even. This verifies the symmetry. Third, if $x - y$ and $y - z$ are even, then $x - z = (x - y) + (y - z)$ is still even. This verifies the transitivity.

More generally, we may fix a nonzero integer n and let $n\mathbb{Z} = \{nk\colon k \in \mathbb{Z}\}$ be all the multiples of n. Then we define $x \sim y$ whenever $x - y \in n\mathbb{Z}$. By $x - x = n0 \in \mathbb{Z}$, we get the reflexivity. By $x - y = nk \in n\mathbb{Z} \implies y - x = n(-k) \in n\mathbb{Z}$, we get the symmetry. By $x - y = nk$ and $y - z = nk' \implies x - z = n(k + k')$, we get the transitivity. Therefore we get an equivalence relation.

On the other hand, define $x \sim y$ when $x - y$ is odd. Then the relation does not satisfy any of the three requirements and is not an equivalence relation.

Example 2.5.2. For real numbers $x, y \in \mathbb{R}$, consider the following relations:

- $x \sim y$ if x and y have the same sign.
- $x \sim y$ if $x - y$ is an integer.
- $x \sim y$ if $x \leq y$.

The first two are equivalence relations. In particular, the three properties for the second relation correspond to "zero is an integer", "negative of an integer is an integer", "sum of two integers is an integer". The third is not an equivalence relation because, although reflexive and transitive, it is not symmetric.

Example 2.5.3. Define two points on the plane to be related if one can be moved to another by rotating around the origin. In other words, $u \sim v$ if $v = R_\theta(u)$ for some θ. The rotation $R_0 = id$ gives us the reflexivity. The inverse rotation $R_\theta^{-1} = R_{-\theta}$ gives the symmetry. Since the composition of two rotations is still a rotation, the relation is also transitive. Therefore we get an equivalence relation.

Example 2.5.4. Consider various relations among all the people in the world.

In sibling relation, $x \sim y$ if a person x and another person y have the same parents. The relation is clearly an equivalence relation.

In friend relation, $x \sim y$ if x is a friend of y. The reflexivity condition means that anybody is his or her own friend. The symmetry condition means that if x is a friend of y, then y is also a friend of x. The transitivity condition means that the friend of a friend is a friend. In an ideal world, the conditions appear to hold and the friend relation becomes an equivalence relation.

In the descendant relation, $x \sim y$ if x is a descendant of y. The relation is neither reflexive nor symmetric, although it is transitive.

In the enemy relation, $x \sim y$ if x is an enemy of y. Here the transitivity condition means that the enemy of an enemy is also an enemy, which is not true (the enemy of an enemy is more likely to be a friend). Therefore the enemy relation is not an equivalent one.

Exercise 2.48. Determine whether the relation is an equivalence relation.

1. $X = \mathbb{R}$, $x \sim y$ if $|x| < 1$ and $|y| < 1$.
2. $X = \mathbb{R}$, $x \sim y$ if $|x - y| < 1$.
3. $X = \mathbb{R}$, $x \sim y$ if $x - y$ is an integer.
4. $X = \mathbb{R}^2$, $(x_1, x_2) \sim (y_1, y_2)$ if $x_1^2 + y_1 = x_2^2 + y_2$.
5. $X = \mathbb{R}^2$, $(x_1, x_2) \sim (y_1, y_2)$ if $x_1y_2 - x_2y_1 = 0$.
6. $X = \mathbb{R}^2$, $(x_1, x_2) \sim (y_1, y_2)$ if $x_1 \le x_2$ or $y_1 \le y_2$.
7. $X = \mathbb{R}^2$, $(x_1, x_2) \sim (y_1, y_2)$ if (x_1, x_2) is obtained from (y_1, y_2) by some rotation around $(1, 1)$.
8. $X = \mathbb{R}^2$, $(x_1, x_2) \sim (y_1, y_2)$ if (x_1, x_2) is a scalar multiple of (y_1, y_2).
9. $X = \mathbb{C}$, $x \sim y$ if $x - y$ is a real number.
10. $X = \mathbb{Z} - \{0\}$, $x \sim y$ if $x = ky$ for some $k \in X$.
11. $X = \mathbb{Q} - \{0\}$, $x \sim y$ if $x = ky$ for some $k \in X$.
12. $X = \mathcal{P}(\{1, 2, \dots, n\})$, $A \sim B$ if $A \cap B \neq \emptyset$.

13. $X = \mathcal{P}(\{1, 2, \dots, n\})$, $A \sim B$ if $A \cap B = \emptyset$.

Exercise 2.49. Suppose a relation $\sim$ on X is reflexive and transitive. Prove that if we force the symmetry by adding $x \sim y$ (new relation) whenever $y \sim x$ (existing relation), then we have an equivalence relation.

Exercise 2.50. Suppose $\sim$ is an equivalence relation on a set X. Suppose $A \subset X$ is a subset. The equivalence relation may be restricted to A by defining $a \sim_A b$ for $a, b \in A$ if $a \sim b$ by considering a and b as elements of X. Show that $\sim_A$ is an equivalence relation on A.

Exercise 2.51. Suppose $\sim_X$ and $\sim_Y$ are equivalence relations on X and Y. Prove that on the product $X \times Y$,

$$(x_1, y_1) \sim (x_2, y_2) \iff x_1 \sim_X x_2, y_1 \sim_Y y_2$$

is an equivalence relation. How are the equivalence classes related?

Let $\sim$ be an equivalence relation on a set X. For any element $x \in X$, the subset of all elements related to x

$$[x] = \{y \colon y \sim x\}$$

is the *equivalence class* determined by x. We also call x a *representative* of the equivalence class.

Example 2.5.5. Consider the equivalence relation on integers $\mathbb{Z}$

$$x \sim y \iff x - y \text{ is even}$$

in Example 2.5.1. We have

$$[0] = \{x \colon x = x - 0 \text{ is even}\} = \{\text{even numbers}\} = \text{Even}.$$

Here we simply denote the equivalence class (a subset of $\mathbb{N}$) by Even. We also have $[x] = \text{Even}$ for any even x. We also have

$$[1] = \{x \colon x - 1 \text{ is even}\} = \{\text{odd numbers}\} = \text{Odd}.$$

Moreover, we also have $[x] = \text{Odd}$ for any odd x. In summary, we have

$$[x] = \begin{cases} \text{Even}, & \text{if } x \text{ is even} \\ \text{Odd}, & \text{if } x \text{ is odd} \end{cases}.$$

We note that the whole set $\mathbb{Z}$ is a disjoint union of the equivalence classes

$$\mathbb{Z} = \text{Even} \sqcup \text{Odd}.$$

More generally, for any fixed integer $n \neq 0$, we have the following *mod n equivalence relation* on integers $\mathbb{Z}$

$$x \sim y \iff x - y \in n\mathbb{Z}.$$

Then

$$[0] = \{x \colon x = x - 0 \in n\mathbb{Z}\} = \{\dots, -2n, -n, 0, n, 2n, \dots\} = n\mathbb{Z}.$$

More generally, we have

$$[x] = \{y \colon y - x \in n\mathbb{Z}\} = \{\dots, -2n + x, -n + x, x, n + x, 2n + x, \dots\} = n\mathbb{Z} + x.$$

We note that, for $n = 2$, we have Even $= 2\mathbb{Z}$ and Odd $= 2\mathbb{Z} + 1$.

For $n = 3$, we have three equivalence classes

$$[x] = \begin{cases} 3\mathbb{Z}, & \text{if } x = 3k \\ 3\mathbb{Z} + 1, & \text{if } x = 3k + 1 \\ 3\mathbb{Z} + 2, & \text{if } x = 3k + 2 \end{cases}.$$

Moreover, the whole set $\mathbb{Z}$ is a disjoint union of the three equivalence classes

$$\mathbb{Z} = 3\mathbb{Z} \sqcup (3\mathbb{Z} + 1) \sqcup (3\mathbb{Z} + 2) = [0] \sqcup [1] \sqcup [2].$$

In general, we have the disjoint union of n mod n equivalence classes

$$\mathbb{Z} = n\mathbb{Z} \sqcup (n\mathbb{Z} + 1) \sqcup \cdots \sqcup (n\mathbb{Z} + (n - 1)) = [0] \sqcup [1] \sqcup \cdots \sqcup [n - 1].$$

Theorem 2.5.1. *Let $\sim$ be an equivalence relation on X. For any $x, y \in X$, there are exactly two mutually exclusive possibilities*

$$\begin{aligned} x \sim y &\iff [x] = [y], \\ x \not\sim y &\iff [x] \cap [y] = \emptyset. \end{aligned}$$

In particular, any two equivalence classes are either identical or disjoint. Moreover, the whole set X is the union of all equivalence classes.

Proof. If $x \sim y$, then

$$\begin{aligned} z \in [x] &\implies z \sim x && \text{(definition of } [x]) \\ &\implies z \sim y && (x \sim y \text{ and transitivity}) \\ &\implies z \in [y]. && \text{(definition of } [y]) \end{aligned}$$

This proves $[x] \subset [y]$. On the other hand, by symmetry, we know $x \sim y$ implies $y \sim x$, which similarly implies $[y] \subset [x]$. This proves $[x] = [y]$.

For $x \not\sim y \implies [x] \cap [y] = \emptyset$, we prove the contrapositive

$$\begin{aligned} z \in [x] \cap [y] &\implies z \sim x, z \sim y && \text{(definition of } [x] \text{ and } [y]) \\ &\implies x \sim z, z \sim y && \text{(symmetry)} \\ &\implies x \sim y. && \text{(transitivity)} \end{aligned}$$

We have proved the following

$$\begin{aligned} x \sim y &\implies [x] = [y], \\ x \not\sim y &\implies [x] \cap [y] = \emptyset. \end{aligned}$$

Since $x \sim y$ and $x \not\sim y$ are all the possibilities, and the two are mutually exclusive, the double $\implies$ actually implies double $\impliedby$.

Here is the proof of $[x] = [y] \implies x \sim y$: Suppose $[x] = [y]$. If $x \sim y$ is not true, then by $x \not\sim y \implies [x] \cap [y] = \emptyset$, we get $[x] \cap [y] = \emptyset$. Combined with $[x] = [y]$, we get $[x] = [y] = \emptyset$. However, by the reflexivity, we have $x \in [x]$, so that $[x]$ cannot be empty. The contradiction implies $x \sim y$.

The proof of $[x] \cap [y] = \emptyset \implies x \not\sim y$ is similar.

By $x \in [x]$, we know any element of X is in some equivalence class. This implies X is the union of all equivalence classes. □

Example 2.5.6. For the relation $x \sim y$ on $\mathbb{R}$ when x and y have the same sign, we have

$$[x] = \begin{cases} (0, \infty) & \text{if } x > 0 \\ (-\infty, 0) & \text{if } x < 0 \\ \{0\} & \text{if } x = 0 \end{cases}$$

Then $\mathbb{R} = (-\infty, 0) \sqcup (0, \infty) \sqcup \{0\}$.

Example 2.5.7. For the relation $x \sim y$ when $x - y$ is an integer, in Example 2.5.2, we have

$$[x] = \mathbb{Z} + x = \{n + x \colon n \in \mathbb{Z}\}.$$

For example, $[0] = \mathbb{Z}$ and $[0.2] = \{\dots, -1.8, -0.8, 0.2, 1.2, 2.2, \dots\}$. Then real numbers are decomposed according to the terms after the decimal point

$$\mathbb{R} = \sqcup_{0 \le x < 1} (\mathbb{Z} + x).$$

Example 2.5.8. For the rotation equivalence relation in Example 2.5.3, the equivalence class is the circle of radius $\|v\|$ ($\|v\| = \|(x, y)\| = \sqrt{x^2 + y^2}$)

$$[v] = \{u \colon \|u\| = \|v\|\} = C_{\|v\|}, \quad C_r = \{(x, y) \colon x^2 + y^2 = r^2\}.$$

Then the plane $\mathbb{R}^2$ is decomposed into concentric circles

$$\mathbb{R}^2 = \sqcup_{r \ge 0} C_r.$$

Exercise 2.52. For the equivalence classes in Exercise 2.48, find the equivalence classes.

A decomposition of a set X into a disjoint union of nonempty subsets is a *partition*

$$X = \sqcup_{i \in I} X_i, \quad X_i \neq \emptyset.$$

We see an equivalence relation leads a partition of X into the disjoint union of equivalence classes

$$X = \sqcup_i [x_i].$$

Here we pick one *representative* x_i from each equivalence class. For example, the partition $\mathbb{Z} = \text{Even} \sqcup \text{Odd} = [0] \sqcup [1]$ in Example 2.5.5 uses 0 and 1 as representatives of even and odd numbers. We may use $1, -1, 0$ as representatives of signs, and express the partition in Example 2.5.6 as $\mathbb{R} = [1] \sqcup [-1] \sqcup [0]$. In Example 2.5.7, we use $0 \leq x < 1$ (i.e., terms after the decimal point) as representatives.

In particular, any equivalence relation on X induces a partition into disjoint union of equivalence classes. For example, the same sign relation induces the partition $\mathbb{R} = (0, \infty) \cup (-\infty, 0) \cup \{0\}$. The even difference relation induces the partition $\mathbb{Z} = 2\mathbb{Z} \cup (2\mathbb{Z} + 1)$. The rotation relation divides the plane into concentric circles centered at the origin.

Conversely, a partition induces a relation:

$$x \sim y \quad \text{if } x \text{ and } y \text{ belong to the same subset } X_i.$$

It is easy to see that this is an equivalence relation, and the equivalence classes are given by

$$[x] = X_i \quad \text{if } x \in X_i.$$

This establishes the one-to-one correspondence between the equivalence relations on X and the partitions of X.

Example 2.5.9. The partition $\mathbb{Z} = 2\mathbb{Z} \sqcup (2\mathbb{Z} + 1)$ gives a relation $x \sim y$ described as follows:

- $x, y \in 2\mathbb{Z}$: Both x and y are even.

- $x, y \in 2\mathbb{Z} + 1$: Both x and y are odd.

The relation is equivalent to $x - y$ being even.

Example 2.5.10. The partition $\mathbb{R} = (0, \infty) \sqcup (-\infty, 0) \sqcup \{0\}$ gives a relation $x \sim y$ described as follows:

- $x, y \in (0, \infty)$: Both x and y are positive.

- $x, y \in (-\infty, 0)$: Both x and y are negative.

- $x, y \in \{0\}$: Both x and y are zero.

This is the same sign relation in Example 2.5.6.

Example 2.5.11. For nonempty X and Y, the product $X \times Y$ can be partitioned into "vertical lines": $X \times Y = \sqcup_{x \in X} x \times Y$. The corresponding equivalence relation is

$$\begin{aligned}(x_1, y_1) \sim (x_2, y_2) &\iff (x_1, y_1), (x_2, y_2) \in x \times Y \text{ for some } x \\ &\iff x_1 = x,\ x_2 = x \text{ for some } x.\end{aligned}$$

Therefore $(x_1, y_1) \sim (x_2, y_2)$ means $x_1 = x_2$.

In a partition $X = \cup_{i \in I} X_i$, the set I of indices serve as labels for the subsets. For example,

1. $\mathbb{R} = (0, \infty) \sqcup (-\infty, 0) \sqcup \{0\}$: $I = \{+, -, 0\}$.
2. $\mathbb{Z} = 2\mathbb{Z} \sqcup (2\mathbb{Z} + 1)$: $I = \{\text{even}, \text{odd}\}$.
3. $\mathbb{R}^2 = \sqcup_{r \geq 0} C_r$: $I = \{\text{radii of circles}\} = [0, \infty)$.
4. $X \times Y = \sqcup_{x \in X} x \times Y$: $I = \{x\text{-coordinates}\} = X$.

Mathematically, the label is nothing but a map

$$q \colon X \to I, \quad q(x) = i \text{ if } x \in X_i.$$

The map is always onto, and is the *quotient map* for the partition. The quotients for the examples above are

1. Sign: $\mathbb{R} \to \{+, -, 0\}$.
2. Parity: $\mathbb{Z} \to \{\text{even}, \text{odd}\}$.
3. Norm: $\mathbb{R}^2 \to [0, \infty)$, $(x, y) \mapsto \sqrt{x^2 + y^2}$.
4. Projection: $X \times Y \to X$, $(x, y) \mapsto x$.

Conversely, an onto map $q \colon X \to I$ induces a partition

$$X = \sqcup_{i \in I} q^{-1}(i).$$

Here X equals the union because the process q can be applied to all elements of x. Moreover, the union is disjoint because output of the process q is not ambiguous. Moreover, $q^{-1}(i)$ are not empty because q is onto. The corresponding equivalence relation is

$$x \sim y \iff q(x) = q(y).$$

Exercise 2.53. Find the equivalence relation and the partition corresponding to the onto map $q(x) = |x|\colon \mathbb{C} \to [0, \infty)$.

Exercise 2.54. Find the partition and the quotient corresponding to the equivalence relation on $\mathbb{C}$ defined by $x \sim y$ if $x = ry$ for some real number $r > 0$.

Now we have three equivalent concepts: equivalence relation, partition, and quotient (i.e., onto) map. We know how to translate from one concept to another, with the only exception of equivalence relation $\implies$ quotient map.

An equivalence relation $\sim$ on X induces a partition of X into equivalence classes $[x]$. In Examples 2.5.7 and 2.5.8, we find natural labels for these equivalence classes. In general, we may simply take the equivalence class $[x]$ as the label of $[x]$ itself. In other words, we may simply take the collection of equivalence classes (called the *quotient set*)

$$X/\!\sim\; = \{[x]\colon x \in X\}$$

as the index set I. The quotient map is the map from x to its equivalence class $[x]$

$$q(x) = [x]\colon X \to X/\!\sim\; .$$

Example 2.5.12. For the equivalence class $x \sim y \iff x - y$ is even in Example 2.5.5, the quotient set $\mathbb{Z}/\sim\; = \{2\mathbb{Z}, 2\mathbb{Z}{+}1\} = \{[0], [1]\}$. The quotient map $\mathbb{Z} \to \mathbb{Z}/\sim$ is the parity map.

In general, for the mod n equivalence relation, we denote the quotient set by $\mathbb{Z}/n\mathbb{Z} = \{[0], [1], \dots, [n-1]\}$. The quotient map is the "remainder map"

$$\mathbb{Z} \to \mathbb{Z}/n\mathbb{Z},\ x \mapsto [r].$$

Specifically, we divide x by n and get

$$x = qn + r, \quad q, r \in \mathbb{Z},\ 0 \le 1 < n.$$

Here q is the quotient of the division, and r is the remainder of x divided by n.

Example 2.5.13. Consider the second equivalence relation $x \sim y \iff x - y$ is an integer on $\mathbb{R}$, in Example 2.5.2. In Example 2.5.7, we identified all equivalence classes, which is the quotient set

$$\mathbb{R}/\mathbb{N} = \{[x] = \mathbb{Z} + x\colon 0 \le x < 1\}.$$

The quotient map $q\colon \mathbb{R} \to \mathbb{R}/\mathbb{N}$ is the "after decimal point map". For example, $q(3.2) = q(-1.8) = [0.2]$, $q(\pi) = [0.1415926\cdots]$.

We note that the notation $\mathbb{R}/\mathbb{N}$ is due to $x - y \in \mathbb{N}$, similar to the notation $\mathbb{Z}/n\mathbb{Z}$ due to $x - y \in n\mathbb{N}$.

The quotient set can be clearly identified with the interval $[0, 1)$. Since $0.999\cdots = 1$, the "right end" of the interval should really be identified with $[1] = [0]$. This means that it is better to identify the quotient set as $\dfrac{[0,1]}{0\sim 1}$, which is the closed interval $[0,1]$ with the two ends 0 and 1 identified. What we get is actually the circle

$$\mathbb{R}/\mathbb{N} \cong \frac{[0,1]}{0\sim 1} \cong S^1 = \{(x,y)\colon x^2+y^2=1\}.$$

Here $\cong$ means identifying two sets.

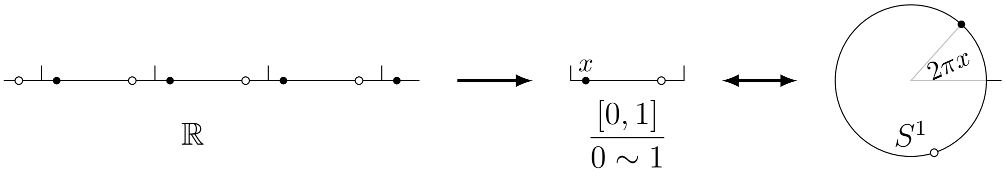

Figure 2.8: Quotient set $\mathbb{R}/\mathbb{N}$.

From Figure 2.8, we see that $[x] \in \mathbb{R}$ is identified with the point $(\cos 2\pi x, \sin 2\pi x)$ of angle $2\pi x$ on the circle. The quotient map $\mathbb{R} \to S^1$ can be understood as a rope $\mathbb{R}$ wrapping around the circle S^1.

Example 2.5.14. For the rotation equivalence in Example 2.5.3, we get know the equivalence classes are concentric circles centered at the origin. The most natural way to label the circles is by their radius r. The quotient set is then the collection $[0,\infty)$ of all radii. The quotient map is the radius function.

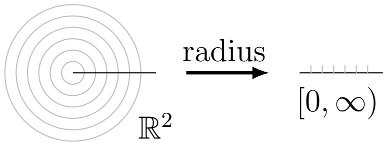

Figure 2.9: Quotient set and quotient map for rotation equivalence.

The more general setting is a group G acting on a set X. Here a group means all the invertible movements of X of certain kind, such that the identity movement (i.e., no move) is included, and composition of movements are also included. Then $x \sim y$ if x can be moved to y by the kind of movement.

For example, let X be the unique sphere $S^2 = \{(x,y,z)\colon x^2+y^2+z^2=1\}$, and the let G be the collection of rotations around the north-south axis. Then the quotient set S^2/G can be identified with the latitudes. The quotient map $S^2 \to S^2/G$ is simply the latitude.

Exercise 2.55. On any set X, define $x \sim y$ if $x = y$. Show that this is an equivalence relation. What are the corresponding partition and the quotient?

Exercise 2.56. Let $F(x) = -x$ be the flip of the line $\mathbb{R}$. In fact, the flip and the identity form a group $\{F, id\}$ that is usually denoted $\mathbb{Z}_2$. This induces an equivalence relation: $x \sim y$ if $x = F(y)$ or $x = id(y) = y$. Draw pictures for the equivalence classes and the quotient set. Then explain the quotient map.

Exercise 2.57. Let $F(x, y) = (x, -y)$ be the flip of the circle S^1 with respect to the x-axis. Again the flip and the identity form a group $\mathbb{Z}_2 = \{F, id\}$ and induce an equivalence relation on the circle. Draw pictures for the equivalence classes and the quotient set. Then explain the quotient map.

Moreover, change F to the antipodal map $F(x, y) = (-x, -y)$, and carry out the similar discussion.

Exercise 2.58. Let $\mathcal{F}$ be the collection of all finite sets. For $A, B \in \mathcal{F}$, define $A \sim B$ if there is a one-to-one correspondence $f \colon A \to B$. Prove that this is an equivalence relation. Moreover, identify the quotient set as the set of non-negative integers and the quotient map as the number of elements in a set. The exercise leads to a general theory of counting.

Exercise 2.59. Let $\sim_1$ and $\sim_2$ be two equivalence relations on X, such that $x \sim_1 y$ implies $x \sim_2 y$. How are the quotient sets $X/{\sim_1}$ and $X/{\sim_2}$ related?

Chapter 3

Number

3.1 Natural Number

The *natural numbers* are $1, 2, 3, 4, \ldots$. These numbers are used in counting. For example, there are four alphabets in the set $\{a, b, c, d\}$. The number 4 is then the *cardinality* of the set $\{a, b, c, d\}$. The natural numbers are also used to indicate the location in an ordered sequence of elements. For example, the alphabet c is the third in the sequence a, b, c, d. The number 3 is then the *ordinality* of c in the sequence.

The natural numbers are rigorously defined by the Peano's axioms[1].

Definition 3.1.1. The natural number is a set $\mathbb{N}$ satisfying the following properties:

1. There is a special element $1 \in \mathbb{N}$.

2. For any $n \in \mathbb{N}$, there is a unique *successor* $n' \in \mathbb{N}$.

3. For any $n \in \mathbb{N}$, we have $n' \neq 1$.

4. If $m' = n'$, then $m = n$.

5. If a subset $S \subset \mathbb{N}$ contains 1 and has the property that $n \in S \implies n' \in S$, then $S = \mathbb{N}$.

The first axiom gives us the starting number, naturally denoted by 1.

The intuition for the second axiom (and the name successor) is obviously $n' = n + 1$. For example, 2 is the successor of 1, 3 is the successor of 2, and 4 is the successor of 3, etc. Note that $n + 1$ is meaningless at the moment because the addition has not been defined. The intention of the first two axioms is to "build

[1] Giuseppe Peano: born 27 Aug 1858 in Cuneo, Piemonte, Italy; died 20 April 1932 in Turin, Italy. The famous axioms were published in *Arithmetices principia, nova methodo exposita* in 1889. Another stunning invention of his was the "space-filling" curves in 1890.

up" all the natural numbers by picking up the starting number 1 and then creat the subsequent ones by repeatedly applying the "successor operation"[2].

The third axiom says that the special number 1 is not a successor. Therefore 1 is the "beginning", and no other natural numbers are "prior" to it.

The fourth axiom says that, if two natural numbers have the same successors, then the two numbers are the same. Therefore we can talk about the *predecessor* of a natural number unambiguously, when the number itself is a successor. For example, 1 is the predecessor of 2, and 2 is the predecessor of 3, etc.

The fifth axiom is the *induction axiom*. Recall that the induction for a sequence of statements $A(1), A(2), A(3), \dots$ involves the verification that $A(1)$ holds, and the proof that $A(n)$ holds implying $A(n')$ also holds. To see how the fifth axiom implies the induction process, we denote

$$S = \{n \in \mathbb{N} \colon A(n) \text{ is true}\}.$$

The two steps in the induction basically mean that $1 \in S$ and $n \in S \implies n' \in S$. Then the fifth axiom would imply that $S = \mathbb{N}$, which means $A(n)$ is true for all n.

The induction axiom has the following implicatiom.

Proposition 3.1.2. *Any natural number other than* 1 *is a successor.*

Proof. Let

$$S = \{1\} \cup \{n' \colon n \in \mathbb{N}\}.$$

Then $1 \in S$. Moreover, for any $n \in \mathbb{N}$, we have $n' \in S$. Then by the fifth axiom, we conclude $S = \mathbb{N}$. □

Definition 3.1.3. The *addition* $m + n$ of two natural numbers is the operation characterized by

- $m + 1 = m'$.
- $m + n' = (m + n)'$.

The definition is consistent with our intuition and only makes use of the knowledge provided by the Peano's axioms. Moreover, it is a typical *inductive definition*. Specifically, we fix the first number m and then induct on the second number n. The first property in the definition means that, for $n = 1$, $m+1$ is the the successor of m. The second property means that, if $m+n$ has been defined, then $m+n'$ is the successor of $m + n$. By the induction axiom, we know that, for any fixed number m, $m + n$ is defined for all n. As a result, $m + n$ is defined for all m and n.

Strictly speaking, we need to verify that the sum, as a map from $\mathbb{N} \times \mathbb{N}$ to $\mathbb{N}$, is *well*-defined by the inductive process. To see the subtle issue involved, all the ambiguities in the inductive process need to be eliminated.

[2] The fifth axiom makes sure that such a construction indeed gives us all the natural numbers.

The first ambiguity is the possibility that the two properties in the definition may overlap. Hypothetically, this may happen if $1 = n'$. Since this is impossible by the third axiom, we know there is no overlap between the two properties.

The second ambiguity is the possibility that the equality in the second item may overlap itself. The equality $m + n' = (m + n)'$ really means that $m + n'$ is defined as $(m + n)'$. Therefore the definition of $m + n'$ depends on the definition of $m + n$, which further depends on the choice of the predecessor n of n'. By the fourth axiom, the predecessor n is uniquely determined by n'. Therefore there is no overlap within the second property.

In summary, the addition is well-defined, thanks to all the axioms.

Proposition 3.1.4. *The addition of natural numbers has the following properties:*

1. *Cancelation:* $m + k = n + k \implies m = n$.
2. *Associativity:* $m + (n + k) = (m + n) + k$.
3. *Commutativity:* $m + n = n + m$.

Proof. We only prove the first two properties. The third is left as an exercise.

The following verifies the cancelation property for $n = 1$:

$$\begin{aligned} m + 1 = n + 1 &\implies m' = n' && \text{(first property in the definition of addition)} \\ &\implies m = n. && \text{(fourth Peano axiom)} \end{aligned}$$

Under the inductive assumption $m + k = n + k \implies m = n$, we have

$$\begin{aligned} m + k' = n + k' &\implies (m + k)' = (n + k)' && \text{(second property)} \\ &\implies m + k = n + k && \text{(fourth Peano axiom)} \\ &\implies m = n. && \text{(inductive hypothesis)} \end{aligned}$$

This completes the inductive proof.

To prove the associativity, we fix m, n and induct on k. The following verifies the associativity for $k = 1$:

$$\begin{aligned} m + (n + 1) &= m + n' && \text{(first property)} \\ &= (m + n)' && \text{(second property)} \\ &= (m + n) + 1. && \text{(first property)} \end{aligned}$$

Under the inductive assumption $m + (n + k) = (m + n) + k$, we have

$$\begin{aligned} m + (n + k') &= m + (n + k)' && \text{(second property)} \\ &= (m + (n + k))' && \text{(second property)} \\ &= ((m + n) + k)' && \text{(inductive assumption)} \\ &= (m + n) + k'. && \text{(second property)} \end{aligned}$$

This completes the inductive proof. □

The associativity implies that the additions $(m+n)+(k+l)$, $(m+(n+k))+l$, $m+(n+(k+l))$ of natural numbers are all equal. Therefore we may write $m+n+k+l$ without any ambiguity. Moreover, the commutativity allows us to freely exchange orders of numbers in an addition, such as $k+n+l+m=m+n+k+l$.

Exercise 3.1. The equality $m+1=1+m$ may be proved by inducting on m. For $m=1$, the equality holds trivially. Next assume $m+1=1+m$. Then

$$\begin{aligned} m'+1 &= (m+1)+1 && \text{(first property)} \\ &= (m+1)' \\ &= (1+m)' \\ &= 1+m'. \end{aligned}$$

Fill in the reason for each step.

Exercise 3.2. The equality $m+n=n+m$ may be proved by inducting on n. For $n=1$, the equality is proved in Exercise 3.1. Next assume $m+n=n+m$. Then

$$\begin{aligned} m+n' &= m+(n+1) \\ &= (m+n)+1 && \text{(associativity)} \\ &= 1+(m+n) \\ &= 1+(n+m) \\ &= (1+n)+m \\ &= (n+1)+m \\ &= n'+m. \end{aligned}$$

Fill in the reason for each step.

Exercise 3.3. Prove the commutativity in Proposition 3.1.4.

Exercise 3.4. Using Proposition 3.1.4, the following proves $(m+n)+(k+l)=(m+k)+(n+l)$.

$$\begin{aligned} (m+n)+(k+l) &= ((m+n)+k)+l && \text{(associativity)} \\ &= (m+(n+k))+l \\ &= (m+(k+n))+l \\ &= ((m+k)+n)+l \\ &= (m+k)+(n+l). \end{aligned}$$

Provide the reason for each step.

Exercise 3.5. Define the order $m \le n$ between natural numbers by $1 \le n$ for any n, and $m \le n \implies m' \le n'$. Show that there is no ambiguity in the definition.

3.2 Integer

The *integers* are $\ldots, -4, -3, -2, -1, 0, 1, 2, 3, 4, \ldots$. We will construct integers from natural numbers, especially the negative integers. The idea is to define integers as differences between natural numbers:

$$-\mathbf{2} = 3 - 5,\ \mathbf{0} = 2 - 2,\ -\mathbf{5} = 1 - 6,\ \mathbf{3} = 5 - 2,$$

where the bold faced numbers are the integers to be constructed, and the normal faced numbers are the natural numbers used for constructing integers. In other words, we attempt to identify -2, 0, -5, 3 with the *ordered pairs* $(3, 5)$, $(2, 2)$, $(1, 6)$, $(5, 2)$. In general, for $m, n \in \mathbb{N}$, the pair (m, n) is intended to represent the integer $m - n$.

Note that we use the pair (m, n) instead of the subtraction notation $m - n$, because the subtraction operation is not yet defined. In fact, we do not have subtraction *within* $\mathbb{N}$. The subtraction can be defined only *after* the whole set $\mathbb{Z}$ of integers is constructed.

There is just one problem with the idea above. The same integer can be expressed as the difference of many pairs of natural numbers. For example, -2 can be represented by $(3, 5)$, by $(4, 6)$, or by $(5, 7)$. Therefore the pairs $(3, 5)$, $(4, 6)$, $(5, 7)$, etc, should be considered as equivalent as far as the integers they represent are concerned. This leads to the definition of integers as equivalence classes of natural number pairs.

Definition 3.2.1. The *integers* is the set $\mathbb{Z}$ of the equivalence classes (i.e., quotient set) of pairs (m, n) of natural numbers $m, n \in \mathbb{N}$, under the equivalence relation

$$(m, n) \sim (k, l) \iff m + l = n + k.$$

An integer is then an equivalence class $[(m, n)]$, $m, n \in \mathbb{Z}$, which we will simply denote by $[m, n]$. For example, $-2 = [3, 5]$, $0 = [2, 2]$, $-5 = [1, 6]$, $3 = [5, 2]$. We also have $[m, n] = [k, l] \iff m + l = n + k$.

Strictly speaking, we should verify the relation in the definition is indeed an equivalence relation:

1. Reflexivity $(m, n) \sim (m, n)$: This means $m + n = n + m$, which is exactly the commutativity in Proposition 3.1.4.

2. Symmetry $(m, n) \sim (k, l) \implies (k, l) \sim (m, n)$: This means $m + l = n + k \implies k + n = l + m$, which follows from the commutativity in Proposition 3.1.4.

3. Transitivity $(m,n)\sim(k,l)$ and $(k,l)\sim(p,q) \implies (m,n)\sim(p,q)$:

$$\begin{aligned}
&(m,n)\sim(k,l),\ (k,l)\sim(p,q)\\
&\iff m+l=n+k,\ k+q=l+p\\
&\implies (m+l)+(k+q)=(n+k)+(l+p)\\
&\implies (m+q)+(k+l)=(n+p)+(k+l) \quad \text{(associativity and commutativity)}\\
&\implies m+q=n+p \quad \text{(cancelation)}\\
&\iff (m,n)\sim(p,q).
\end{aligned}$$

The addition of natural numbers can be extended to integers. Based on the expectation $(m-n)+(k-l)=(m+k)-(n+l)$, we may define the addition of two integers by

$$[m,n]+[k,l]=[m+k,n+l].$$

Specifically, the addition is the following process: For any integers $a,b\in\mathbb{Z}$, we express them as $a=[m,n]$ and $b=[k,l]$ for some $m,n,k,l\in\mathbb{N}$. Then $a+b=[m+k,n+l]$.

By the definition of integers, the process always work for any pair of integers $a,b\in\mathbb{Z}$. On the other hand, we need the process has to be unambiguous, due to the choices of m,n,k,l. Consider two ways of representing a and b

$$a=[m_1,n_1]=[m_2,n_2],\quad b=[k_1,l_1]=[k_2,l_2].$$

We need to verify

$$[m_1+k_1,n_1+k_1]=[m_2+k_2,n_2+k_2].$$

By $[m_1,n_1]=[m_2,n_2]$ and $[k_1,l_1]=[k_2,l_2]$, we get $m_1+n_2=n_1+m_2$ and $k_1+l_2=l_1+k_2$. Then by Proposition 3.1.4, this implies $(m_1+k_1)+(n_2+k_2)=(n_1+k_1)+(m_2+k_2)$. The equality means $[m_1+k_1,n_1+k_1]=[m_2+k_2,n_2+k_2]$. This proves that the addition in $\mathbb{Z}$ is well-defined.

Proposition 3.2.2. *The addition of integers has the following properties:*

1. *Associativity: $a+(b+c)=(a+b)+c$.*

2. *Commutativity: $a+b=b+c$.*

3. *Zero: There is a unique integer 0 satisfying $a+0=0+a=a$.*

4. *Negative: For any integer a, there is a unique integer $-a$ satisfying $a+(-a)=(-a)+a=0$.*

Proof. The first and the second properties follow directly from the corresponding properties in Proposition 3.1.4 and the definition of addition of integers.

For integers $a = [m, n]$ and $z = [k, l]$, we have $a + z = [m + k, n + l]$, and

$$\begin{aligned} a + z = a &\iff m + k + n = n + l + m && \text{(definition of equivalence class)} \\ &\iff k = l. && \text{(cancelation in Proposition 3.1.4)} \end{aligned}$$

The argument shows the existence of 0, which must be of the form $z = [k, k]$. Moreover, we have $[k, k] = [l, l]$ for any $k, l \in \mathbb{N}$. Therefore the special integer 0 is unique.

For integers $a = [m, n]$ and $b = [k, l]$, we have $a + b = [m + k, n + l]$, and by using $0 = [1, 1]$,

$$\begin{aligned} a + b = 0 &\iff m + k + 1 = n + l + 1 && \text{(definition of equivalence class)} \\ &\iff m + k = n + l && \text{(fourth Peano axiom)} \\ &\iff b = [n, m]. && \text{(definition of equivalence class)} \end{aligned}$$

The argument shows the existence of $-a = -[m, n]$, which must be of the form $b = [n, m]$. In particular, the negative $-a$ of an integer a is unique. □

Exercise 3.6. Prove the first and second properties in Proposition 3.2.2.

Exercise 3.7. The following is another way of proving the third property in Proposition 3.2.2.

1. Verify that $a + [1, 1] = a$ for any integer a. This gives the existence of 0.

2. Next we prove the uniqueness. Suppose 0 and $\bar{0}$ are two candidate integers for zero. Then we have $a + 0 = 0 + a = a$ and $a + \bar{0} = \bar{0} + a = a$. By taking $a = 0$ and $a = \bar{0}$ in these equalities, show that $0 = \bar{0}$.

Exercise 3.8. The following is another way of proving the fourth property in Proposition 3.2.2.

1. Verify that $[m, n] + [n, m] = [1, 1]$ for any natural numbers m, n. This gives the existence of the nagative.

2. Next we prove the uniqueness. Suppose b and c are two candidate integers for $-a$. Then we have $a + b = b + a = 0$ and $a + c = c + a = 0$. This implies

$$b = b + 0 = b + (a + c) = (b + a) + c = 0 + c = c.$$

Provide reason for each step.

As remarked after the proof of Proposition 3.1.4, the associativity and the commutativity imply that we may write the addition of integers such as $a + b + c + d$ without any ambiguity, and we may freely change the order of terms in the addition.

Compared with natural numbers, the zero and negative are the new ingredients for integers. The existence of negative allows us to introduce the subtraction operation for integers, defined as

$$a - b = a + (-b).$$

Then the expressions such as $a - b + c - d$ make sense for integers. Moreover, the property in Proposition 3.2.2 regarding the negative becomes $a - a = 0$. We also get the *cancelation law* for integers by subtracting c:

$$a + c = b + c \implies a = b.$$

The subtraction also has the following properties:

1. $-(a + b) = -a - b$, where $-a - b$ really means $(-a) + (-b)$.

2. $a = b \iff a - b = 0$.

3. $-(-a) = a$.

The proofs are left as exercises.

Exercise 3.9. Prove $[m, n] = m - n$. In other words, $[m, n] = [m + 1, 1] - [n + 1, 1]$.

Exercise 3.10. Let $a, b \in \mathbb{Z}$, explain each step in the following computation by properties in Proposition 3.2.2:

$$(a + b) + ((-a) + (-b)) = (b + a) + ((-a) + (-b)) = ((b + a) + (-a)) + (-b)$$

$$= (b + (a + (-a))) + (-b) = (b + 0) + (-b) = b + (-b) = 0.$$

Then use the uniqueness of negative to conclude that $-(a + b) = -a - b$.

Exercise 3.11. Use properties in Proposition 3.2.2 and the cancelation law to prove the uniqueness of negative: $a + b = 0 \implies b = -a$.

Exercise 3.12. Explain why $-(-a) = a$.

Exercise 3.13. Using propositions and earlier exercises, provide reason for each step of the following proof of $-(a - b) = b - a$:

$$-(a - b) = -(a + (-b)) = -((-b) + a) = -(-b) - a = b - a.$$

3.3 Order

We know the natural number $\mathbb{N}$ is part of the integer $\mathbb{Z}$. This means that we may identify $\mathbb{N}$ with a subset of $\mathbb{Z}$. By $n = (n+1) - 1$, we introduce a map

$$f(n) = [n+1, 1] \colon \mathbb{N} \to \mathbb{Z}.$$

The map is one-to-one because $f(m) = f(n)$ means $m + 1 + 1 = 1 + n + 1$. By the cancelation property for integers, this implies $m = n$. It is also easy to verify that the map also preserves the sum

$$f(m+n) = f(m) + f(n).$$

The one-to-one map f identifies the natural numbers with the subset $f(\mathbb{N})$ of $\mathbb{Z}$. Therefore we may simply write $n = [n+1, 1]$ and call integers of the form $[n+1, 1]$ natural numbers. Based on this understanding, we may describe all the integers as follows.

Proposition 3.3.1. *Any integer is either zero, or a natural number, or the negative of a natural number. The three cases are mutually exclusive.*

Proof. First we prove that any integer $a = [m, n]$ can be written as either $[r, 1]$ or $[1, r]$ for some $r \in \mathbb{N}$. We rephrase the claim as the following (which depends on $n \in \mathbb{N}$)

- $A(n)$: For any $m \in \mathbb{N}$, there is $r \in \mathbb{N}$, such that either $[m, n] = [r, 1]$ or $[m, n] = [1, r]$.

First, we know $A(1)$ is true for trivial reason. Next, we assume $A(n)$ is true and try to prove $A(n+1)$. By Proposition 3.1.2, any $m \in \mathbb{N}$ has two possibilities:

- $m = 1$. In this case, $[m, n] = [1, r]$ for $r = n$.
- $m = k+1$ for some $k \in \mathbb{N}$. In this case, we have $[m, n+1] = [k, n] + [1, 1] = [k, n]$ because $[1, 1] = 0$. Applying $A(n)$ to $[k, n]$, we get $A(n)$ for $[m, n+1]$.

This completes the inductive proof of $A(n)$.

By Proposition 3.1.2, in the statement $A(n)$ above, either $r = 1$ or $r = s + 1$ for some $s \in \mathbb{N}$. Therefore we have three possibilities for any integer $a \in \mathbb{Z}$:

- $a = [1, 1] = 0$.
- $a = [s+1, 1] \in \mathbb{N}$ is a natural number.
- $a = [1, s+1] = -[s+1, 1]$, where the second equality appeared in the proof of Proposition 3.1.4. In particular, this means that a is the negative of a natural number.

Finally, we need to prove that the three cases are mutually exclusive. First, if $[s+1,1]=[1,1]$, then $(s+1)+1=1+1$. By the fourth Peano axiom, we have $s+1=1$, contradicting the third Peano axiom. Therefore $[s+1,1]\neq[1,1]$. Similarly, $[1,s+1]\neq[1,1]$. Moreover, if $[s+1,1]=[1,r+1]$, then $(s+1)+(r+1)=1+1$, which leads to similar contradiction. This completes the proof. □

By the proposition, we may use the notations $\dots,-4,-3,-2,-1,0,1,2,3,4,\dots$ to denote integers, without any ambiguity. The proposition can also be used to introduce order among integers.

Definition 3.3.2. We say an integer a is *bigger* than another integer b, and denote $a>b$, if $a-b\in\mathbb{N}$. We also say b is *smaller* than a and denote $b<a$.

Proposition 3.3.3. *The order among integers has the following properties:*

1. *For any two integers a and b, one of the following mutually exclusive cases happen: $a=b$, $a>b$, $a<b$.*

2. *$a>b$ and $b>c \implies a>c$.*

3. *$a>b \implies a+c>b+c$.*

4. *$a>b \implies -a<-b$.*

5. *1 is the smallest natural number.*

We write $a\geq b$ for $a>b$ or $a=b$. The first property says that $a\geq b$ is the opposite of $a<b$. We may similarly introduce $a\leq b$, which is the opposite of $a>b$.

By the first property, we may also define $\max\{a,b\}$ and $\min\{a,b\}$.

Proof. By Proposition 3.3.1, there are three mutually exclusive possibilities for the integer $a-b$:

1. $a-b=0$: This means $a=b$.

2. $a-b\in\mathbb{N}$: This means $a>b$.

3. $a-b=-n$ for some $n\in\mathbb{N}$: We have $b-a=n\in\mathbb{N}$. This means $a<b$.

This proves the first property.

The second property follows from $a-c=(a-b)+(b-c)$ and the fact the addition of natural numbers is a natural number (consistency of the addition of integers in $\mathbb{N}$ and $\mathbb{Z}$). The third property follows from

$$(a+c)-(b+c)=a+c-b-c=a-b.$$

The fourth property follows from

$$(-b) - (-a) = -b + a = a - b.$$

The fifth property follows from Proposition 3.1.2: Any natural number other than 1 is of the form $n+1$, and $n+1 > 1$ because $(n+1) - 1 = n \in \mathbb{N}$. □

Exercise 3.14. Prove that $m \leq n$ in the sense of Exercise 3.5, if and only if $m < n$ (in the sense of Definition 3.3.2) or $m = n$.

Exercise 3.15. Prove that $a > b$ and $c > d$ imply $a + b > c + d$.

Exercise 3.16. Prove the properties of $a \geq b$:

1. Reflexivity: $a \geq a$.

2. Antisymmetry: $a \geq b$ and $b \geq a$ implies $a = b$.

3. Transitivity: If $a \geq b$ and $b \geq c$, then $a \geq c$.

Moreover, find more properties such as $a \geq b$ and $b > c$ imply $a > c$.

Exercise 3.17. An *(total) order* on a set X is a relation $x < y$ among pairs of elements, satisfying the following properties:

1. Trichotomy: For any x, y, one and only one fo the following holds: $x > y$, $x < y$, $x = y$.

2. Transitivity: If $x > y$ and $y > z$, then $x > z$.

Show that the relation $a < b$ for integers is an order.

Exercise 3.18. Prove $\max\{a, \max\{b, c\}\} = \max\{\max\{a, b\}, c\} = \max\{\max\{a, c\}, b\}$.

3.4 Multiplication

We first define the multiplication of natural numbers. The definition is inductive, just like the definition of the addition of integers.

Definition 3.4.1. The *multiplication* mn of two natural numbers is the operation characterized by

- $m1 = m$.

- $mn' = mn + m$.

Similar to addition, the first thing we need to do after the definition is to show multiplication is well-defined. The argument is similar to the addition, and is left as exercise.

Proposition 3.4.2. *The multiplication of natural numbers has the following properties:*

1. *Distributivity:* $(m+n)k = mk + nk$.
2. *Associativity:* $m(nk) = (mn)k$.
3. *Commutativity:* $mn = nm$.

Proof. We only prove the first and third properties. The second is left as an exercise.

To prove the distributivity, we fix m, n and induct on k. The case $k = 1$ follows from the first property in the definition of multiplication:

$$(m+n)1 = m + n = m1 + n1.$$

Under the inductive assumption $(m+n)k = mk + nk$, we have

$$\begin{aligned}
(m+n)k' &= (m+n)k + (m+n) && \text{(second property in definition)}\\
&= (mk+nk) + (m+n) && \text{(inductive assumption)}\\
&= (mk+m) + (nk+n) && \text{(Proposition 3.1.4)}\\
&= mk' + nk'. && \text{(second property)}
\end{aligned}$$

This completes the inductive proof.

The proof of the commutativity is a double induction. We first prove $m1 = 1m$ by inducting on m. For $m = 1$, the equality holds trivially. Moreover, under the assumption $m = 1m$, we have

$$\begin{aligned}
m'1 &= m' && \text{(first property)}\\
&= m + 1 && \text{(definition of addition)}\\
&= m1 + 1 && \text{(first property)}\\
&= 1m + 1 && \text{(inductive assumption)}\\
&= 1m'. && \text{(second property)}
\end{aligned}$$

This completes the inductive proof of $m1 = 1m$.

Next, under the inductive (on n) assumption $mn = nm$, we have

$$\begin{aligned}
mn' &= mn + m && \text{(second property)}\\
&= nm + m && \text{(inductive assumption)}\\
&= nm + m1 && \text{(first property)}\\
&= nm + 1m && \text{($m1 = 1m$, just proved)}\\
&= (n+1)m && \text{(distributivity, just proved)}\\
&= n'm.
\end{aligned}$$

This completes the inductive proof of the commutativity. □

Note that the distributivity and the commutativity imply the other distributivity

$$k(m+n) = km + kn.$$

Moreover, the associativity tells us $(mn)(kl) = (m(nk))l = m(n(kl))$. Therefore we may write $mnkl$ without any ambiguity. The commutativity further allows us to freely exchange orders of numbers in a multiplication, such as $knlm = nmkl$.

Exercise 3.19. Explain the multiplication of natural numbers is well-defined.

Exercise 3.20. Prove the associativity in Proposition 3.4.2 by inducting on k and using the other distributivity.

Next we extend the multiplication to integers. Based on the expectation $(m-n)(k-l) = (mk+nl)-(ml+nk)$, we may define

$$[m,n][k,l] = [mk+nl, ml+nk].$$

Similar to the addition of integers, we need to verify that this is well-defined. The verification is left as an exercise. The following are the properties of the product.

Proposition 3.4.3. *The multiplication of integers has the following properties:*

1. *The multiplication is consistent with the multiplication of natural numbers.*
2. *Distributivity:* $(a+b)c = ac+bc$ *and* $a(b+c) = ab+ac$.
3. *Associativity:* $a(bc) = (ab)c$.
4. *Commutativity:* $ab = ba$.
5. *One:* $a1 = 1a = 1$.
6. *Zero:* $ab = 0 \iff a = 0$ *or* $b = 0$.
7. *Negative:* $(-a)b = -ab = a(-b)$.
8. *Order: If* $a > 0$, *then* $b > c \iff ab > ac$.

Proof. The first property means that the map $f(n) = [n+1,1] \colon \mathbb{N} \to \mathbb{Z}$ satisfies $f(mn) = f(m)f(n)$. By

$$\begin{aligned} f(mn) &= [mn+1,1], \\ f(m)f(n) &= [m+1,1][n+1,1] = [(m+1)(n+1)+1, (m+1)1+1(n+1)], \end{aligned}$$

the problem becomes the verification of

$$mn+1+(m+1)1+1(n+1)=1+(m+1)(n+1)+1.$$

By Propositions 3.1.4 and 3.4.2, this can be easily done.

The second, third, fourth and fifth properties can be routinely verified similar to the first property, and are left as exercises.

Next, we verify the $\Longleftarrow$ direction of the sixth property: $a0=0$. By commutativity, this is the same as $0a=0$. We may verify $[m,n][1,1]=[1,1]$ in a routine way. Alternatively, we multiply a to $0+0=0$ and use the distributivity to get $a0+a0=a0$. Then we may use the cancelation law for integers to get $a0=0$.

Now we can prove the seventh property. We have

$$\begin{aligned}
ab+(-a)b &= (a+(-a))b && \text{(distributivity)}\\
&= 0b && \text{(definition of negative)}\\
&= 0. && \text{(just proved)}
\end{aligned}$$

Then we get $(-a)b=-ab$ by the uniqueness of the negative of ab. The proof of $a(-b)=-ab$ is similar.

Now we can prove the $\Longrightarrow$ direction of the sixth property: $ab=0$ implies $a=0$ or $b=0$. We will actually prove the contrapositive statement: $a\neq 0$ and $b\neq 0$ imply $ab\neq 0$. By Proposition 3.3.1, the assumption $a\neq 0$ and $b\neq 0$ means $a=\pm m$ and $b=\pm n$ for some natural numbers m,n and suitable signs $\pm$. By using the just proved formula $(-a)b=-ab=a(-b)$ if necessary, we find $ab=\pm mn$. Then by Proposition 3.3.1, we get $ab\neq 0$.

Finally, we prove the last property. We let $x=b-c$ and note that

$$\begin{aligned}
b>c &\iff x=b-c>0,\\
ab>ac &\iff ax=a(b-c)=ab-ac>0.
\end{aligned}$$

Therefore the problem becomes the following: If $a>0$, then $x>0$ if and only if $ax>0$. By Proposition 3.3.1, we have three mutually exclusive possibilities of x:

1. $x=0$: By the sixth property, we get $ax=0$.

2. $x>0$: Both a,x are natural numbers. By the first property, we know ax is also a natural number. Therefore $ax>0$.

3. $x<0$: We have $x=-n$ for a natural number n. By the seventh property, we get $ax=-an$. Since an is a natural number, we get $ax<0$.

Since the first and the third possibilities contradict $ax>0$, we conclude the second possibility is equivalent to $ax>0$. This proves the last property. □

The development so far justifies all of our usual treatment of integers, particularly in terms of the addition, multiplication, order and sign. From now on, we will abandon those provisional notations such as $[m,n]$, $f(n)$. We can safely use our everyday life notations for the integers and manipulate them in our usual way. For example, we may freely apply the formulae such as $(a-b)c = ac - bc$ and $(a+b)(a-b) = a^2 - b^2$ to integers.

Exercise 3.21. Verify that the multiplication of integers is well defined. In other words, prove that $m_1 + n_2 = n_1 + m_2$ and $k_1 + l_2 = l_1 + k_2$ imply

$$(m_1k_1 + n_1l_1) + (m_2l_2 + n_2k_2) = (m_1l_1 + n_1k_1) + (m_2k_2 + n_2l_2).$$

Exercise 3.22. Prove the second, third, fourth and fifth properties in Proposition 3.4.3.

Exercise 3.23. Prove that if $c < 0$, then $a > b \iff ac < bc$.

Exercise 3.24. Prove the multiplicative cancelation law: If $a \neq 0$, then $ab = ac \implies b = c$.

3.5 Rational Number

Rational numbers are quotients of integers, with nonzero denominator. However, the choice of numerator and denominator is not unique, similar to the non-uniqueness of representing integers as differences of natural numbers. The solution to the problem is again through equivalence classes. Here the pair (a,b), $a,b \in \mathbb{Z}$, $b \neq 0$, is used to represent the quotient $\dfrac{a}{b}$.

Definition 3.5.1. The *rational numbers* is the set $\mathbb{Q}$ of the equivalence classes of pairs (a,b) of integers $a,b \in \mathbb{Z}$, $b \neq 0$, under the equivalence relation

$$(a,b) \sim (c,d) \iff ad = bc.$$

Instead of the notation $[a,b]$ for the equivalence classes used before, we will use the more conventional notation $\dfrac{a}{b}$ for the equivalence classes. Then

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

The important thing to remember here is that we cannot yet think of $\dfrac{a}{b}$ as a divided as b, because the division operation is not yet defined. At the moment, $\dfrac{a}{b}$ is simply an *integrated* notation for the equivalence class.

Similar to $\mathbb{N} \subset \mathbb{Z}$, we regard intergers as part of rational numbers through a map

$$g(a) = \frac{a}{1} \colon \mathbb{Z} \to \mathbb{Q}.$$

We emphasis again that $\frac{a}{1}$ means the equivalence class of $(a, 1)$, not yet a divided by 1. The following verifies g is one-to-one:

$$g(a) = g(b) \iff (a, 1) \sim (b, 1) \iff a = a1 = 1b = b.$$

Then we may identify integers $\mathbb{Z}$ with the subset $g(\mathbb{Z}) \subset \mathbb{Q}$ by writing $a = \frac{a}{1}$ for $a \in \mathbb{Z}$. For example, $0 = \frac{0}{1}$ and $1 = \frac{1}{1}$ are also rational numbers.

The addition and multiplication can be extended to rational numbers in the obvious way:

$$\frac{a}{b} + \frac{c}{d} = \frac{ad + bc}{bd}, \quad \frac{a}{b}\frac{c}{d} = \frac{ac}{bd}.$$

Proposition 3.5.2. *The addition and multiplication of rational numbers have the following properties:*

1. *The operations are consistent with the operations of integers.*
2. *Associativity: $r + (s + t) = (r + s) + t$, $r(st) = (rs)t$.*
3. *Commutativity: $r + s = s + r$, $rs = sr$.*
4. *Distributivity: $(r + s)t = rt + st$, $r(s + t) = rs + rt$.*
5. *Zero: The integer 0 is the unique rational number satisfying $r + 0 = 0 + r = r$.*
6. *Negative: For any rational number r, there is a unique rational number $-r$ satisfying $r + (-r) = (-r) + r = 0$.*
7. *One: The integer 1 is the unique rational number such that $r1 = 1r = r$.*
8. *Reciprocal: For any rational number $r \neq 0$, there is a unique rational number r^{-1} satisfying $rr^{-1} = r^{-1}r = 1$.*

Proof. The following verifies the first property

$$\frac{a}{1} + \frac{b}{1} = \frac{a1 + 1b}{1 \cdot 1} = \frac{a + b}{1}, \quad \frac{a}{1}\frac{b}{1} = \frac{ab}{1 \cdot 1} = \frac{ab}{1}.$$

The second, third and fourth properties can be similarly verified, and are left as exercises.

For the fifth property, we have

$$\frac{a}{b}+\frac{0}{1}=\frac{a1+b0}{b1}=\frac{a+0}{b}=\frac{a}{b}.$$

Similarly, we have $\frac{0}{1}+\frac{a}{b}=\frac{a}{b}$. Therefore $\frac{0}{1}$ can be used as (rational) zero. To prove the uniqueness of zero, we assume both rational numbers 0 and $\bar{0}$ satisfy

$$r+0=r=0+r, \quad r+\bar{0}=r=\bar{0}+r.$$

Then by taking $r=\bar{0}$ in the first equality and taking $r=0$ in the second equality, we get

$$\bar{0}+0=\bar{0}=0+\bar{0}, \quad 0+\bar{0}=0=\bar{0}+0.$$

Therefore $\bar{0}=0+\bar{0}=0$. This proves the uniqueness of zero.

The proof of the sixth, seventh and eighth properties is similar. For $r=\frac{a}{b}$, we have $-r=\frac{-a}{b}$ and $r^{-1}=\frac{b}{a}$. We also have $1=\frac{1}{1}$. Here we only remark that $r=\frac{a}{b}\neq 0=\frac{0}{1}$ means $a=a1\neq 0b=0$. □

Similar to integers, we may define subtraction of rational numbers by using the negative:

$$r-s=r+(-s).$$

Likewise, we may define the division by using the reciprocal:

$$r\div s=rs^{-1}.$$

For integers a,b, the following shows that $\frac{a}{b}$ is indeed the division of a by b

$$a\div b=\frac{a}{1}\left(\frac{b}{1}\right)^{-1}=\frac{a}{1}\cdot\frac{1}{b}=\frac{a1}{1b}=\frac{a}{b}.$$

Here the second equality makes use of the formula for r^{-1} in the proof of Proposition 3.5.2. Therefore we may also write $\frac{r}{s}$ for the division $r\div s$ of rational numbers.

Exercise 3.25. Prove the second, third and fourth properties in Proposition 3.5.2.

Exercise 3.26. Prove the uniqueness in the sixth, seventh and eighth properties in Proposition 3.5.2.

Exercise 3.27. Use Proposition 3.5.2 to prove the cancelation laws for rational numbers:

1. $r+t=s+t$ imply $r=s$.

2. $rt = st$ and $t \neq 0$ imply $r = s$.

Exercise 3.28. Prove that $rs = 0$ if and only if $r = 0$ or $s = 0$.

Exercise 3.29. Prove properties of the division of rational numbers

$$\frac{0}{t} = 0, \quad \frac{r+s}{t} = \frac{r}{t} + \frac{s}{t}, \quad \frac{r-s}{t} = \frac{r}{t} - \frac{s}{t}, \quad \frac{rt}{st} = \frac{r}{s}, \quad \left(\frac{r}{s}\right)^{-1} = \frac{s}{r}.$$

It remains to define the order among rational numbers: We define $r > s$, if $r - s = \dfrac{a}{b} > 0$ for some $a > 0$ and $b > 0$.

By $\dfrac{a}{b} = \dfrac{-a}{-b}$, we may always write $r - s = \dfrac{a}{b}$ for some $b > 0$. Then there are three possibilities for a, corresponding to three possibilities for $r - s$:

1. $a = 0$: We have $r - s = \dfrac{0}{b} = 0$. Therefore $r = s$.

2. $a > 0$: By $a, b > 0$, we get $r > s$.

3. $a < 0$: We have $s - r = -(r - s) = \dfrac{-a}{b}$. By $-a, b > 0$, we get $s > r$.

The discussion is independent of the choice of the expression $\dfrac{a}{b}$ with $b > 0$, for the following reason: Suppose $\dfrac{a}{b} = \dfrac{c}{d}$ with $b > 0$ and $d > 0$. Then $ad = bc$, and by the last property in Proposition 3.4.3, we get

$$a > 0 \iff ad > 0 \iff bc > 0 \iff c > 0.$$

Similarly, we get $a < 0 \iff c < 0$. By the sixth property in the proposition, we also get $a = 0 \iff c = 0$. This proves the first property in the following.

Proposition 3.5.3. *The order of rational numbers has the following properties:*

1. *For any two rational numbers r and s, one of the following mutually exclusive cases happens: $r = s$, $r > s$, $r < s$.*

2. *$r > s$ and $s > t \implies r > t$.*

3. *$r > s \implies r + t > s + t$.*

4. *$r > s \implies -r < -s$.*

5. *If $r > 0$, then $s > t \iff rs > rt$.*

6. *If $r, s > 0$, then $r > s \iff \dfrac{1}{r} < \dfrac{1}{s}$.*

7. *For any $r > s$, there is t satisfying $r > t > s$.*

8. *For any $r > 0$, there is a natural number n satisfying $n > r > \frac{1}{n}$.*

Proof. For the second property, the assumption $r > s$ and $s > t$ means

$$r - s = \frac{a}{b} \text{ and } s - t = \frac{c}{d}, \text{ where } a, b, c, d > 0.$$

Then

$$r - t = (r - s) + (s - t) = \frac{ad + bc}{bd}.$$

By $a, b, c, d > 0$, we get $ad + bc > 0$ and $bd > 0$. Therefore $r > t$.

The third, fourth, fifth and sixth properties can be routinely verified similar to the second property, and are left as exercises.

For the seventh property, we may choose $t = \frac{r+s}{2}$. Then $r-t = t-s = \frac{1}{2}(r-s)$. By the fifth property, we get $\frac{1}{2}(r - s) > 0$. Then by the third property, we get $r > t$ and $t > s$.

For the eighth property, the assumption means $r = \frac{a}{b}$ for some $a, b > 0$. By $\frac{a}{b} = \frac{2a}{2b}$, we may also assume that $a, b > 1$. Then we take $n = \max\{a, b\}$ (see the comment after Proposition 3.3.3). By $a, b > 1$ and the fifth and sixth properties, we get

$$n \geq a > r = \frac{a}{b} > \frac{1}{b} \geq \frac{1}{n}.$$

This proves the eighth property. □

Similar to the remark after Proposition 3.3.3, we may introduce $r \geq s$, $r \leq s$, $\max\{r, s\}$ and $\min\{r, s\}$ for rational numbers. We also define the *absolute value* of a rational number

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

The absolute value has the following properties.

$$|r + s| \leq |r| + |s|, \quad |rs| = |r||s|, \quad |r| < s \iff -s < r < s.$$

The proof is left as an exercise.

We have established the four arithmetic operations and the order for the rational numbers. We also proved all the usual properties. From now on, we may freely manipulate rational numbers just as we do in everyday life.

Exercise 3.30. Prove the third, fourth, fifth and sixth properties in Proposition 3.5.3.

Exercise 3.31. Prove properties of the absolute values

$$|r+s| \le |r|+|s|, \quad |rs| = |r||s|, \quad |r| < s \iff -s < r < s.$$

Exercise 3.32. Prove $\max\{r,s\}+\min\{r,s\} = r+s$ and $\max\{r,s\}-\min\{r,s\} = |r-s|$.

3.6 Real Number

The length of the diagonal of a square of side 1 is $\sqrt{2}$, which we have proved is not rational. The ratio π between the circumference of a circle and its diameter is also not rational. Therefore it is necessary to further expand the set of rational numbers.

We often write the real numbers in infinite decimal expansions, such as

$$\begin{aligned} \frac{1}{3} &= 0.33333333\cdots, \\ \sqrt{2} &= 1.41421356\cdots, \\ \pi &= 3.14158265\cdots. \end{aligned}$$

The irrational numbers are those expansions that are not "periodic". If we try use these expressions as rigorous definitions of real numbers, however, we have to deal with lots of problems. How do you add or multiply such expressions together? Would you consider $0.99999999\cdots$ and $1.00000000\cdots$ to be the same? How do you express $\sqrt{2}$ (defined as the number a satisfying $x^2 = 2$) as a decimal expansion, so that the number fits into the definition in terms of decimal expansion?

A better idea is to consider the actual meaning of the expansion. By the expansion, we mean that, by including more and more decimal digits, we are getting closer and closer to the number. For example, $\sqrt{2} = 1.41421356\cdots$ means that $\sqrt{2}$ is the *limit* of the sequence of *rational* numbers

$$1, 1.4, 1.41, 1.414, 1.4142, 1.41421, 1.414213, 1.4142135, 1.41421356, \dots.$$

This suggests that it might be possible to define real numbers as the limits of *convergent* sequences of rational numbers. Such sequences are called *Cauchy sequences*, and it is indeed possible to construct real numbers in this way. This is the Cauchy[3] method.

Another approach is based on the observation that the a real number can be described as the *supremum* (i.e., least upper bound) of some set of rational numbers (i.e., subsets of $\mathbb{Q}$). For example, $\sqrt{2}$ is the smallest real number that is bigger than all the numbers (i.e., an upper bound) in

$$X = \{1, 1.4, 1.41, 1.414, 1.4142, 1.41421, 1.414213, 1.4142135, 1.41421356, \dots\} \subset \mathbb{Q}.$$

[3] Augustin Louis Cauchy: born 21 Aug 1789 in Paris, France; died 23 May 1857 in Sceaux (near Paris), France. Many fundamental results in real and complex analysis are due to Cauchy and bear his name: Cauchy integral theorem, Cauchy-Kovalevskaya theorem, the Cauchy-Riemann equations, Cauchy sequences.

However, $\sqrt{2}$ is also the supremum of the set

$$Y = \{1, 1.41, 1.4142, 1.414213, 1.41421356, \dots\} \subset \mathbb{Q}.$$

To deal with the problem of many choices of subsets of $\mathbb{Q}$ suitable for defining $\sqrt{2}$, we may use equivalence relation, or simply choose the biggest subset for which $\sqrt{2}$ is a supremum

$$Z = \{r \in \mathbb{Q} \colon r < \sqrt{2}\} \subset \mathbb{Q}.$$

Alternatively, we may also choose the biggest subset for which $\sqrt{2}$ is a infimum (greatest lower bound)

$$Z' = \{r \in \mathbb{Q} \colon r > \sqrt{2}\} = \{r \in \mathbb{Q} \colon r^2 > 2\} \subset \mathbb{Q}.$$

This later approach is called the Dedekind[4] cut.

We will pursue the Dedekind cut method. This means we need to describe subsets of $\mathbb{Q}$ similar to Z'. The tricky thing here is that the description should not refer to real numbers.

Definition 3.6.1. A *real number* is a nonempty subset $X \subset \mathbb{Q}$ of rational numbers satisfying the following:

1. If $s > r \in X$, then $s \in X$.
2. If $r \in X$, then there is $s \in X$, such that $r > s$.
3. There is $p \in \mathbb{Q}$, such that $r > p$ for any $r \in X$. This means X has a lower bound.

The collection of all real numbers is denoted $\mathbb{R}$.

Let us understand the motivation behind the three conditions in the definition. The first condition implies that there are three possibilities for X:

- $X = (x, \infty) \cap \mathbb{Q} = \{r \in \mathbb{Q} \colon r > x\}$ for some real number x.
- $X = [x, \infty) \cap \mathbb{Q} = \{r \in \mathbb{Q} \colon r \geq x\}$ for some real number x.
- $X = \mathbb{Q}$.

The third condition implies that only the first two happens.

Next, we wish to unambiguously pick from the first two possibilities. If x is irrational, then $(x, \infty) \cap \mathbb{Q} = [x, \infty) \cap \mathbb{Q}$, i.e., the two possibilities are the same.

[4] Julius Wihelm Richard Dedekind: born 6 October 1831 in Braunschweig, Germany; died 12 Feb 1916 in Braunschweig, Germany. Dedekind made a number of highly significant contributions to mathematics, particularly in algebraic number theory. Dedekind came up with the idea of cut on 24 November 1858.

If x is rational, however, the two subsets $(x,\infty)\cap\mathbb{Q}$ and $[x,\infty)\cap\mathbb{Q}$ are different. The second condition is not satisfied by $[x,\infty)\cap\mathbb{Q}$ for rational x. Therefore the condition means we choose the first possibility.

It is also possible to develop the theory of real numbers by choosing the second possibility. This means modifying the second condition, and will cause some complication in later argument.

Our theory of real numbers will work with special subsets of rational numbers. Therefore we use capital letter X, Y, Z, etc., when we think of real numbers as subsets. On the other hand, if we forget about the definition in terms of subsets and just think about real numbers, then it is more comfortable to use lower case letters x, y, z, etc. In what follows, we will use both lower and capital letters according to the context, and will keep in mind that $x = X$, $y = Y$, $z = Z$, etc.

The next step of developing real numbers is to consider rational numbers as part of real numbers through the map:

$$h(r) = \{s\in\mathbb{Q}\colon s > r\}\colon \mathbb{Q}\to\mathbb{R}.$$

The seventh property in Proposition 3.5.3 implies that the map is one-to-one. Therefore we may think of $\mathbb{Q}$ as a subset of $\mathbb{R}$.

The following says that any real number is approximated by rational numbers.

Lemma 3.6.2. *For any real number X and any rational number $\epsilon > 0$, there are rational numbers $r\in X$ and $s\notin X$, such that $r - s = \epsilon$.*

Proof. Since X is not empty, we have rational $p\in X$. By the third condition in Definition 3.6.1, we have rational $q\notin X$. Applying the eighth property in Proposition 3.5.3 to $-\dfrac{p}{\epsilon}$ and $\dfrac{q}{\epsilon}$, we get natural number m,n satisfying $\dfrac{p}{\epsilon} > -m$ and $n > \dfrac{q}{\epsilon}$.

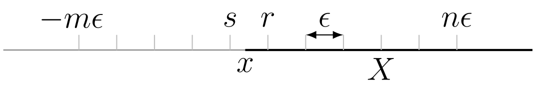

Figure 3.1: Approximate real numbers by rational numbers.

By $-m\epsilon < p \notin X$, and $n\epsilon > q\in X$, and the first condition in Definition 3.6.1, we get $-m\epsilon\notin X$ and $n\epsilon\in X$. By adding ϵ repeatedly to $-m\epsilon$, we get an increase sequence of rational numbers

$$-m\epsilon,\ (-m+1)\epsilon,\ (-m+2)\epsilon,\ \dots,\ (n-1)\epsilon,\ n\epsilon.$$

Since the sequence begins with $-m\epsilon\notin X$ and ends with $n\epsilon\in X$, there are two adjacent terms in the sequence, say $s = k\epsilon$ and $r = (k+1)\epsilon$, such that $s\notin X$ and $r\in X$, and the difference $r - s = \epsilon$. □

Exercise 3.33. For any real number X, explain that $r\in X$ and $\notin X$ imply $r > s$. Also explain that $r\notin X$ and $s < r$ imply $s\notin X$.

Exercise 3.34. For any real number X and any rational number $\epsilon > 0$, prove that there are rational numbers $r \in X$ and $s \notin X$, such that $r - s < \epsilon$.

Exercise 3.35. If we try to use sets similar to Z to define real numbers, how would you modify Definition 3.6.1? Moreover, does $\{r \in \mathbb{Q} \colon r^2 < 2\} \subset \mathbb{Q}$ satisfy the modified definition?

Define the addition of two real numbers

$$X + Y = \{r + s \colon r \in X,\ s \in Y\}.$$

We need verify that $X + Y$ satisfies the conditions Definition 3.6.1:

1. Suppose a rational number $t > r + s$ for some $r \in X$ and $s \in Y$. By $t - r > s$ and the first condition for Y, we get $t - r \in Y$ and $t = r + (t - r) \in X + Y$.

2. Suppose $t = r + s \in X + Y$ for some $r \in X$ and $s \in Y$. By the second condition for X, there is $r' \in X$ satisfying $r > r'$. Then $t' = r' + s \in X + Y$ satisfies $t > t'$.

3. If p and q are lower bounds for X and Y, then $p + q$ is a lower bound for $X + Y$.

Next we verify $h(r + s) = h(r) + h(s)$. This is an equality of two subsets. A number inside $h(r) + h(s)$ is a sum $r' + s'$ of rational numbers, such that $r' > r$ and $s' > s$. Then $r' + s' > r + s$, and $r' + s'$ is inside $h(r + s)$. This proves $h(r + s) \supset h(r) + h(s)$.

On the other hand, a number inside $h(r+s)$ is a rational number $t > r+s$. Then $t - r > s$. By the seventh property in Proposition 3.5.3, there is rational number s', such that $t - r > s' > s$. Then $t = r' + s'$ for $r' = t - s' > r$ and $s' > s$. We have $r' \in h(r)$, $s' \in h(s)$, and $t \in h(r) + h(s)$. This proves $h(r + s) \subset h(r) + h(s)$.

Proposition 3.6.3. *The addition of real numbers has the following properties:*

1. *Associativity:* $x + (y + z) = (x + y) + z$.

2. *Commutativity:* $x + y = y + x$.

3. *Zero: There is a unique real number* 0 *satisfying* $x + 0 = 0 + x = x$.

4. *Negative: For any real number* x*, there is a unique real number* $-x$ *satisfying* $x + (-x) = (-x) + x = 0$.

Proof. For the first property, we note that

$$X + (Y + Z) = \{r + (s + t) \colon r \in X,\ s \in Y,\ t \in Z\},$$
$$(X + Y) + Z = \{(r + s) + t \colon r \in X,\ s \in Y,\ t \in Z\}.$$

By $r+(s+t)=(r+s)+t$, we get $X+(Y+Z)=(X+Y)+Z$. This proves the first property. The second property can be proved in the similar way.

For the third property, we have $h(0)=\{s\in\mathbb{Q}\colon s>0\}$ and

$$X+h(0)=\{r+s\colon r\in X,\ s\in\mathbb{Q},\ s>0\}.$$

By $r+s>r\in X$ and the first condition in Definition 3.6.1, we get $r+s\in X$. This proves $X+h(0)\subset X$. On the other hand, for $r\in X$, by the second condition in Definition 3.6.1, there is $s\in X$ satisfying $r>s$. Then $r-s\in h(0)$, and $r=s+(r-s)\in X+h(0)$. This proves $X+h(0)\supset X$. This proves $X+h(0)=X$.

For the uniqueness of zero, the proof for rational numbers (fifth property in Proposition 3.5.2) can be adopted without change.

For the fourth property, we note that the negative Y of a real number X satisfies

$$X+Y=\{r+t\colon r\in X,\ t\in Y\}=h(0).$$

Therefore it is tempting to construct the negative as

$$Y=\{t\in\mathbb{Q}\colon r+t>0 \text{ for all } r\in X\}.$$

However, for $X=(x,+\infty)\cap\mathbb{Q}$, the construction gives $Y=[-x,+\infty)\cap\mathbb{Q}$. On the other hand, what we really want is $(-x,+\infty)\cap\mathbb{Q}$, which is different from Y in case x is a rational number. So we further modify Y to get $(-x,+\infty)\cap\mathbb{Q}$:

$$\begin{aligned}Z&=\{s\in\mathbb{Q}\colon s>t \text{ for some } t\in Y\}\\&=\{s\in\mathbb{Q}\colon \text{There is } t\in\mathbb{Q} \text{ such that } s>t,\ \text{and } r+t>0 \text{ for all } r\in X\}.\end{aligned}$$

The following verifies that Z satisfies the three conditions in Definition 3.6.1:

1. Suppose $s'>s\in Z$. Then by the second property in Proposition 3.5.3, we get $s>t$ for some $t\in Y$ implies $s'>t$ for the same $t\in Y$. This implies $s'\in Z$.

2. Suppose $s\in Z$. Then by the second condition in Definition 3.6.1, we have $s>t$ for some $t\in Y$. By the seventh property in Proposition 3.5.3, there is $s'\in\mathbb{Q}$ satisfying $s>s'>t$. Then $s'>t$ implies $s'\in Z$, and we also have $s>s'$.

3. Fix any $r\in X$. For any $s\in Z$, we have $s>t$ for some $t\in Y$. Then $r+s>r+t>0$. This implies $s>-r$. Therefore $-r$ is a lower bound of Z.

Next we prove $X+Z=h(0)$. For $r\in X$ and $s\in Z$, we have $s>t$ for some $t\in Y$. Then $r+s>r+t>0$. This proves $X+Z\subset h(0)$. For the other inclusion $X+Z\supset h(0)$, we need to show that any $\epsilon\in h(0)$ can be expressed as $r+s$ for some $r\in X$ and $s\in Z$. Note that $\epsilon\in h(0)$ means $\epsilon>0$. By Lemma 3.6.2, there are rational numbers $r\in X$ and $r'\notin X$, such that $r-r'=\dfrac{\epsilon}{2}<\epsilon$. Then $\epsilon=r+s$ with $s=\epsilon-r$. We know $r\in X$, and only need to show $s\in Z$. We also know $s=\epsilon-r>-r'$, and the problem is reduced to showing $-r'\in Y$. This means that for any $s\in X$, we need to show $s+(-r')>0$, or $s>r'$. By the first condition in Definition 3.6.1, $s\in X$ and $r'\notin X$ indeed imply $s>r'$. □

Exercise 3.36. Suppose Y is a non-empty subset of rational numbers with lower bound. Prove that $Z = \{s \in \mathbb{Q} \colon s > t \text{ for some } t \in Y\}$ satisfies the three conditions in Definition 3.6.1. The fact is used in the proof of Proposition 3.6.3. In fact, the real number Z is the *infimum* of the subset Y.

Exercise 3.37. What are the rational numbers in $-\sqrt{2}$? What are the rational numbers in $1 - \sqrt{2}$?

Define the order of real numbers

$$X \geq Y \iff X \subset Y.$$

The definition is the same as

$$X > Y \iff X \subset Y \text{ and } X \neq Y.$$

Suppose $r \in X - Y$. Then by applying the first condition in Definition 3.6.1 to Y, we get $r < s$ for all $s \in Y$. Applying the first condition again to $r \in X$, we know $s \in X$. This proves $X \supset Y$. By $X \neq Y$ (due to $X - Y \neq \emptyset$), we get $X < Y$.

We proved $X - Y \neq \emptyset \implies X < Y$. Therefore for any two real numbers X and Y, there are three mutually exclusive possibilities:

1. $X = Y$.

2. $X - Y \neq \emptyset$: $X < Y$.

3. $Y - X \neq \emptyset$: $X > Y$.

This is the first property in Proposition 3.6.5.

Next we verify the consistency of orders in $\mathbb{Q}$ and $\mathbb{R}$. If $r > s$ in $\mathbb{Q}$, then $t > r$ implies $t > s$. Therefore $h(r) \subset h(s)$. Moreover, by the seventh property in Proposition 3.5.3, there is t satisfying $r > t > s$. Then $t \in h(s)$ and $t \notin h(r)$. Therefore $h(r) \neq h(s)$. This proves $r > s \implies h(r) > h(s)$. Then by the first property in Proposition 3.5.3, and three mutually exclusive order relations between $h(s)$ and $h(r)$, we conclude $r > s \iff h(r) > h(s)$.

The following is the criterion for comparing rational numbers and real numbers.

Lemma 3.6.4. *Suppose $r \in \mathbb{Q}$ and $X \in \mathbb{R}$. Then $r > X$ if and only if $r \in X$. In other words, $X = \{r \colon r \in \mathbb{Q},\ r > X\}$.*

Proof. By the definition, $r > X$ means the following:

1. $h(r) \subset X$: $s > r$ implies $s \in X$.

2. $h(r) \neq X$: There is $t \in X$ satisfying $t \leq r$.

By the first condition in Definition 3.6.1, the second statement means $r \in X$. Conversely, if $r \in X$, then the first statement holds by the first condition in Definition 3.6.1, and the second statement holds with $t = r$. □

Proposition 3.6.5. *The order of real numbers has the following properties:*

1. *For any two real numbers x and y, one of the following mutually exclusive cases happens: $x = y$, $x > y$, $x < y$.*
2. *$x > y$ and $y > z \implies x > z$.*
3. *$x > y \implies x + z > y + z$.*
4. *$x > y \implies -x < -y$.*
5. *For any $x > y$, there is a rational number r satisfying $x > r > y$.*

Proof. We already proved the first property.

The second property follows from $X \subset Y$ and $Y \subset Z$ imply $X \subset Z$.

The third property follows from $X \subset Y$ implies $X + Z \subset Y + Z$.

The fourth property is proved as follows

$$x > y \implies -y = x - (x + y) > y - (x + y) = -x.$$

For the fifth property, we note that $x > y$ implies $Y - X \neq \emptyset$. Then there is rational number s satisfying $s \notin X$ and $s \in Y$. By Lemma 3.6.4, we get $X \geq s > Y$. By the second condition in Definition 3.6.1, there is another rational number $r \in Y$ satisfying $s > r$. Then by Lemma 3.6.4, we get $X > r > Y$. □

The definition of real numbers is motivated by the supremum and the infimum. Naturally we need to verify that the motivation is fulfilled.

Theorem 3.6.6. *Any set of real numbers bounded above has a real number as the supremum. Any set of real numbers bound below has a real number as has the infimum.*

The *infimum* of a set of numbers A is the greatest lower bound. More specifically, $p = \inf A$ is characterized by two properties:

1. p is a lower bound: $x \geq p$ for all $x \in A$.
2. Any number bigger than p is not lower bound: If $q > p$, then there is $x \in A$ satisfying $x < q$.

The *supremum* $q = \sup A$ is characterized by two properties:

1. q is an upper bound: $x \leq q$ for all $x \in A$.

2. Any number smaller than q is not upper bound: If $p < q$, then there is $x \in A$ satisfying $x > p$.

We clearly have

$$\sup A = -\inf(-A) \text{ for } -A = \{-x \colon x \in A\}.$$

Proof. Let A be a set of real numbers with a lower bound p. Define

$$Z = \{r \in \mathbb{Q} \colon r > x \text{ for some } x \in A\}.$$

We verify the three conditions in Definition 3.6.1:

1. Suppose $s > r \in Z$. Then by the second property in Proposition 3.6.5, we get $s > x$. Therefore $s \in Z$.

2. Suppose $r \in Z$. Then by the fifth property in Proposition 3.6.5, there is a rational number s satisfying $r > s > x$. Then $r > s$ and $s \in Z$.

3. By the second property in Proposition 3.6.5, we know all $r \in Z$ satisfy $r > p$. Let p' be a rational number satisfying $p' \le p$. Then all $r \in Z$ satisfy $r > p'$.

Therefore Z is a real number.

Let $X \in A$. By Lemma 3.6.4, we have

$$r \in X \implies r > X \implies r \in Z.$$

Therefore $X \subset Z$. This means $X \ge Z$. This proves Z is a lower bound of A.

To show that Z is the biggest lower bound, we consider another real number $W > Z$. We have $r \in \mathbb{Q}$ satisfying $r \in Z$ and $r \notin W$. By $r \in Z$, we have $r > X$ for some $X \in A$. By Lemma 3.6.4, we get $r \in X$. Then $r \in X$ and $r \notin W$ imply $X \not\subset W$, which means $X \not\ge W$. Therefore W is not a lower bound of A.

We conclude Z is the infimum of A. The existence of the supremum can be proved by applying the negative to everything. □

Exercise 3.38. For any $x \in \mathbb{R}$, prove that $x = \inf\{r \colon r \in \mathbb{Q},\ r > x\}$. This recovers the original idea of constructing real numbers as the infima of rational numbers.

Exercise 3.39. Give an alternative proof of Proposition 3.6.6 by using $Z = \cup\{X \colon X \in A\}$.

Next we try to define the multiplication of real numbers. Since multiplying negative numbers exchanges the supremum and the infimum, we can only define the multiplication first for positive real numbers. Then we extend the multiplication to all real numbers.

We remark that $X \ge 0$ means $X \subset h(0)$, or all the rational numbers in X are positive.

Lemma 3.6.7. *Suppose X is a positive real number.*

1. *There is a natural number n satisfying $n > X > \frac{1}{n}$.*
2. *For any rational number $\epsilon > 0$, there are rational numbers $r \in X$ and $s \notin X$, such that $s > 0$ and $\frac{r}{s} < 1 + \epsilon$.*

The first statement is similar to the last property in Proposition 3.5.3. The second statement is the multiplicative version of Lemma 3.6.2. Note that $r \in X$ and $s \notin X$ already imply $r > s$.

Proof. For the first statement, by $X > 0$ and the fifth property in Proposition 3.6.5, there is a rational number t satisfying $X > t > 0$. Pick any $r \in X$. Then by Lemma 3.6.4, we get $r > X$. Then $r > X > t > 0$.

By the eighth property in Proposition 3.5.3, there are natural numbers n_1, n_2 satisfying $n_1 > r > \frac{1}{n_1}$ and $n_2 > t > \frac{1}{n_2}$. Then the natural number $n = \max\{n_1, n_2\}$ satisfies $n > r > X > t > \frac{1}{n}$.

For the second statement, take $t \in \mathbb{Q}$ above satisfying $X > t > 0$. By Lemma 3.6.2, there are $r \in X$ and $u \notin X$ satisfying $r - u < \epsilon t$. By Lemma 3.6.4 and $t < X$, we get $t \notin X$. Then $s = \max\{u, t\} \notin X$. By $s \geq u$, $s \geq t$ and $\epsilon > 0$, we get $r - s \leq r - u < \epsilon t \leq \epsilon s$. By $s \geq t > 0$ and the fifth property in Proposition 3.5.3, we may divide s and get $\frac{r}{s} - 1 < \epsilon$. □

Define the multiplication of *non-negative* real numbers $X, Y \geq 0$ by

$$XY = \{rs\colon r \in X,\ s \in Y\}.$$

We note that all rational numbers in X and Y are positive. The following verifies the three conditions in Definition 3.6.1:

1. Suppose a rational number $t > rs$ for some $r \in X$ and $s \in Y$. Then $r > 0$ and we have $\frac{t}{r} > s$. By $s \in Y$, this implies $\frac{t}{r} \in Y$. Then $t = r\frac{t}{r} \in XY$.
2. Suppose $t = rs \in XY$ with $r \in X$ and $s \in Y$. Then there are $r' \in X$ and $s' \in Y$ satisfying $r > r'$ and $s > s'$. By $r', s' > 0$, we get $t' = r's' \in XY$ satisfying $t > t'$.
3. 0 is a lower bound of XY.

Next, we verify $h(rs) = h(r)h(s)$ for non-negative rational numbers $r, s \geq 0$. This means the following are equivalent for $t \in \mathbb{Q}$:

$$t > rs \iff t = r's' \text{ for some rational } r', s' \in \mathbb{Q} \text{ satisfying } r' > r \text{ and } s' > s.$$

The $\Longleftarrow$ direction is obvious. For the $\Longrightarrow$ direction, we note that $t > rs$ implies $\dfrac{t}{r} > s$. By the seventh property in Proposition 3.5.3, there is s' satisfying $\dfrac{t}{r} > s' > s$. Then $r' = \dfrac{t}{s'} > r$ and we have $t = r's'$.

Proposition 3.6.8. *The multiplication of non-negative real numbers has the following properties:*

1. *Associativity:* $x(yz) = (xy)z$.
2. *Commutativity:* $xy = yx$.
3. *Distributivity:* $(x + y)z = xz + yz$, $x(y + z) = xy + xz$.
4. *Zero:* $x0 = 0x = 0$.
5. *One:* $x1 = 1x = x$.
6. *Reciprocal: For any real number* $x > 0$, *there is a unique real number* x^{-1} *satisfying* $xx^{-1} = x^{-1}x = 1$.
7. $x > 0$ *and* $y > 0$ *imply* $xy > 0$.

Proof. For the first property, we note that

$$X(YZ) = \{r(st) \colon r \in X,\ s \in Y,\ t \in Z\},$$
$$(XY)Z = \{(rs)t \colon r \in X,\ s \in Y,\ t \in Z\}.$$

By $r(st) = (rs)t$, we get $X(YZ) = (XY)Z$. This proves the first property. The second property can be proved in the similar way.

A rational number in $(X + Y)Z$ is of the form $(r + s)t$ with $r \in X$, $s \in Y$, $t \in Z$. By $(r + s)t = rt + st \in XZ + YZ$, we get $(X + Y)Z \subset XZ + YZ$. Conversely, a rational number in $XZ + YZ$ is of the form $rt_1 + st_2$ with $r \in X$, $s \in Y$, $t_1 \in Z$, $t_2 \in Z$. Then $t = \min\{t_1, t_2\} \in Z$, and we have $rt_1 + st_2 \geq (r + s)t \in (X + Y)Z$. By the first condition in Definition 3.6.1 for the real number $(X + Y)Z$, we get $rt_1 + st_2 \in (X + Y)Z$. This proves $XZ + YZ \subset (X + Y)Z$, and completes the proof of the third property.

A rational number in $Xh(0)$ is of the form rs with $r \in X$ and $s > 0$. By $X \geq 0$, we get $r > 0$. Therefore $rs > 0$, and $rs \in h(0)$. This proves $Xh(0) \subset h(0)$. On the other hand, fix $r \in X$. Then $r > 0$. Any $s \in h(0)$ also satisfies $s > 0$. Then $s = r\dfrac{s}{r} \in Xh(0)$. This proves $Xh(0) \supset h(0)$, and completes the proof of $x0 = 0$.

A rational number in $Xh(1)$ is of the form rs with $r \in X$ and $s > 1$. Then $rs > r \in X$ implies $rs \in X$. This proves $Xh(1) \subset X$. On the other hand, for any $r \in X$, there is $s \in X$ satisfying $r > s$. Then $\dfrac{r}{s} > 1$ implies $\dfrac{r}{s} \in h(1)$, and we have $r = s\dfrac{r}{s} \in Xh(1)$. This proves $Xh(1) \supset X$, and completes the proof of $x1 = x$.

For positive X, construct the reciprocal by

$$X^{-1} = \{s \in \mathbb{Q}\colon \text{There is } t \in \mathbb{Q} \text{ satisfying } s > t, \text{ and } rt > 1 \text{ for any } r \in X\}.$$

The construction is similar to the construction of the negative in the proof of Proposition 3.6.3. It can be similarly verified that X^{-1} satisfies the three conditions in Definition 3.6.1. In particular, we note that 0 is a lower bound of X^{-1}.

For $r \in X$ and $x \in X^{-1}$, we have $rs > rt > 1$. Therefore $rs \in h(1)$. This proves $XX^{-1} \subset h(1)$. On the other hand, to prove $XX^{-1} \supset h(1)$, we consider $u \in h(1)$. We have $u > 1$. By Lemma 3.6.7, there are $v \in X$ and $w \notin X$ satisfying $\frac{v}{w} < u$. Then $u = v\frac{u}{v}$, with $v \in X$. It remains to prove $s = \frac{u}{v} \in X^{-1}$. By $\frac{v}{w} < u$, we have $s = \frac{u}{v} > t = \frac{1}{w}$. Moreover, for any $r \in X$, by $w \notin X$, we get $r > w$. Therefore $rt = \frac{r}{w} > 1$. This proves $s \in X^{-1}$, and completes the proof of $XX^{-1} = h(1)$.

Finally, suppose $x > 0$ and $y > 0$. Then $xy \geq 0$ by the definition of multiplication. If $xy = 0$, then by multiplying the reciprocal of y (which exists because $y > 0$), we get $x = x(yy^{-1}) = (xy)y^{-1} = 0y^{-1} = 0$. The contradiction implies $xy > 0$. □

Now we extend the multiplication to all real numbers

$$xy = \begin{cases} xy, & \text{if } x \geq 0,\ y \geq 0 \\ -(-x)y, & \text{if } x \leq 0,\ y \geq 0 \\ -x(-y), & \text{if } x \geq 0,\ y \leq 0 \\ (-x)(-y), & \text{if } x \leq 0,\ y \leq 0 \end{cases}.$$

It can be easily verified that in the overlapping cases, the multiplication is always 0.

Proposition 3.6.9. *The multiplication of real numbers has the following properties:*

1. *The map $h\colon \mathbb{Q} \to \mathbb{R}$ preserves the product.*

2. *Associativity: $x(yz) = (xy)z$.*

3. *Commutativity: $xy = yx$.*

4. *Distributivity: $(x + y)z = xz + yz$, $x(y + z) = xy + xz$.*

5. *Zero: $x0 = 0x = 0$.*

6. *One: $x1 = 1x = x$.*

7. *Reciprocal: For any real number $x \neq 0$, there is a unique real number x^{-1} satisfying $xx^{-1} = x^{-1}x = 1$.*

8. *If $x > 0$, then $y > z \iff xy > xz$.*

Proof. We need to consider various possibilities of signs. We illustrate the idea by proving some cases of the distributivity. We already know the distributivity in case $x, y, z \geq 0$ from Proposition 3.6.8. For the case $x + y \geq 0$, $y \leq 0$ and $z \geq 0$, we have $x + y, -y, z \geq 0$. By Proposition 3.6.8, we have

$$(x + y)z + (-y)z = [(x + y) + (-y)]z = xz.$$

Adding $-(-y)z = yz$ (the equality is the definition of yz) to both sides, we get $(x + y)z = xz + yz$. For the case $x + y \leq 0$, $x \geq 0$, $y \leq 0$ and $z \geq 0$, we have $-(x + y), x, -y, z \geq 0$. By Proposition 3.6.8, we have

$$xz + (-(x + y))z = [x - (x + y)]z = (-y)z.$$

We also have $-(-(x+y))z = (x+y)z$ and $-(-y)z = yz$. Adding the three equalities together, we get $xz + yz = (x + y)z$.

The rest of the proof are left as an exercise. □

Exercise 3.40. Complete the proof of Proposition 3.6.9.

Exercise 3.41. Prove that, if $x, y > 0$, then $x > y$ implies $x^{-1} < y^{-1}$.

Exercise 3.42. Another way of extending the multiplication to all real numbers is to show that any real number is a difference between two positive real numbers, and then define the multiplication of $x = x_1 - x_2$ and $y = y_1 - y_2$ with $x_1, x_2, y_1, y_2 \geq 0$ by

$$xy = x_1y_1 + x_2y_2 - x_1y_2 - x_2y_1.$$

Carry out the details of this approach.

3.7 Exponential

We develop the definition and properties for the exponential x^y, for $x > 0$ and any y. We start with natural number y, and gradually extend to more sophisticated y.

For a real number x (no need to be positive) and a natural number n, define x^n to be multiplying n copies of x together. Strictly speaking, the definition is given by the following inductive process.

- $x^1 = x$.
- $x^{n+1} = x^n x$.

Proposition 3.7.1. *The natural number exponential x^n has the following properties:*

$$(xy)^n = x^n y^n, \quad x^{m+n} = x^m x^n, \quad x^{mn} = (x^m)^n, \quad x > y > 0 \implies x^n > y^n.$$

Proof. We prove $x^{m+n} = x^m x^n$ by fixing m and inducting on n. The other properties can be similarly proved by induction.

For $n = 1$, we have $x^{m+1} = x^m x = x^m x^1$ by the inductive definition. Next assume $x^{m+n} = x^m x^n$. Then $x^{m+n+1} = x^{m+n}x = (x^m x^n)x = x^m(x^n x) = x^m x^{n+1}$. This completes the inductive proof of $x^{m+n} = x^m x^n$. □

Exercise 3.43. Prove the other properties in Proposition 3.7.1.

Next, we define the exponential for $y = \frac{1}{n}$, $n \in \mathbb{N}$. This is actually the n-th root.

Proposition 3.7.2. *For any $x > 0$ and $n \in \mathbb{N}$, there is a unique $x^{\frac{1}{n}} > 0$ satisfying $(x^{\frac{1}{n}})^n = x$.*

Proof. By using Theorem 3.6.6, we may construct the expected n-th root

$$w = \inf\{z \colon z > 0,\ z^n > x\}.$$

The infimum is defined because 0 is a lower bound of the subset. In fact, $u = \min\{1, x\}$ satisfies $u^n = uu^{-1} \le x1^{n-1} = x < z^n$. By the last property in Proposition 3.7.1, we cannot have $u \ge z$. Therefore u is a lower bound, and the greatest lower bound $w \ge u > 0$.

We need to verify $w^n = x$. This means both $w^n > x$ and $w^n < x$ will lead to contradictions.

Suppose $w^n > x$. Then $w^n - x > 0$. For $0 < z < w$, by Example 1.5.3, we have

$$w^n - z^n \le n(w - z)w^{n-1}.$$

We wish to find z very close to w, such that $n(w - z)w^{n-1} < w^n - x$. Then we get $w^n - z^n < w^n - x$. This implies $z^n > x$. Then w is not a lower bound, a contradiction.

To find z, we note that $n(w-z)w^{n-1} < w^n - x$ means $w - z < \dfrac{w^n - x}{nw^{n-1}}$. Therefore we may take z to satisfy $w - z = \dfrac{w^n - x}{2nw^{n-1}}$. In other words, we take $z = w - \dfrac{w^n - x}{2nw^{n-1}}$.

Suppose $w^n < x$. Then $x - w^n > 0$. For $w < v < 2w$, by Example 1.5.3, we have

$$v^n - w^n \le n(v - w)v^{n-1} < n(v - w)2^{n-1}w^{n-1}.$$

We wish to find v very close to w, such that $n(v - w)2^{n-1}w^{n-1} < x - w^n$. Then we get $v^n - w^n < x - w^n$. This implies $v^n < x < z^n$. By the last property in Proposition 3.7.1, we get $v < z$ for all $z > 0$ satisfying $z^n > x$. Therefore v is a bigger lower bound than w, a contradiction.

Similar to the case $w^n > x$, we find v to satisfy $n(v - w)2^{n-1}w^{n-1} < \dfrac{x - w^n}{2}$ and $w < v < 2w$. We may take $v = w + \min\left\{w, \dfrac{x - w^n}{n2^n w^{n-1}}\right\}$.

By the last property in Proposition 3.7.1, we have $w_1 > w_2 > 0$ implies $w_1^n > w_2^n$. Therefore $w_1 \neq w_2$ implies $w_1^n \neq w_2^n$. This proves the uniqueness of $x^{\frac{1}{n}}$. □

For $x > 0$ and rational $r > 0$, we write $r = \dfrac{m}{n}$ with $m, n \in \mathbb{N}$ and define $x^r = (x^{\frac{1}{n}})^m$. To show this is well defined, we assume $r = \dfrac{m_1}{n_1} = \dfrac{m_2}{n_2}$. Then $m_1 n_2 = m_2 n_1$. By Proposition 3.7.1, we have

$$\begin{aligned} ((x^{\frac{1}{n_1}})^{m_1})^{n_1 n_2} &= (x^{\frac{1}{n_1}})^{m_1 n_1 n_2} = ((x^{\frac{1}{n_1}})^{n_1})^{m_1 n_2} = x^{m_1 n_2}, \\ ((x^{\frac{1}{n_2}})^{m_2})^{n_1 n_2} &= (x^{\frac{1}{n_2}})^{m_2 n_1 n_2} = ((x^{\frac{1}{n_2}})^{n_2})^{m_2 n_1} = x^{m_2 n_1}. \end{aligned}$$

By $m_1 n_2 = m_2 n_1$, we get $((x^{\frac{1}{n_1}})^{m_1})^{n_1 n_2} = ((x^{\frac{1}{n_2}})^{m_2})^{n_1 n_2}$. Then by the uniqueness in Proposition 3.7.2, we get $(x^{\frac{1}{n_1}})^{m_1} = (x^{\frac{1}{n_2}})^{m_2}$.

For $x > 0$ and rational r, we write $r = r_1 - r_2$ with rational $r_1, r_2 > 0$, and define $x^r = \dfrac{x^{r_1}}{x^{r_2}}$. Again we need to show this is well defined. Assume $r_1 - r_2 = s_1 - 2_2$, with $r_1, r_2, s_1, s_2 > 0$. Then $r_1 + s_2 = r_2 + s_1$. If we can show $x^{r+s} = x^r x^s$ for $r, s > 0$, then $x^{r_1} x^{s_2} = x^{r_1 + s_2} = x^{r_2 + s_1} = x^{r_2} x^{s_1}$. This further implies $\dfrac{x^{r_1}}{x^{r_2}} = \dfrac{x^{s_1}}{x^{s_2}}$, and proves x^r is well defined.

To show $x^{r+s} = x^r x^s$ for $r, s > 0$, we write $r = \dfrac{m}{n}$ and $s = \dfrac{k}{l}$, where m, n, k, l are natural numbers. Then we need to show $x^{\frac{m}{n} + \frac{k}{l}} = x^{\frac{m}{n}} x^{\frac{k}{l}}$. Using Proposition 3.7.1, we can verify $(x^{\frac{m}{n} + \frac{k}{l}})^{nl} = (x^{\frac{m}{n}} x^{\frac{k}{l}})^{nl}$. Then by the uniqueness in Proposition 3.7.2, we get $x^{\frac{m}{n} + \frac{k}{l}} = x^{\frac{m}{n}} x^{\frac{k}{l}}$.

By $0 = r - r$, we get $x^0 = 1$. Moreover, for $r > 0$, we write $-r = 1 - (1 + r)$ and get $x^{-r} = \dfrac{x^1}{x^{1+r}} = \dfrac{x}{x x^r} = \dfrac{1}{x^r}$. Therefore x^{-r} is the reciprocal of x^r. This also justifies the earlier use of the notation x^{-1} for the reciprocal.

Proposition 3.7.3. *The rational exponential x^r has the following properties:*

$$(xy)^r = x^r y^r, \quad x^{r+s} = x^r x^s, \quad x^{rs} = (x^r)^s,$$

$$x > 1,\ r > s \implies x^r > x^s, \quad x > y > 0 \implies \begin{cases} x^r > y^r, & \text{if } r > 0 \\ x^r < y^r, & \text{if } r < 0 \end{cases}.$$

Proof. We already proved $x^{r+s} = x^r x^s$ for $r, s > 0$. By $x^0 = 1$, we may easily verify the equality in cases one of $r + s, r, s$ is 0. If $r + s < 0$, and $r < 0$, and $s > 0$, then $-r = s + (-(r + s))$ is a sum of positive rational numbers, and we get $x^{-r} = x^{s + (-(r+s))} = x^s x^{-(r+s)}$. Then using x^{-r} being the reciprocal of x^r, the equality implies $x^{r+s} = x^r x^s$. The other cases can be proved in the similar way.

The other equalities $(xy)^r = x^r y^r$ and $x^{rs} = (x^r)^s$ can also be proved in the similar way. We may first prove the equalities for positive exponentials. Then for the case some exponentials are negative, we use reciprocal to convert into equivalent equalities with positive exponentials.

Next we turn to inequalities.

Suppose $x > 1$ and $r > s$. Then $r - s = \dfrac{m}{n}$ with $m, n \in \mathbb{N}$. By $x > 1$, we get $(x^{\frac{m}{n}})^n = x^m > 1 = 1^n$. Then by the last property in Proposition 3.7.1, we cannot have $x^{\frac{m}{n}} \le 1$. In other words, we have $x^{\frac{m}{n}} > 1$. On the other hand, by the equalities we already proved, we get $x^{r-s} = \dfrac{x^r}{x^s}$. Then $x^{r-s} = x^{\frac{m}{n}} > 1$ means $x^r > x^s$.

Finally, the second inequality follows from

$$x > y > 0,\ r > 0 \implies \frac{x}{y} > 1,\ r > 0 \implies x^r = \left(\frac{x}{y}\right)^r y^r > \left(\frac{x}{y}\right)^0 y^r = y^r. \qquad \square$$

Exercise 3.44. The proof above discusses one case of $x^{r+s} = x^r x^s$. What are the other cases, and how do you prove the equality for the other cases?

Exercise 3.45. Prove $(xy)^r = x^r y^r$ and $x^{rs} = (x^r)^s$ in Proposition 3.7.3.

We will need the following technical result.

Lemma 3.7.4. *For any $x > 1$ and $\epsilon > 0$, there is $n \in \mathbb{N}$ satisfying $x^{\frac{1}{n}} - 1 < \epsilon$.*

Proof. Let $z_n = x^{\frac{1}{n}} - 1$. Then $x > 1$ implies $z_n > 0$. Moreover,

$$x = (1 + z_n)^n = 1 + \binom{n}{1} z_n + \binom{n}{2} z_n^2 + \cdots > \binom{n}{1} z_n = n z_n.$$

By Lemma 3.6.7, there is $n \in \mathbb{N}$ satisfying $\dfrac{x}{\epsilon} < n$. Then $x^{\frac{1}{n}} - 1 = z_n < \dfrac{x}{n} < \epsilon$. $\square$

Now we define the general exponential. For any $x > 1$ and $y \in \mathbb{R}$, define the exponent x^y by

$$x^y = \inf\{x^r \colon r \in \mathbb{Q},\ r > y\}.$$

By Lemma 3.6.4, if y is given by the set Y of rational numbers, then

$$x^y = \inf\{x^r \colon r \in Y\}.$$

We need to verify the consistency with the rational exponent. This means that, for any rational $y = s \in \mathbb{Q}$, the rational exponent x^s defined earlier should have the following properties

1. x^s is a lower bound: $r > s \implies x^r > x^s$.

2. Any number bigger than x^s is not a lower bound: For any $\epsilon > 0$, there is $r > s$, such that $x^r \le x^s + \epsilon$.

The first follows from Proposition 3.7.3. For the second, we apply Lemma 3.7.4. We have $x > 1$, and we take ϵ in the lemma to be ϵx^{-s}. Then we get $n \in \mathbb{N}$ satisfying $x^{\frac{1}{n}} - 1 < \epsilon x^{-s}$. Then $r = s + \dfrac{1}{n} > s$ satisfies $x^r = x^s x^{\frac{1}{n}} < x^s(1 + \epsilon x^{-s}) = x^s + \epsilon$.

Next we try to extend the exponent x^y from $x > 1$ to any positive x. Express any $x > 0$ as $x = \dfrac{x_1}{x_2}$ with $x_1, x_2 > 1$. This can be done, for example, by finding a rational number $r = \dfrac{m}{n}$ satisfying $x > r > 0$ and writing $x = \dfrac{nx}{n}$. Then we define $x^y = \dfrac{x_1^y}{x_2^y}$. To show this is well defined, we make use of the properties of the infimum.

Lemma 3.7.5. *Suppose A and B are lower bounded sets of real numbers.*

1. *If $A + B = \{x + y \colon x \in A,\ y \in B\}$, then $\inf(A + B) = \inf A + \inf B$.*
2. *If A and B contain only positive real numbers and $AB = \{xy \colon x \in A,\ y \in B\}$, then $\inf AB = \inf A \inf B$.*

Proof. A number in $A + B$ is $x + y$, with $x \in A$ and $y \in B$. We have $x \geq \inf A$ and $y \geq \inf B$. Then $x + y \geq \inf A + \inf B$, and $\inf A + \inf B$ is a lower bound of $A + B$.

On the other hand, if $z > \inf A + \inf B$, then $z = z_A + z_B$ for some $z_A > \inf A$ and $z_B > \inf B$ (take z_A to be any number satisfying $z - \inf B > z_A > \inf A$ and take $z_B = z - z_A$). Then there are $x \in A$ and $y \in B$ satisfying $x < z_A$ and $y < z_B$. This implies $x + y < z_A + z_B = z$. Therefore z is not a lower bound of $A + B$.

This completes the proof that $\inf A + \inf B$ is the infimum of $\inf(A + B)$.

The proof for the second statement is similar. □

For $x, y > 1$, by Lemma 3.7.5, we have $xy > 1$, and

$$\begin{aligned} (xy)^z &= \inf\{(xy)^r \colon r > z\} = \inf\{x^r y^r \colon r > z\}, \\ x^z y^z &= \inf\{x^r \colon r > z\} \inf\{y^s \colon s > z\} = \inf\{x^r y^s \colon r, s > z\}. \end{aligned}$$

Since $x^r y^s$ lies between $x^r y^r$ and $x^s y^s$, the two infimums are the same. This proves $(xy)^z = x^z y^z$ for $x, y > 1$.

Suppose $x = \dfrac{x_1}{x_2} = \dfrac{x_1'}{x_2'}$, with $x_1, x_2, x_1', x_2' > 1$. Then $x_1 x_2' = x_1' x_2$. By what we just proved, we get $x_1^y x_2'^y = (x_1 x_2')^y = (x_1' x_2)^y = x_1'^y x_2^y$. This implies $\dfrac{x_1^y}{x_2^y} = \dfrac{x_1'^y}{x_2'^y}$, and proves that x^y is well defined for any $x > 0$ and $y \in \mathbb{R}$.

Proposition 3.7.6. *The real exponential x^y has the following properties:*

$$(xy)^z = x^z y^z, \quad x^{y+z} = x^y x^z, \quad x^{yz} = (x^y)^z,$$

$$x > 1,\ y > z \implies x^y > x^z, \quad x > y > 0 \implies \begin{cases} x^z > y^z, & \text{if } z > 0 \\ x^z < y^z, & \text{if } z < 0 \end{cases}.$$

Proof. We already proved $(xy)^z = x^z y^z$ for $x, y > 1$. To prove $(xy)^z = x^z y^z$ for general $x, y > 0$, we write $x = \dfrac{x_1}{x_2}$ and $y = \dfrac{y_1}{y_2}$, with $x_1, x_2, y_1, y_2 > 1$. Then $xy = \dfrac{x_1 y_1}{x_2 y_2}$, with $x_1 y_1, x_2 y_2 > 1$, and

$$(xy)^z = \frac{(x_1 y_1)^z}{(x_2 y_2)^z} = \frac{x_1^z y_1^z}{x_2^z y_2^z} = \frac{x_1^z}{x_2^z}\frac{y_1^z}{y_2^z} = x^z y^z.$$

For the equality $x^{y+z} = x^y x^z$, the following proves the case $x > 1$:

$$\begin{aligned}
x^{y+z} &= \inf\{x^r \colon r > y + z\} && \text{(definition of } x^{y+z}\text{)}\\
&= \inf\{x^r \colon r \in Y + Z\} && \text{(Lemma 3.6.4)}\\
&= \inf\{x^{s+t} \colon s \in Y,\ t \in Z\} && \text{(definition of } Y + Z\text{)}\\
&= \inf\{x^s x^t \colon s \in Y,\ t \in Z\} && \text{(Proposition 3.7.3)}\\
&= \inf\{x^s \colon s \in Y\} \inf\{x^t \colon t \in Z\} && \text{(Lemma 3.7.5)}\\
&= x^y x^z. && \text{(definition of } x^y, x^z\text{)}
\end{aligned}$$

For the general $x > 0$, we have

$$x^{y+z} = \frac{x_1^{y+z}}{x_2^{y+z}} = \frac{x_1^y x_1^z}{x_2^y x_2^z} = \frac{x_1^y}{x_2^y}\frac{x_1^y}{x_2^y} = x^y x^z.$$

Next we prove that $x > 1$ and $y > z$ imply $x^y > x^z$. By $x^y = x^{y-z} x^z$ and the last property in Proposition 3.6.9, it is sufficient to prove $x^{y-z} > 1$. We have

$$x^{y-z} = \inf\{x^s \colon s \in \mathbb{Q},\ s > y - z\}.$$

By the last property in Proposition 3.6.5, there is $r \in \mathbb{Q}$ satisfying $y - z > r > 0$. Then s in the subset above satisfies $x^s > x^r$. Therefore by Proposition 3.7.3, we get $x^{y-z} \geq x^r > 1$.

For $x > y > 0$ and $z > 0$, we have $\dfrac{x}{y} > 1$. By what we just proved, we get $x^z = \left(\dfrac{x}{y}\right)^z y^z > \left(\dfrac{x}{y}\right)^0 y^z = y^z$.

Finally, we prove the equality $x^{yz} = (x^y)^z$.

For the case $z = n \in \mathbb{N}$, the equality $x^{yn} = (x^y)^n$ follows from the property $x^{y+z} = x^y x^z$ and the induction on n. Next, for the case $z = -n$, with $n \in \mathbb{N}$, we have

$$x^{y(-n)} = x^{(-y)n} = (x^{-y})^n = \left(\frac{1}{x^y}\right)^n = \frac{1}{(x^y)^n} = (x^y)^{-n}.$$

The case $z = 0$ is trivial, and we conclude $x^{yz} = (x^y)^z$ for integers z.

For a rational number $z = r = \dfrac{a}{n}$, with $n \in \mathbb{N}$ and $a \in \mathbb{Z}$, we have

$$((x^y)^r)^n = (x^y)^{rn} = (x^y)^a = x^{ya} = x^{yrn} = (x^{yr})^n.$$

By the uniqueness of the n-th root in Proposition 3.7.2, we get $(x^y)^r = x^{yr}$.

Now assume $x > 1$ and $y > 0$. We have $x^y > 1$ by the inequality just proved. By the definition of exponential x^y in case $x > 1$, we have

$$\begin{aligned}(x^y)^z &= \inf\{(x^y)^r \colon r \in \mathbb{Q},\ r > z\},\\ x^{yz} &= \inf\{x^s \colon r \in \mathbb{Q},\ s > yz\}.\end{aligned}$$

For any $s > yz$, there is rational r satisfying $\dfrac{s}{y} > r > z$. By $x > 1$, we get $(x^y)^r = x^{yr} < x^s$. This proves $(x^y)^z \le x^{yz}$. On the other hand, for any $r > z$, there is rational s satisfying $yr > s > yz$. Then we get $(x^y)^r = x^{yr} > x^s$. This proves $(x^y)^z \ge x^{yz}$. This completes the proof of $x^{yz} = (x^y)^z$ for the case $x > 1$ and $y > 0$.

Finally, for $x, y > 0$, we write $x = \dfrac{x_1}{x_2}$, with $x_1, x_2 > 1$. Then

$$x^{yz} = \frac{x_1^{yz}}{x_2^{yz}} = \frac{(x_1^y)^z}{(x_2^y)^z} = \left(\frac{x_1^y}{x_2^y}\right)^z = (x^y)^z.$$

Here the second equality makes use of $x_1^y, x_2^y > 1$, and the thid equality follows from $(xy)^z = x^z y^z$. The equality can be further extended to any $x > 0$ and all y by using $x^{-y} = \dfrac{1}{x^y}$. □

3.8 Complex Number

A *complex number* is of the form $z = x + yi$, with $x, y \in \mathbb{R}$ and $i = \sqrt{-1}$ satisfying $i^2 = -1$. The addition and multiplication of complex numbers are

$$\begin{aligned}(x + yi) + (u + vi) &= (x + u) + (y + v)i,\\ (x + yi)(u + vi) &= (xu - yv) + (xv + yu)i.\end{aligned}$$

It can be easily verified that the operations satisfy the usual properties (such as commutativity, associativity, distributivity) of arithmetic operations. In particular, the subtraction is

$$(x + yi) - (u + vi) = (x - ui) + (y - vi),$$

and the division is

$$\frac{x + yi}{u + vi} = \frac{(x + yi)(u - vi)}{(u + vi)(u - vi)} = \frac{(xu + yv) + (-xv + yu)i}{u^2 + v^2}.$$

The following are some concrete examples

$$\begin{aligned}
&(1+2i)+(3+4i)=(1+3)+(2+4)i=4+6i,\\
&(1+2i)-(3+4i)=-2-2i,\\
&(1+2i)\times(3+4i)=1\cdot 3+1\cdot 4i+2\cdot 3i+2\cdot 4i^2=(3-8)+(4+6)i=-5+10i,\\
&(1+2i)\div(3+4i)=\frac{1+2i}{3+4i}=\frac{(1+2i)(3-4i)}{(3+4i)(3-4i)}\\
&\qquad\qquad\qquad\;=\frac{1\cdot 3-1\cdot 4i+2\cdot 3i-2\cdot 4i^2}{3^2+4^2}=\frac{11}{25}+\frac{2}{25}i.
\end{aligned}$$

The *conjugation* of a complex number is $\bar{z}=\overline{x+iy}=x-iy$. It is compatible with the four arithmetic operations

$$\overline{z_1+z_2}=\bar{z}_1+\bar{z}_2,\quad \overline{z_1-z_2}=\bar{z}_1-\bar{z}_2,\quad \overline{z_1z_2}=\bar{z}_1\bar{z}_2,\quad \overline{\left(\frac{z_1}{z_2}\right)}=\frac{\bar{z}_1}{\bar{z}_2}.$$

Moreover, we have

$$\bar{\bar{z}}=z,\quad z\bar{z}=x^2+y^2.$$

Exercise 3.46. Show that $z+\bar{z}=2x$ and $z-\bar{z}=2yi$. Then explain that z is a real number if and only if $\bar{z}=z$.

A major application of complex number is to solve quadratic equations.

Example 3.8.1. To solve $z^2-2z+5=0$, we first eliminate the first order term $-2z$ by *completing the square*

$$z^2-2z+5=(z^2-2z+1)+4=(z-1)^2+4.$$

Then the quadratic equation is $(z-1)^2=-4$. Taking the square root, we get $z-1=\pm 2\sqrt{-1}=\pm 2i$. Therefore the solutions are $z=1\pm 2i$.

The general process of completing the square is

$$\begin{aligned}
az^2+bz+c&=a\left(z^2+\frac{b}{a}z\right)+c\\
&=a\left(z^2+2\frac{b}{2a}z+\frac{b^2}{(2a)^2}\right)-a\frac{b^2}{(2a)^2}+c\\
&=a\left(z+\frac{b}{2a}\right)^2-a\frac{b^2-4ac}{4a^2}.
\end{aligned}$$

Then the quadratic equation $az^2+bz+c=0$, with $a\neq 0$, is the same as

$$\left(z+\frac{b}{2a}\right)^2=\frac{D}{4a^2},\quad D=b^2-4ac.$$

Here D is the *discriminant* of the quadratic equation. If $D \geq 0$, then taking the square root gives $z+\dfrac{b}{2a}=\pm\dfrac{\sqrt{D}}{2a}$, and we get the solution $z=\dfrac{-b\pm\sqrt{D}}{2a}$. If $D<0$, then we get complex solution $z=\dfrac{-b\pm\sqrt{-D}i}{2a}$.

We remark that completing the square is the most basic technique for treating quadratic function. You should always derive the results by the technique, instead of memorizing the formula.

Exercise 3.47. Solve the equation $z^2+z+1=0$ by completing the square. Then solve the equation $z^3=1$.

All complex numbers $\mathbb{C}$ can be identified with the Euclidean plane $\mathbb{R}^2$, with the *real part* $\mathrm{Re}(x+yi)=x$ as the first coordinate and the *imaginary part* $\mathrm{Im}(x+yi)=y$ as the second coordinate. The corresponding real vector $(x,y)\in\mathbb{R}^2$ has norm r and angle θ (i.e., polar coordinate), and we have

$$x+yi=r\cos\theta+ir\sin\theta=re^{i\theta}.$$

The first equality is trigonometry, and the second equality uses the expansion (the theoretical explanation is the complex analytic continuation of the exponential function of real numbers)

$$\begin{aligned}
e^{i\theta}&=1+\frac{1}{1!}i\theta+\frac{1}{2!}(i\theta)^2+\cdots+\frac{1}{n!}(i\theta)^n+\cdots\\
&=\left(1-\frac{1}{2!}\theta^2+\frac{1}{4!}\theta^4+\cdots\right)+i\left(\theta-\frac{1}{3!}\theta^3+\frac{1}{5!}\theta^5+\cdots\right)\\
&=\cos\theta+i\sin\theta.
\end{aligned}$$

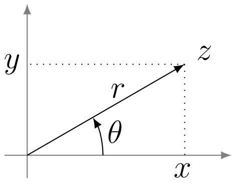

Figure 3.2: $z=x+yi=r\cos\theta+ir\sin\theta$.

We call $r=\sqrt{x^2+y^2}=\sqrt{z\bar{z}}$ the *modulus* of z, and call θ the *argument* of z. We also denote the m,odulus by $|z|$ because it is really the "absolute value". Moreover, we note that the argument is unique only up to adding multiples of 2π.

The multiplication of $z_1=r_1(\cos\theta_1+i\sin\theta_1)$ and $z_2=r_2(\cos\theta_2+i\sin\theta_2)$ is

$$\begin{aligned}
z_1z_2&=r_1r_2(\cos\theta_1+i\sin\theta_1)(\cos\theta_2+i\sin\theta_2)\\
&=r_1r_2[(\cos\theta_1\cos\theta_2-\sin\theta_1\sin\theta_2)+i(\sin\theta_1\cos\theta_2+\cos\theta_1\sin\theta_2)]\\
&=r_1r_2(\cos(\theta_1+\theta_2)+i\sin(\theta_1+\theta_2)).
\end{aligned}$$

This means that the modulus of z_1z_2 is $r_1r_2 = |z_1|\,|z_2|$, and the argument of z_1z_2 is $\theta_1 + \theta_2$. In terms of $re^{i\theta}$, this means

$$r_1e^{i\theta_1}\ r_2e^{i\theta_2} = r_1r_2e^{i(\theta_1+\theta_2)}.$$

Geometrically, this means that multiplying $re^{i\theta}$ on $\mathbb{C}$ is scaling by r and rotation by θ.

Geometrically, the complex conjugation is the flip with respect to the x-axis. This preserves the modulus r, and changes the argument θ to $-\theta$. Therefore

$$\overline{re^{i\theta}} = re^{-i\theta}.$$

Exercise 3.48. Explain the following properties of the modulus:

1. $|z| = 0 \iff z = 0$.

2. $|z_1 + z_2| \le |z_1| + |z_2|$.

3. $|z_1z_2| = |z_1|\,|z_2|$, and $\left|\dfrac{z_1}{z_2}\right| = \dfrac{|z_1|}{|z_2|}$.

4. $|\bar{z}| = |-z| = |z|$.

5. $|z_1 + z_2|^2 + |z_1 - z_2|^2 = 2(|z_1|^2 + |z_2|^2)$.

Example 3.8.2. The equation $e^{i(\alpha+\beta)} = e^{i\alpha}e^{i\beta}$ means

$$\begin{aligned}\cos(\alpha+\beta) + i\sin(\alpha+\beta) &= (\cos\alpha + i\sin\alpha)(\cos\beta + i\sin\beta)\\ &= (\cos\alpha\cos\beta - \sin\alpha\sin\beta) + i(\cos\alpha\sin\beta + \sin\alpha\cos\beta).\end{aligned}$$

Comparing the two sides, we get

$$\begin{aligned}\cos(\alpha+\beta) &= \cos\alpha\cos\beta - \sin\alpha\sin\beta,\\ \sin(\alpha+\beta) &= \cos\alpha\sin\beta + \sin\alpha\cos\beta.\end{aligned}$$

The equation $e^{i3\theta} = (e^{i\theta})^3$ means

$$\cos 3\theta + i\sin 3\theta = (\cos\theta + i\sin\theta)^3.$$

Expanding the right side, we get

$$\cos 3\theta = \cos^3\theta - 3\cos\theta\sin^2\theta = \cos^3\theta - 3\cos\theta(1 - \cos^2\theta) = 4\cos^3\theta - 3\cos\theta.$$

By comparing the imaginary parts, we can similarly get the formula for $\sin 3\theta$.

Example 3.8.3. To solve the equation $z^3 = 1$, we let $z = re^{i\theta}$. Then $z^3 = r^3e^{i3\theta}$ and the equation becomes $r^3e^{i3\theta} = e^{i2n\pi}$. Therefore $r^3 = 1$ and $3\theta = 2n\pi$. In other words, the solution is

$$z = e^{i\frac{2n}{3}\pi} = \cos\frac{2n}{3}\pi + i\sin\frac{2n}{3}\pi.$$

By $e^{i\theta} = e^{1(\theta+2\pi)}$, we actually get three solutions:

$$\begin{aligned}
&n = 0\colon z = e^0 = 1,\\
&n = 1\colon z = e^{i\frac{2}{3}\pi} = \cos\frac{2}{3}\pi + i\sin\frac{2}{3}\pi = \frac{-1+\sqrt{3}i}{2},\\
&n = 2\colon z = e^{i\frac{4}{3}\pi} = \cos\frac{4}{3}\pi + i\sin\frac{4}{3}\pi = \frac{-1-\sqrt{3}i}{2}.
\end{aligned}$$

In general, for each natural number n, there are n complex complex solutions of $z^n = 1$. The solutions are the n-th roots of unity

$$\cos\frac{2k}{n}\pi + i\sin\frac{2k}{n}\pi = \left(\cos\frac{2}{n}\pi + i\sin\frac{2}{n}\pi\right)^k = \xi_n^k, \quad k = 0, 1, \dots, n-1.$$

Here $\xi_n = e^{i\frac{2}{n}\pi}$ is the n-th *primitive root of unity.*

Example 3.8.4. To calculate $(1-i)^{10}$, we write the complex number is polar form (the vector $(1,-1)$ has length $\sqrt{2}$ and angle $-\frac{1}{4}\pi$): $1 - i = \sqrt{2}e^{-i\frac{1}{4}\pi}$. Then

$$(1-i)^{10} = (\sqrt{2})^{10}e^{-i\frac{10}{4}\pi} = 2^5e^{-i\frac{5}{2}\pi} = 32e^{-i\frac{1}{2}\pi} = -32i.$$

Example 3.8.5. The equation $|z - c| = R$ is the circle of radius r centered at c. Using complex conjugation, the equation is the same as

$$R^2 = (z-c)(\bar{z}-\bar{c}) = z\bar{z} - c\bar{z} - \bar{c}z + c\bar{c} = |z|^2 + |c|^2 - 2\mathrm{Re}(\bar{c}z).$$

For the special case $c = R$, the equation is $|z|^2 = 2R\,\mathrm{Re}(z)$.

Exercise 3.49. Solve the equation $(z+1)^4 + i(z+2)^4 = 0$.

Exercise 3.50. Solve the equation $z^2 = \bar{z}$.

Exercise 3.51. Solve the equation $|z^2 - 1| = |z|^2 + 1$.

Exercise 3.52. Calculate $(1+\sqrt{3}i)^{10}$.

Exercise 3.53. Suppose $|z_1| = |z_2| = |z_3| = |z_1 + z_2 + z_3| = 1$. Find $\left|\frac{1}{z_1} + \frac{1}{z_2} + \frac{1}{z_3}\right|$.

Exercise 3.54. Suppose $|z+1| \leq 2$, find the maximum of $|z+3|$.

A major difference between $\mathbb{R}$ and $\mathbb{C}$ is that the polynomial t^2+1 has no root in $\mathbb{R}$ but has a pair of roots $\pm i$ in $\mathbb{C}$. In fact, complex numbers has the following so called *algebraically closed* property.

Theorem 3.8.1 (Fundamental Theorem of Algebra). *Any non-constant complex polynomial has roots.*

The real number $\mathbb{R}$ is not algebraically closed.

Chapter 4

Integer and Polynomial

4.1 Quotient and Remainder

In high school, we did long divisions of natural numbers such as

$$\begin{array}{r}
1873\\
13\overline{\smash{)}24357}\\
\underline{13000}\\
11357\\
\underline{10400}\\
957\\
\underline{910}\\
47\\
\underline{39}\\
8
\end{array}$$

The result means $24357 = 1873 \cdot 13 + 8$. In general, we have the following result when we try to divide one integer by another.

Lemma 4.1.1. *For any $a, b \in \mathbb{Z}$ satisfying $b \neq 0$, there are unique integers q and r such that*

$$a = qb + r, \quad 0 \leq r < |b|.$$

The integer q is called the *quotient* of the division, and r is the *remainder*. The integer a is *divisible* by b if and only if $r = 0$.

Proof. Assume $b > 0$. By the eighth property in Proposition 3.5.3, for the rational number $\left|\frac{a}{b}\right|$, there is a natural number n, such that $\left|\frac{a}{b}\right| < n$. Thus $-n < \frac{a}{b} < n$. Then among the increasing sequence of finitely many integers $-n < -n+1 < -n+2 < \cdots < n-2 < n-1 < n$, there is a biggest integer q such that $q \leq \frac{a}{b}$.

Being the biggest, we must also have $q+1 > \dfrac{a}{b}$. Thus

$$q \le \frac{a}{b} < q+1.$$

Since $b > 0$, we get $qb \le a < qb + q$. Let $r = a - qb$. We then have $a = qb + r$ and $0 \le r < b = |b|$.

Assume $b < 0$. Then $-b > 0$. By the lemma just proved for $b > 0$, we have $a = q(-b) + r = (-q)b + r$ with $0 \le r < -b = |b|$. The quotient of a by b is then $-q$, with the same r as the remainder.

For the uniqueness of q and r, let us assume

$$a = q_1 b + r_1 = q_2 b + r_2, \quad 0 \le r_1 < |b|, \quad 0 \le r_2 < |b|.$$

Then $(q_1 - q_2)b = r_2 - r_1$. Suppose $q_1 \ne q_2$. Then by the fourth property in Proposition 3.3.3, $|q_1 - q_2| \ge 1$. By the sixth property in Proposition 3.4.3, we get

$$|(q_1 - q_2)b| = |(q_1 - q_2)||b| \ge 1|b| = |b|.$$

On the other hand,

$$0 \le r_1 < |b|, 0 \le r_2 < |b| \implies -|b| = 0 - |b| < r_1 - r_2 < |b| - 0 = |b| \iff |r_1 - r_2| < |b|.$$

This contradicts with $(q_1 - q_2)b = r_2 - r_1$. The contradiction shows that $q_1 = q_2$, which futher implies $r_1 = r_2$.

□

A *polynomial* is a function of the form

$$p(t) = a_n t^n + a_{n-1} t^{n-1} + \cdots + a_2 t^2 + a_1 t + a_0.$$

If $a_n \ne 0$, then n is called the *degree* $\deg p(t)$ of the polynomial. For example, $t^2 + 2t - 1$ and $7 + 3t - 2t^4$ are polynomials of degrees 2 and 4, respectively. The zero polynomial, in which all the coefficients are zero, has degree $-\infty$. We clearly have

$$\deg(p(t) + q(t)) \le \max\{\deg p(t), \deg q(t)\}, \quad \deg(p(t)q(t)) = \deg p(t) + \deg q(t).$$

The long division can also be applied to polynomials (with rational or real coefficients). The following computation

$$
\begin{array}{r|rrrrrr}
 & & & t^3 & +t^2 & +t & +2 \\ \cline{2-7}
t^2-3t+2 & t^5 & -2t^4 & & +t^2 & & +3 \\
 & -t^5 & +3t^4 & -2t^3 & & & \\ \cline{2-4}
 & & t^4 & -2t^3 & +t^2 & & \\
 & & -t^4 & +3t^3 & -2t^2 & & \\ \cline{3-5}
 & & & t^3 & -t^2 & & \\
 & & & -t^3 & +3t^2 & -2t & \\ \cline{4-6}
 & & & & 2t^2 & -2t & +3 \\
 & & & & -2t^2 & +6t & -4 \\ \cline{5-7}
 & & & & & 4t & -1
\end{array}
$$

means that $t^5 - 2t^4 + t^2 + 3 = (t^3 - 3t + 2)(t^2 - 3t + 2) + (4t - 1)$. In other words, the division of $t^5 - 2t^4 + t^2 + 3$ by $t^2 - 3t + 2$ has the quotient $t^3 - 3t + 2$ and the remainder $4t - 1$.

Lemma 4.1.2. *For any (rational or real) polynomials $a(t)$ and $b(t)$ satisfying $b(t) \neq 0$, there are unique polynomials $q(t)$ and $r(t)$ such that*

$$a(t) = q(t)b(t) + r(t), \quad \deg r(t) < \deg b(t).$$

The polynomial $q(t)$ is called the *quotient* of the division, and $r(t)$ is the *remainder*. The polynomial $a(t)$ is *divisible* by $b(t)$ if and only if $r(t) = 0$.

Proof. First we note that if $\deg b(t) > \deg a(t)$, then the lemma holds with $q(t) = 0$ and $r(t) = a(t)$.

Now let us fix $b(t)$ and induct on the degree $d = \deg a(t)$. Suppose $d = 0$, then $a(t) = a_0$ is a constant. If $\deg b(t) > 0$, then the lemma holds by the remark above. If $\deg b(t) = 0$, then $b(t) = b_0$ is a nonzero constant, and we have $a_0 = \dfrac{a_0}{b_0} b_0 + 0$. Therefore $a(t) = q(t)b(t) + r(t)$ for $q(t) = \dfrac{a_0}{b_0}$ and $r(t) = 0$. This completes the verification for $d = 0$.

Suppose the lemma holds for all $a(t)$ of degree $< d$. Now consider a polynomial

$$a(t) = a_d t^d + a_{d-1} t^{d-1} + \cdots + a_2 t^2 + a_1 t + a_0$$

of degree d. Let

$$b(t) = b_n t^n + b_{n-1} t^{n-1} + \cdots + b_2 t^2 + b_1 t + b_0, \quad b_n \neq 0.$$

If $n > d$, then we can find $q(t)$ and $r(t)$ as in the lemma, by the remark at the beginning of the proof. If $n \leq d$, then we have

$$a(t) = \left(\frac{a_d}{b_n} t^{d-n}\right) b(t) + \tilde{a}(t),$$

where

$$\tilde{a}(t) = \left(a_{d-1} - \frac{b_{n-1}}{b_n} a_d\right) t^{d-1} + \left(a_{d-2} - \frac{b_{n-2}}{b_n} a_d\right) t^{d-2} + \cdots$$

has degree $< d$. By applying the inductive assumption to the polynomial $\tilde{a}(t)$, we get

$$\tilde{a}(t) = \tilde{q}(t)b(t) + \tilde{r}(t), \quad \deg \tilde{r}(t) < \deg b(t).$$

Then

$$a(t) = \left(\frac{a_d}{b_n} t^{d-n} + \tilde{q}(t)\right) b(t) + \tilde{r}(t).$$

We conclude that the proposition holds for a with

$$q(t) = \frac{a_d}{b_n}t^{d-n} + \tilde{q}(t), \quad r(t) = \tilde{r}(t).$$

For the uniqueness of $q(t)$ and $r(t)$, let us assume

$$a(t) = q_1(t)b(t) + r_1(t) = q_2(t)b(t) + r_2(t),$$

$$0 \le \deg r_1(t) < \deg b(t), \quad 0 \le \deg r_2(t) < \deg b(t).$$

Then $(q_1(t) - q_2(t))b(t) = r_2(t) - r_1(t)$. If $q_1(t) \ne q_2(t)$, then

$$\deg[(q_1(t) - q_2(t))b(t)] = \deg(q_1(t) - q_2(t)) + \deg b(t) \ge degb(t),$$

which contradicts to

$$\deg(r_2(t) - r_1(t)) \le \max\{\deg r_1(t), \deg r_2(t)\} < \deg b(t).$$

The contradiction shows that $q_1(t) = q_2(t)$, which further implies $r_1(t) = r_2(t)$. □

Exercise 4.1. Find the quotient and the remainder of the division of integers.

$456 \div 123, \quad 123 \div 456, \quad (-456) \div 123, \quad (-456) \div (-123), \quad (-123) \div 456, \quad 1221 \div 33.$

Exercise 4.2. Find the quotient and the remainder of the division of polynomials.

$$(t^4 + 2t^3 - 3t + 1) \div (t + 2), \quad (3t^5 + 4t^3 - 2t + 5) \div (t^2 + t + 2).$$

Exercise 4.3. Suppose the divisions of a_1 and a_2 by b and b have respective remainders r_1 and r_2. What can you say about the remainder of the division of $a_1 + a_2$ by b?

Exercise 4.4. Prove that the remainder of the division of a polynomial $a(t)$ by $b(t) = t - t_0$ is $a(t_0)$.

4.2 Decimal Expansion

In decimal expression, the number 24357 really means

$$24357 = 20000 + 4000 + 300 + 50 + 7 = 2 \cdot 10^4 + 4 \cdot 10^3 + 3 \cdot 10^2 + 5 \cdot 10 + 7.$$

In general, the we have the *decimal expansion* for a natural number is

$$n = m_k \cdot 10^k + m_{k-1} \cdot 10^{k-1} + \cdots + m_2 \cdot 10^2 + m_1 \cdot 10 + m_0, \quad 0 \le m_i < 10, \quad m_k \ne 0.$$

In this case, the decimal expression for the number is $n = m_k m_{k-1} \cdots m_2 m_1 m_0$.

The number 10 is the base of the decimal expression. The other numbers can also be used as the base of the expression. To get the expression of 24357 based on 13, for example, we carry the following divisions:

$$\begin{aligned} 24357 &= 1873 \cdot 13 + 8, \\ 1873 &= 144 \cdot 13 + 1, \\ 144 &= 11 \cdot 13 + 1. \end{aligned}$$

Then we combine these and get

$$24357 = (144{\cdot}13{+}1){\cdot}13{+}8 = ((11{\cdot}13{+}1){\cdot}13{+}1){\cdot}13{+}8 = 11{\cdot}13^3{+}1{\cdot}13^3{+}1{\cdot}13^3{+}8,$$

which gives rise to the expression

$$24357 = \widehat{11}118_{[13]}.$$

The subscript "[13]" indicates the base of the expansion, and $\widehat{11}$ indicates *one digit* in the expansion.

Strictly speaking, we should write $24357_{[10]}$ for 24357. The base 10 is omitted because our usual writing system is based on 10.

Note that the expansion above may be computed by making repeated use of Lemma 4.1.1. The method above also inspires the proof of the following general result.

Proposition 4.2.1. *For any $n, b \in \mathbb{N}$ satisfying $b > 1$, we have*

$$n = m_k b^k + m_{k-1} b^{k-1} + \cdots + m_2 b^2 + m_1 b + m_0,$$

for unique natural number k and integers $m_0, m_1, m_2, \ldots, m_{k-1}, m_k$ satisfying

$$0 \le m_i < b, \quad m_k \neq 0.$$

Proof. For the existence of the expansion, let us fix b and induct on n. For $n = 1$, the statement holds for $k = 0$, $m_0 = 1$, because $b > 1$. Now assume the expansion exists for all natural numbers $< n$. Then by Lemma 4.1.1, we have $n = qb + r$ for some integer q and r satisfying $0 \le r < b$.

Since $qb = n - r > 0 - b = -b$, multiplying $b^{-1} > 0$ on both sides gives us $q > -1$. Since q is an integer, and -1 is the biggest negative integer (by third and fourth properties in Proposiiton 3.3.3), we have $q \ge 0$.

If $q = 0$, then $n = r < b$, and we have the expansion with $k = 0$ and $m_0 = n$. If $q > 0$, then $q \ge 1$ and by $b > 1$, $n = qb + r \ge qb > q$. Therefore the inductive assumption may be applied to q, and we get

$$q = m_l b^l + m_{l-1} b^{l-1} + \cdots + m_2 b^2 + m_1 b + m_0.$$

This further gives us the expansion

$$n = qb + r = m_l b^{l+1} + m_{l-1} b^l + \cdots + m_2 b^3 + m_1 b^2 + m_0 b + r$$

for n, where we note that $0 \le r < b$.

The uniqueness can also be proved by induction on n. For $n = 1$, we must have $k = 0$ and $m_0 = 1$ because if $k > 0$, then

$$m_k b^k + m_{k-1} b^{k-1} + \cdots + m_2 b^2 + m_1 b + m_0 \ge m_k b^k > m_k \ge 1 = n.$$

Now assume the uniqueness of the expansion for all natural numbers $< n$. Then an expansion of n can be rewritten as

$$n = qb + r, \quad q = m_k b^{k-1} + m_{k-1} b^{k-2} + \cdots + m_2 b + m_1, \quad r = m_0.$$

This satisfies the conditions in Lemma 4.1.1, which implies that q and m_0 are uniquely determined by n. Moreover, we have $n \ge qb > q$, so that the inductive assumption may be applied to q. The result is that q further uniquely determines $k - 1$ and $m_1, m_2, \ldots, m_{k-1}, m_k$.

□

The natural number $b > 1$ is the *base* of the expansion. The decimal expansion takes $b = 10$. The *binary expansion* is based on $b = 2$. For example,

$$6 = 2^2 + 2, \quad 15 = 2^3 + 2^2 + 2 + 1, \quad 80 = 2^6 + 2^4$$

yield the binary expressions

$$6 = 110_{[2]}, \quad 15 = 1111_{[2]}, \quad 80 = 101000_{[2]}.$$

The binary expansion is of fundamental importance to the computer science. Another frequently used system in the computer science is the *hexadecimal expansion*, which makes use of the base $b = 16$. The system usually uses the following notations for the digits between 10 and 15

$$A = \widehat{10}, \quad B = \widehat{11}, \quad C = \widehat{12}, \quad D = \widehat{13}, \quad E = \widehat{14}, \quad F = \widehat{15}.$$

Therefore

$$\begin{aligned} 2C0_{[13]} &= 2\widehat{12}0_{[13]} = 2 \cdot 16^2 + 12 \cdot 16 = 704, \\ ABCD_{[13]} &= \widehat{10}\widehat{11}\widehat{12}\widehat{13}_{[13]} = 10 \cdot 16^3 + 11 \cdot 16^2 + 12 \cdot 16 + 13 = 28621. \end{aligned}$$

Exercise 4.5. Express the following numbers in binary and hexadecimal forms.

$$6, \quad 20, \quad ,200, \quad 1024, \quad 12345.$$

Exercise 4.6. Express the following binary and hexadecimal numbers in decimal form.

$$1000_{[2]},\quad 11011_{[2]},\quad 101010_{[2]};\quad 37_{[16]},\quad 3\widehat{12}_{[16]},\quad 1000_{[16]},\quad 3\widehat{10}\widehat{11}_{[16]}.$$

Exercise 4.7. Express the following binary numbers in the form based on 4. Do you a general method for the conversion?

$$1000_{[2]},\quad 11011_{[2]},\quad 101010_{[2]},\quad 110001111_{[2]}.$$

Exercise 4.8. Carry out the long operaitons:

1. long additions $1101_{[2]} + 110_{[2]}$ and $120_{[3]} + 21_{[3]}$.
2. long substractions $1101_{[2]} - 110_{[2]}$, $120_{[3]} - 21_{[3]}$.
3. long multiplications $1101_{[2]} \times 110_{[2]}$ and $120_{[3]} \times 21_{[3]}$.
4. long divisions $10111_{[2]} \div 11_{[2]}$ and $11202_{[3]} \div 12_{[3]}$.

Exercise 4.9. Prove that a natural number is divisible by 5 $\iff$ The last digit in the decimal expression is 0 or 5.

Exercise 4.10. Prove that $10^k - 1$ is always divisible by 9. Then prove the following.

1. A natural number is divisible by 3 $\iff$ The sum of its digits in the decimal expression is divisible by 3.
2. A natural number is divisible by 9 $\iff$ The sum of its digits in the decimal expression is divisible by 9.

Exercise 4.11. A natural number is divisible by 11 $\iff$ In the decimal expression, the difference between the sum of digits in the odd positions and the sum of digits in the even positions is divisible by 11. For example, 12345 is not divisible by 11 because $(1 + 3 + 5) - (2 + 4) = 3$ is not divisible by 11, while 19261 is divisible by 111 because $(1 + 2 + 1) - (9 + 6) = -11$ is divisible by 11.

Exercise 4.12. In hexadecimal expression, prove the following.

1. A natural number is divisible by 3 $\iff$ The sum of its digits is divisible by 3.
2. A natural number is divisible by 4 $\iff$ The last digit is 0, 4, 8, or 12.
3. A natural number is divisible by 5 $\iff$ The sum of its digits is divisible by 5.
4. A natural number is divisible by 8 $\iff$ The last digit is 0 or 8.

5. A natural number is divisible by 15 $\iff$ The sum of its digits is divisible by 15.

6. A natural number is divisible by 17 $\iff$ The difference between the sum of digits in the odd positions and the sum of digits in the even positions is divisible by 17.

Can you find similar criteria for expressions of natural numbers based on 5?

Exercise 4.13. Because the numerical expansion is based on Lemma 4.1.1, which has an analogous counterpart the Lemma 4.1.2 for polynomials, the expansion can also be applied to polynomials.

1. Expand $t^5 - 2t^4 + t^2 + 3$ as a sum of powers of $t^2 - 3t + 2$.

2. For polynomials, write down a proposition similar to Proposition 4.2.1.

3. Prove the proposition.

4. Expand $t^5 - 2t^4 + t^2 + 3$ as a sum of powers of $s = t - 2$, by substituting $t = s + 2$ into the polynomial.

Exercise 4.14. What is the pun in the following statement: There are 10 kinds of people in the world, those who know binary numbers and those who don't.

4.3 Greatest Common Divisor

Let a and b be integers, with $b \neq 0$. If $a = qb$ for some integer q, then we say a is *divisible* by b, or b is a *divisor* of a. In this case, we denote $b \mid a$. If a is not divisible by b, then we denote $b \nmid a$. The following properties of the divisibility is easy to verify.

Proposition 4.3.1. *The divisibility has the following properties:*

1. $a \mid 0$ *and* $1 \mid a$ *for any* a.

2. $a \mid b, b \mid c \implies a \mid c$.

3. $a \mid b, a \mid c \implies a \mid b + c$.

4. $a \mid b, c \mid d \implies ac \mid bd$.

5. $a \mid b, b \mid a \iff a = b$ *or* $a = -b$.

Two integers may share divisors. For example, both 204 and 90 are divisible by 2, 3, 6. In other words, 2, 3, 6 are *common* divisors of 204 and 90. In general, for integers $a_1, a_2, \ldots, a_k$, let

$$D(a_1, a_2, \ldots, a_k) = \{n \in \mathbb{N} \colon n \mid a_1, n \mid a_2, \ldots, n \mid a_k\}$$

be the set of all positive common divisors.

Proposition 4.3.2. *$D(a_1, a_2, \ldots, a_k)$ is not changed under the following modifications.*

1. *Change order and sign.*
2. *If one number is a multiple of another, drop the multiple. In particular, zeros may be dropped.*
3. *Add a multiple of one number to another number.*

Examples of the first change are $D(a, b, c) = D(b, c, a) = D(-a, -b, c)$, of the second are $D(a, ab, c) = D(a, c) = D(a, 0, c)$, of the third are $D(a, b, c) = D(a, b + qa, c)$ for any $q \in \mathbb{Z}$.

Proof. The definition of D does not depend on the order of $a_1, a_2, \ldots, a_k$. The equivalence $n \mid a \iff n \mid -a$ implies that D is not changed if some of the signs are changed. The equivalence $n \mid a \iff n \mid -aextandn \mid ab$ implies that D is not changed if multiples are dropped. It is also easy to prove the equivalence $n \mid aextandn \mid b \iff n \mid aextandn \mid qa + b$ for $q \in \mathbb{Z}$, which implies that D is not changed if an integer multiple of one is added to another.

□

The modifications in Proposition 4.3.2 may be used to simplify the collection $\{a_1, a_2, \ldots, a_k\}$ and help us to find the common divisors. To find the common divisors of two numbers a and b, for example, we may change the order or the sign so that $a, b \in \mathbb{N}$ and $a \geq b$. If $a = b$, then $D(a, b) = D(a)$ is the set of divisors of a. If $a > b$, then by Lemma 4.1.1, a may be divided by b:

$$a = qb + a_1, \quad 0 \leq a_1 < b.$$

This implies $D(a, b) = D(qb + a_1, b) = D(a_1, b)$. Since $\max\{a_1, b\} = b < a = \max\{a, b\}$, the problem of finding $D(a_1, b)$ is easier. The same idea may be repeated and yield $D(a_1, b) = D(a_1, b_1)$ with even smaller $\max\{a_1, b_1\} = a_1 < b = \max\{a_1, b\}$. The process continues until the numbers are small enough for us to find the common divisors. The process is called the *Euclidean*[1] *algorithm.*

[1] Euclid of Alexandria: born about 325 BC; died about 265 BC in Alexandria, Egypt. Best known for his treatise on mathematics The Elements, which consists of 13 books. Books 7 to 9 deal with number theory, and the Euclidean algorithm is contained in book 7. “It is sometimes said that, next to the Bible, the ”Elements” may be the most translated, published, and studied of all the books produced in the Western world.” - B. L. van der Waerden.

Let us carry out the Euclidean algorithm for 204 and 90.

$$\begin{aligned}
204 &= 2\cdot 90 + 24 && \Longrightarrow D(204,90) = D(90,24),\\
90 &= 3\cdot 24 + 18 && \Longrightarrow D(90,24) = D(24,18),\\
24 &= 1\cdot 18 + 6 && \Longrightarrow D(24,18) = D(18,6),\\
18 &= 3\cdot 6 && \Longrightarrow D(18,6) = D(6).
\end{aligned}$$

The series of computations give us

$$D(204,90) = D(6) = \{n \in \mathbb{N} \colon n \mid 6\} = \{1,2,3,6\}.$$

This means

$$n \mid 204, n \mid 90 \iff n \mid 6.$$

Note that 6 is the biggest of the common divisors. By tracing back the computation above, 6 can be expressed in terms of 204 and 90:

$$\begin{aligned}
6 &= 24 - 1\cdot 18\\
&= 24 - 1\cdot(90 - 3\cdot 24) = 4\cdot 24 - 1\cdot 90\\
&= 4\cdot(204 - 2\cdot 90) - 1\cdot 90 = 4\cdot 204 - 9\cdot 90.
\end{aligned}$$

The Euclidean algorithm suggests the idea for the proof of the following general result.

Theorem 4.3.3. *If $a_1, a_2, \dots, a_k \in \mathbb{Z}$ are not all zero, then there is a unique natural number n such that $d \mid a_1, d \mid a_2, \dots, d \mid a_k \iff d \mid n$. Moreover,*

$$n = u_1 a_1 + u_2 a_2 + \cdots + u_k a_k$$

for some integrs $u_1, u_2, \dots, u_k$.

The property $d \mid a_1, d \mid a_2, \dots, d \mid a_k \iff d \mid n$ means exactly $D(a_1, a_2, \dots, a_k) = D(n)$. The number n is the *greatest common divisor* of $a_1, a_2, \dots, a_k$, and is denoted

$$n = \gcd(a_1, a_2, \dots, a_k).$$

The greatest common divisor is not changed under the modifications in Proposition 4.3.2.

Proof. Since common divisors are independent of the order and signs, we will assume a_i are natural numbers, and the proof is by inducting on $\max\{a_1, a_2, \dots, a_k\}$.

If $\max\{a_1, a_2, \dots, a_k\} = 1$, then $a_i = 1$, and $D(a_1, a_2, \dots, a_k) = D(1)$. We may also choose $u_1 = 1, u_2 = \cdots = u_k = 0$.

Assume the theorem holds for the cases $\max\{a_1, a_2, \dots, a_k\} < a$. Now consider the case $\max\{a_1, a_2, \dots, a_k\} = a$. After dropping zeros among $a_1, a_2, \dots, a_k$, without loss of generality, we may assume that $a_1, a_2, \ldots, a_k$ are nonzero and a_1 is the smallest among $a_1, a_2, \ldots, a_k$. We consider two cases $a_1 = a$ and $a_1 < a$.

If $a_1 = a$, then $a_2 = \cdots = a_k = a$, $D(a_1, a_2, \ldots, a_k) = D(a)$. We may also choose $u_1 = 1, u_2 = \cdots = u_k = 0$.

If $a_1 < a$, then we divide all the other a_i by a_1 and get

$$\begin{aligned}
a_2 &= q_2 a_1 + a_2', & a_2' &< a_1, \\
a_3 &= q_3 a_1 + a_3', & a_3' &< a_1, \\
&\vdots & &\vdots \\
a_k &= q_k a_1 + a_k', & a_k' &< a_1.
\end{aligned}$$

This implies

$$D(a_1, a_2, \ldots, a_k) = D(a_1, q_2 a_1 + a_2', \ldots, q_k a_1 + a_k') = D(a_1, a_2', \ldots, a_k').$$

Since $\max\{a_1, a_2', \ldots, a_k'\} = a_1 < a$, we may apply the inductive assumption to $a_1, a_2', \ldots, a_k'$ and find unique $n \in \mathbb{N}$ such that $D(a_1, a_2', \ldots, a_k') = D(n)$. Moreover, we also have

$$n = u_1 a_1 + u_2' a_2' + \ldots + u_k' a_k'$$

for some integrs $u_1, u_2', \ldots, u_k'$. Then we conclude $D(a_1, a_2, \ldots, a_k) = D(n)$, with

$$n = (u_1 - q_2 u_2' - \cdots - q_k u_k') a_1 + u_2' a_2 + \ldots + u_k' a_k.$$

□

The Euclidean algorithm can be applied to several numbers. For example, to find the greatest common divisor of $-36, 204, 90, -114$, we use the divisions

$$\begin{aligned}
204 &= 5 \cdot 36 + 24, & 36 &= 3 \cdot 12, \\
90 &= 2 \cdot 36 + 18, & 24 &= 2 \cdot 12, \\
114 &= 3 \cdot 36 + 12. & 18 &= 1 \cdot 12 + 6.
\end{aligned}$$

and the fact that 12 is a multiple of 6 to get

$$\gcd(-36, 204, 90, -114) = \gcd(36, 204, 90, 114) = \gcd(36, 24, 18, 12) = \gcd(0, 0, 6, 12) = 6.$$

Moreover,

$$\begin{aligned}
6 &= 18 - 1 \cdot 12 = (90 - 2 \cdot 36) - 1 \cdot (114 - 3 \cdot 36) \\
&= 90 - 1 \cdot 114 + 1 \cdot 36 = (-1) \cdot (-36) + 0 \cdot 204 + 1 \cdot 90 + 1 \cdot (-114).
\end{aligned}$$

Since the division of polynomials is similar to the division of integers, the discussion about divisors and greatest common divisors also applies to polynomials.

Let $a(t)$ and $b(t) \neq 0$ be polynomials. If $a(t) = q(t)b(t)$ for some polynomial $q(t)$, then we say $a(t)$ is *divisible* by $b(t)$, or $b(t)$ is a *divisor* of $a(t)$. In this case, we denote $b(t) \mid a(t)$. If $a(t)$ is not divisible by $b(t)$, then we denote $b(t) \nmid a(t)$. The divisibility of polynomials has the properties similar to that of integers. The only modification is the last one:

$$a(t) \mid b(t), b(t) \mid a(t) \iff a(t) = rb(t) \text{ for some number } r \neq 0.$$

Similar to integers, define the set

$$D(a_1(t), a_2(t), \dots, a_k(t)) = \{p(t) \colon p(t) \mid a_1(t), p(t) \mid a_2(t), \dots, p(t) \mid a_k(t)\}$$

of all common divisors of polynomials $a_1(t), a_2(t), \dots, a_k(t)$. The set is not changed under modifications as in Proposition 4.3.2 (multiple means multiplying by a polynomial). This allows us to carry out the Euclidean algorithm for polynomials and prove the following general result.

Theorem 4.3.4. *If $a_1(t), a_2(t), \dots, a_k(t)$ are polynomials, not all zero, then there is a polynomial $p(t)$ such that $d(t) \mid a_1(t), d(t) \mid a_2(t), \dots, d(t) \mid a_k(t) \iff d(t) \mid p(t)$. Moreover, $p(t)$ is unique up to multiplication by a nonzero number and*

$$p(t) = u_1(t)a_1(t) + u_2(t)a_2(t) + \cdots + u_k(t)a_k(t)$$

for some polynomials $u_1(t), u_2(t), \dots, u_k(t)$.

The theorem may be proved by inducting on the maximum degree of $a_1(t), a_2(t), \dots, a_k(t)$. The polynomial $p(t)$ is the *greatest common divisor* because it has the greatest degree among the common divisors. The greatest common divisor is denoted

$$p(t) = \gcd(a_1(t), a_2(t), \dots, a_k(t)).$$

The following is the application of the Euclidean algorithm to finding the greatest common divisor of $t^4 - t^3 + t^2 - 1$, $t^3 - 1$, $-4t^4 + 4t^3 + t^2 + 1$. First the divisions

$$\begin{aligned} t^4 - t^3 + t^2 - 1 &= (t-1)(t^3 - 1) + (t^2 + t - 2), \\ -4t^4 + 4t^3 + t^2 - 1 &= (-4t + 4)(t^3 - 1) + (t^2 - 4t + 3), \end{aligned}$$

imply that

$$\gcd(t^4 - t^3 + t^2 - 1, t^3 - 1, -4t^4 + 4t^3 + t^2 - 1) = \gcd(t^2 + t - 2, t^3 - 1, t^2 - 4t + 3).$$

The the divisions

$$\begin{aligned} t^3 - 1 &= (t-1)(t^2 + t - 2) + (3t - 3), \\ t^2 - 4t + 3 &= 1(t^2 + t - 2) + (-5t + 5), \end{aligned}$$

imply that

$$\gcd(t^2+t-2, t^3-1, t^2-4t+3) = \gcd(-5t+5, 3t-3, t^2-4t+3) = \gcd(t-1, t^2-4t+3).$$

Finally, since $t^2-4t+3=(t-3)(t-1)$ is a multiple of $t-1$, we conclude that

$$\gcd(t^4-t^3+t^2-1, t^3-1, -4t^4+4t^3+t^2-1)=t-1.$$

The greatest common divisor can be expressed in terms of the original polynomials by tracing back the series of divisions:

$$\begin{aligned}
t-1 &= \frac{1}{3}(t^2-4t+3)-\frac{1}{3}(t^2+t-2)\\
&= \frac{1}{3}[(-4t^4+4t^3+t^2-1)-(-4t+4)(t^3-1)]-\frac{1}{3}[(t^4-t^3+t^2-1)-(t-1)(t^3-1)]\\
&= -\frac{1}{3}(t^4-t^3+t^2-1)+(t-1)(t^3-1)+\frac{1}{3}(-4t^4+4t^3+t^2-1).
\end{aligned}$$

Exercise 4.15. Find the greatest common divisors and express the results in terms of the original numbers.

1. 1053, 390.

2. 1053, −390, 247.

3. 1053, −390, 247, −500.

Exercise 4.16. Find the greatest common divisors and express the results in terms of the original polynomials.

1. $t^5-t^3+t^2-1$, t^7-t^3.

2. $t^5-t^3+t^2-1$, t^7-t^3, t^4-2t+1.

3. $t^5-t^3+t^2-1$, t^7-t^3, t^4-2t+1, t^4+1.

Exercise 4.17. Prove $D(a,ab,c)=D(a,c)$, $D(a,b)=D(a,b+qa)$.

Exercise 4.18. Prove $\gcd(ac,bc)=\gcd(a,b)c$, $\gcd(a,b,c)=\gcd(a,\gcd(b,c))$.

Exercise 4.19. Prove Theorem 4.3.4 by inducting on the maximum degree of the polynomials.

4.4 Prime and Factorization

Any natural number is divisible by 1 and itself. If a natural number $p>1$ satisfies

$$n\in\mathbb{N}, n\mid p \implies n=1 \text{ or } n=p,$$

i.e., p is divisible *only* by 1 and itself, then p is called a *prime number*. An integer may also be defined as prime if its absolute value is prime. But in this course we will stick to natural numbers for prime numbers.

If a natural number $n \geq 2$ is not a prime number, then we have $n = n_1 n_2$ for some natural numbers $n_1, n_2 > 1$. In this case, n is a *composite number*. Thus composite numbers are multiples of 2, multiples of 3, multiples of 5, etc. (multiples of 4 are also multiples of 2). After all such multiples are taken away, the remaining ones are the prime numbers. For example, in the following table of natural numbers between 2 and 100, m_n indicates that m is a multiple of n. The 25 remaining natural numbers (the ones without subscripts) are all the prime numbers between 2 and 100.

	2	3	4_2	5	6_2	7	8_2	9_3	10_2
11	12_2	13	14_2	15_5	16_2	17	18_2	19	20_2
21_3	22_2	23	24_2	25_5	26_2	27_3	28_2	29	30_2
31	32_2	33_3	34_2	35_5	36_2	37	38_2	39_3	40_2
41	42_2	43	44_2	45_5	46_2	47	48_2	49_7	50_2
51_3	52_2	53	54_2	55_5	56_2	57_3	58_2	59	60_2
61	62_2	63_3	64_2	65_5	66_2	67	68_2	69_3	70_2
71	72_2	73	74_2	75_5	76_2	77_7	78_2	79	80_2
81_3	82_2	83	84_2	85_5	86_2	87_3	88_2	89	90_2
91_7	92_2	93_3	94_2	95_5	96_2	97	98_2	99_3	100_2

Proposition 4.4.1. *Let p be a prime number and let n be a natural number. Then*

$$\gcd(n,p) = \begin{cases} 1 & \textit{if } p \nmid n \\ p & \textit{if } p \mid n \end{cases}.$$

Proof. Since p is a prime number and $\gcd(n,p)$ is a divisor of p, $\gcd(n,p)$ must be either 1 or p. It remains to show that $\gcd(n,p) = p \iff p \nmid n$. If n is divisible by p, then $\gcd(n,p) = p$ by the second property in Proposition 4.3.2. Conversely, if $\gcd(n,p) = p$, then n is divisible by p because $\gcd(n,p)$ is a divisor of both n and p. □

Proposition 4.4.2. *Let p be a prime number and let m, n be natural numbers. Then $p \mid mn \implies p \mid m$ or $p \mid n$.*

Proof. We prove the equivalent statement that $p \mid mn, p \nmid m \implies p \mid n$. By Proposition 4.4.1, since $p \nmid m$, we have $\gcd(m,p) = 1$. Then by proposiition, 4.3.3, we have $um + vp = 1$ for some integers u, v. Then $umn + vpn = n$. Since $p \mid mn$ by the assumption, and $p \mid vpn$ always holds, we conclude that $p \mid n$.

□

Theorem 4.4.3. *Any natural number > 1 can be expressed as a product of prime numbers. Moreover, the expression is unique up to permutation.*

The theorem shows that the prime numbers are the (multiplicative) building blocks of all natural numbers. For example,

$$6 = 2 \cdot 3, \quad 60 = 2 \cdot 2 \cdot 3 \cdot 5 = 2^2 \cdot 3 \cdot 5, \quad 600 = 2 \cdot 2 \cdot 2 \cdot 3 \cdot 5 \cdot 5 = 2^3 \cdot 3 \cdot 5^2.$$

Proof. First, the natural number 2 is the product of one prime number 2. The expression is clearly unique.

Next assume the theorem holds for all natural numbers > 1 and $< n$. Consider the theorem for n. If n is a prime number, then n is the product of one prime number n. Given any other expression $n = p_1 p_2 \cdots p_k$, the fact that $p_1 \mid n$ and the assumptions n is prime imply that $p_1 = n$. This implies $k = 1$, so that the expression is unique.

If n is not a prime number, then $n = n_1 n_2$ for two natural numbers $n_1, n_2 > 1$. This implies $n_1, n_2 < n$. By the inductive assumption, both can be written as products of prime numbers, so that n is also a product of prime numbers. As for the uniqueness, let

$$n = p_1 p_2 \cdots p_k = p'_1 p'_2 \cdots p'_{k'}$$

be two expressions of n as products of prime numbers. Then by Proposition 4.4.2 and $p_1 \mid n = p'_1 p'_2 \cdots p'_{k'}$, p_1 is a divisor of one of p'_i. Up to permutation, we may assume that $p_1 \mid p'_1$. Since p'_1 is prime and $p_1 > 1$, we get $p'_1 = p_1$. Thus $m = \dfrac{n}{p_1} < n$ is a natural number and

$$m = p_2 \cdots p_k = p'_2 \cdots p'_{k'}$$

are two expressions of m as products of prime numbers. By inductive assumption, the primes $p_2, \cdots, p_k$ is the same as the primes $p'_2, \cdots, p'_{k'}$ up to permutation. By adding $p'_1 = p_1$ to the lists, we conclude that the primes $p_1, p_2, \cdots, p_k$ is the same as the primes $p'_1, p'_2, \cdots, p'_{k'}$ up to permutation.

□

An immediate consequence of Proposition 4.4.3 is that any natural number has prime divisors. This leads to the following important result, for which the proof appeared in Euclid's The Elements.

Theorem 4.4.4. *There are infinitely many prime numbers.*

Proof. Suppose there are only finitely many prime numbers $p_1, p_2, \ldots, p_k$. Then consider the number $p = p_1 p_2 \cdots p_k + 1$. Since $p_1, p_2, \ldots, p_k$ is the list of *all* prime numbers, by Proposition 4.4.3, one of them, say p_i, must be a divisor of p. Since $p_1 p_2 \cdots p_k$ is also divisible by p_i, we conclude that p divides 1, a contradiction. Therefore the number of prime numbers must be infinite.

□

By combining the prime factors that appear repeatedly, Theorem 4.4.3 yields the expression

$$n = p_1^{e_1} p_2^{e_2} \cdots p_k^{e_k},$$

where $p_1, p_2, \ldots, p_k$ are distinct prime numbers and $e_i \in \mathbb{N}$. Since $p^0 = 1$ for any nonzero number p, we may also write

$$n = \prod_{\text{prime } p} p^e.$$

The non-negative integer $e = e_p(n)$ is the *exponent* of p in n. The exponent is zero for all except that finitely many primes that actually divide the number. For example, $60 = 2^2 \cdot 3 \cdot 5 \cdot 7^0 \cdot 11^0$, and $e_2(60) = 2$, $e_3(60) = 1$, $e_5(60) = 1$, $e_7(60) = e_{11}(60) = \cdots = 0$.

We have $e_p(1) = 0$ and define $e_p(0) = -\infty$ for all p. The exponent for zero is consistent with the expectation that $p^{-\infty} = 0$.

The exponent may be extended to integers by $e_p(a) = e_p(|a|)$ for $a \neq 0$. It can be further extended to rational numbers by $e_p\left(\dfrac{a}{b}\right) = e_p(a) - e_p(b)$. Then for nonzero rational numbers we have

$$r = \text{sign}(r) \prod_{\text{prime } p} p^{e_p(r)}.$$

Proposition 4.4.5. *The exponent has the following properties.*

1. *For $r, s \in \mathbb{Q}$, $e_p(rs) = e_p(r) + e_p(s)$.*
2. *For $r, s \in \mathbb{Q}$ satisfying $r + s \neq 0$, $e_p(r + s) \geq \min\{e_p(r), e_p(s)\}$.*
3. *For $a, b \in \mathbb{Z} - \{0\}$, $a \mid b \iff e_p(a) \leq e_p(b)$.*
4. *For $a_1, a_2, \ldots, a_k \in \mathbb{Z} - \{0\}$, $e_p(\gcd\{a_1, a_2, \ldots, a_k\}) = \min\{e_p(a_1), e_p(a_2), \ldots, e_p(a_k)\}$.*

The proof is left as an exercise.

Next we turn to polynomials. If a *non-constant* polynomial $p(t)$ satisfies

$$a(t) \mid p(t) \implies a(t) = \text{constant or } a(t) = rp(t) \text{ for some constant } r,$$

i.e., p is divisible *only* by constants and constant multiples of itself, then p is called an *irreducible polynomial.*

The allowance of the (nonzero) constant is due to the fact that for any number $r \neq 0$:

$$ra(t) \mid b(t) \iff a(t) \mid b(t) \iff a(t) \mid rb(t).$$

Therefore multiplying (and dividing) a nonzero constant does not change the divisibility properties of polynomials. In fact, this introduces some ambiguity among irreducible polynomials. We should think of an irreducible polynomial $p(t)$ and its constant multiple $rp(t)$ as essentially the same "prime polynomial". We will fix the ambiguity by additionally insisting that irreducible polynomials must have leading coefficient 1: $p(t) = t^k + p_{k-1}t^{k-1} + p_{k-2}t^{k-2} + \cdots$.

The issue with regard to the nonzero constant polynomials also appears for the prime numbers. For integers, multiplying -1 does not change the divisibility properties of integers:

$$-a \mid b \iff a \mid b \iff a \mid -b.$$

As a result, we should think of a prime number p and its negative $-p$ as of the same kind. Of course the ambiguity was solved by insisting prime numbers to be positive.

In general, the integers and the polynomials form algebraic systems called commutative monoid, in which products are defined and satisfy the usual properties. One may always try to introduce prime elements. However, the definition must allow the prime element to be multiplied by an "invertible" element (the elements for which reciprocals exist). Thus the elements x of the monoid are divided into the following mutually exclusive classes:

1. Invertible: There is y, such that $xy = 1 = yx$.
2. Prime/Irreducible: x is not invertible and $x = yz \implies y$ or z is invertible.
3. Composite: $x = yz$ with both y and z not invertible.

For integers, the invertibles are 1 and -1 (which is why 1 is not considered as a prime number). For polynomials, the invertibles are nonzero constants.

With similar proof, Propositions 4.4.1 and 4.4.2 hold for polynomials:

1. For an irreducible polynomial $p(t)$ and any nonzero polynomial $a(t)$, we have

   $$\gcd(a(t), p(t)) = \begin{cases} 1 & \text{if } p(t) \nmid a(t) \\ p(t) & \text{if } p(t) \mid a(t) \end{cases}.$$

   Note that the greatest common divisor is unique only up to multiplying a nonzero constant.

2. For an irreducible polynomial $p(t)$ and any nonzero polynomials $a(t), b(t)$, we have

   $$p(t) \mid a(t)b(t) \implies p(t) \mid a(t) \text{ or } p(t) \mid b(t).$$

Based on these and by inducting on the degree of polynomials, the extension to Theorem 4.3.2 to polynomials may be proved.

Theorem 4.4.6. *Any nonzero polynomial can be expressed as a product of a constant and irreducible polynomials. Moreover, the expression is unique up to permutation.*

Consider $t^4 - 1 = (t^2 + 1)(t^2 - 1) = (t^2 + 1)(t + 1)(t - 1)$. By $\deg(a(t)b(t)) = \deg a(t) + \deg b(t)$, it is easy to deduce that first degree polynomials $p(t) = t - t_0$ are irreducible. On the other hand, if $t^2 + 1$ were *reducible* (i.e., not irreducible), then $t^2 + 1 = (t - t_1)(t - t_2)$ must be a product of two degree one polynomials. In particular, we have $t_1^2 + 1 = (t_1 - t_1)(t_1 - t_2) = 0$ and similarly $t_2^2 + 1 = 0$. However, there is not *real* (or rational) number t_1 satisfying $t_1^2 + 1 = 0$. Therefore $t^2 + 1$ is also irreducible, and $t^4 - 1 = (t^2 + 1)(t^2 - 1) = (t^2 + 1)(t + 1)(t - 1)$ is a factorization into a product of irreducible polynomials.

Note that the impossibility of $t_1^2 + 1 = 0$ is due to the fact that we are restricted to real numbers only. If we are allowed to use complex numbers, then $t^2 + 1 = (t + i)(t - i)$, $i = \sqrt{-1}$, is a further factorization. Therefore $t^2 + 1$ is not irreducible *as a complex polynomial*, although it is irreducible as real or rational polynomials.

Thus the irreducibility of polynomials depend on the numbers allowed to be the coefficients. For another example, let us consider $t^2 + 2t - 2$. It the polynomial were reducible, then we must have $t^2 + 2t - 2 = (t - t_1)(t - t_2)$. This implies t_1 and t_2 are two roots of $t^2 + 2t - 2$. The roots are $-1 \pm \sqrt{3}$, which are real but not rational. Therefore $t^2 + 2t - 2$ is irreducible as a rational polynomial and is reducible as a real polynomial.

We may define the exponent of irreducible polynomials similar to the exponent of prime numbers. The definition can be extended to the quotient of polynomials, called *rational functions*. Proposition 4.4.5 still holds for polynomials.

Exercise 4.20. Prove that if $n > 1$ is a natural number such that $p \nmid n$ for any prime $p \leq \sqrt{n}$, then n is a prime number.

Exercise 4.21. Factor 1053, -390, 247, -500 into products of primes. Then use the result to find the greatest common divisor.

Exercise 4.22. Prove Proposition 4.4.5.

Exercise 4.23. Let n be a natural number.

1. Prove that $3 \mid n, 5 \mid n \implies 15 \mid n$.

2. Is it true that $3 \mid n, 20 \mid n \implies 60 \mid n$? Explain.

3. Is it true that $6 \mid n, 10 \mid n \implies 60 \mid n$? Explain.

Exercise 4.24. The following extends the fact that $\sqrt{2}$ is not a rational number.

1. Prove that a natural number is the square of another natural number if and only if all the exponents are even.

2. Extend the result of the first part to rational numbers.

3. Prove that for distinct primes p and q, $\sqrt{p}$ and $\sqrt{pq}$ are not rational numbers.

4. Extend the result of the third part to cube root and higher roots.

Exercise 4.25. Prove that the exponent of a prime p in an integer $a \neq 0$ is the biggest non-negative number such that $p^e \mid a$.

Exercise 4.26. Using proposition 4.4.5, do the following.

1. Prove that the product of any three consecutive natural numbers is divisible by 6.

2. Prove that the product of any four consecutive natural numbers is divisible by 24.

3. Prove that the product of any three consecutive odd natural numbers is divisible by 15.

4. What number can divide the product of any three consecutive even natural numbers? Explain.

Exercise 4.27. The following introduces a concept that is somehow the dual of greatest common divisor.

1. For nonzero integers $a_1, a_2, \dots, a_k$, let m be the natural number such that $e_p(m) = \max\{e_p(a_1), e_p(a_2), \dots, e_p(a_k)\}$. Prove that for any integer b,

$$a_1 \mid b, a_2 \mid b, \dots, a_k \mid b \iff m \mid b.$$

   The number m is the *least common multiple* and is denoted $\operatorname{lcm}(a_1, a_2, \dots, a_k)$.

2. Find the least common multiple of 1053, -390, 247, -500.

3. Prove that for any two nonzero natural numbers m, n, $\gcd(m,n)\operatorname{lcm}(m,n) = mn$.

4. Is the least common multiple changed by the modifications in Proposition 4.3.2.

Exercise 4.28. Prove Theorem 4.4.6.

Exercise 4.29. Consider the irreducibility of rational polynomials.

1. Factor t^6-1, t^4-2t+1, t^4+4 into products of irreducible rational polynomials.

2. Use the result above to find $\gcd(t^6-1, t^4-2t+1)$ and $\gcd(t^6-1, t^4-2t+1, t^4+4)$.

3. What is the condition for a degree two rational polynomial to be irreducible?

4. Prove that $t^4 + 3t^2 + 4$ is an irreducible rational polynomial.

Exercise 4.30. Consider the irreducibility of degree one polynomials.

1. Prove that any degree one polynomial $t - t_0$ is irreducible.

2. The *Fundamental Theorem of Algebra* says that any nonconstant complex polynomial must have complex roots. Use this to prove that any complex polynomial is a product of a constant and degree one polynomials.

3. Use the second part to prove that the only complex irreducible polynomials are the degree one polynomials.

Exercise 4.31. Consider the irreducibility of degree two polynomials.

1. Show that $t^2 + 2t - 3$ and $2t^2 + 4t - 3$ are reducible as real polynomials. Are they irreducible as rational polynomials?

2. Prove that a real degree two polynomial is irreducible if and only if its roots are real.

3. Prove that if $\alpha + i\beta$ is a complex root of a real polynomial, then the complex conjugate $\alpha - i\beta$ is also a root of the polynomial. Use this to prove that any real polynomial is a product of real polynomials of degree one or two.

4. What are the irreducible real polynomials?

4.5 Congruence

Let n be a fixed natural number. Two integers a, b are *congruent modulo* n, denoted

$$a \equiv b \pmod{n},$$

if $a - b$ is divisible by n. It is easy to verify that the congruence modulo n is an equivalence relation. The equivalence classes are called the congruence classes and denoted

$$\bar{a} = \{\dots, a - 2n, a - n, a, a + n, a + 2n, \dots\}.$$

The collection of all congruence classes modulo n is denoted $\mathbb{Z}/n\mathbb{Z}$, or $\mathbb{Z}_n$ for short. For example,

$$\mathbb{Z}_1 = \{\bar{0}\}$$

because modulo 1, $\bar{0} = \mathbb{Z}$ is the only equivalence class. Moreover,

$$\mathbb{Z}_2 = \{\bar{0}, \bar{1}\}$$

where modulo 2, $\bar{0} = 2\mathbb{Z}$ is the collection of all even integers and $\bar{1} = 2\mathbb{Z} + 1$ is the collection of all odd integers. In general, there are n congruence classes modulo n:

$$\mathbb{Z}_n = \{\bar{0}, \bar{1}, \dots, \overline{n-1}\}.$$

Define the sum and product of congruence classes by

$$\bar{a} + \bar{b} = \overline{a+b}, \quad \bar{a}\bar{b} = \overline{ab}.$$

Strictly speaking, the following needs to be verified in order for the operations to be well defined:

$$a_1 \equiv a_2, b_1 \equiv b_2 \pmod{n} \implies a_1 + b_1 \equiv a_2 + b_2, a_1 b_1 \equiv a_2 b_2 \pmod{n}.$$

The verification is left as an exercise. The following are some examples of the sum and product of congruence classes in $\mathbb{Z}_{16}$:

$$\begin{aligned}
\overline{13} + \overline{10} &= \overline{23} = \overline{1 \cdot 16 + 7} = \bar{7}.\\
\overline{13} - \overline{10} &= \bar{3}.\\
\overline{10} - \overline{13} &= \overline{-3} = \overline{(-1) \cdot 16 + 13} = \overline{13}.\\
\overline{13} \times \overline{10} &= \overline{130} = \overline{8 \cdot 16 + 2} = \bar{2}.\\
\overline{13} \times \bar{5} &= \overline{65} = \overline{5 \cdot 16 + 1} = \bar{1}.
\end{aligned}$$

Since the sum and the product in $\mathbb{Z}_n$ are derived form the similar operations for integers, the associativity, commutativity, and distributivity still holds for the operations of congruence classes.

It is also easy to see that the classes $\bar{0}$ and $\bar{1}$ are the unique classes that behave like 0 and 1 for integers:

$$\bar{a} + \bar{0} = \bar{a} = \bar{0} + \bar{0}, \quad \bar{a}\bar{1} = \bar{a} = \bar{1}\bar{0}.$$

Moreover, $\overline{-a}$ behaves like the negative of $\bar{a}$ with respect to the sum:

$$\bar{a} + \overline{-a} = \bar{0} = \overline{-a} + \bar{a}.$$

Therefore we denote $-\bar{a} = \overline{-a}$.

As far as the sum and product are concerned the system $\mathbb{Z}_n$ is very much like the integer system. However, there is a major difference between the two: There is no order for $\mathbb{Z}_n$. Note that instead of $n > 0$ in $\mathbb{Z}$, we have $\bar{n} = \bar{0}$ in $\mathbb{Z}_n$. Because of this, the following property for the product of integers

$$ab = 0 \iff a = 0 \text{ or } b = 0$$

no longer holds for the product of the congruence classes (note that the property was proved by making use of the order).

The lack of order makes $\mathbb{Z}_n$ appears to be "inferior" to $\mathbb{Z}$. However, $\mathbb{Z}_n$ has the advantage in the existence of reciprocal for many, if not all, of the congruence classes. For example, because $\overline{13} \times \bar{5} = \bar{1}$ in $\mathbb{Z}_{16}$, $\bar{5}$ is the reciprocal of $\overline{13}$, whereas 13 has no reciprocal in $\mathbb{Z}$. In general, we have the following result.

Proposition 4.5.1. *For $\bar{a} \in \mathbb{Z}_n$, there is $\bar{b}$ such that $\bar{a}\bar{b} = \bar{1}$ if and only if* $\gcd(a, n) = 1$.

Proof. If there is $\bar{b}$ such that $\bar{a}\,\bar{b} = \bar{1}$, then $ab - 1 = kn$ for some $k \in \mathbb{Z}$. Therefore if d divides a and n, then d divides 1. In other words, the only integers dividing both a and n are ±1. Thus $\gcd(a, n) = 1$.

Conversely, if $\gcd(a, n) = 1$, then by Theorem 4.3.3, we have $ua + vn = 1$ for some $u, v \in \mathbb{Z}$. This implies $\bar{a}\bar{u} = \overline{au} = \overline{1 - vn} = \bar{1}$. Thus $\bar{u}$ is the reciprocal of a. □

The only invertible class in $\mathbb{Z}_2$ is $\bar{1}$. The invertible classes in $\mathbb{Z}_3$ are $\bar{1}$ and $\bar{2}$, with $\bar{1}^{-1} = \bar{1}$ and $\bar{2}^{-1} = \bar{2}$. The following are the invertible congruence classes in $\mathbb{Z}_{12}$ and their inverses.

$$\bar{1}^{-1} = \bar{1}, \bar{5}^{-1} = \bar{5}, \bar{7}^{-1} = \bar{7}, \overline{11}^{-1} = \overline{11}.$$

The following are the invertible congruence classes in $\mathbb{Z}_{15}$ and their inverses.

$$\bar{1}^{-1} = \bar{1}, \bar{2}^{-1} = \bar{8}, \bar{4}^{-1} = \bar{4}, \bar{7}^{-1} = \overline{13}, \bar{8}^{-1} = \bar{2}, \overline{11}^{-1} = \overline{11}, \overline{13}^{-1} = \bar{7}, \overline{14}^{-1} = \overline{14}.$$

A collection of integers $a_1, a_2, \dots, a_k$ are said to be *coprime* if $\gcd(a_1, a_2, \dots, a_k) = 1$. In other words, the only natural number dividing all $a_1, a_2, \dots, a_k$ is 1. By Proposition 4.3.3, for the coprime integers $a_1, a_2, \dots, a_k$, we have

$$u_1a_1 + u_2a_2 + \cdots + u_ka_k = 1, \quad \text{for some integers } u_1, u_2, \dots, u_k.$$

Conversely, given the above, any natural number dividing all $a_1, a_2, \dots, a_k$ must also divide the combination, which is 1. Therefore the only common natural number divisor is 1, and $a_1, a_2, \dots, a_k$ are coprime.

Since the common divisors are not changed by the modifications in Proposition 4.3.2, the modifications also preserve the coprime property. In particular, the coprime property can be determined by the Euclidean algorithm (coprime if we evnetually get 1). The coprime property can also be seen from the exponent. By the last property in Proposition 4.4.5, we know $a_1, a_2, \dots, a_k$ are coprime if and only if any prime does not divide at least one of $a_1, a_2, \dots, a_k$.

By Proposition 4.5.1, the invertible congruence classes in $\mathbb{Z}_n$ are given by $\bar{a}$ with a and n coprime. It is easy to see that if $\bar{a}$ and $\bar{b}$ are invertible, then $\bar{a}^{-1}\bar{b}^{-1}$ is the inverse of $\overline{ab} = \bar{a}\bar{b}$. In particular, we have

$$a, n \text{ coprime}, extandb, n \text{ coprime} \implies ab, n \text{ coprime}.$$

Exercise 4.32. Find invertible elements in $\mathbb{Z}_8$ and compute the products between them.

Exercise 4.33. List all the invertible elements in $\mathbb{Z}_{30}$.

Exercise 4.34. Let p be a prime number. Prove that for any $a \in \mathbb{N}$ satisfying $0 < a < p$, $\bar{a} \in \mathbb{Z}_p$ is invertible.

Exercise 4.35. Let p and q be distinct prime numbers. Let e and f be natural numbers. How many invertible congruence classes are there in $\mathbb{Z}_{p^e q^f}$?

Exercise 4.36. Prove that $\bar{a} \in \mathbb{Z}_n$ has the property that $\bar{a}\bar{b} = \bar{0} \iff \bar{b} = \bar{0}$ if and only if $\bar{a}$ is invertible. Does the same statement hold with $\mathbb{Z}$ in place of $\mathbb{Z}_n$?

Exercise 4.37. Prove that $a_1, a_2, \ldots, a_k$ are coprime $\implies$ $a_1, a_2, \ldots, a_k, a_{k+1}$ are coprime.

Exercise 4.38. Prove that $a, a_1, a_2, \ldots, a_k$ are coprime, and $b, a_1, a_2, \ldots, a_k$ are coprime $\implies$ $ab, a_1, a_2, \ldots, a_k$ are coprime.

Exercise 4.39. Prove that $a_1, a_2, a_3, \ldots, a_k$ are coprime $\iff$ $\gcd(a_1, a_2), a_3 \ldots, a_k$ are coprime.

Chapter 5

Counting