Foundation of Mathematics

Min Yan

December 2, 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 opposite of a statement A is simply "not A". The statement "not A" is often further rephrased in a more familiar way, using the opposite words in English. For example, the following statements

- n is even.
- The sum of two odd numbers is even.
- $x > y$ (i.e., x is strictly bigger than y).
- A dog has four legs.

have the following opposite statements

- n odd.
- There are odd numbers m and n, such that $m + n$ is still odd.
- $x \leq y$ (i.e., x is strictly smaller than y, or x is equal to y).
- Some dog does not have four legs.

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

Exercise 1.3. Rewrite the following statements and their opposites into equivalent ones using mathematical symbol expressions.

1. n is an odd number.
2. $\sqrt{2}$ is an irrational number.
3. The square of an even number is even.
4. n 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.

Exercise 1.4. True or false.

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.

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.

Combining "x is positive", and "y is positive", and "z is positive", we get

- x, y, z are positive.

Combining "A dog has four legs" and "A dog has a tail", we 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, $n = 2, 3, 4, 6, 8, 9, \ldots$ fit the statement, while $n = 1, 5, 7, 11, 13, \ldots$ do not fit. Combining "x is positive", or "y is positive", or "z is positive", we get

- One of x, y, z is positive.

The following are other examples of disjunction statements

- An integer is either even or odd.
- A number is either positive, or negative or is zero.

Exercise 1.5. 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 divisible by 2 and 3 and 6.
7. n is not divisible by 3.
8. n is not divisible by 2 and not by 3.
9. n is not divisible by 2 or not by 3.

We know "A and B" means both A and B happen. Therefore the opposite of "A and B" is either A does not happen, or B does not happen

$$\text{not}(A \text{ and } B) = (\text{not } A) \text{ or } (\text{not } B).$$

Similarly, the opposite of "A or B" is A does not happen, and B does not happen

$$\text{not}(A \text{ or } B) = (\text{not } A) \text{ and } (\text{not } B).$$

For example, the opposites of

- m and n are even numbers.
- $x > 3$ and $x \leq 5$ (i.e., $3 < x \leq 5$).
- x is positive or zero.

- $x > y > z$ (i.e., $x > y$ and $y > z$).

are

- m or n is an odd number.
- $x \leq 3$ or $x > 5$.
- x is not positive and not zero (i.e., x is negative).
- $x \leq y$ (i.e., not $x > y$) or $y \leq z$ (i.e., not $y > z$).

The rule can be applied repeatedly. For example, the opposites of

- $x > 0$ and $y > 0$, or $x < 0$ and $y < 0$.

are

- $x \leq 0$ or $y \leq 0$, and $x \geq 0$ or $y \geq 0$.

Example 1.1.1. The following statement basically says m and n have the same parity

- m and n are even, or m and n are odd.

The opposite statement is

- m or n is odd, and m or n is even.

The opposite statement is a bit hard to understand. "m or n is odd" means one of m, n is odd. "m or n is even" means one of m, n is odd. Therefore the opposite statement is equivalent to

- one of m, n is odd, and one of m, n is even.

Since we only have two integers, the statement is the same as m, n have different parity.

Example 1.1.2. The opposite of

- $x > 0$ or $y > 0$, and $x < 0$ or $y > 0$.

is

- $x \leq 0$ and $y \leq 0$, and $x \geq 0$ or $y \leq 0$.

The statement "$x \leq 0$ and $y \leq 0$" means the third quadrant of the plane $\mathbb{R}^2$, including the boundary. The statement "$x \geq 0$ and $y \geq 0$" means the first quadrant, also including the boundary. The two statements together means the intersection of the two quadrants, which is exactly the origin. Therefore the opposite is the origin

Figure 1.1: $x > 0$ or $y > 0$, and $x < 0$ or $y > 0$; and the opposite.

- $x = 0$ and $y = 0$.

Example 1.1.3. The following statement

- ($x < 1$ or ($x > 2$ and $x < 4$)) and $x < 3$.

means x satisfy both of the following two conditions

- $x < 1$, or $x > 2$ and $x < 4$.
- $x < 3$.

The first condition means the union of the two intervals $(-\infty, 1)$ and $(2, 4)$. The second means the interval $(-\infty, 3)$. The whole statement means the union of the two intervals $(-\infty, 1)$ and $(2, 3)$. This is the following statement

- $x < 1$ or $2 < x < 3$.

  The opposite statement is

- ($x \ge 1$ and ($x \le 2$ or $x \ge 4$)) or $x \ge 3$.

This means x satisfy one of the following two conditions

- $x \ge 1$, and $x \le 2$ or $x \ge 4$.
- $x \ge 3$.

The first condition means the union of $[1, 2]$ and $[4, \infty)$. The second condition means $[3, \infty)$. To say one of the conditions is satisfied means the union of $[1, 2]$ and $[3, \infty)$. This is the following statement

- $1 \le x \le 2$ or $x \ge 3$.

Exercise 1.6. 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.
3. There are no integers m and n, such that $10 = m^2 + n^2$.
4. $x > y$ or $x < y$.
5. l, m, n are odd, or l, m, n are even.
6. one of l, m, n is odd, and one of l, m, n is even.
7. A non-negative number has either one or two square roots.
8. $x > y > z$, or $x < y < z$.
9. $x > y$, and $x > z$, and $y > z$.
10. $x > y$, or $x > z$, or $y > z$.
11. $x > y$, and ($x > z$ or $y > z$).

Exercise 1.7. True or false.

1. If A is equivalent to B, and C is equivalent to D, then "A and C" is equivalent to "B and D".
2. If A is equivalent to B, and C is opposite to D, then "A and C" is equivalent to "B or D".
3. If A is opposite to B, and C is opposite to D, then "A and C" is opposite to "B and D".
4. If A is opposite to B, and C is opposite to D, then "A or C" is opposite to "B and C".

Exercise 1.8. Which are the same? Which are opposite?

1. A and (B and C).
2. A and (B or C).
3. A or (B and C).
4. A or (B or C).
5. (A and B) and C.
6. (A and B) or C.
7. (A or B) and C.
8. (A or B) or C.
9. (A and B) and (A and C).
10. (A and B) or (A and C).
11. (A or B) and (A or C).
12. (A or B) or (A or C).

Exercise 1.9. Which are the same? Which are opposite?

1. A and (not B).
2. A or (not B).
3. (not A) and B.
4. (not A) or B.
5. (not A) and (not B).
6. (not A) or (not B).
7. not (A and B).
8. not (A or B).

1.2 Implication

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

- A: n is divisible by 10.

then the statement

- B: n is an even integer.

is also true. We say the first statement *implies* the second statement, and denote $A \Longrightarrow B$ or $A \Longleftarrow B$. The implication can be made into the following statements

- If n is divisible by 10, then n is even.
- n is even, if n is divisible by 10.

In other words, we can say "if A, then B", or "B, if A".

For another example, the statement

- Bob studies hard.

implies the statement

- Bob gets good grade.

The whole implication can be made into the following statements

- Bob gets good grade because he studies hard.
- Bob gets good grade, if he studies hard.

Exercise 1.10. Determine which one implies which one, and write the implication as "if A, then B".

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.

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

The word "if" means certain condition happens, and is similar to "when". Let us compare the following two if statements

- B, if A: good grade, if study hard.

- B, only if A: good grade, only if study hard.

where the only difference is "only".

The first statement means that, when study hard happens, then good grade also happens. In other words, study hard is enough (or sufficient) for getting good grade. This does not exclude other ways to get good grade, such as cheating. In general, in saying "B, if A" (or "if A, then B"), we mean A is one situation that B happens, and this allows the possibility of some other situation that B also happens. Therefore we say A is a *sufficient condition* for B.

The second statement means that, good grade happens only when study hard happens. Here "only" emphasizes that study hard is the only situation good grade happens. This excludes all other ways to get good grade. In other words, if we see somebody gets good grade, then the person must (or necessarily) study hard. In general, in saying "B, only if A", we mean A is the only situation that B happens, and no other situation makes B happen. Therefore we have "if B, then A", and we say A is a *necessary condition* for B.

In the following statements,

- If n is divisible by 10, then n is even.

- $m + n$ is even, if m and n are odd.

- n is even, only if n^2 is even.

- Given the rain, we bring umbrellas.

- Bigger investment implies bigger return.

- We take medicine if we get ill.

- We take medicine only if we get ill.

"n is divisible by 10", "m and n are odd", "n is even", "rain", "bigger investment", "get ill", "take medicine" are sufficient conditions, and "n is even", "$m + n$ is even", "bring umbrellas", "bigger return", "take medicine", "get ill" are necessary conditions.

By common sense, we know the equivalence of A and B means $A \implies B$ and $B \implies A$. Therefore we may write $A \iff B$ for equivalent statements. We also say "A if and only if B", or "A is necessary and sufficient for B". Here are some examples:

- Even is equivalent to multiple of 2.
- $m > n$ is the necessary and sufficient condition for $-m < -n$.
- An integer is a multiple of 6 if and only if 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.

Exercise 1.11. Rewrite the statements into "A, if B" and "A, only if B". Moreover, determine the sufficient condition and the necessary condition.

1. Even plus even is even.
2. Positive multiplying positive is positive.
3. If $x = 2$, then $x^2 - 3x + 2 = 0$.
4. If $x^2 - 3x + 2 = 0$, then $x = 1$ or 2.
5. It is getting dark. Therefore we turn on the light.
6. To become healthy, we must exercise.
7. If you exercise, then you will be healthy.
8. We will go when we are ready.
9. We will go only when we are ready.
10. No water, no fish.

Suppose $A \implies B$. Then when B does not happen, we know A cannot happen. Therefore we get (not B) $\implies$ (not A). In other words, we know "$A \implies B$" implies "(not B) $\implies$ (not A)". Then "(not B) $\implies$ (not A)" implies "(not(not A)) $\implies$ (not(not B))", which is $A \implies B$. We conclude that $A \implies B$ is the same as (not B) $\implies$ (not A).

We say (not B) $\implies$ (not A) is the *contrapositive* of $A \implies B$. For example, the statements

- If n is divisible by 10, then n is an even integer.
- If n^2 is even, then n is even.
- If we are healthy, then we will not take medicine.
- If $x = 2$, then $x^2 - 3x + 2 = 0$.
- If you study hard, then you get good grade.
- If we are ill, then we will take medicine.

are equivalent to the following contrapositives

- If a number is odd, then it is not divisible by 10.
- If n is odd, then n^2 is odd.
- If $x^2 - 3x + 2 \neq 0$, then $x \neq 2$.
- If your grade is poor, then you do not study hard.
- We take medicine only if we are ill.
- If we are ill, then we take medicine.

Exercise 1.12. State the contrapositives of the statements in Exercise 1.11.

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:

- An integer is even or odd.
- If an integer is not even, then it is odd.
- If an integer is not odd, then it is even.

The following are also equivalent:

- If $m + n$ is odd, then either m is odd, or n is odd.
- If $m + n$ is odd, and m is even, then n is odd.
- If $m + n$ is odd, and m is odd, then n is even.

In contrast, we know "$A \implies (B$ and $C)$" is the same as "$A \implies B$, and $A \implies C$".

Exercise 1.13. Rephrase the statements.

1. If $xy > 0$, then $x > 0$ and $y > 0$, or $x < 0$ and $y < 0$.
2. If $x^2 - 3x + 2 = 0$, and $x \neq 1$, then $x = 2$.
3. A (real) number is positive, or negative, or zero.
4. A person is male or female.
5. If it is too expansive, then we get more money or not buy it.
6. If we get ill and cannot find a doctor, then we take medicine by ourselves.
7. If we want to travel, and there is no train and no plane, then we take bus.

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 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 $x_m > x_n$.
- If you study hard, then you get good grade.
- If it rains, then we bring umbrellas.

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 $x_{m_0} \leq x_{n_0}$.
- You may study hard but still get bad grade.
- We do not bring umbrellas even in case of rain.

Exercise 1.14. 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 1.15. 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 1.16. True or false.

1. "$A \iff B$" is the same as "($A \implies B$) and ((not A) $\implies$ (not B))".
2. "($A \implies B$) and ($A \implies C$)" is the same as "$A \implies$ (B and C)".
3. "($A \implies C$) and ($B \implies C$)" is the same as "(A and B) $\implies C$".
4. "(A and B) $\implies C$" implies "$A \implies C$".
5. "(A or B) $\implies C$" implies "$A \implies C$".
6. "$A \implies$ (B and C)" implies "$A \implies B$".
7. "$A \implies$ (B or C)" implies "$A \implies B$".

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

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

10. "(A and B and C) $\implies$ D" is the same as "(A and B) $\implies$ ((not C) 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 quantifier (all, every, any) and existential quantifier (some, there is, at least one).

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

- For all even n, n^2 is even.
- For all x, x^2 is non-negative.
- For every rational number r, there are integers m and n, such that $r = \dfrac{m}{n}$.
- 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.

Note that we also say "if n is even, then n^2 is even". The universal quantifier is implicit in the statement. For more example, the 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.

Exercise 1.17. Determine which statement is true.

1. There is an integer k such that $10 = 3k + 1$.
2. $20 = 3k + 1$ for some integer k.
3. For every rational number r, $2r$ is an integer.
4. For some rational number r, $2r$ is an integer.
5. $(m + n)^2 = m^2 + 2mn + n^2$ for any numbers m and n.
6. $(m + n)^2 = m^2 + n^2$ for some nonzero numbers m and n.
7. $(m + n)^3 = m^3 + n^3$ for some nonzero numbers m and n.

A statement may involve several quantifiers.

Example 1.3.1. The statement

- For any x, there is y, such that $x < y$.

has universal quantifier for x, followed by existential quantifier for y. By taking $y = x + 1$, we indeed get $x < y$. Therefore the statement is correct.

If we exchange the quantifiers, then we get

- There is x, such that for any y, we have $x < y$.

The statement means x is the lower bound for arbitrary number y. By taking $y = x - 1$, we know the statement is wrong.

Example 1.3.2. By $\lim \frac{1}{n} = 0$, we mean the following: For any number $\epsilon > 0$, there is a number N, such that $n > N$ implies $\frac{1}{n} < \epsilon$.

The following verifies the statement: For any $\epsilon > 0$, take $N = \frac{1}{\epsilon}$. By $\epsilon > 0$, we get $N > 0$. Then $n > N = \frac{1}{\epsilon}$ implies $n > 0$. By $n, \epsilon > 0$, we know $n > \frac{1}{\epsilon}$ further implies $\epsilon > \frac{1}{n}$.

The choice of $N = \frac{1}{\epsilon}$ in the verification is inspired by $\frac{1}{n} < \epsilon \iff n > \frac{1}{\epsilon}$. In fact, we only need $\frac{1}{n} < \epsilon \Longleftarrow n > N$ (the $\Longrightarrow$ direction is unnecessary). We know $n > \frac{2}{\epsilon}$ implies $\frac{1}{n} < \frac{\epsilon}{2} < \epsilon$. Therefore we may also take $N = \frac{2}{\epsilon}$ in the verification.

Example 1.3.3. A function $f(x)$ is continuous at a, if for any $\epsilon > 0$, there is $\delta > 0$, such that $|x - a| < \delta$ implies $|f(x) - f(a)| < \epsilon$.

The function is continuous, if it is continuous at all a. By adding "for all a" to the definition of continuity at a above, we get the following definition of continuous function: For any a and any $\epsilon > 0$, there is $\delta > 0$, such that $|x - a| < \delta$ implies $|f(x) - f(a)| < \epsilon$.

Note that we combine 'for all a" and "for any $\epsilon > 0$", and write "for any a and any $\epsilon > 0$".

Exercise 1.18. Determine which statement is true.

1. For any integer k, we have $10 = 3k + 1$.
2. There is an integer k, such that $10 = 3k + 1$.
3. For any integer n, there an integer k, such that either $n = 2k$ or $n = 2k + 1$.
4. There is an integer n, such that for any integer k, we have $n = 2k$ or $n = 2k+1$.
5. There is $\epsilon > 0$, such that for any N, we have $n > N$ implies $\frac{1}{n} < \epsilon$.
6. There is $\epsilon > 0$, such that for any $N > 1$, we have $n > N$ implies $\frac{1}{n} < \epsilon$.
7. For any $\epsilon > 0$, there is N, such that $n \geq N$ implies $\frac{1}{n} < \epsilon$.

Exercise 1.19. Determine whether the statement is true. Prove the true one and explain the false one.

1. For any $0 \leq x < 1$, there is $0 \leq y < 1$, such that $x < y$.
2. There is $0 \leq x < 1$, such that for any $0 \leq y < 1$, we have $x < y$.
3. There are $0 \leq x < 1$ and $0 \leq y < 1$, such that $x < y$.
4. For any $0 \leq x < 1$ and $0 \leq y < 1$, we have $x < y$.

Next we discuss the opposite of quantified statements. The opposite of

- For all n, $x_n > y_n$.

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 choice of n. Therefore the opposite statement is

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

You may replace n_0 by n. Here we use n_0 is emphasize one choice.

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. For example, the following are opposite:

- 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:

- 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 $m < n$.
- There are $m \neq n$ satisfying $m^2 = n^2$.
- There is b, such that all $x \in X$ satisfies $x \leq b$ (X is bounded above).
- For any x and $\epsilon > 0$, there is a rational number r, such that $|x - r| < \epsilon$.

are

- There are m and n, such that $m \leq n$, and $m \neq n$, and $m \geq n$.
- For all $m \neq n$, we have $m^2 \neq n^2$.
- For all b, there is $x \in X$ satisfying $x > b$ (X is unbounded above).
- There are x and $\epsilon > 0$, such that for any rational number r, we have $|x - r| \geq \epsilon$.

Example 1.3.4. The opposite of the first statement in Example 1.3.1 is

- There is x, such that for any y, we have $x \geq y$.

By taking $y = x + 1$, we see the statement is wrong. In fact, this has to be wrong because its opposite statement in Example 1.3.1 is correct.

The opposite of the second statement in Example 1.3.1 is

- For any x, there is y, such that $x \geq y$.

By taking $y = x - 1$, we find the statement to be correct. In fact, this is the same reason that the second statement in Example 1.3.1 is wrong.

Example 1.3.5. A sequence x_n converges to l, if for any $\epsilon > 0$, there is N, such that for any $n > N$, we have $|x_n - l| < \epsilon$.

For the sequence not converging to l, we need to change "for any $\epsilon > 0$" to "for some $\epsilon > 0$", change "there is N" to "for any N", change "for any $n > N$" to "there is $n > N$", and change "$|x_n - l| < \epsilon$" to "$|x_n - l| \geq \epsilon$". The whole definition is the following: A sequence x_n does not converge to a, if there is $\epsilon > 0$, such that for all N, there is $n > N$ satisfying $|x_n - l| \geq \epsilon$.

We remark that x_n not converging to l means two possibilities: (1) x_n converges to k, and $k \neq l$; (2) x_n does not converge to any k, i.e., x_n diverges.

Therefore x_n diverges if for any k, we know x_n does not converge to k. We replace "x_n does not converge to k" by the opposite statement we found above, and get the following definition: A sequence x_n diverges, if for any k, there is $\epsilon > 0$, such that for all N, there is $n > N$ satisfying $|x_n - k| \geq \epsilon$.

Example 1.3.6. We defined the continuity of $f(x)$ at a in Example 1.3.3. The function is not continuous at a, if there is $\epsilon > 0$, such that for any $\delta > 0$, the implication "$|x - a| < \delta \implies |f(x) - f(a)| \geq \epsilon$" is wrong. The opposite of the implication is that there is x satisfying $|x - a| < \delta$ and $|f(x) - f(a)| \geq \epsilon$. Therefore we get the full definition: A function $f(x)$ is discontinuous at a, if there is $\epsilon > 0$, such that for any $\delta > 0$, there is x satisfying $|x - a| < \delta$ and $|f(x) - f(a)| \geq \epsilon$.

The function is not continuous, if it is discontinuous at some a. Here is the full definition: A function $f(x)$ is not continuous, if there is a and $\epsilon > 0$, such that for any $\delta > 0$, there is x satisfying $|x - a| < \delta$ and $|f(x) - f(a)| \geq \epsilon$.

Exercise 1.20. A function $f(x)$ is uniformly continuous, if for any $\epsilon > 0$, there is $\delta > 0$, such that $|x - y| < \delta$ implies $|f(x) - f(y)| < \epsilon$. State the definition that the function is not uniformly continuous.

Exercise 1.21. A sequence x_n is Cauchy, if for any $\epsilon > 0$, there is N, such that $m, n > N$ implies $|x_m - x_n| < \epsilon$. State the definition that the sequence is not Cauchy.

Exercise 1.22. 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. You can fool some of the people all of the time.

4. Any people makes some mistake some of the time.

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

6. In every lecture, some students do not show up.

7. All students in the class love mathematics.

8. All students in the class are 19 years old and born in Hong Kong.

9. Some students in the class are from Beijing or Shanghai.

Exercise 1.23. 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.

1.4 Proof

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.

The following are straightforward proofs.

Example 1.4.1. We prove the following statement: The sum of two consecutive integers is odd.

The two consecutive integers are n and $n+1$. Then $n+(n+1) = 2n+1$ is odd.

Exercise 1.24. Prove the following. Write down all the details, with all the key words included.

1. n is even if and only if $n+3$ is odd.
2. For any integer n, $n(n+1)$ is even.
3. If n is odd, then $n^2 = 8m+1$ for some integer m.

Example 1.4.2. 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.

Exercise 1.25. For $m \le n$, prove

$$m + (m+1) + (m+2) + \cdots + (n-1) + n = \frac{(n-m+1)(m+n)}{2}.$$

A legendary story in mathematics is that Gauss[1] 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.

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

The statement in Example 1.4.2 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, $n(2n-5) \geq 0$.

is true for all $n > 2$. However, the statement does not hold for $n = 1$ and $n = 2$. The single counterexample for the case $n = 1$ already 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*.

Example 1.4.3. We prove the following statement: n is even if and only if n^2 is even.

Suppose n is even. Then $n = 2k$ for some integer k. Then $n^2 = 4k^2 = 2 \cdot 2k^2$ is even.

To prove the converse, that n^2 even implies n even, we prove the equivalent contrapositive, that n is odd implies n^2 is odd. Suppose n is odd. Then $n = 2k+1$ for some integer k. Then $n^2 = 4k^2 + 4k + 1 = 4(k^2 + k) + 1$ is odd.

Example 1.4.4. We prove the following statement: If n^3 is divisible by 5, then n is divisible by 5.

We prove the equivalent statement, that n not divisible by 5 implies n^3 not divisible by 5. Suppose n is not divisible by 5. We have $n = 5k + r$, for integers k and r satisfying $r = 1, 2, 3, 4$. Then

$$n^3 = (5k+r)^3 = 5^3k^3 + 3 \cdot 5^2k^2r + 3 \cdot 5kr^2 + r^3 = 5m + r^3,$$

where $m = 25k^3 + 15k^2r + 3kr^2$. Moreover, we know

$$r^3 = \begin{cases} 1, & r = 1 \\ 8, & r = 2 \\ 27, & r = 3 \\ 64, & r = 4 \end{cases}$$

is not divisible by 5. This implies $n^3 = 5m + r^3$ is not divisible by 5.

The proofs in Examples 1.4.3 can also be done by using the fact that 2 and 5 are prime numbers. However, this makes use of the theory of prime numbers, which takes some effort to develop in Section 4.4. Moreover, the statements in the examples are also true for non-primes such as 6.

Exercise 1.26. Prove the following. Write down all the details, with all the key words included.

1. n is divisible by 3 if and only if n^2 is divisible by 3.
2. n is divisible by 3 if and only if n^2 is divisible by 9.
3. n is even if and only if n^3 is even.
4. If n^2 is divisible by 6, then n is divisible by 6.
5. $n^2 + 1$ is not divisible by 6.

Example 1.4.5. We prove the following statement: If n is divisible by 2 and 3, then n is divisible by 6. (The reverse direction is easy.)

Any integer n can be expressed as $n = 6k + r$, where k, r are integers satisfying $0 \le r \le 5$.

Since n is divisible by 2, we know $n = 2p$ for some integer p. Then $r = n - 6k = 2(p - 3k)$ is even. By $0 \le r \le 5$, we get $r = 0, 2, 4$.

Since n is divisible by 3, we know $n = 3q$ for some integer q. Then $r = n - 6k = 3(q - 2k)$ is divisible by 3. By $r = 0, 2, 4$, this implies $r = 0$. Therefore $n = 6k$ is divisible by 6.

Exercise 1.27. Prove the following. Write down all the details, with all the key words included.

1. n is divisible by 35 if and only if n is divisible by 5 and 7.
2. n is divisible by 30 if and only if n is divisible by 2, 3 and 5.
3. n is divisible by 30 if and only if n is divisible by 6 and 15.

One strategy of proving a statement A is to starts 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 1.4.6. We prove $\sqrt{2}$ is an irrational number[2].

Suppose $\sqrt{2}$ is a rational number. Then $\sqrt{2} = \dfrac{m}{n}$ for some natural numbers m and n. If both m and n are even, then we may write $\sqrt{2} = \dfrac{m/2}{n/2}$, where $m/2, n/2$ are

[2]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.

still natural numbers, but are strictly smaller than m, n. By applying the process repeatedly, we eventually get $\sqrt{2} = \dfrac{m}{n}$, in which either m or n is odd[3].

Therefore we may assume $\sqrt{2} = \dfrac{m}{n}$, in which either m or n is odd. Then $2n^2 = m^2$. Therefore m^2 is even. By Example 1.4.3, this implies m is even. Then $m = 2k$ for some k. Then $2n^2 = m^2 = 4k^2$ implies $n^2 = 2k^2$. Therefore n^2 is even. By Example 1.4.3 again, this implies n is even. We conclude that both m and n are even, which contradicts the 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.

Exercise 1.28. Prove the following. Write down all the details, with all the key words included.

1. $\sqrt[3]{5}$ is irrational.
2. $\sqrt[3]{2}$ is irrational.
3. $\sqrt{6}$ is irrational.
4. $\sqrt{\frac{3}{2}}$ is irrational.
5. $\sqrt{2} + \sqrt{3}$ is irrational.

Exercise 1.29. Prove the following. Write down all the details, with all the key words included.

1. For integers m and n, mn is odd if and only if both m and n are odd.
2. For any integer n, $n^3 - n$ is a multiple of 6.
3. The square of any integer is either of the form $4k$ or the form $4k + 1$, where k is an integer.
4. The sum of the square of three consecutive integers cannot be of the form $12k - 1$ for some integer k.

---

[3]There is a bit jump in the argument. More detailed argument is the following: Assume m, n are even. Then $\sqrt{2} = \frac{m'}{n'}$ for natural numbers $m' = \frac{m}{2} < m$ and $n' = \frac{m}{2} < n$. If m', n' are still even, then we may repeat the process. Since each repeat produces strictly smaller m, n, the process stops after finitely many steps. The stop of the process means that the assumption m, n are even is not true. In other words, we eventually get $\sqrt{2} = \frac{m}{n}$ in which one of m, n is odd.

The most rigorous argument actually uses induction. You are not required to write any of these details. The details in the example is enough.

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 1.4.2 by induction.

By

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

the equality is verified for $n = 1$.

Next, assume the equality 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)

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

Exercise 1.30. Prove the following. Write down all the details, with all the key words included.

1. $1 + 3 + 5 + \cdots + (2n-1) = n^2$.

2. $1 \cdot 2 + 2 \cdot 3 + 3 \cdot 4 + \cdots + (n-1)n = \frac{1}{3}n(n-1)(n+1)$.
3. $1^2 + 2^2 + 3^2 + \cdots + n^2 = \frac{1}{6}n(n+1)(2n+1)$.
4. $1^2 + 3^2 + 5^2 + \cdots + (2n-1)^2 = \frac{1}{3}n(2n-1)(2n+1)$.
5. $1^3 + 2^3 + 3^3 + \cdots + n^3 = \frac{1}{4}n^2(n+1)^2$.

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.

The second step for the induction can also be changed to "if $A(n)$ is true, then $A(n+1)$ is also true". For example, we may assume $2^n > n^2$. Then $2^{n+1} = 2 \cdot 2^n > 2n^2$. By $2n^2 - (n+1)^2 = n^2 - 2n - 1 = (n-1)^2 - 2 > 0$ for $n \geq 5$, we get $2^{n+1} > (n+1)^2$.

Exercise 1.31. Find N, such that the inequality holds for $n \geq N$. Then prove the inequality. Write down all the details, with all the key words included.

1. $3^n > n^2$.
2. $2^n > n^3$.
3. $2^n > 2n^2$.

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 $n \geq 2$.

Exercise 1.32. For $x > y > 0$, prove $x^n - y^n \geq n(x-y)y^{n-1}$.

Exercise 1.33. Prove the sum of internal angles in an n-sided polygon is $(n-2)\pi$.

Exercise 1.34. For any odd number n, prove that $n^4 - 1$ is divisible by 16. In fact, $n^{2^k} - 1$ is divisible by 2^{k+2}.

Exercise 1.35. Let a sequence a_n be defined by $a_0 = 1$, $a_1 = 2$, and $a_{n+2} = 3a_{n+1} - 2a_n$.

1. Compute a_2, a_3, a_4, a_5 and make a guess on the general formula for a_n.
2. Prove the general formula.

Exercise 1.36. The Fibonacci sequence a_n is defined by $a_0 = 0$, $a_1 = 1$, and $a_{n+2} = a_{n+1} + a_n$. Prove that

$$a_n = \frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^n - \frac{1}{\sqrt{5}}\left(\frac{1-\sqrt{5}}{2}\right)^n.$$

Exercise 1.37. Prove that the Fibonacci sequence satisfies the following equalities:

1. $a_1 + a_2 + a_3 + \cdots + a_n = a_{n+2} - 1$.
2. $a_1 - a_2 + a_3 - \cdots + (-1)^n a_{n+1} = (-1)^n a_n + 1$.
3. $a_1 + a_3 + a_5 + \cdots + a_{2n-1} = a_{2n} - 1$.
4. $a_1^2 + a_2^2 + a_3^2 + \cdots + a_n^2 = a_n a_{n+1}$.

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 first step

- $A(1)$ is true.

Taking $n = 2$ in the second step, the only natural number $k < 2$ 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 statement holds for all natural numbers $< n$. Now consider two possibilities for n.

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

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

Example 1.5.5. The following fact appeared in Example 1.4.6: Any rational number is of the form $\frac{m}{n}$, where m, n are integers, and one of m, n is odd. The proof earlier is not the most rigorous. The rigorous proof can be given by induction.

Suppose $r = \frac{m}{n}$, where m, n are integers. By $r = \frac{-m}{-n}$, we may further assume that n is a natural number. We use induction on n to prove that $r = \frac{m'}{n'}$ for some integers m', n', one of which is odd.

For $n = 1$, we have $r = \frac{m}{1}$, and one of $m, 1$ is odd. The statement is verified.

Suppose the statement is true for all natural number denominators $< n$. For $r = \frac{m}{n}$, we consider two possibilities.

If one of m, n is odd, then $r = \frac{m}{n}$ verifies the statement.

If both m, n are even, then $m = 2m'$ and $n = 2n'$ for some integer m' and natural number n'. Then $r = \frac{m'}{n'}$. By $n' < n$ and the inductive assumption, we get $r = \frac{m''}{n''}$, where one of m'', n'' is odd. Therefore the statement is also verified.

Exercise 1.38. For any natural number n, prove $n = 10^k m$ for some integer $k \geq 0$ and integer m not divisible by 10.

Exercise 1.39. Prove that any rational number is of the form $\frac{m}{n}$, where m, n are integers, and one of m, n is not divisible by 3.

Exercise 1.40. Use Example 1.5.4 to prove the statement in Example 1.5.5.

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 1.5.6. 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, induct on n, which means the following two steps:

1. $N(m,1)$ is divisible $1!$.

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

The first step is trivial because any integer is divisible by $1! = 1$. For the second step, we relate $N(m,n)$ and $N(m,n-1)$

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

The inductive assumption is that $N(m,n-1)$ is divisible by $(n-1)!$. Therefore $N(m,n-1)n$ is divisible by $(n-1)!n = n!$. To prove $N(m,n)$ is divisible by $n!$, therefore, all we need is that $N(m-1,n)$ is divisible by $n!$. However, the precess of going from $N(m-1,n)$ to $N(m,n)$ can be achieved by inducting on m.

Therefore we set up the following double induction proof. The following is the formal proof.

We first induct on m. For $m = 1$, we have $N(1,n) = n!$ divisible by $n!$ (for all n). Next assume $N(m-1,n)$ is divisible by $n!$ (for the given m and all n), and we want to prove $N(m,n)$ is divisible by $n!$. For this purpose, we further induct[4] on n.

[4]Remember the second induction is under the first inductive assumption, that $N(m-1,n)$ is divisible by $n!$. Here m is already fixed in the first induction.

For $n = 1$, we have $N(m, 1) = m$, which is indeed divisible by $1! = 1$. Next assume $N(m, n-1)$ is divisible by $(n-1)!$. Then

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

By the first inductive assumption, we know $N(m-1, n)$ is divisible by $n!$. By the second inductive assumption, we know $N(m, n-1)n$ is divisible by $(n-1)!n = n!$. Therefore $N(m, n)$ is divisible by $n!$.

This completes the second induction on n, and proves the second part of the first induction. Therefore the first induction on m is also complete.

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

1. First prove $A(1, n)$.
2. Next, under the assumption $A(m-1, n)$ for arbitrarily given m and all n, we want to prove $A(m, n)$.
3. Prove $A(m, 1)$ for any m (for the problem in the step 2, start induction on n).
4. Under the assumption of $A(m-1, n)$ and $A(m, n-1)$, prove $A(m, n)$ (second step of induction on n).

There are various other ways of doing double inductions.

The proof in Example 1.5.6 seems rather complicated. The first half of the example is the analysis of the problem. It is always very important to analyze your problem before writing proof. In the analysis, you always need 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 1.41. Prove the statement in Example 1.5.6 by inducting on $m+n$, starting with $m + n = 2$.

Exercise 1.42. Prove that the multiplication of $n_1 + n_2 + n_3$ consecutive integers is divisible by $n_1!n_2!n_3!$. In general, the multiplication of $n_1 + n_2 + \cdots + n_k$ consecutive integers is divisible by $n_1!n_2! \cdots n_k!$.

Exercise 1.43. Can you prove the statement in Example 1.5.6 by reversing m and n in the double induction?

Exercise 1.44. 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.
- Real polynomials $\mathbb{R}[t] = \{a_0 + a_1 t + a_1 t^2 + \cdots + a_n t^n \colon a_i \in \mathbb{R}\}$.
- 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 $\mathbb{N} = \{n \colon n \text{ is obtained by repeatedly adding 1 to itself}\}$.
- Rational numbers $\mathbb{Q} = \{r \colon r = \dfrac{a}{b} \text{ for some integers } a, b\}$.
- Prime numbers $\{p \in \mathbb{N} \colon p \mid mn \text{ for } m, n \in \mathbb{N} \text{ implies } p \mid m \text{ or } p \mid n\}$.
- Irreducible numbers $\{p \in \mathbb{N} \colon mn \mid p \text{ for } m, n \in \mathbb{N} \text{ implies } m = 1 \text{ or } n = 1\}$.

- Open interval $(a, b) = \{x \colon a < x < b\}$.
- Closed interval $[a, b] = \{x \colon a \leq x \leq b\}$.
- Unit sphere $S^2 = \{(x_1, x_2, x_3) \colon x_1^2 + x_2^2 + x_3^2 = 1\}$ in $\mathbb{R}^3$.
- Unit ball $B^3 = \{(x_1, x_2, x_3) \colon x_1^2 + x_2^2 + x_3^2 \leq 1\}$ in $\mathbb{R}^3$.
- $\{1, 2, 3, \ldots, n\} = \{x \colon x \in \mathbb{N}, x \leq n\}$.
- $\{3, -2\} = \{x \in \mathbb{R} \colon x^2 - x - 6 = 0\}$.

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 B^n in $\mathbb{R}^n$.
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. $\{(x, y) \colon x = 0\}$.
2. $\{(x, y) \colon x = y\}$.
3. $\{(x, y) \colon x^2 + y^2 = 4\}$.
4. $\{(x, y) \colon x^2 + y^2 > 4\}$.
5. $\{(x, y) \colon y < 0\}$.
6. $\{(x, y) \colon x^2 + 4y^2 = 4\}$.
7. $\{(x, y) \colon |x| + |y| < 1\}$.
8. $\{(x, y) \colon |x| < 1,\ |y| < 1\}$.

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[1] $X \subset Y$ (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 subset related to the implication in logic. If $X = \{x\colon x \text{ satisfies } A\}$ and $Y = \{y\colon y \text{ satisfies } B\}$. Then $X \subset Y$ means exactly A implies B. For example, we have

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

By this interpretation, the transitivity corresponds to the fact that, if A implies B, and B implies C, then A implies C. Moreover, the third property corresponds to the fact that, A is equivalent B if and only if A implies B and B implies A.

Exercise 2.4. What are the subsets of the empty set $\emptyset$?

Exercise 2.5. We say X is a *proper subset* of Y if $X \subset Y$ and $X \neq Y$. 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?

Example 2.1.1. For $r > 0$, let

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

Geometrically, B_r is the ball of radius r, and S_r is the square of side length $2r$ (both including the interior).

From the geometric meaning, we have $S_{r'} \subset B_r$ if and only if $r' \leq \dfrac{r}{\sqrt{2}}$, and $B_r \subset S_{r'}$ if and only if $r \leq r'$.

Exercise 2.6. We know $S_{r'} \subset B_r$ means the implication

$$|x| \leq r' \text{ and } |y| \leq r' \implies x^2 + y^2 \leq r^2.$$

[1] In this course, the notation $X \subset Y$ allows the possibility that $X = Y$. If $X \subset Y$ and $X \neq Y$, i.e., X is a proper subset of Y (see Exercise 2.4), then we denote $X \subsetneq Y$. In some other textbooks, our $X \subset Y$ is denoted $X \subseteq Y$, and our $X \subsetneq Y$ is denoted $X \subset Y$.

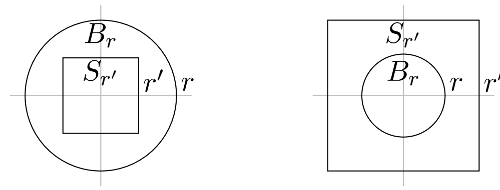

Figure 2.1: $S_{r'} \subset B_r$ and $B_r \subset S_{r'}$.

Directly argue (i.e., without using picture) that the implication holds if and only if $r' \le \dfrac{r}{\sqrt{2}}$. Then make the similar argument for the case $B_r \subset S_{r'}$.

Exercise 2.7. For $r > 0$, let

$$B_r = \{(x, y) \in \mathbb{R}^3 \colon x^2 + y^2 + z^3 \le r^2\},$$
$$S_r = \{(x, y) \in \mathbb{R}^3 \colon |x| \le r, |y| \le r, |z| \le r\}.$$

Find the meaning of $S_{r'} \subset B_r$ and $B_r \subset S_{r'}$ in terms of relation between r and r'.

What is the answer to the similar problem in higher dimension?

Exercise 2.8. 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?

Exercise 2.9. Find the exact conditions for inclusions of intervals.

1. $(a, b) \subset (c, d)$.
2. $(a, b) \subset [c, d]$.
3. $[a, b] \subset (c, d)$.
4. $[a, b] \subset [c, d]$.
5. $(a, b) \subset [c, d)$.
6. $[a, b) \subset (c, d]$.

Exercise 2.10. 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 $S_{m,n} \subset S_{m',n'}$?

Example 2.1.2. 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 2.1.3. 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 2.11. 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 2.12. 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 2.13. Given $\delta > 0$, find $\epsilon > 0$, such that

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

Exercise 2.14. Find a number $n \in \mathbb{N}$, such that

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

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 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.

Example 2.1.4. It is not unusual to see elements of some set are sets themselves. For example, the set Parity= $\{\text{Even}, \text{Odd}\}$ appears to be just a set of two elements. However, if try to define the two elements, you find they are actually sets

$$\text{Even} = \{0, \pm 2, \pm 4, \dots\} = \{2a\colon a \in \mathbb{Z}\} = 2\mathbb{Z},$$
$$\text{Odd} = \{\pm 1, \pm 3, \dots\} = \{2a+1\colon a \in \mathbb{Z}\} = 2\mathbb{Z}+1.$$

Example 2.1.5. The power set of the power set of $\{1, 2\}$ is

$$\mathcal{P}(\mathcal{P}\{1,2\}) = \mathcal{P}\left(\{\emptyset,\ \{1\},\ \{2\},\ \{1,2\}\}\right)$$
$$= \left\{ \begin{array}{c} \emptyset, \quad \{\emptyset\},\ \{\{1\}\},\ \{\{2\}\},\ \{\{1,2\}\}, \\ \{\emptyset,\{1\}\},\ \{\emptyset,\{2\}\},\ \{\emptyset,\{1,2\}\}, \\ \{\{1\},\{2\}\},\ \{\{1\},\{1,2\}\},\ \{\{2\},\{1,2\}\}, \\ \{\emptyset,\{1\},\{2\}\},\ \{\emptyset,\{1\},\{1,2\}\},\ \{\emptyset,\{2\},\{1,2\}\},\ \{\{1\},\{2\},\{1,2\}\}, \\ \{\emptyset,\{1\},\{2\},\{1,2\}\} \end{array} \right\}$$

The first row consists of subsets of zero or one "element" (which are subsets of $\{1, 2\}$). The second and third rows consist of subsets of two "elements". The fourth row consists of subsets of three "elements". The last row consists of the only subset of four "elements". We also remark that $\emptyset$ and $\{\emptyset\}$ are different subsets of $\mathcal{P}\{1, 2\}$. The first is empty, and the second is not.

Exercise 2.15. List all elements in $\mathcal{P}\{1, 2, 3, 4\}$ that do not include 2.

The following explains the reason for the notation 2^X of the power set.

Theorem 2.1.1. *If X has n elements, then $\mathcal{P}(X)$ has 2^n elements.*

Proof. Each subset of X can be considered as independent choices of Yes and No for all elements in X. More precisely, suppose $Y \subset X$ is a subset. For each x, we assign Yes to x if $x \in Y$, and we assign No if $x \notin Y$. Moreover, the choices are independent between different elements in x. Therefore the elements of $\mathcal{P}(X)$ correspond to such choices. Since we have 2 choices for each element of x, the total number of choices, which is the same as the total number of elements in $\mathcal{P}(X)$, is 2^n. $\square$

Exercise 2.16. Suppose X has n elements. How many subsets of X contains even number of elements? How many contains odd number of elements? There are two ways of doing this.

1. Let $f(n)$ be the number of even subsets, and let $g(n)$ be the number of odd subsets. Then $f(n) + g(n) = 2^n$. Moreover, pick one $x \in X$. Then an even subset is either an even subset of $X - x$, or an odd subset of $X - x$ plus x. Therefore $f(n) = f(n-1) + g(n-1)$. By the same reason, we have $g(n) = g(n-1) + f(n-1)$. Now we can derive $f(n)$ and $g(n)$ from these equalities.

2. A subset of k elements in X is the choice of k elements from n elements. Therefore the number of even subsets is

$$\sum_{\text{even } k \le n} \binom{n}{k} = \binom{n}{0} + \binom{n}{2} + \binom{n}{4} + \cdots,$$

the number of odd subsets is

$$\sum_{\text{odd } k\le n} \binom{n}{k} = \binom{n}{1} + \binom{n}{3} + \binom{n}{5} + \cdots,$$

We also have the equality

$$(a+b)^n = \sum_{k\le n} \binom{n}{k} a^{n-k}b^k = \binom{n}{0} a^n b^0 + \binom{n}{1} a^{n-1}b^1 + \binom{n}{2} a^{n-2}b^2 + \cdots.$$

By suitable choices and a and b, we may get the two sums we want.

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.

If $X = \{x \colon x \text{ satisfies } A\}$ and $Y = \{x \colon x \text{ satisfies } B\}$, then

$$X \cup Y = \{x \colon x \text{ satisfies } A \text{ or } B\}.$$

Therefore union corresponds to "or" for statements:

- Any integer is either even or odd: $\mathbb{Z} = \text{Even} \cup \text{Odd} = 2\mathbb{Z} \cup (2\mathbb{Z}+1)$.
- A mathematician does research either in pure mathematics, or in applied mathematics:

$$\{\text{mathematician}\} = \{\text{pure mathematician}\} \cup \{\text{applied mathematician}\}.$$

The second union allows a mathematician to do research in both fields.

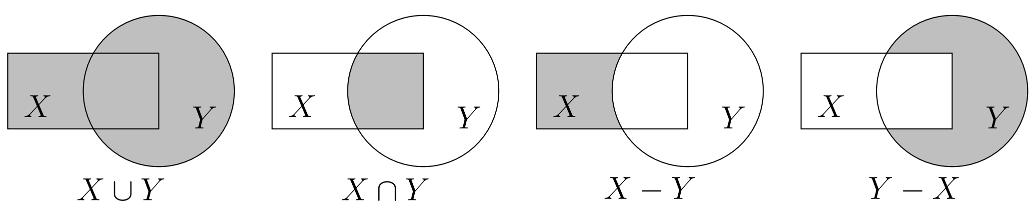

Figure 2.2: Union, intersection, and difference.

The *intersection* of two sets is

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

If X and Y are characterized by properties A and B, then

$$X \cap Y = \{x \colon x \text{ satisfies } A \text{ and } B\}.$$

Therefore intersection corresponds to "and" for statements:

- A number is divisible by 6 if and only if it is divisible by 2 and 3: $6\mathbb{Z} = 2\mathbb{Z} \cap 3\mathbb{Z}$.
- The students in "Foundation of Mathematics" are first year math students:

$$\text{Foundation of Mathematics class} \subset \{\text{first year students}\} \cap \{\text{math students}\}.$$

If $X \cap Y = \emptyset$, then there is no common elements shared by X and Y, and we say X and Y are *disjoint*. We note that disjoint property means "if $x \in X$, then $x \notin Y$":

- If an integer is even, then it is not odd: $\text{Even} \cap \text{Odd} = \emptyset$.
- If a person is a male, then the person is not female: $\text{Male} \cap \text{Female} = \emptyset$.

For disjoint X and Y, we call the union $X \cup Y$ a *disjoint union*, and we may emphasise the disjoint property by writing the union as $X \sqcup Y$. For example, we have $\mathbb{Z} = \text{Even} \sqcup \text{Odd} = 2\mathbb{Z} \sqcup (2\mathbb{Z}+1)$ and $\text{People} = \text{Male} \sqcup \text{Female}$.

Next, 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:

- The complement of even numbers is odd numbers: $\mathbb{Z} - 2\mathbb{Z} = 2\mathbb{Z} + 1$.
- The complement of odd numbers is even numbers: $\mathbb{Z} - (2\mathbb{Z} + 1) = 2\mathbb{Z}$.

We note that there are two possible differences $X - Y$ and $Y - X$, and the two are different sets.

Exercise 2.17. What are $X \cup \emptyset$, $X \cap \emptyset$, $X - \emptyset$, $\emptyset - X$, $X \cup X$, $X \cap X$, $X - X$?

Exercise 2.18. Express as union of disjoint intervals.

1. $(-1,5) \cap [-5,4) - (-\infty,0) \cup (1,2]$.
2. $(-1,5) \cup [-5,4) - (-\infty,0) \cup (1,2]$.
3. $(\,(-1,5) \cap [-5,4) - (-\infty,0)\,) \cup (1,2]$.
4. $(\,(-1,5) - (-\infty,0)\,) \cup (\,[-5,4) - (1,2]\,)$.
5. $(\,(-1,5) - (-\infty,0)\,) \cap (\,[-5,4) - (1,2]\,)$.
6. $(-1,5) \cap [-5,4) - \{1,2,3\}$.
7. $(-1,5) \cup [-5,4) - 2\mathbb{Z}$.

Exercise 2.19. What are the following sets?

1. $\{a\colon a \text{ is even}\} \cap \{a\colon a \text{ is divisible by } 5\}$.
2. $\{x\colon x \text{ is positive}\} \cup \{x\colon x \text{ is negative}\} \cup \{0\}$.
3. $\{a\colon a \text{ is even}\} \cup \{a \in \mathbb{Z}\colon |a| < 10\} - \{x\colon x \neq 2\} \cap \{x\colon x^2 \neq 6x - 8\}$.

Exercise 2.20. Find all the unions and intersections among $A = \{\emptyset\}$, $B = \{\emptyset, A\}$, $C = \{\emptyset, A, B\}$.

Exercise 2.21. 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. $\{x\colon x \in X, \text{ and } x \notin Y, \text{ and } x \in Z\}$.

Example 2.2.1. We prove $X, Y \subset Z$ implies $X \cup Y \subset Z$.

If $x \in X \cup Y$, then $x \in X$ or $x \in Y$. If $x \in X$, then by $X \subset Z$, we have $x \in Z$. If $x \in Y$, then by $Y \subset Z$, we have $x \in Z$. Therefore $x \in X \cup Y$ implies $x \in Z$. In other words, we have $X \cup Y \subset Z$.

Exercise 2.22. Prove the following, in the style of Example 2.2.1.

1. If $Z \subset X, Y$, then $Z \subset X \cap Y$.
2. If $X \subset Y$ and $Z \subset W$, then $X \cup Y \subset Y \cup W$ and $X \cap Y \subset Y \cap W$.
3. If $X \subset Y \cup Z$, then $X - Y \subset Z$.
4. $X \cap Y$ and $X - Y$ are disjoint.
5. $X - Y = \emptyset$ if and only if $X \subset Y$.
6. $X - Y = X$ if and only if X and Y are disjoint.

Exercise 2.23. 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 2.24. Which statement is 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. $Z \subset X$ or $Z \subset Y \implies Z \subset X \cap Y$.

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

Example 2.2.2. From Figure 2.2, we can see $X = (X \cap Y) \cup (X - Y)$. Let us prove the equality by definition. This means the two sets contain each other.

Suppose $x \in X$. Then either $x \in Y$ or $x \notin Y$. If $x \in Y$, then by $x \in X$ and $x \in Y$, we get $x \in X \cap Y$. If $x \notin Y$, then by $x \in X$ and $x \notin Y$, we get $x \in X - Y$. This proves $X \subset (X \cap Y) \cup (X - Y)$.

On the other hand, by the definition of intersection and difference, we have $X \cap Y \subset X$ and $X - Y \subset X$. By Example 2.2.1, we get $(X \cap Y) \cup (X - Y) \subset X$.

Exercise 2.25. Prove $X - (X - Y) = X \cap Y$. If $Y \subset X$, then the equality becomes $X - (X - Y) = Y$, which means complement of complement is the original.

Exercise 2.26. Suppose $X = Y \sqcup Z$. Prove that $Z = X - Y$ and $Y = X - Z$.

In Example 2.2.2, we see the proof of set equalities can be rather tedious. In practise, if you suspect some set theoretical equality holds, then you may verify by drawing pictures like Figure 2.2, called *Venn diagram*. In this course, we will emphasise argument of set equality by Venn diagram, and will not require argument by definition.

Example 2.2.3. In Figure 2.2, we see the disjoint union $X = (X \cap Y) \sqcup (X - Y)$ (see Example 2.2.2 and Exercise 2.22.4). We also see (see Exercise 2.26)

$$X - (X \cap Y) = X - Y, \quad X - (X - Y) = X \cap Y.$$

The second equality is Exercise 2.25, and is illustrated by the Venn diagram in Figure 2.2.

Exercise 2.27. Use Venn diagram to verify the equality.

1. Commutativity: $Y \cup X = X \cup Y$, and $Y \cap X = X \cap Y$.

2. Associativity: $(X \cup Y) \cup Z = X \cup (Y \cup Z)$, and $(X \cap Y) \cap Z = X \cap (Y \cap Z)$.

3. Distributivity: $(X \cup Y) \cap Z = (X \cap Z) \cup (Y \cap Z)$, and $(X \cap Y) \cup Z = (X \cup Z) \cap (Y \cup Z)$.

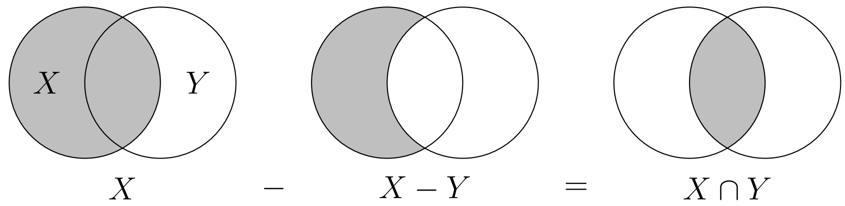

Figure 2.3: Venn diagram for seeing $X - (X - Y)$.

4. $(X \cup Y) - Z = (X - Z) \cup (Y - Z)$, and $(X \cap Y) - Z = (X - Z) \cap (Y - Z)$.
5. $(X - Y) \cup (Y - X) = X \cup Y - X \cap Y$.
6. $(X - Y) \cap (Y - X) = \emptyset$.

Exercise 2.28. Use Venn diagram to show $X \cup Y = (X - Y) \sqcup (X \cap Y) \sqcup (Y - X)$. Then identify the unions of some of the three disjoint subsets. How many such union subsets are there?

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

Example 2.2.4. The *deMorgan's Law* is the following

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

The first equality is illustrated by the Venn diagram in Figure 2.4.

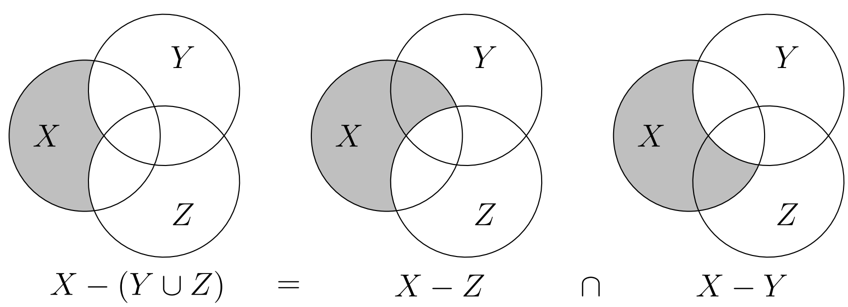

Figure 2.4: deMorgan's law.

The law means that the operation $X-?$ exchanges between $\cup$ and $\cap$. On the other hand, Exercise 2.27.4 show that the operation $? - Z$ preserves $\cup$ and $\cap$.

Exercise 2.29. Use Venn diagram to show $X - (Y \cap Z) = (X - Y) \cup (X - Z)$.

Exercise 2.30. Which statement is true?

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

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

Finally, the (cartesian) *product* of two sets X and Y is

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

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 finally the difference $-$ is taken. For example,

$$x \in (X - Y \times Z) \cap W - U = (\,(\,(X - (Y \times Z)\,)\cap W\,) - U$$

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 arithmetic computations.

Example 2.2.5. The product of $X = \{1, 2, 3\}$ and $Y = \{4, 5\}$ consists of six elements

$$X \times Y = \{(1, 4), (2, 4), (3, 4), (1, 5), (2, 5), (3, 5)\}.$$

Figure 2.2 shows the product set like the cartesian coordinates of $\mathbb{R}^2$.

5	$(1,5)$	$(2,5)$	$(3,5)$
4	$(1,4)$	$(2,4)$	$(3,4)$
	1	2	3

Figure 2.5: $\{1, 2, 3\} \times \{4, 5\}$.

Exercise 2.31. What is the condition for a pair of real numbers (x, y) to be in the set?

1. $\mathbb{R} \times \mathbb{Q} \cup \mathbb{R} \times \mathbb{Q}$.
2. $\mathbb{R} \times \mathbb{Q} \cap \mathbb{R} \times \mathbb{Q}$.

3. $\mathbb{R} \times \mathbb{R} - \mathbb{Q} \times \mathbb{Q}$.

4. $\mathbb{R} \times \mathbb{Q} - \mathbb{Q} \times \mathbb{R}$.

5. $(\mathbb{R} - \mathbb{Q}) \times (\mathbb{R} - \mathbb{Q})$.

Exercise 2.32. Draw picture of the set.

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

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

Exercise 2.33. What is the meaning that $X \times Y$ and $Z \times W$ are disjoint?

Exercise 2.34. Prove $X \subset Z$ and $Y \subset W$ if and only if $X \times Y \subset Z \times W$.

Example 2.2.6. We may get set identities involving products by looking at the cartesian like picture. For example, we see $(X - Y) \times Z = X \times Z - Y \times Z$ from Figure 2.2.6.

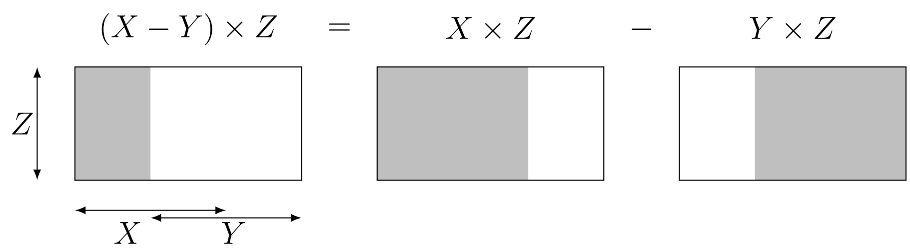

Figure 2.6: $(X - Y) \times Z = X \times Z - Y \times Z$.

The following is the proof by definition, which is less important for this course.

Suppose $(x, y) \in (X - Y) \times Z$. Then we have $x \in X - Y$ and $z \in Z$. By $x \in X - Y$, we have $x \in X$ and $x \notin Y$. By $x \in X$ and $y \in Z$, we get $(x, y) \in X \times Z$. By $x \notin Y$, we get $(x, y) \notin Y \times Z$. Therefore $(x, y) \in X \times Z - Y \times Z$.

Suppose $(x, y) \in X \times Z - Y \times Z$. Then we have $(x, y) \in X \times Z$ and $(x, y) \notin Y \times Z$. By $(x, y) \in X \times Z$, we get $x \in X$ and $y \in Z$. By $(x, y) \notin Y \times Z$, we have $x \notin Y$ or $y \notin Z$. Since we already know $y \in Z$ (derived from $(x, y) \in X \times Z$), we know $x \notin Y$. Then by $x \in X$ and $x \notin Y$, we get $x \in X - Y$. Combined with $y \in Z$, we get $(x, y) \in (X - Y) \times Z$.

We remark that $(x, y) \notin X \times Y$ means $x \notin X$ *or* $y \notin Y$. Therefore some extra argument is needed in the second part of the proof above.

Exercise 2.35. Express as union of disjoint rectangles.

1. $(1, 3) \times [2, 4]$.

2. $(1,3) \times [2,4] \cup (2,4) \times [3,5]$.
3. $(1,3) \times [2,4] \cap (2,4) \times [3,5]$.
4. $(1,3) \times [2,4] - (2,4) \times [3,5]$.

Exercise 2.36. Determine whether equality holds by drawing picture.

1. $(X \cup Y) \times Z = X \times Z \cup Y \times Z$.
2. $(X \cap Y) \times Z = X \times Z \cap Y \times Z$.
3. $X \times Y \cup Z \times W = (X \cup Z) \times (Y \cup W)$.
4. $X \times Y \cap Z \times W = (X \cap Z) \times (Y \cap W)$.
5. $(X - Y) \times (Z - W) = X \times Z - Y \times W$.

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 $f(x) \in Y$.
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*.

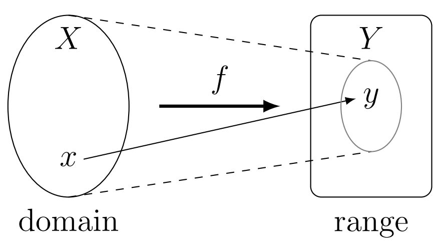

Figure 2.7: Domain, range, and image.

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 subtract 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 2.8, 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.

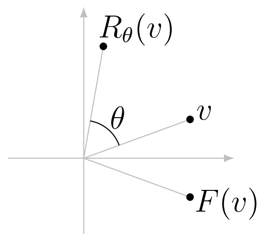

Figure 2.8: Flip and rotation of $\mathbb{R}^2$.

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 $\to$ Number.
- ID_s: Student $\to$ Number.
- ID_p: Professor $\to$ Number.
- Instructor: Course $\to$ Professor.
- Maker: Product $\to$ Manufacturer.
- Capital City: Country $\to$ City.
- Population: City $\to$ Number.

Exercise 2.37. 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 $y \in (-\infty, -2] \cup [1, 2)$, such that $y^2 = x$. Then $f(x) = y$.

Exercise 2.38. Describe the processes that define the maps.

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

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

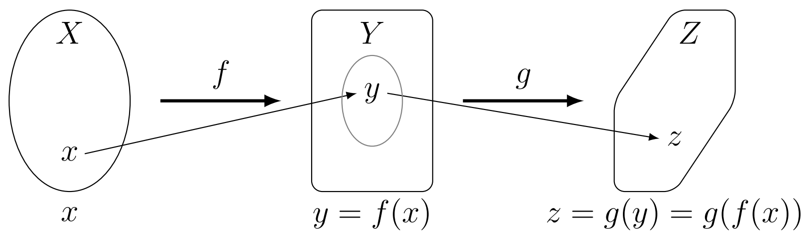

Figure 2.9: Composition of maps.

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. 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 $R_{\theta_2} R_{\theta_1} = R_{\theta_1 + \theta_2}$.

Example 2.3.10. 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".

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

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

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 $y \in Y$, we get the preimage of a point

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

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 $f(x) = x^2 \colon \mathbb{R} \to \mathbb{R}$. The following are some examples of images

$$\begin{aligned} f(\{1\}) &= f(\{1,-1\}) = \{1\}, & f(\{2\}) &= f(\{2,-2\}) = \{4\}, \\ f(\{1,-2,3\}) &= f(\{-1,2,-2,-3\}) = \{1,4,9\}, & f(\{a\}) &= f(\{a,-a\}) = \{a^2\}. \end{aligned}$$

Moreover, we have

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

The following are some examples of preimages

$$\begin{aligned} f^{-1}(\{1\}) = f(\{1,-1\}) &= \{1,-1\}, \\ f^{-1}(\{4\}) = f(\{4,-1,-2\}) &= \{2,-2\}, \\ f^{-1}(\{1,4\}) &= \{1,-1,2,-2\}, \end{aligned}$$

Moreover, we have

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

Example 2.3.12. 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

$$\begin{aligned} f[0,\infty) = f(-\infty,0] &= [-1,\infty), \\ f[-1,\infty) = f(-\infty,1] &= [-1,\infty), \\ f[2,\infty) = f(-\infty,-2] &= [7,\infty). \end{aligned}$$

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

$$\begin{aligned} f^{-1}[0,\infty) &= (-\infty,-\tfrac{1}{\sqrt{2}}]\cup[\tfrac{1}{\sqrt{2}},\infty),\\ f^{-1}[1,\infty) &= (-\infty,-1]\cup[1,\infty),\\ f^{-1}[-1,\infty) = f^{-1}[-5,\infty) &= \mathbb{R}. \end{aligned}$$

In general, we have

$$f^{-1}[b,\infty)=\begin{cases}(-\infty,\sqrt{\frac{b+1}{2}}]\cup[\sqrt{\frac{b+1}{2}},\infty), & \text{if } b>-1\\ \mathbb{R}, & \text{if } b\le -1\end{cases}.$$

Example 2.3.13. 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 2.3.14. 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

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

Exercise 2.39. 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 2.40. 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 2.41. 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 2.42. 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 2.43. 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 2.44. 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 2.45. 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?

Now we consider properties of the image and preimage. First the simplest properties

$$f(\emptyset) = \emptyset, \quad f^{-1}(\emptyset) = \emptyset, \quad f(X) \subset Y, \quad f^{-1}(Y) = X.$$

We note that $f(X)$ may not be equal to Y. In fact, $f(X) = Y$ means f is onto. We see that image behaves s bit worse than the preimage.

Both the image and preimage behave nicely with respect to the inclusion:

$$A_1 \subset A_2 \implies f(A_1) \subset f(A_2), \quad B_1 \subset B_2 \implies f^{-1}(B_1) \subset f^{-1}(B_2).$$

The also behave nicely with respect to the composition:

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

The following are properties with regard to the set operations

$$\begin{aligned} f(A_1 \cup A_2) &= f(A_1) \cup f(A_2), & f^{-1}(B_1 \cup B_2) &= f^{-1}(B_1) \cup f^{-1}(B_2), \\ f(A_1 \cap A_2) &\subset f(A_1) \cap f(A_2), & f^{-1}(B_1 \cap B_2) &= f^{-1}(B_1) \cap f^{-1}(B_2), \\ f(A_1 - A_2) &\supset f(A_1) - f(A_2), & f^{-1}(B_1 - B_2) &= f^{-1}(B_1) - f^{-1}(B_2). \end{aligned}$$

Again $\subset$ above may not be equality.

Example 2.3.15. We consider the intersection property.

By $A_1 \cap A_2 \subset A_1$ and $A_1 \cap A_2 \subset A_2$, we get $f(A_1 \cap A_2) \subset f(A_1)$ and $f(A_1 \cap A_2) \subset f(A_2)$. Therefore $f(A_1 \cap A_2) \subset f(A_1) \cap f(A_2)$. For an example of not equality, consider $f(x) = x^2$ and $A_1 = \{2\}$, $A_2 = \{-2\}$. We have $A_1 \cap A_2 = \emptyset$ and $f(A_1) = f(A_2) = \{4\}$. Therefore $f(A_1 \cap A_2) = \emptyset$ and $f(A_1) \cap f(A_2) = \{4\}$ are not equal.

The following is the proof of the equality for preimage

$$\begin{aligned} x \in f^{-1}(B_1 \cap B_2) &\iff f(x) \in B_1 \cap B_2 \\ &\iff f(x) \in B_1 \text{ and } f(x) \in B_2 \\ &\iff x \in f^{-1}(B_1) \text{ and } x \in f^{-1}(B_2) \\ &\iff x \in f^{-1}(B_1) \cap f^{-1}(B_2). \end{aligned}$$

Exercise 2.46. Explain why for any map $f\colon X \to Y$, we have $f^{-1}(Y) = X$.

Exercise 2.47. Prove $f(A_1 \cup A_2) = f(A_1) \cup f(A_2)$ and $f^{-1}(B_1 \cup B_2) = f^{-1}(B_1) \cup f^{-1}(B_2)$.

Exercise 2.48. Prove $f(A_1 - A_2) \supset f(A_1) - f(A_2)$. Then show inequality may happen by finding an example, such that $A_2 \subset A_1$, $A_2 \neq A_1$, and $f(A_1) = f(A_2)$.

Exercise 2.49. Prove $f^{-1}(B_1 - B_2) = f^{-1}(B_1) - f^{-1}(B_2)$.

Exercise 2.50. Prove $f(f^{-1}(B)) = B \cap f(X)$, where X is the domain of f.

Exercise 2.51. Prove $f^{-1}(f(A)) \supset A$. However, the equality may not happen.

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 if two elements have the same image, then the two elements are 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.

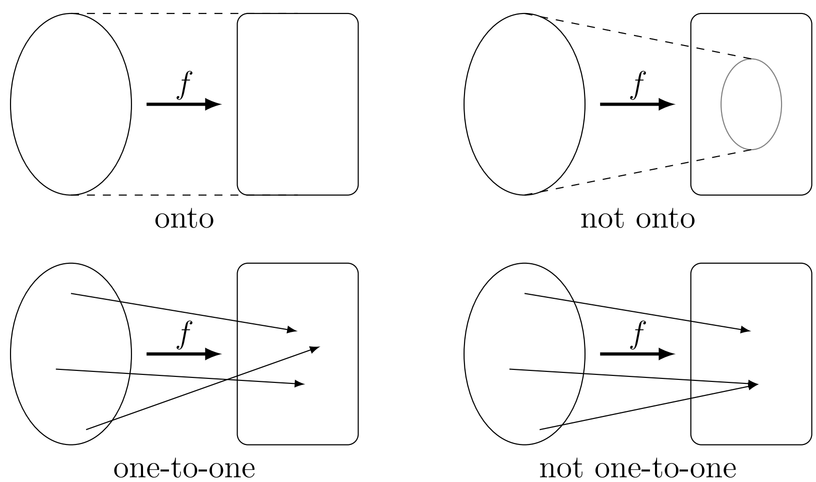

Figure 2.10: Onto and one-to-one.

The following interprets the concepts from the viewpoint of solving the equation $f(x) = y$. Here the right side $y \in Y$ is given, 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 x.*
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 x.*

For any $y \in Y$, there are three possibilities for the solution of the equation:

- $0 \to 1$: 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 $f(x) = y$.

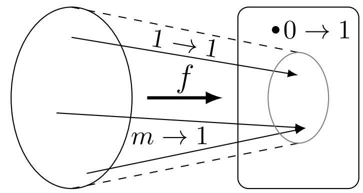

Figure 2.11: Possibilities for the solution of $f(x) = y$.

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.

Example 2.4.1. Consider the function $f(x) = x^2 \colon \mathbb{R} \to \mathbb{R}$. The map is $0 \to 1$ for $y < 0$, and $1 \to 1$ for $y = 0$, and $2 \to 1$ for $y > 0$.

To make the function onto, we may remove the $0 \to 1$ case. This means modifying the function to $f(x) = x^2 \colon \mathbb{R} \to [0, +\infty)$. In general, for any map $f \colon X \to Y$, we may reduce the range to $f(X) \subset Y$ and get an onto map $f \colon X \to f(X)$. Although we use the same notation f, we should regard $f \colon X \to Y$ and $f \colon X \to f(X)$ as two different maps.

To make the function one-to-one, we may restrict the choice in the $2 \to 1$ case. The two choices are actually two square roots x for $y > 0$, one positive and one negative. For example, if we choose positive x, then we get a one-to-one function $f(x) = x^2 \colon [0, \infty) \to \mathbb{R}$. We may also choose negative x an get another one-to-one function $f(x) = x^2 \colon (-\infty, 0] \to \mathbb{R}$. Again we should regard $x^2 \colon \mathbb{R} \to \mathbb{R}$, $x^2 \colon [0, \infty) \to \mathbb{R}$, $x^2 \colon (-\infty, 0] \to \mathbb{R}$ as three different maps.

If we combine the onto and one-to-one modifications, then we get one-to-one correspondences $f(x) = x^2 \colon [0, \infty) \to [0, \infty)$ and $f(x) = x^2 \colon (-\infty, 0] \to [0, \infty)$.

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 infinitely many different angles, all differ by multiples of 2π. By restricting the angles to $[0, 2\pi)$, we modify the map to $[0, \infty) \times [0, 2\pi) \to \mathbb{R}^2$. The new Polar map is almost a one-to-one correspondence, with the origin as the only exception. If we further fix the angle of the origin to be 0, then we get a one-to-one correspondence

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

Example 2.4.3. For the map Instructor: Course $\to$ 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 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 2.52. 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 $\to$ Natural Number.
10. Population: City $\to$ Natural Number.
11. Price: Book $\to$ Positive Number.

Exercise 2.53. Determine the cases of how many to 1 for $f(x) = 2x^2 - 1\colon \mathbb{R} \to \mathbb{R}$. Then modify to suitable onto, or one-to-one, or one-to-one correspondence maps.

Exercise 2.54. 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 2.55. After the domain is made smaller or the range is made larger, how are the onto and one-to-one properties changed?

Exercise 2.56. If gf is onto, prove that g is onto. If gf is one-to-one, prove that f is one-to-one.

Exercise 2.57. 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 2.58. In Example 2.4.1, we modify any map $f\colon X \to Y$ to an onto map $\hat{f}\colon X \to f(X)$. When is $\hat{f}$ a one-to-one correspondence?

Exercise 2.59. Let $f\colon X \to Y$ be a map. When is the image map $A \mapsto f(A)\colon \mathcal{P}(X) \to \mathcal{P}(Y)$ onto or one-to-one? What about the preimage map $B \mapsto f^{-1}(B)\colon \mathcal{P}(Y) \to \mathcal{P}(X)$?

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, for all $x \in X$ and $y \in Y$, we have

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

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 correspondence, 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 not ambiguous.

Therefore g is indeed a map. It remains to verify $gf(x) = x$ and $fg(y) = y$. First, by following the definition of g, we get $gf(x) = g(f(x)) = g(y) = x$. Second, for $y \in Y$, by f onto, we may write $y = f(x)$. Then $fg(y) = f(g(f(x))) = f(x) = y$. Here we use $g(f(x)) = x$ in the second equality. $\square$

Example 2.4.5. The function $\sin\colon \mathbb{R} \to \mathbb{R}$ is strictly increasing (and therefore one-to-one) on the interval $[-\frac{1}{2}\pi, \frac{1}{2}\pi]$, and has values filling the interval $[-1, 1]$. Therefore the function $\sin\colon [-\frac{1}{2}\pi, \frac{1}{2}\pi] \to [-1, 1]$ is a one-to-one correspondence. The corresponding inverse map is the inverse sine function $\arcsin\colon [-1, 1] \to [-\frac{1}{2}\pi, \frac{1}{2}\pi]$.

Similarly, the cosine function becomes a one-to-one correspondence when we restrict the range and domain to $\cos\colon [0, \pi] \to [-1, 1]$. The corresponding inverse cosine function is $\arccos\colon [-1, 1] \to [0, \pi]$.

Example 2.4.6. In Example 2.4.1, we modify the function x^2 to one-to-one correspondences

$$f_1(x) = x^2\colon [0, \infty) \to [0, \infty), \quad f_2(x) = x^2\colon (-\infty, 0] \to [0, \infty).$$

The corresponding inverse maps are

$$f_1^{-1}(y) = \sqrt{y}\colon [0, \infty) \to [0, \infty), \quad f_2^{-1}(y) = -\sqrt{y}\colon [0, \infty) \to (-\infty, 0]$$

In fact, we also have one-to-one correspondence

$$f_3(x) = x^2\colon (-2, 0] \cup [2, \infty) \to [0, \infty).$$

The corresponding inverse map is

$$f_3^{-1}(y) = \begin{cases} \sqrt{y}, & \text{if } y \geq 4 \\ -\sqrt{y}, & \text{if } 0 \leq y < 4 \end{cases} : [0, \infty) \to (-2, 0] \cup [2, \infty).$$

Example 2.4.7. 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 $f^{-1}(y_1, y_2) = (-3y_1 + 2y_2 + 5, 2y_1 - y_2 - 3)$.

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 2.4.8. The flip F with respect to the x-axis in Example 2.3.9 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 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 2.4.9. 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 2.60. 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 2.61. Consider $f(x) = |x^2 - 1|\colon \mathbb{R} \to \mathbb{R}$.

1. What are the possible numbers of elements in $f^{-1}(y)$? Moreover, for each possible number, find the corresponding range of y.
2. Modify f to become a one-to-one correspondence, such that the domain contains $[-1, 0]$, and the range is $f(\mathbb{R}) = [0, \infty)$. Then find the formula for the inverse of the one-to-one correspondence.

Exercise 2.62. For the one-to-one correspondences you obtain in Exercise 2.53, find the inverse map.

Exercise 2.63. In Example 2.4.2, we improved the polar map to become a one-to-one correspondence. Find the inverse of this improved polar map.

Exercise 2.64. 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 2.65. 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 2.66. 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 $a, b \in \mathbb{Z}$, define $a \sim b$ if $a - b$ is even. This is an equivalence relation: First, by $a - a = 0$ being even, we get the reflexivity. Second, if $a - b$ is even, then $b - a = -(a - b)$ is also even. This verifies the symmetry. Third, if $a - b$ and $b - c$ are even, then $a - c = (a - b) + (b - c)$ is still even. This verifies the transitivity.

More generally, we may fix a natural number n and let $n\mathbb{Z} = \{nk : k \in \mathbb{Z}\}$ be all the multiples of n. Then we define $a \sim b$ whenever $a - b \in n\mathbb{Z}$, i.e., $a - b$ is divisible by n. By $a - a = n0 \in \mathbb{Z}$, we get the reflexivity. By $a - b = nk \in n\mathbb{Z} \implies b - a = n(-k) \in n\mathbb{Z}$, we get the symmetry. By $a - b = nk$ and $b - c = nk' \implies a - c = n(k + k')$, we get the transitivity. Therefore we get an equivalence relation.

On the other hand, define $a \sim b$ when $a - b$ 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. Specifically, 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 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 2.67. 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 not 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. $X = \mathbb{Z}$, $a \sim b$ if $a - b$ is a multiple of 3 and not a multiple of 5.
10. $X = \mathbb{Z} - \{0\}$, $a \sim b$ if $a = xb$ for some $x \in X$.
11. $X = \mathbb{Q} - \{0\}$, $r \sim s$ if $r = xs$ for some $x \in X$.
12. $X = \mathcal{P}(\{1, 2, \ldots, n\})$, $A \sim B$ if $A \cap B \neq \emptyset$.
13. $X = \mathcal{P}(\{1, 2, \ldots, n\})$, $A \sim B$ if $A \cap B = \emptyset$.

Exercise 2.68. 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 get an equivalence relation.

Exercise 2.69. Suppose a relation is reflexive, and has the property that $x \sim y$ and $x \sim z$ imply $y \sim z$. Prove that the relation is an equivalence relation.

Exercise 2.70. The following purports to prove that the reflexivity condition is unnecessary. In other words, it can be derived from symmetry and transitivity: Suppose $x \sim y$. Then by symmetry, we get $y \sim x$. Then by $x \sim y$, and $y \sim x$, and the transitivity, we get $x \sim x$.

Is the argument valid? Why?

Exercise 2.71. 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 $x \sim_A y$ for $x, y \in A$ if $x \sim y$ by considering x and y as elements of X. Show that $\sim_A$ is an equivalence relation on A.

Exercise 2.72. Suppose $\sim_X$ and $\sim_Y$ are equivalence relations on X and Y. Prove that on the product $X \times Y$, the following is an equivalence relation

$$(x_1, y_1) \sim (x_2, y_2) \iff x_1 \sim_X x_2 \text{ and } y_1 \sim_Y y_2.$$

How about defining the relation by $x_1 \sim_X x_2$ or $y_1 \sim_Y y_2$?

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

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

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

Example 2.5.5. Consider the equivalence relation on integers $\mathbb{Z}$ in Example 2.5.1

$$a \sim b \iff a - b \text{ is even}.$$

We have

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

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

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

and $[a] = \text{Odd}$ for any odd a. In summary, we have

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

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 natural number n, we have the following *mod n equivalence relation* on integers $\mathbb{Z}$

$$a \sim b \iff a - b \in n\mathbb{Z}.$$

Then

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

More generally, we have

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

In particular, 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

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

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 $\Longleftarrow$.

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 $x \sim y$ on $\mathbb{R}$ 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 $\|v\| = \|(x, y)\| = \sqrt{x^2 + y^2}$

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

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

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

Exercise 2.73. For the equivalence relations in Exercise 2.67, find the equivalence classes.

Exercise 2.74. For the equivalence relation in Exercise 2.72, find the equivalence classes.

Exercise 2.75. Find the smallest equivalence relation on $\{1, 2, 3, 4, 5\}$, such that $1 \sim 2$, $1 \sim 4$, $3 \sim 3$. Moreover, find the equivalence classes.

Exercise 2.76. Find the smallest equivalence relation on integers $\mathbb{Z}$, such that if $a = nb$ or $b = na$ for some natural number n, then $a \sim b$. Moreover, find the equivalence classes.

Exercise 2.77. Suppose $\sim_1$ and $\sim_2$ are equivalence relations. Define $x \sim y$ if $x \sim_1 y$ and $x \sim_2 y$. Prove that $\sim$ is an equivalence relation. Moreover, find the equivalence class of $\sim$ in terms of the equivalence classes of $\sim_1$ and $\sim_2$.

Moreover, what if we define $x \sim y$ if $x \sim_1 y$ or $x \sim_2 y$?

Exercise 2.78. Suppose $f \colon X \to Y$ is a map, and $\sim_Y$ is an equivalence relation on Y. Define a relation $x_1 \sim_X x_2$ on X if $f(x_1) \sim_Y f(x_2)$.

1. Prove that $\sim_X$ is an equivalence relation.
2. Describe the equivalence classes for $\sim_X$ in terms of equivalence classes for $\sim_Y$.

Exercise 2.79. Suppose $f \colon X \to Y$ is a map, and $\sim_X$ is an equivalence relation on Y. Define a relation $y_1 \sim_Y y_2$ on Y if there are $x_1, x_2 \in X$ satisfying $f(x_1) = y_1$, $f(x_2) = y_2$, and $x_1 \sim_X x_2$. Is $\sim_Y$ an equivalence relation?

We see an equivalence relation on X decomposes 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, in $\mathbb{Z} = \text{Even} \sqcup \text{Odd} = [0] \sqcup [1]$, we use 0 and 1 as representatives of even and odd numbers. In general, a *partition* of a set X is a decomposition into a disjoint union of nonempty subsets

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

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. Moreover, 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 $a \sim b$ described as follows:

- $a, b \in 2\mathbb{Z}$: Both a and b are even.
- $a, b \in 2\mathbb{Z}+1$: Both a and b are odd.

The relation is equivalent to $a - b$ 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 as the 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 called 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 the output of the process q is unique. 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 2.80. Find the equivalence relation and the partition corresponding to the onto map $q(x, y) = x^2 + y^2 \colon \mathbb{R}^2 \to [0, \infty)$.

Exercise 2.81. Find the partition and the quotient corresponding to the equivalence relation on $\mathbb{R}^2$ defined by $u \sim v$ if $u = rv$ for some real number $r > 0$.

It remains to explain how to get the quotient map from an equivalence relation. 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 define the *quotient set* to be the collection of all equivalence classes

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

and use the quotient set as the index set I. Then the quotient map take x to its equivalence class $[x]$

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

Example 2.5.12. For the equivalence relation $a \sim b \iff a - b$ 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 $\mathbb{Z}/n\mathbb{Z} = \{[0], [1], \ldots, [n-1]\}$. The quotient map is the "remainder map"

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

Specifically, we divide a by n and get

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

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

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

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

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 $a - b \in n\mathbb{N}$.

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

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

Here $\cong$ means identifying two sets.

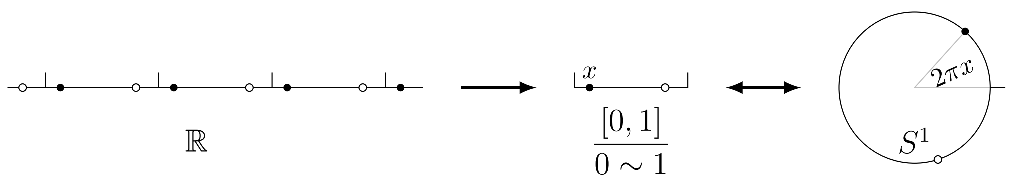

Figure 2.12: Quotient set $\mathbb{R}/\mathbb{N}$.

From Figure 2.12, 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 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.

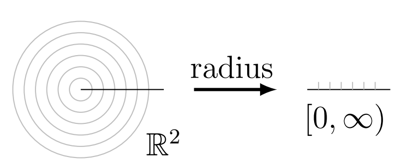

Figure 2.13: Quotient set and quotient map for rotation equivalence.

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 in G.

For example, let X be the unique sphere $S^2 = \{(x, y, z) \colon x^2 + y^2 + z^2 = 1\}$, and 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 2.82. 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 2.83. 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 2.84. 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 2.85. 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 2.86. 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 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 the *ordinality* of c in the sequence.

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

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 initial number, naturally denoted as 1.

The intuition for the second axiom 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 up" all the natural numbers by starting

[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.

with the initial number 1 and then creating the subsequent ones by repeatedly applying the "successor operation"[2].

The third axiom says that the special number 1 is not a successor. In other words, there is no natural number prior to the initial 1.

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 $A(1), A(2), A(3), \dots$ means 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 implies that $S = \mathbb{N}$, which means $A(n)$ is true for all n.

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 $S = \mathbb{N}$. Therefore a natural number m not in $\{1\}$, i.e., $m \neq 1$, must be in $\{n' \colon n \in \mathbb{N}\}$, i.e., $m = n'$ for some $n \in \mathbb{N}$. □

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 axioms. Moreover, it is a typical *inductive definition.* Specifically, we fix the first number m and define $m+n$ by inducting 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 $m + n$. Then by the induction axiom, for any fixed 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 addition, as a map from $\mathbb{N} \times \mathbb{N}$ to $\mathbb{N}$, is well defined by the inductive process. The map is the following process:

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

- If $n = 1$, then $m + n = m'$.
- If $n = k'$, then $m + n = (m + k)'$.

By Proposition 3.1.2, any $n \in \mathbb{N}$ is in one of the two above cases. Therefore the process works for all n (the second case requires induction). Then we need to verify that the outcome is unique. By the third axiom, the two cases are mutually exclusive. Therefore we only need to verify each case has the unique outcome. In the first case, the outcome is the unique (by the second axiom) successor. In the second case, we need to choose a predecessor k. By the fourth axion, we know k is unique. Therefore the outcome $(m + k)'$ is unique.

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

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

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. $\square$

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. 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 3.4. Define the order $m \geq n$ between natural numbers by $n \geq 1$ for any n, and $m \geq n \implies m' \geq n'$. Show that the order is well defined.

3.2 Integer

The *integers* are

$$\dots, -4, -3, -2, -1, 0, 1, 2, 3, 4, \dots.$$

We will construct integers (especially negative integers) from natural numbers. The idea is to define integers as subtraction of natural numbers:

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

Here 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 intergers -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 $(m,n) \in \mathbb{N} \times \mathbb{N}$ 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 subtraction 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. Therefore we introduce the relation on $\mathbb{N} \times \mathbb{N}$

$$(m,n) \sim (k,l), \text{ if } m + l = k + n.$$

The following shows this is an equivalence relation:

1. Reflexivity: We need to verify $(m,n) \sim (m,n)$. This means $m + n = m + n$. Therefore the property is verified.

2. Symmetry: We need to verify $(m,n) \sim (k,l)$ implies $(k,l) \sim (m,n)$. The relation $(m,n) \sim (k,l)$ means $m + l = k + n$. The relation $(k,l) \sim (m,n)$ means $k + n = m + l$. Therefore the property is verified.

3. Transitivity: We need to verify $(m,n) \sim (k,l)$ and $(k,l) \sim (p,q)$ imply $(m,n) \sim (p,q)$. The relation $(m,n) \sim (k,l)$ means $m + l = k + n$. The relation $(k,l) \sim (p,q)$ means $k + q = p + l$. Adding the two equalities, we get

   $$(m + l) + (k + q) = (k + n) + (p + l).$$

   By the associativity and commutativity in Proposition 3.1.4, this means

   $$(m + q) + (k + l) = (p + n) + (k + l).$$

   Then by the cancelation property in Proposition 3.1.4, we get $m + q = p + n$. This means $(m,n) \sim (p,q)$.

Now we may define integers to be the quotient set by the equivalence relation

$$\mathbb{Z} = \mathbb{N} \times \mathbb{N} / \sim .$$

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

An integer is an equivalence class $[(m, n)]$, 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 $[m, n] = [k, l]$ if and only if $m + l = k + n$.

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 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 $a, b \in \mathbb{Z}$. Then we need to verify the choices of m, n, k, l will not change the outcome. In other words, given 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

$$[m_1 + k_1, n_1 + l_1] = [m_2 + k_2, n_2 + l_2].$$

By $[m_1, n_1] = [m_2, n_2]$ and $[k_1, l_1] = [k_2, l_2]$, we get

$$m_1 + n_2 = m_2 + n_1, \quad k_1 + l_2 = k_2 + l_1.$$

Then by Proposition 3.1.4, this implies

$$(m_1+k_1)+(n_2+l_2) = (m_1+n_2)+(k_1+l_2) = (m_2+n_1)+(k_2+l_1) = (m_2+k_2)+(n_1+l_1).$$

The equality means $[m_1 + k_1, n_1 + l_1] = [m_2 + k_2, n_2 + l_2]$.

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:* $a + b = b + a$.

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

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

We remark that the unique 0 in the third property is given by $0 = [1, 1]$. Moreover, in the fourth property, the unique negative of $a = [m, n]$ is given by $-a = [n, m]$.

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.

We have $[m, n]+[1, 1] = [m+1, n+1]$. By Proposition 3.1.4, we have $(m+1)+n = m + (n + 1)$. This means $[m + 1, n + 1] = [m, n]$. This proves $0 = [1, 1]$ satisfies $a + 0 = a$. By the second (commutative) property, we also get $0 + a = a$.

We have proved the existence of 0 satisfying the third property. For the uniqueness, let $\bar{0}$ be another zero satisfying the third property. Then we have

$$a + 0 = a = 0 + a, \quad a + \bar{0} = a = \bar{0} + a.$$

Taking $a = \bar{0}$ in the right of the first equality, and taking $a = 0$ in the left of the second equality, we get

$$0 = 0 + \bar{0} = \bar{0}.$$

This proves the uniqueness of zero.

We have $[m, n] + [n, m] = [m + n, n + m]$. By Proposition 3.1.4, we have $(m + n) + 1 = 1 + (n + m)$. This means $[m + n, n + m] = [1, 1] = 0$. This proves that, for $a = [m, n]$, by taking $-a = [n, m]$, we get $a + (-a) = 0$. By the second (commutative) property, we also get $(-a) + a = 0$.

We have proved the existence of negative satisfying the fourth property. For the uniqueness, let $-\bar{a}$ be another negative satisfying the fourth property. Then we have

$$a + (-a) = 0 = (-a) + a, \quad a + (-\bar{a}) = 0 = (-\bar{a}) + a.$$

Then we get

$$-a = (-a) + 0 = (-a) + (a + (-\bar{a})) = ((-a) + a) + (-\bar{a}) = 0 + (-\bar{a}) = -\bar{a}.$$

The first and fifth equalities follow from the third property. The second and fourth equalities follow from the two equalities above. The third equality follows from the third (associativity) property. This proves the uniqueness of the negative. $\square$

Exercise 3.5. Prove the first and second properties in Proposition 3.2.2.

Exercise 3.6. The following is another way of proving the third property in Proposition 3.2.2.

1. For integers $a = [m, n]$ and $b = [k, l]$, prove that $a + b = a$ if and only if $k = l$.
2. For the uniqueness, prove that $[k, k] = [l, l]$ for any $k, l \in \mathbb{N}$.

Exercise 3.7. Prove the fourth property in Proposition 3.2.2 by showing that $[m, n]+[k, l] = [1, 1]$ if and only if $[k, l] = [n, m]$.

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. We may define *subtraction of integers* by using the negative

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

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

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

The subsequent exercises give a number of properties for the subtraction.

Exercise 3.8. Prove $[m, n] = m - n$. In other words, $[m, n] = [m + 1, 1] - [n + 1, 1]$.

Exercise 3.9. Let $a, b \in \mathbb{Z}$, explain each step in the following computation by properties in Proposition 3.2.2:

$$\begin{aligned}(a + b) + ((-a) + (-b)) &= (b + a) + ((-a) + (-b)) = ((b + a) + (-a)) + (-b) \\ &= (b + (a + (-a))) + (-b) = (b + 0) + (-b) = b + (-b) = 0.\end{aligned}$$

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

Exercise 3.10. Use properties in Proposition 3.2.2 and the cancelation law to prove the uniqueness of negative: $a + b = 0 \implies b = -a$.

Exercise 3.11. Explain $-(-a) = a$.

Exercise 3.12. Explain $a = b$ if and only if $a - b = 0$.

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

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

Exercise 3.14. Explain $(a + c) - (b + c) = a - b$ and $(-b) - (-a) = a - b$.

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 $\mathbb{Z}$. Motivated 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 $m+1+1 = n+1+1$. Applying the fourth Peano axiom twice, this implies $m = n$. It is also easy to verify that the map preserves the sum

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

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

Exercise 3.15. Prove $f(m+n) = f(m) + f(n)$. In other words, $[m+n+1, 1] = [m+1, 1] + [n+1, 1]$.

Exercise 3.16. Prove $[m, n] = m - n$. In other words, $[m, n] = [m+1, 1] - [n+1, 1]$.

Definition 3.3.1. An integer a is *bigger* than another integer b, and denoted $a > b$, if $a - b \in \mathbb{N}$. We also say b is *smaller* than a and denote $b < a$.

Proposition 3.3.2. *The order among integers has the following properties:*

1. *Trichotomy: For any two integers a and b, one of the following mutually exclusive cases happen: $a = b$, $a > b$, $a < b$.*
2. *Transitivity: $a > b$ and $b > c \implies a > c$.*
3. *Compatible with Addition: $a > b \implies a + c > b + c$.*
4. *Compatible with Negative: $a > b \implies -a < -b$.*
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 $a > b$.

We remark that, if we take $b = 0$ in the first property, then any integer a can be one and only one of the following: $a = 0$, $a > 0$, $a < 0$. By the definition of order, $a > 0$ means $a = a - 0 \in \mathbb{N}$, and $a < 0$ means $-a = 0 - a \in \mathbb{N}$. Moreover, by $-(-a) = a$ (see Exercise 3.11), we know $a < 0$ means $a = -n$ for some $n \in \mathbb{N}$. Therefore the first property can be rephrased as

$$\mathbb{Z} = \mathbb{N} \sqcup \{0\} \sqcup (-\mathbb{N}). \tag{3.3.1}$$

Then we may use the decomposition to define the *sign* $+, 0, -$ of integers.

Proof. We leave the first property to the last.

For the second property, we note that $a > b$ and $b > c$ mean $a - b = f(m)$ and $b - c = f(n)$ for some $m, n \in \mathbb{N}$. Then we get $a - c = (a - b) + (b - c) = f(m) + f(n) = f(m + n) \in \mathbb{N}$. This means $a > c$.

The third property follows from $(a+c)-(b+c) = a-b$, and the fourth property follows from $(-b) - (-a) = a - b$. See Exercise 3.14.

For the fifth property means that if a natural number is not 1, then it is bigger than 1. By Proposition 3.1.2, the number is $n' = n + 1$ for some $n \in \mathbb{N}$. Then by $n' - 1 = n \in \mathbb{N}$, we get $n' > 1$.

Now we turn to the first property. Let $c = a - b$. Then $-c = b - a$ (see Exercise 3.13). The three cases of order between a, b corresponds to the three cases for c:

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

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

3. $a < b$: This means $-c = b - a \in \mathbb{N}$.

Then the first property is the same as the following: For any $c \in \mathbb{Z}$, one of the following mutually exclusive cases happen: $c = 0$, $c \in \mathbb{N}$, $-c \in \mathbb{N}$.

To prove the property for c, we first claim that $c = [k, 1]$ or $[1, k]$ for some $k \in \mathbb{N}$. This means that, for any $m, n \in \mathbb{N}$, there is k, such that $[m, n] = [k, 1]$ or $[m, n] = [1, k]$. We fix m and prove the claim by inducting on n.

For $n = 1$, we have $[m, n] = [k, 1]$ for $k = m$.

Suppose the claim is true for n. Then for $[m, n + 1]$, by Proposition 3.1.2, we have either $m = 1$ or $m = l + 1$ for some $l \in \mathbb{N}$. If $m = 1$, then $[m, n+1] = [1, k]$ for $k = n + 1 \in \mathbb{N}$. If $m = l + 1$, then we can verify $[m, n + 1] = [l + 1, n + 1] = [l, n]$. Applying the inductive assumption to $[l, n]$, we get $[m, n + 1] = [l, n] = [k, 1]$ or $[1, k]$. This completes the inductive proof of the claim.

By Proposition 3.1.2, we have $k = 1$ or $k = l + 1$ for some $l \in \mathbb{N}$. Then $c = [k, 1] = [1, 1] = 0$, or $c = [l + 1, 1] = l$, or $c = [1, k] = [1, 1] = 0$, or $-c = -[1, l + 1] = [l + 1, 1] = l$. Therefore there are three possibilities for c: $c = 0$, $c \in \mathbb{N}$, $-c \in \mathbb{N}$.

Finally, we need to show that the three possibilities are mutually exclusive. First, if $[k + 1, 1] = [1, 1]$, then $(k + 1) + 1 = 1 + 1$. By the fourth Peano axiom, we get $k + 1 = 1$, contradicting the third Peano axiom. Therefore $[k + 1, 1] \neq [1, 1]$. By the same reason, we get $[1, k + 1] \neq [1, 1]$. Moreover, if $[k + 1, 1] = [1, l + 1]$, then $(k + 1) + (l + 1) = 1 + 1$, which leads to similar contradiction. This completes the proof. $\square$

Exercise 3.17. Prove that $m \leq n$ in the sense of Exercise 3.4, if and only if $m < n$ (in the sense of Definition 3.3.1) or $m = n$.

Exercise 3.18. Prove that $a > b$ and $c > d$ imply $a + c > b + d$.

Exercise 3.19. 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 3.20. 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 of the following holds: $x = y$, $x > y$, $y > x$.

2. Transitivity: If $x > y$ and $y > z$, then $x > z$.

Show that the relation $a > b$ for integers is an order.

Exercise 3.21. Use the trichotomy property in Proposition 3.3.2 to define $\max\{a, b\}$ and $\min\{a, b\}$. Then prove

$$\begin{aligned} \max\{a,b\} &= \max\{b,a\}, & \max\{a,\max\{b,c\}\} &= \max\{\max\{a,b\},c\}, \\ \min\{a,b\} &= \min\{b,a\}, & \min\{a,\min\{b,c\}\} &= \min\{\min\{a,b\},c\}. \end{aligned}$$

Exercise 3.22. Compare the following quantities:

$$\max\{a,\min\{b,c\}\},\ \min\{a,\max\{b,c\}\},\ \max\{\min\{a,b\},c\},\ \min\{\max\{a,b\},c\}.$$

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 verify is that the multiplication is well-defined. The argument is similar to the addition, and is left as an 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.

We prove the commutativity by double induction. We first prove $m1 = 1m$ by inducting on m. By the first property in the definition, this is the same as $1m = m$.

First, we get $1 \cdot 1 = 1$ by the first property in the definition. Moreover, under the assumption $1m = m$, we have

$$\begin{aligned} 1m' &= 1m + 1 && \text{(second property in definition)} \\ &= m + 1 && \text{(inductive assumption)} \\ &= m'. && \text{(first property in Definition 3.1.3)} \end{aligned}$$

This completes the inductive proof of $m1 = 1m$.

Next, under the inductive (on n) assumption $mn = nm$, we have

$$\begin{aligned} n'm &= (n+1)m && \text{(first property in Definition 3.1.3)} \\ &= nm + 1m && \text{(distributivity, just proved)} \\ &= mn + m && \text{(inductive assumption and } 1m = m\text{, just proved)} \\ &= mn' && \text{(second property in definition)} \end{aligned}$$

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 3.23. Explain the multiplication of natural numbers is well-defined.

Exercise 3.24. 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:* $a1 = 1 = 1a$.
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 $a0 = 0$.

In the seventh property, $(-a)b = -ab$ means $ab + (-a)b = 0$. We may prove this by using the $\Longleftarrow$ direction of the sixth property

$$\begin{aligned} ab + (-a)b &= (a + (-a))b && \text{(distributivity)} \\ &= 0b && \text{(definition of negative)} \\ &= 0. && \text{(just proved)} \end{aligned}$$

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 the remark after Proposition 3.3.2, especially (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 seventy property $(-a)b = -ab = a(-b)$ if necessary, we find $ab = \pm mn$. By the just proved first property, we know the natural number $mn > 0$. Then by (3.3.1) again, this implies $ab \neq 0$.

Finally, we prove the last property. We let $d = b - c$ and note that

$$\begin{aligned} b > c &\iff d = b - c > 0, \\ ab > ac &\iff ad = a(b - c) = ab - ac > 0. \end{aligned}$$

Therefore the problem becomes the following: If $a > 0$, then $d > 0$ if and only if $ad > 0$. By (3.3.1), we have three mutually exclusive possibilities of d:

1. $d = 0$: By the sixth property, we get $ad = 0$.
2. $d > 0$: By $a, d \in \mathbb{N}$ and the first property, we know $ad \in \mathbb{N}$. Therefore $ad > 0$.
3. $d < 0$: We have $d = -n$ for some $n \in \mathbb{N}$. By the seventh property, we get $ad = -an$. By $a, n \in \mathbb{N}$ and the first property, we get $an \in \mathbb{N}$. Then we get $ad = -an < 0$.

Since the first and the third possibilities contradict $ad > 0$, we conclude the second possibility is equivalent to $ad > 0$. This proves the last property. □

The development so far justifies all of our usual treatment of integers, particularly in terms of the addition, subtraction, 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 3.25. 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 3.26. Prove the second, third, fourth and fifth properties in Proposition 3.4.3.

Exercise 3.27. Prove that if $c < 0$, then $a > b \iff ac < bc$.

Exercise 3.28. 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 $\frac{a}{b}$.

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

$$(a, b) \sim (c, d) \iff ad = cb.$$

Instead of the notation $[a, b]$ for the equivalence classes used before, we will use the more conventional notation $\frac{a}{b}$ 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 $\frac{a}{b}$ as a divided as b, because the division operation is not yet defined. At the moment, $\frac{a}{b}$ 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:

$$g(a) = g(b) \iff (a, 1) \sim (b, 1) \iff a = a1 = b1 = b.$$

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:

$$\frac{a}{b}+\frac{c}{d}=\frac{ad+cb}{bd},\quad \frac{a}{b}\frac{c}{d}=\frac{ac}{bd}.$$

Of course we need to verify the operations are well defined. The verification is left as exercise.

By Proposition 3.2.2, we have $(ac)b = a(bc)$. This implies

$$\frac{ac}{bc}=\frac{a}{b}. \tag{3.5.1}$$

Then we get

$$\frac{a}{b}+\frac{c}{b}=\frac{ab+cb}{bb}=\frac{(a+c)b}{bb}=\frac{a+c}{b}. \tag{3.5.2}$$

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* $r+0 = r = 0+r$.
6. *Negative: For any rational number* r*, there is a unique rational number* $-r$ *satisfying* $r + (-r) = 0 = (-r) + r$.
7. *One: The integer* 1 *is the unique rational number satisfying* $r1 = r = 1r$.
8. *Reciprocal: For any rational number* $r \neq 0$*, there is a unique rational number* r^{-1} *satisfying* $rr^{-1} = 1 = r^{-1}r$.

Proof. The following verifies the first property (for the first, see (3.5.2))

$$\frac{a}{1}+\frac{b}{1}=\frac{a+b}{1},\quad \frac{a}{1}\frac{b}{1}=\frac{ab}{1\cdot 1}=\frac{ab}{1}.$$

The second, third and fourth properties can be similarly verified, and are left as exercises.

For the fifth property, we have

$$\frac{a}{b}+\frac{0}{1}=\frac{a1+0b}{b1}=\frac{a+0}{b}=\frac{a}{b}.$$

By the second (commutative) property, we also have $\frac{0}{1}+\frac{a}{b}=\frac{a}{b}$. Therefore $g(0)=\frac{0}{1}$ can be used as the rational zero.

The uniqueness of zero can be proved in exactly the same way as the proof of Proposition 3.2.2. Suppose 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 $0=0+\bar{0}=\bar{0}$.

For the sixth property, we have (for the last equality, see (3.5.1))

$$\frac{a}{b}+\frac{-a}{b}=\frac{a+(-a)}{b}=\frac{0}{b}=\frac{0b}{1b}=\frac{0}{1}.$$

By the second (commutative) property, this means that, for $r=\frac{a}{b}$, the number $-r=\frac{-a}{b}$ satisfies the sixth property. This proves the existence of the negative. The uniqueness of the negative can be proved in exactly the same way as the proof of Proposition 3.2.2.

For the seventh property, the following (and the commutative property) shows $g(1)$ can be used as 1

$$\frac{a}{b}\frac{1}{1}=\frac{a1}{b1}=\frac{a}{b}.$$

The uniqueness of 1 can be proved similar to the uniqueness of 0: If $r1=r=1r$ and $r\bar{1}=r=\bar{1}r$, then $1=1\cdot\bar{1}=\bar{1}$.

For the eighth property, we first note that $\frac{a}{b}=0=\frac{0}{1}$ means $a=a1=b0=0$. Therefore $r=\frac{a}{b}\neq 0$ means $a\neq 0$. Therefore $\frac{b}{a}$ is also a rational number, and the following (and the commutative property) shows this can be the reciprocal ((3.5.1) used in the last equality)

$$\frac{a}{b}\frac{b}{a}=\frac{ab}{ba}=\frac{1ab}{1ab}=\frac{1}{1}.$$

The uniqueness of the reciprocal can be proved similar to the uniqueness of the negative. □

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, with $b \neq 0$, the following shows that $\frac{a}{b}$ is indeed the division of a by b

$$a \div b = \frac{a}{1}\left(\frac{b}{1}\right)^{-1} = \frac{a}{1}\frac{1}{b} = \frac{a1}{1b} = \frac{a}{b}.$$

Here the second equality makes use of the formula for r^{-1} in the proof of Proposition 3.5.2. Therefore we will also write $\frac{r}{s}$ for the division $r \div s$ of rational numbers. In particular, we have

$$\frac{1}{s} = 1 \div s = 1s^{-1} = s^{-1}.$$

Exercise 3.29. Prove the addition and multiplication of rational numbers are well defined.

Exercise 3.30. Prove the second, third and fourth properties in Proposition 3.5.2.

Exercise 3.31. Prove the uniqueness of negative and reciprocal for rational numbers.

Exercise 3.32. Use Proposition 3.5.2 to prove the cancelation laws for rational numbers:

1. $r + t = s + t$ implies $r = s$.
2. $rt = st$ and $t \neq 0$ imply $r = s$.

Exercise 3.33. Prove that $rs = 0$ if and only if $r = 0$ or $s = 0$.

Exercise 3.34. 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 discuss the order among rational numbers: We define $r > s$, if $r - s = \frac{a}{b} > 0$ for some $a > 0$ and $b > 0$.

By $\frac{a}{b} = \frac{-a}{-b}$, we may always write $r - s = \frac{a}{b}$ for some $b > 0$. Then there are three possibilities for a, corresponding to three possibilities for $r - s$:

1. $a = 0$: We have $r - s = \frac{0}{b} = 0$. Therefore $r = s$.
2. $a > 0$: By $a, b > 0$, we get $r > s$.
3. $a < 0$: We have $s - r = -(r - s) = \frac{-a}{b}$. By $-a, b > 0$, we get $s > r$.

The discussion is independent of the choice of the expression $\frac{a}{b}$ with $b > 0$, for the following reason: Suppose $\frac{a}{b} = \frac{c}{d}$ 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 $r > s \iff r^{-1} < s^{-1}$.*

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, by $r > s$ and $s > t$, we have

$$r - s = \frac{a}{b} \text{ and } s - t = \frac{c}{d}, \text{ where } a, b, c, d > 0.$$

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 property is due to $(r + t) - (s + t) = r - s$.

The fourth property is due to $(-s) - (-r) = r - s$.

For the fifth property, we first prove the special case that $r, s > 0$ implies $rs > 0$ (taking $t = 0$ in the fifth property), and the property $r > 0$ implies $\frac{1}{r} = r^{-1} > 0$. By $r > 0$ and $s > 0$, we have

$$r = \frac{a}{b} \text{ and } s = \frac{c}{d}, \text{ where } a, b, c, d > 0.$$

Then

$$rs = \frac{ac}{bd}, \quad r^{-1} = \frac{b}{a}.$$

By $a, b, c, d > 0$, we get $ac > 0$ and $bd > 0$. Therefore $rs > 0$. We also get $b, a > 0$. Therefore $r^{-1} > 0$.

For the general case of the fifth property, we note that $rs - rt = r(s - t)$. By the third property, we know $s > t$ if and only if $s - t > 0$, and $rs > rt$ if and only if $rs - rt > 0$. Therefore for $u = s - t$, the property becomes that, if $r > 0$, then $u > 0$ if and only if $ru > 0$. From the special case, we already know $r > 0$ and $u > 0$ imply $ru > 0$. Conversely, suppose $r > 0$ and $ru > 0$. By the special case proved above, we know $r^{-1} > 0$. Then applying the special case to $r^{-1} > 0$ and $ru > 0$, we get $u = r^{-1}(ru) > 0$. This completes the proof of the fifth property.

The sixth property can be obtained by repeatedly applying the fifth property

$$s^{-1} > r^{-1} \iff 1 = ss^{-1} > sr^{-1} \iff r = 1r > sr^{-1}r = s.$$

For the seventh property, we may choose $t = \dfrac{r+s}{2}$. Then $r - t = t - s = \dfrac{r-s}{2}$. By the fifth property, we get $\dfrac{r-s}{2} = (r - s) \cdot 2^{-1} > 0$. Then by the third property, we get $r > t$ and $t > s$.

For the eighth property, the assumption means $r = \dfrac{a}{b}$ for some $a, b > 0$. By $\dfrac{a}{b} = \dfrac{2a}{2b}$, we may also assume that $a, b > 1$. Then we take $n = \max\{a, b\}$ (see Exercise 3.21). 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 3.3.2, we may introduce $r \geq s$. Exercise 3.21 about $\max\{r, s\}$ and $\min\{r, s\}$ can also be extended to rational numbers. We also define the *absolute value* of a rational number

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

Exercise 3.35 gives important properties about absolute value.

$$|r + s| \leq |r| + |s|, \quad |rs| = |r||s|, \quad |r| < s \iff -s < r < s.$$

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 3.35. Prove properties of the absolute values

$$|r+s| \le |r|+|s|, \quad |rs| = |r||s|, \quad |r| < s \iff -s < r < s.$$

Exercise 3.36. 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 real numbers in infinite decimal expansions, such as

$$\begin{aligned} \frac{1}{3} &= 0.33333333\cdots, \\ \sqrt{2} &= 1.41421356\cdots, \\ \pi &= 3.14159265\cdots. \end{aligned}$$

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. There is $l \in \mathbb{Q}$, such all $r \in X$ satisfy $r > l$.
2. If $s > r \in X$, then $s \in X$.
3. If $r \in X$, then there is $s \in X$, such that $r > s$.

The collection of all real numbers is denoted $\mathbb{R}$.

The first condition means X has lower bound. The condition is needed if we wish to use the lower bound of X to define real numbers.

The second condition means X is maximal. It implies that X is the rational number interval $(x, \infty) \cap \mathbb{Q}$ or $[x, \infty) \cap \mathbb{Q}$. In fact, a set of real numbers satisfy the second condition if and only if it is (x, ∞) or $[x, \infty)$.

The third condition is means unambiguously picking $(x, \infty) \cap \mathbb{Q}$ from the two possible ways of representing x. In fact, if x is rational, then $x \notin (x, \infty) \cap \mathbb{Q}$, and $x \in [x, \infty) \cap \mathbb{Q}$. The two ways are different.

We use capital letter X, Y, Z, etc., when we think of real numbers as subsets. If we forget about the subsets and just think about real numbers in the usual sense, then it is more comfortable to use lower case letters x, y, z, etc. In what follows,

[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.

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.

We develop operations and properties of real numbers by using rational numbers. Unlike the algebraic relation such as $-2 = [1, 3]$ for $\mathbb{N}$ and $\mathbb{Z}$ used before, the relation between $\mathbb{R}$ and $\mathbb{Q}$ is approximation. The following is the key technical lemma about such approximation.

Lemma 3.6.2. *For any real number X and any rational number $\epsilon > 0$, there are $r \in X$ and $s \notin X$, such that $r - s = \epsilon$.*

The lemma means that, for any real number x and rational number $\epsilon > 0$, there are rational numbers r, s, such that $r > x > s$ and $r - s = \epsilon$. Note that the numbers in $r - s = \epsilon$ must be rational, since we have not yet defined arithmetic operations in $\mathbb{R}$.

The lemma implies $|x - r| < \epsilon$ and $|x - s| < \epsilon$, which means approximations of x by rational numbers r and s. Again the inequalities do not yet make sense, because arithmetic operations and order have not been defined in $\mathbb{R}$.

Proof. Since X is not empty, we have rational $p \in X$. By the first condition in Definition 3.6.1, we also have rational lower bound l of X. Applying the eighth property in Proposition 3.5.3 to $\dfrac{p}{\epsilon}$ and $-\dfrac{l}{\epsilon}$, we get natural number m, n satisfying $\dfrac{p}{\epsilon} < m$ and $-\dfrac{l}{\epsilon} < n$. Then $p < m\epsilon$ and $l > -n\epsilon$. By $p \in X$ and the second condition in Definition 3.6.1, we know $m\epsilon \in X$. By $l > -n\epsilon$ and the lower bound l of X, we know $-n\epsilon \notin X$.

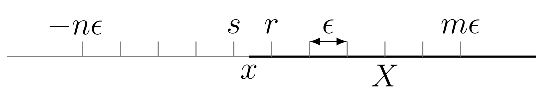

Figure 3.1: Approximate real numbers by rational numbers.

We get an decreasing sequence of rational numbers from $m\epsilon$ to $-n\epsilon$

$$m\epsilon,\ (m-1)\epsilon,\ (m-2)\epsilon,\ \dots,\ (-n+1)\epsilon,\ -n\epsilon.$$

Since $m\epsilon \in X$ to $-n\epsilon \notin X$, there are two adjacent terms in the sequence, say $r = k\epsilon$ and $s = (k-1)\epsilon$, such that $r \in X$ and $s \notin X$. Moreover, we have $r - s = \epsilon$. □

Exercise 3.37. For any real number X, explain that $r \in X$ and $s \notin X$ imply $r > s$. Also explain that $r \notin X$ and $s < r$ imply $s \notin X$.

Exercise 3.38. 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 3.39. 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?

We consider rational numbers as part of real numbers, through the following map

$$h(r) = \{s \in \mathbb{Q} \colon s > r\} \colon \mathbb{Q} \to \mathbb{R}.$$

By the seventh property in Proposition 3.5.3, the map is one-to-one.

Define the addition of two real numbers

$$X + Y = \{r + s \colon r \in X,\ s \in Y\}.$$

The following verifies that $X + Y$ satisfies the conditions Definition 3.6.1 (which we call $\mathbb{R}$-conditions):

1. Suppose l_X and l_Y are lower bounds of X and Y. Then for any $r \in X$ and $s \in Y$, we have $r > l_X$ and $s > l_Y$. Then $r + s > l_X + l_Y$. Therefore $X + Y$ satisfies the first $\mathbb{R}$-condition for $l = l_X + l_Y$.

2. Suppose a rational number $t > r+s$ for some $r \in X$ and $s \in Y$. 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 $r' = t - s' > r$, and $t = r' + s'$. By $r' > r \in X$ and $s' > s \in Y$, and the second $\mathbb{R}$-condition, we get $r' \in X$ and $s' \in Y$. Therefore $t = r' + s' \in X + Y$.

3. Consider $t = r + s \in X + Y$ for some $r \in X$ and $s \in Y$. By the third $\mathbb{R}$-condition for X and Y, there are $r' \in X$ and $s' \in Y$, such that $r > r'$ and $s > s'$. Then $t' = r' + s' \in X + Y$ satisfies $t > t'$.

Proposition 3.6.3. *The addition of real numbers is consistent with the addition of rational numbers, and 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* $x + (-x) = 0 = (-x) + x$.

Proof. The consistency with the addition of rational numbers means $h(r + s) = h(r) + h(s)$. This is an equality of two subsets.

A number inside $h(r)+h(s)$ is the addition $r'+s'$ of rational numbers, such that $r' > r$ and $s' > s$. Then $r' + s' > r + s$. This means $r' + s' \in h(r + s)$. This proves $h(r) + h(s) \subset h(r + s)$.

A number inside $h(r+s)$ is a rational number $t > r+s$. By the same argument as the verification above of the second condition for $X+Y$, we get $t = r'+s'$ for some rational numbers $r' > r$ and $s' > s$. Then $r' \in h(r)$, $s' \in h(s)$, and $t = r'+s' \in h(r)+h(s)$. This proves $h(r+s) \subset h(r)+h(s)$.

This completes the proof of $h(r+s) = h(r)+h(s)$.

For the first and second properties, we have

$$\begin{aligned} X+(Y+Z) &= \{r+(s+t) \colon r \in X,\ s \in Y,\ t \in Z\}, \\ (X+Y)+Z &= \{(r+s)+t \colon r \in X,\ s \in Y,\ t \in Z\}, \\ X+Y &= \{r+s \colon r \in X,\ s \in Y\}, \\ Y+X &= \{s+r \colon s \in Y,\ r \in X\}. \end{aligned}$$

By $r+(s+t) = (r+s)+t$ and $r+s = s+r$, we get $X+(Y+Z) = (X+Y)+Z$ and $X+Y = Y+X$. This completes the proof of the two properties.

For the third property, by $h(0) = \{s \in \mathbb{Q} \colon s > 0\}$, we have

$$X + h(0) = \{r+s \colon r \in X,\ s \in \mathbb{Q},\ s > 0\}.$$

By $r+s > r \in X$ and the second $\mathbb{R}$-condition, we get $r+s \in X$. This proves $X+h(0) \subset X$. On the other hand, for $r \in X$, by the third $\mathbb{R}$-condition, 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 $X \subset X+h(0)$. The uniqueness of zero can be proved in exactly the same way as the proof of Proposition 3.2.2. This completes the proof of the third property.

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 actually gives $Y = [-x, +\infty) \cap \mathbb{Q}$. So we further modify Y to get $(-x, +\infty) \cap \mathbb{Q}$

$$Z = \{s \in \mathbb{Q} \colon s > t \text{ for some } t \in Y\}.$$

The following verifies that Z satisfies the three $\mathbb{R}$-conditions:

1. 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.

2. Suppose $s' > s \in Z$. We have $s > t$ for some $t \in Y$. Then $s' > t$ for the same $t \in Y$. Therefore $s' \in Z$.

3. Suppose $s \in Z$. Then $s > t$ for some $t \in Y$. By the seventh property in Proposition 3.5.3, there is $s' \in \mathbb{Q}$ satisfying $s > s' > t$. By $s' > t \in Y$, we get $s' \in Z$, which also satisfies $s > s'$.

It remains to 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 $\epsilon \in h(0)$, we have $\epsilon > 0$. By Lemma 3.6.2, there are $r \in X$ and $s \notin X$, such that $r - s = \frac{\epsilon}{2}$. We claim $-s + \frac{\epsilon}{2} \in Z$. By $-s + \frac{\epsilon}{2} > -s$, we only need to prove $-s \in Y$. This means that for any $r' \in X$, we need to prove $r + (-s) = r - s > 0$. By $r' \in X$, and $s \notin X$, and the second $\mathbb{R}$-condition (see Exercise 3.37), we get $r' > s$. This means $r' - s > 0$. Then the claim is proved. Now we have $\epsilon = r + (-s + \frac{\epsilon}{2})$, with $r \in X$ and $-s + \frac{\epsilon}{2} \in Z$. Therefore $\epsilon \in X + Z$. This proves $h(0) \subset X + Z$.

We conclude $X + Z = h(0)$. The uniqueness of the negative can be proved in exactly the same way as the proof of Proposition 3.2.2. This completes the proof of the fourth property. □

Exercise 3.40. 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 3.41. What are the rational numbers in $-\sqrt{2}$? What are the rational numbers in $1 - \sqrt{2}$?

Next we introduce order in $\mathbb{R}$. We define $X \geq Y$ if $X \subset Y$. This is the same as

$$X > Y \iff X \subset Y \text{ and } X \neq Y.$$

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 $h(r) > h(s)$.

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, $X = \{r \in \mathbb{Q} \colon r > X\}$.*

Proof. We know $r > X$ 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 $\mathbb{R}$-condition, the second statement means $r \in X$. Conversely, if $r \in X$, then the first statement holds by the first $\mathbb{R}$-condition, 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. Suppose $X - Y \neq \emptyset$. Then we have $r \in X - Y$. For any $s \in Y$, by $r \notin Y$ and the second $\mathbb{R}$-condition (see Exercise 3.37), we get $r < s$. Applying the second $\mathbb{R}$-condition again to $r \in X$, we get $s \in X$. This proves $Y \subset X$. By $X - Y \neq \emptyset$, we also get $X \neq Y$. Therefore $X < Y$.

We just proved $X - Y \neq \emptyset$ implies $X < Y$. Therefore there are three mutually exclusive possibilities for any two real numbers X and Y:

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.

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 following proves the fourth property

$$x > y \implies -y = x - (x + y) > y - (x + y) = -x.$$

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 $\mathbb{R}$-condition, 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.

The *infimum* of a number set A is the greatest lower bound. More specifically, $l = \inf A$ is characterised by two properties:

1. l is a lower bound: $x \geq l$ for all $x \in A$.
2. Any number bigger than l is not lower bound: If $l' > l$, then there is $x \in A$ satisfying $x < l'$.

The *supremum* $k = \sup A$ is the least upper bound, and is characterised by two properties:

1. k is an upper bound: $x \le k$ for all $x \in A$.
2. Any number smaller than k is not upper bound: If $k' < k$, then there is $x \in A$ satisfying $x > k'$.

We clearly have

$$\sup A = -\inf(-A) \text{ for } -A = \{-x \colon x \in A\}.$$

Theorem 3.6.6. *Any set of real numbers with upper bound has a real number as the supremum. Any set of real numbers with lower bound has a real number as the infimum.*

Proof. Let A be a set of real numbers with a lower bound l. Define

$$Z = \{r \in \mathbb{Q} \colon r > x \text{ for some } x \in A\}.$$

We verify the three $\mathbb{R}$-conditions (similar to Exercise 3.40):

1. We know all $x \in A$ satisfy $x \ge l$. Then $r > x$ and $x \ge l$ imply $r > l$. Therefore l is a lower bound of Z.
2. Suppose $s > r \in Z$. We know there is $x \in A$ satisfying $r > x$. Then by $s > r$, we get $s > x$ for the same $x \in A$. Therefore $s \in Z$.
3. 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 $s \in Z$.

Therefore Z is a real number. We will verify that Z is the infimum of A.

Let $X = x \in A$. 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.

Let a real number $W > Z$. Then we have $r \in Z - W$ (see the proof of the first property in Proposition 3.6.5). 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 $X - W \neq \emptyset$. Therefore $X < W$, ad 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. $\square$

Exercise 3.42. 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 3.43. Give an alternative proof of Proposition 3.6.6 by constructing $Z = \cup\{X\colon X \in A\}$.

Next we 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.

We remark that $X \geq 0$ 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 $n > X > \dfrac{1}{n}$.*
2. *For any rational number $\epsilon > 0$, there are rational numbers $r \in X$ and $s \notin X$, such that $s > 0$ and $\dfrac{r}{s} < 1 + \epsilon$.*

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 $r > s$. Therefore $\dfrac{r}{s} > 1$.

Proof. For the first statement, by $X > 0$ and the fifth property in Proposition 3.6.5, there is $t \in \mathbb{Q}$ 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 $n_1 > r > \dfrac{1}{n_1}$ and $n_2 > t > \dfrac{1}{n_2}$. Then the natural number $n = \max\{n_1, n_2\}$ satisfies $n > r > X > t > \dfrac{1}{n}$.

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 $\dfrac{r}{s} - 1 < \epsilon$. $\square$

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 the rational numbers in X and Y are positive. The following verifies the three $\mathbb{R}$-conditions:

1. 0 is a lower bound of XY.
2. Suppose a rational number $t > rs$ for some $r \in X$ and $s \in Y$. Then $r > 0$ and $\dfrac{t}{r} > s$. By $s \in Y$, this implies $\dfrac{t}{r} \in Y$. Then $t = r\dfrac{t}{r} \in XY$.

3. Consider $rs \in XY$ 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 $rs > r's'$, and $r's' \in XY$.

Proposition 3.6.8. *The multiplication of non-negative real numbers is compatible with the multiplication of non-negative rational numbers, and 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:* $x0 = 0 = 0x$.
5. *One:* $x1 = x = 1x$.
6. *Reciprocal: For any real number* $x > 0$, *there is a unique real number* x^{-1} *satisfying* $xx^{-1} = 1 = x^{-1}x$.
7. $x > 0$ *and* $y > 0$ *imply* $xy > 0$.

Proof. The consistency with rational numbers means $h(rs) = h(r)h(s)$ for rational numbers $r, s \geq 0$. This means the following are equivalent for $t \in \mathbb{Q}$:

$$t > rs \iff t = r's' \text{ for some } r', s' \in \mathbb{Q} \text{ satisfying } r' > r \text{ and } s' > s.$$

The $\Longleftarrow$ direction is obvious. For the $\Longrightarrow$ direction, we note that $t > rs$ implies $\frac{t}{r} > s$. By the seventh property in Proposition 3.5.3, there is s' satisfying $\frac{t}{r} > s' > s$. Then we have $t = r's'$ for $r' = \frac{t}{s'} > r$ and $s' > s$.

The first property follows from $r(st) = (rs)t$, and

$$X(YZ) = \{r(st) \colon r \in X,\ s \in Y,\ t \in Z\},$$
$$(XY)Z = \{(rs)t \colon r \in X,\ s \in Y,\ t \in Z\}.$$

The second property can be proved in the similar way.

A rational number in $(X + Y)Z$ is $(r + s)t$ with $r \in X$, $s \in Y$, $t \in Z$. By $(r + s)t = rt + st \in XZ + YZ$, we get $(X + Y)Z \subset XZ + YZ$. On the other hand, 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 $\mathbb{R}$-condition, 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 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 $\frac{s}{r} > 0$, and $s = r\frac{s}{r} \in Xh(0)$. This proves $h(0) \subset Xh(0)$, and completes the proof of the fourth property.

A rational number in $Xh(1)$ is 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 $\frac{r}{s} > 1$, and $r = s\frac{r}{s} \in Xh(1)$. This proves $h(1) \subset Xh(1)$, and completes the proof of the fifth property.

We introduce the reciprocal similar to the negative in Proposition 3.6.3. First, we let

$$Y = \{t \in \mathbb{Q} \colon rt > 1 \text{ for all } r \in X\}.$$

Then we further modify Y to get

$$Z = \{s \in \mathbb{Q} \colon s > t \text{ for some } t \in Y\}.$$

The positivity $X > 0$ is needed to get $Y \neq \emptyset$ (and then $Z \neq \emptyset$). By the fifth property in Proposition 3.6.5, there is rational l satisfying $X > l > 0$. Let $t = \frac{1}{l}$. Then any $r \in X$ satisfies $r > X > l$, and we get $rt = \frac{r}{l} > 1$.

Similar to the proof of Proposition 3.6.3 (and see Exercise 3.40), we can verify that Z satisfies the three $\mathbb{R}$-conditions.

For $r \in X$ and $s \in Z$, we have $rs > rt > 1$. Therefore $rs \in h(1)$. This proves $XZ \subset h(1)$. On the other hand, any $u \in h(1)$ satisfies $u > 1$. By Lemma 3.6.7, there are $r \in X$ and $v \notin X$ satisfying $\frac{r}{v} < u$. Then $u = r\frac{u}{r}$. By $r \in X$, if we can show $\frac{u}{r} \in Z$, then we get $u \in XZ$, and this proves $h(1) \subset XY$.

By $\frac{r}{v} < u$, we get $\frac{u}{r} > \frac{1}{v}$. Therefore to show $\frac{u}{r} \in Z$, we only need to show $\frac{1}{v} \in Y$. For any $r' \in X$, by $v \notin X$, we have $r' > v$. Therefore we get $r'\frac{1}{v} = \frac{r'}{v} > 1$. This verifies $\frac{1}{v} \in Y$, and completes the proof of $XZ = h(1)$.

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 $xy > 0$. $\square$

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 result is always 0.

Proposition 3.6.9. *The multiplication of real numbers is compatible with the multiplication of rational numbers, and 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:* $x0 = 0 = 0x$.
5. *One:* $x1 = x = 1x$.
6. *Reciprocal: For any real number* $x \neq 0$, *there is a unique real number* x^{-1} *satisfying* $xx^{-1} = 1 = x^{-1}x$.
7. *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.

For the case $x + y \geq 0$, $y \leq 0$ and $z \geq 0$, we have $x + y, -y, z \geq 0$. By the non-negative distributivity in Proposition 3.6.8, we have

$$(x + y)z + (-y)z = [(x + y) + (-y)]z = xz.$$

On the other hand, for $y \leq 0$ and $z \geq 0$, we have $-(-y)z = yx$ by the definition. Adding the equality to the equality above, 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. $\square$

Exercise 3.44. For $x, y \leq 0$ and $z \geq 0$, prove $(xy)z = x(yz)$ and $xy = yx$.

Exercise 3.45. For $x, y, z \leq 0$, prove $(x + y)z = xz + yz$.

Exercise 3.46. Prove that, if $x, y > 0$, then $x > y$ implies $x^{-1} < y^{-1}$.

Exercise 3.47. Define x^{-1} for the case $x < 0$, and prove the sixth property in Proposition 3.6.9.

Exercise 3.48. Prove the seventh property in Proposition 3.6.9.

Exercise 3.49. 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

$$xy = x_1 y_1 + x_2 y_2 - x_1 y_2 - x_2 y_1.$$

Carry out the details of this approach.

3.7 Exponential

In general, the exponential x^y is defined 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

$$x^n = \underbrace{x \cdot x \cdots x}_{n}.$$

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.

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 $x^{m+n} = x^m x^n$. $\square$

Exercise 3.50. Prove the other properties in Proposition 3.7.1.

Next, we define the exponential for $y = \dfrac{1}{n}$, $n \in \mathbb{N}$. We actually get 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 $(x^{\frac{1}{n}})^n = x$.*

We remark that, by Proposition 3.7.1, we have

$$(x^{\frac{1}{n}})^{mn} = ((x^{\frac{1}{n}})^n)^m = x^m. \tag{3.7.1}$$

We also remark that $0^{\frac{1}{n}} = 0$. Moreover, for $x < 0$ and odd n, we may define $x^{\frac{1}{n}} = -|x|^{\frac{1}{n}} = -(-x)^{\frac{1}{n}}$. The extended definition still satisfies $(x^{\frac{1}{n}})^n = x$.

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 subset has infimum because 0 is a lower bound. In fact, $u = \min\{1, x\}$ satisfies $u^n = uu^{n-1} \le x1^{n-1} = x < z^n$. 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 verify $w^n = x$, by showing that both $w^n > x$ and $w^n < x$ 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 any v satisfying $w < v < 2w$, 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 $v^n < x$. By the last property in Proposition 3.7.1, for all $z > 0$ satisfying $z^n > x$, we get $v < z$. 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 $v = w + \min\left\{w, \dfrac{x - w^n}{n2^n w^{n-1}}\right\}$.

This proves the existence of $w = x^{\frac{1}{n}}$. For the uniqueness, by the last property in Proposition 3.7.1, we know $w_1 > w_2 > 0$ implies $w_1^n > w_2^n$. Therefore $w_1 \ne w_2$ implies $w_1^n \ne w_2^n$. $\square$

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 $m_1n_2 = m_2n_1$. By (3.7.1), we get $((x^{\frac{1}{n_1}})^{m_1})^{n_1n_2} = (x^{\frac{1}{n_1}})^{m_1n_1n_2} = x^{m_1n_2}$. Similarly, we get $((x^{m_2})^{\frac{1}{n_2}})^{n_1n_2} = x^{m_2n_1}$. Then by $m_1n_2 = m_2n_1$, we get $((x^{\frac{1}{n_1}})^{m_1})^{n_1n_2} = ((x^{\frac{1}{n_2}})^{m_2})^{n_1n_2}$. Then by the uniqueness in Proposition 3.7.2, we get $(x^{\frac{1}{n_1}})^{m_1} = (x^{\frac{1}{n_2}})^{m_2}$.

Next, we show $x^{r+s} = x^r x^s$ for $r, s > 0$. Let $r = \dfrac{m}{n}$ and $s = \dfrac{k}{l}$, with $m, n, k, l \in \mathbb{N}$. Then by Proposition 3.7.1 and (3.7.1), we get

$$\begin{aligned}(x^{\frac{m}{n}+\frac{k}{l}})^{nl} &= (x^{\frac{ml+kn}{nl}})^{nl} = ((x^{\frac{1}{nl}})^{ml+kn})^{nl} = (x^{\frac{1}{nl}})^{(ml+kn)nl} = x^{ml+kn} \\ &= x^{ml}x^{kn} = (x^{\frac{1}{n}})^{mnl}(x^{\frac{1}{l}})^{knl} = ((x^{\frac{1}{n}})^m)^{nl}((x^{\frac{1}{l}})^k)^{nl} = (x^{\frac{m}{n}}x^{\frac{k}{l}})^{nl}.\end{aligned}$$

Then by the uniqueness in Proposition 3.7.2, we get $x^{\frac{m}{n}+\frac{k}{l}} = x^{\frac{m}{n}}x^{\frac{k}{l}}$.

Next we extend to x^r for any rational number r. We write $r = r_1 - r_2$ with rational $r_1, r_2 > 0$, and define $x^r = \dfrac{x^{r_1}}{x^{r_2}}$. To see this is well defined, we assume $r_1 - r_2 = s_1 - s_2$, with $r_1, r_2, s_1, s_2 > 0$. 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.

By $0 = r - r$, we get $x^0 = 1$. In fact, we also define $0^0 = 1$.

For $r > 0$, we write $-r = 1-(1+r)$ and get $x^{-r} = \dfrac{x^1}{x^{1+r}} = \dfrac{x}{xx^r} = \dfrac{1}{x^r}$. 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:*

$$(xy)^r = x^r y^r, \quad x^{r+s} = x^r x^s, \quad x^{rs} = (x^r)^s,$$

$$x > 1,\ r > s \implies x^r > x^s,$$

$$x > y > 0 \implies \begin{cases} x^r > y^r, & \text{if } r > 0 \\ x^r < y^r, & \text{if } r < 0 \end{cases}.$$

Proof. We already proved $x^{r+s} = x^r x^s$ for $r, s > 0$. In general, let $r = r_1 - r_2$ and $s = s_1 - s_2$ with rational $r_1, r_2, s_2, s_2 > 0$. Then

$$x^{r+s} = x^{(r_1+s_1)-(r_2+s_2)} = \frac{x^{r_1+s_1}}{x^{r_2+s_2}} = \frac{x^{r_1}x^{s_1}}{x^{r_2}x^{s_2}} = \frac{x^{r_1}}{x^{r_2}}\frac{x^{s_1}}{x^{s_2}} = x^r x^s.$$

The equality $(xy)^r = x^r y^r$ can be proved in similar way. First, we prove the equality for $r > 0$. Let $r = \dfrac{m}{n}$. Then by Proposition 3.7.1 and (3.7.1), we get

$$((xy)^{\frac{m}{n}})^n = (xy)^m = x^m y^m = (x^{\frac{m}{n}})^n (y^{\frac{m}{n}})^n = (x^{\frac{m}{n}}y^{\frac{m}{n}})^n.$$

Then by the uniqueness in Proposition 3.7.2, we get $(xy)^{\frac{m}{n}} = x^{\frac{m}{n}}y^{\frac{m}{n}}$. Next, for $r = r_1 - r_2$ with rational $r_1, r_2 > 0$, we get

$$(xy)^r = \frac{(xy)^{r_1}}{(xy)^{r_2}} = \frac{x^{r_1}y^{r_1}}{x^{r_2}y^{r_2}} = \frac{x^{r_1}}{x^{r_2}}\frac{y^{r_1}}{y^{r_2}} = x^r y^r.$$

The proof of $x^{rs} = (x^r)^s$ is also similar, and is left as an exercise.

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 $x^{\frac{m}{n}} \leq 1$. Therefore we have $x^{r-s} = x^{\frac{m}{n}} > 1$. Then by the second equality that we already proved, we get $x^r = x^{r-s}x^s > 1x^s = x^s$.

Finally, suppose $x > y > 0$ and $r > 0$. Then $\dfrac{x}{y} > 1$. Then by the inequality that we already proved, we have $\left(\dfrac{x}{y}\right)^r > \left(\dfrac{x}{y}\right)^0 = 1$. Then by the first equality that we already proved, we get $y^r < \left(\dfrac{x}{y}\right)^r y^r = \left(\dfrac{x}{y}y\right)^r = x^r$. $\square$

Exercise 3.51. Prove $x^{rs} = (x^r)^s$ in Proposition 3.7.3.

Next we extend the exponential x^y to any real y. First, we consider the case $x > 1$, and define

$$x^y = \inf\{x^r : r \in \mathbb{Q},\ r > y\} = \inf\{x^r : r \in Y\}.$$

The second equality follows from Lemma 3.6.4.

We need to verify the consistency with the rational exponential. This means that, for any rational $y = s \in \mathbb{Q}$, the rational exponential 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 $l > x^s$, there is $r > s$, such that $x^r < l$.

The first follows from Proposition 3.7.3. For the second, we 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 $z_n > 0$. By Exercise 1.32, we have

$$x - 1 = (1 + z_n)^n - 1^n \geq nz_n1^{n+1} = nz_n.$$

By Lemma 3.6.7, there is $n \in \mathbb{N}$ satisfying $\dfrac{x-1}{\epsilon} < n$. Then we get $x^{\frac{1}{n}} - 1 = z_n \leq \dfrac{x-1}{n} < \epsilon$. $\square$

In the second property, we take $\epsilon = \dfrac{l}{x^s} - 1$. By $l > x^s$, we have $\epsilon > 0$. Then there is $n \in \mathbb{N}$ satisfying $x^{\frac{1}{n}} - 1 < \dfrac{l}{x^s} - 1$. This means $r = s + \dfrac{1}{n} > s$ satisfies $x^r = x^s x^{\frac{1}{n}} < l$.

To further extend x^y from $x > 1$ to any $x > 0$, we 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 need the following properties of the infimum.

Lemma 3.7.5. *Suppose A and B are sets of real numbers with lower bounds.*

1. *If $A + B = \{x + y : x \in A,\ y \in B\}$, then $\inf(A + B) = \inf A + \inf B$.*
2. *If both A and B contain only positive real numbers, and $AB = \{xy : x \in A,\ y \in B\}$, then $\inf AB = \inf A \inf B$.*

Proof. For $x \in A$ and $y \in B$, we have $x \geq \inf A$ and $y \geq \inf B$. Then $x + y \geq \inf A + \inf B$. This proves that $\inf A + \inf B$ 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 satisfy $z - \inf B > z_A > \inf A$, and then 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 $A + B$. The proof for the second statement is similar. □

For $x, y > 1$, by Lemma 3.7.5, we have $xy > 1$, and

$$\begin{aligned}(xy)^z &= \inf\{(xy)^r : r > z\} = \inf\{x^r y^r : r > z\}, \\ x^z y^z &= \inf\{x^r : r > z\} \inf\{y^s : s > z\} = \inf\{x^r y^s : r, s > z\}.\end{aligned}$$

Since $x^r y^s$ lies between $x^r y^r$ and $x^s y^s$, the two infima 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:*

$$(xy)^z = x^z y^z, \quad x^{y+z} = x^y x^z, \quad x^{yz} = (x^y)^z,$$

$$x > 1,\ y > z \implies x^y > x^z,$$

$$x > y > 0 \implies \begin{cases} x^z > y^z, & \text{if } z > 0 \\ x^z < y^z, & \text{if } z < 0 \end{cases}.$$

Proof. We already proved the first equality for $x, y > 1$. In general, let $x = \dfrac{x_1}{x_2}$ and $y = \dfrac{y_1}{y_2}$, with $x_1, x_2, y_1, y_2 > 1$. Then $xy = \dfrac{x_1y_1}{x_2y_2}$, with $x_1y_1, x_2y_2 > 1$, and

$$(xy)^z = \frac{(x_1y_1)^z}{(x_2y_2)^z} = \frac{x_1^z y_1^z}{x_2^z y_2^z} = \frac{x_1^z}{x_2^z}\frac{y_1^z}{y_2^z} = x^z y^z.$$

For the second equality, the following proves the case $x > 1$:

$$\begin{aligned}
x^{y+z} &= \inf\{x^r \colon r \in Y + Z\} && \text{(definition of } x^{y+z}\text{)}\\
&= \inf\{x^{s+t} \colon s \in Y,\ t \in Z\} && \text{(definition of } Y + Z\text{)}\\
&= \inf\{x^s x^t \colon s \in Y,\ t \in Z\} && \text{(Proposition 3.7.3)}\\
&= \inf\{x^s \colon s \in Y\}\inf\{x^t \colon t \in Z\} && \text{(Lemma 3.7.5)}\\
&= x^y x^z. && \text{(definition of } x^y, x^z\text{)}
\end{aligned}$$

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 the first inequality. By $x^y = x^{y-z}x^z$ and the last property in Proposition 3.6.9, we only need 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 $s > r$. By Proposiiton 3.7.3, we get $x^s > x^r$. This implies the infimum $x^{y-z} \geq x^r > 1$.

For the second inequality, 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 third equality $x^{yz} = (x^y)^z$. For $z = n \in \mathbb{N}$, we have

$$(x^y)^n = \underbrace{x^y \cdot x^y \cdots x^y}_{n} = x^{y+y+\cdots+y} = x^{yn}.$$

Then for $z = -n$, we have

$$x^{y(-n)} = x^{(-y)n} = (x^{-y})^n = \left(\frac{1}{x^y}\right)^n = \frac{1}{(x^y)^n} = (x^y)^{-n}.$$

We also have $x^{y0} = 1 = (x^y)^0$. Therefore we have $x^{yz} = (x^y)^z$ for any integer 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 for $x > 1$, we have ($(x^y)^r = x^{yr}$ used in the second equality of the first line)

$$\begin{aligned}
(x^y)^z &= \inf\{(x^y)^r \colon r \in \mathbb{Q},\ r > z\} = \inf\{x^{yr} \colon r \in \mathbb{Q},\ r > z\},\\
x^{yz} &= \inf\{x^s \colon s \in \mathbb{Q},\ s > yz\}.
\end{aligned}$$

For $r > z$, we have $yr > yz$. By the last property in Proposition 3.6.5, there is $s \in \mathbb{Q}$ satisfying $yr > s > yz$. Then by the inequality just proved, we get $x^{yr} > x^s$. On the other hand, for $s > yz$, by the last property in Proposition 3.6.5, there is $r \in \mathbb{Q}$ satisfying $\dfrac{s}{y} > r > z$. Then $s > yr$. By the inequality just proved, we get $x^s > x^{yr}$. Therefore the two infima are the same, and we get $(x^y)^z = x^{yz}$ for $x > 1$ and $y > 0$.

For $x > 1$ and any y, z, we write $y = y_1 - y_2$, with $y_1, y_2 > 0$. Then

$$x^{yz} = x^{y_1 z - y_2 z} = \frac{x^{y_1 z}}{x^{y_2 z}} = \frac{(x^{y_1})^z}{(x^{y_2})^z} = \left(\frac{x^{y_1}}{x^{y_2}}\right)^z = (x^y)^z.$$

Finally, for $x > 0$ and any y, z, we write $x = \dfrac{x_1}{x_2}$, with $x_1, x_2 > 1$. Then

$$x^{yz} = \frac{x_1^{yz}}{x_2^{yz}} = \frac{(x_1^y)^z}{(x_2^y)^z} = \left(\frac{x_1^y}{x_2^y}\right)^z = (x^y)^z. \qquad \square$$

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

$$\begin{aligned}
(1+2i)+(3+4i) &= (1+3)+(2+4)i = 4+6i,\\
(1+2i)-(3+4i) &= -2-2i,\\
(1+2i)\times(3+4i) &= 1\cdot 3+1\cdot 4i+2\cdot 3i+2\cdot 4i^2\\
&= (3-8)+(4+6)i = -5+10i,\\
(1+2i)\div(3+4i) &= \frac{1+2i}{3+4i} = \frac{(1+2i)(3-4i)}{(3+4i)(3-4i)}\\
&= \frac{1\cdot 3-1\cdot 4i+2\cdot 3i-2\cdot 4i^2}{3^2+4^2} = \frac{11}{25}+\frac{2}{25}i.
\end{aligned}$$

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 3.52. 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 $z-1 = \pm\sqrt{-4} = \pm 2i$. 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 real solution $z = \dfrac{-b \pm \sqrt{D}}{2a}$. If $D < 0$, then we get the complex solution $z = \dfrac{-b \pm \sqrt{-D}i}{2a}$.

We remark that completing the square is the most basic technique for treating quadratic functions. You should always derive the results by the technique, instead of memorizing the formula.

Exercise 3.53. 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)

$$\begin{aligned} e^{i\theta} &= 1 + \frac{1}{1!}i\theta + \frac{1}{2!}(i\theta)^2 + \cdots + \frac{1}{n!}(i\theta)^n + \cdots \\ &= \left(1 - \frac{1}{2!}\theta^2 + \frac{1}{4!}\theta^4 + \cdots\right) + i\left(\theta - \frac{1}{3!}\theta^3 + \frac{1}{5!}\theta^5 + \cdots\right) \\ &= \cos\theta + i\sin\theta. \end{aligned}$$

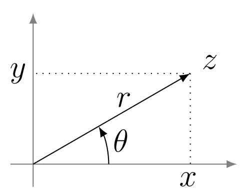

Figure 3.2: $z = x + yi = r\cos\theta + ir\sin\theta$.

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 modulus by $|z|$ because it is really the "absolute value". Moreover, we note that the argument is unique only up to adding multiples of 2π.

Exercise 3.54. Explain the following properties of the modulus:

1. $|z| = 0 \iff z = 0$.

2. $|z_1 + z_2| \leq |z_1| + |z_2|$.
3. $|z_1 z_2| = |z_1|\,|z_2|$, and $\left|\dfrac{z_1}{z_2}\right| = \dfrac{|z_1|}{|z_2|}$.
4. $|\bar{z}| = |-z| = |z|$.
5. $|z_1 + z_2|^2 + |z_1 - z_2|^2 = 2(|z_1|^2 + |z_2|^2)$.

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

$$\overline{re^{i\theta}} = re^{-i\theta}.$$

The exponential is characterized by the equalty $e^{z_1+z_2} = z^{z_1}z^{z_2}$. The multiplication of complex numbers in polar form is

$$r_1e^{i\theta_1}\ r_2e^{i\theta_2} = r_1r_2e^{i\theta_1}e^{i\theta_2} = r_1r_2e^{i(\theta_1+\theta_2)}.$$

Geometrically, this means the moduli are multiplied, and the arguments are added.

The complex exponential is

$$e^z = e^{x+iy} = e^x e^{iy}.$$

Here e^x is the modulus of e^z, and y is the argument of e^z. For any $a > 0$, we also have

$$a^z = e^{z\log a} = e^{x\log a}e^{iy\log a}.$$

Example 3.8.2. We have

$$\begin{aligned} e^{i\theta_1}e^{i\theta_2} &= (\cos\theta_1 + i\sin\theta_1)(\cos\theta_2 + i\sin\theta_2) \\ &= (\cos\theta_1\cos\theta_2 - \sin\theta_1\sin\theta_2) + i(\sin\theta_1\cos\theta_2 + \cos\theta_1\sin\theta_2). \end{aligned}$$

Then the equality $e^{i\theta_1}e^{i\theta_2} = e^{i\theta_1}$ means

$$\begin{aligned} \cos(\theta_1 + \theta_2) &= \cos\theta_1\cos\theta_2 - \sin\theta_1\sin\theta_2, \\ \sin(\theta_1 + \theta_2) &= \sin\theta_1\cos\theta_2 + \cos\theta_1\sin\theta_2. \end{aligned}$$

Example 3.8.3. 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 get the similar formula for $\sin 3\theta$.

Example 3.8.4. To solve the equation $z^3 = 1$, we let $z = re^{i\theta}$. Then $z^3 = r^3 e^{i3\theta}$ and the equation becomes $r^3 e^{i3\theta} = 1e^{i2n\pi}$. 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 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*.

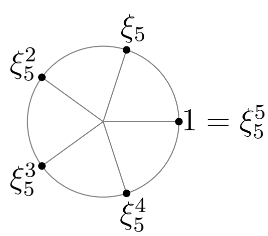

Figure 3.3: Five solutions of $z^5 = 1$.

Example 3.8.5. To calculate $(1-i)^{10}$, we write the complex number in 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

$$(1-i)^{10} = (\sqrt{2})^{10} e^{-i\frac{10}{4}\pi} = 2^5 e^{-i\frac{5}{2}\pi} = 32e^{-i\frac{1}{2}\pi} = -32i.$$

Example 3.8.6. 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 3.55. Solve the equation $(z+1)^4 + i(z+2)^4 = 0$.

Exercise 3.56. Solve the equation $z^2 = \bar{z}$.

Exercise 3.57. Prove that $|z_1 + z_2| = |z_1| + |z_2|$ if and only if z_1, z_2 are related by multiplying a non-negative real number. Then solve the equation $|z^2 - 1| = |z|^2 + 1$.

Exercise 3.58. Calculate $(1 + \sqrt{3}i)^{10}$.

Exercise 3.59. 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 3.60. Suppose $|z+1| \le 2$, 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

$$\begin{array}{r@{\,}r}
 & 1898 \\ \cline{2-2}
13\,\big) & 24681 \\
 & \underline{13000} \\
 & 11681 \\
 & \underline{10400} \\
 & 1281 \\
 & \underline{1170} \\
 & 111 \\
 & \underline{104} \\
 & 7
\end{array}$$

The result means $24681 = 1898 \cdot 13 + 7$. In general, we have the following.

Lemma 4.1.1. *For any $a, b \in \mathbb{Z}$ satisfying $b \neq 0$, there are unique integers q and r, such that*

$$a = qb + r, \quad 0 \leq r < |b|.$$

In the equality in the lemma, a is the *dividend*, b is the *divisor*, q is the *quotient*, and r is the *remainder*. We say a is *divisible* by b if $r = 0$. In other words, we have $a = qb$ for some integer q.

Proof. Suppose $b > 0$. 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 $\left|\frac{a}{b}\right| < n$. This means $-n < \frac{a}{b} < n$. Then among the increasing sequence of finitely many integers

$$-n < -n+1 < -n+2 < \cdots < n-2 < n-1 < n,$$

there is the biggest integer q, such that $q \leq \frac{a}{b}$. Being the biggest, we must also have $q+1 > \frac{a}{b}$. Then

$$q \leq \frac{a}{b} < q+1.$$

By $b > 0$, we get $qb \leq a < qb + q$. Let $r = a - qb$. We then have $a = qb + r$ and $0 \leq r < b = |b|$.

Suppose $b < 0$. Then $-b > 0$. By what we just proved, we have $a = q(-b) + r = (-q)b + r$, with $0 \leq r < -b = |b|$. 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

$$a = q_1 b + r_1 = q_2 b + r_2, \quad 0 \leq r_1 < |b|, \quad 0 \leq r_2 < |b|.$$

Then $(q_1 - q_2)b = r_2 - r_1$. Suppose $q_1 \neq q_2$. Then by the fourth property in Proposition 3.3.2, we get $|q_1 - q_2| \geq 1$. By the eighth property in Proposition 3.4.3, we get

$$|(q_1 - q_2)b| = |(q_1 - q_2)||b| \geq 1|b| = |b|.$$

On the other hand, by $0 \leq r_1 < |b|$ and $0 \leq r_2 < |b|$, we get

$$-|b| = 0 - |b| < r_1 - r_2 < |b| - 0 = |b|.$$

This means $|r_1 - r_2| < |b|$, and contradicts $|(q_1 - q_2)b| \geq |b|$. The contradiction proves the uniqueness. $\square$

Exercise 4.1. Find the quotient and the remainder.

1. $456 \div 123$.
2. $123 \div 456$.
3. $(-456) \div 123$.
4. $(-456) \div (-123)$.
5. $(-123) \div 456$.
6. $1221 \div 33$.

Exercise 4.2. Suppose the divisions of a_1 and a_2 by 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?

A *polynomial* is a function of the form

$$p(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_2 x^2 + a_1 x + a_0.$$

If $a_n \neq 0$, then n is the *degree* $\deg p(x)$ of the polynomial. For example, $x^2 + 2x - 1$ and $7 + 3x - 2x^4$ are polynomials of degrees 2 and 4. The zero polynomial, in which all the coefficients are zero, has degree $-\infty$. We clearly have

$$\deg(p(x) + q(x)) \leq \max\{\deg p(x), \deg q(x)\},$$
$$\deg(p(x)q(x)) = \deg p(x) + \deg q(x).$$

The long division can also be applied to polynomials. The following computation

$$
\begin{array}{r|rrrrrr}
 & & & x^3 & +x^2 & +x & +2 \\ \cline{2-7}
x^2-3x+2 & x^5 & -2x^4 & & +x^2 & & +3 \\
 & -x^5 & +3x^4 & -2x^3 & & & \\ \cline{2-4}
 & & x^4 & -2x^3 & +x^2 & & \\
 & & -x^4 & +3x^3 & -2x^2 & & \\ \cline{3-5}
 & & & x^3 & -x^2 & & \\
 & & & -x^3 & +3x^2 & -2x & \\ \cline{4-6}
 & & & & 2x^2 & -2x & +3 \\
 & & & & -2x^2 & +6x & -4 \\ \cline{5-7}
 & & & & & 4x & -1
\end{array}
$$

means that

$$x^5 - 2x^4 + x^2 + 3 = (x^3 - 3x + 2)(x^2 - 3x + 2) + (4x - 1).$$

Lemma 4.1.2. *For any polynomials $a(x)$ and $b(x)$, such that $b(x)$ is not constant, there are unique polynomials $q(x)$ and $r(x)$, such that*

$$a(x) = q(x)b(x) + r(x), \quad \deg r(x) < \deg b(x).$$

By $b(x)$ not being constant, we mean $\deg b(x) \geq 1$. The polynomials $a(x), b(x), q(x), r(x)$ are the *dividend*, *divisor*, *quotient*, *remainder*. The polynomial $a(x)$ is *divisible* by $b(x)$ if $r(x) = 0$. In other words, we have $a(x) = q(x)b(x)$ for some polynomial $q(x)$.

Proof. We fix $b(x)$ and induct on the degree $n = \deg a(x)$.

If $n = 0$, then $a(x) = a_0$ is a constant. Then we take $q(x) = 0$ and $r(x) = a_0$ to get

$$a_0 = 0b(x) + a_0, \quad \deg a_0 = 0 < \deg b(x).$$

Next, we make the inductive assumption that the lemma holds for all $a(x)$ of degree $< n$. Now consider a polynomial of degree n

$$a(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_2x^2 + a_1x + a_0, \quad a_n \neq 0.$$

Let

$$b(x) = b_k x^k + b_{k-1}x^{k-1} + \cdots + b_2x^2 + b_1x + b_0, \quad b_k \neq 0.$$

If $n < k$, then we take $q(x) = 0$ and $r(x) = a(x)$ to get

$$a(x) = 0b(x) + a(x), \quad \deg a(x) = n < k = \deg b(x).$$

If $n \geq k$, then we have

$$a(x) = \left(\frac{a_n}{b_k}x^{n-k}\right) b(x) + c(x).$$

Here

$$c(x) = \left(a_{n-1} - \frac{b_{k-1}}{b_k}a_n\right)x^{n-1} + \left(a_{n-2} - \frac{b_{k-2}}{b_k}a_n\right)x^{n-2} + \cdots$$

satisfies $\deg c(x) < n$. Applying the inductive assumption to the polynomial $c(x)$, we get

$$c(x) = \tilde{q}(x)b(x) + r(x), \quad \deg r(x) < \deg b(x).$$

Then

$$a(x) = \left(\frac{a_n}{b_k}x^{n-k} + \tilde{q}(x)\right)b(x) + r(x).$$

We conclude that the proposition holds for $a(x)$, with $q(x) = \dfrac{a_n}{b_k}x^{n-k} + \tilde{q}(x)$.

For the uniqueness of $q(x)$ and $r(x)$, let us assume

$$a(x) = q_1(x)b(x) + r_1(x) = q_2(x)b(x) + r_2(x), \quad \deg r_1(x), \deg r_2(x) < \deg b(x).$$

Then $(q_1(x) - q_2(x))b(x) = r_2(x) - r_1(x)$. If $q_1(x) \neq q_2(x)$, then

$$\deg[(q_1(x) - q_2(x))b(x)] = \deg(q_1(x) - q_2(x)) + \deg b(x) \geq \deg b(x).$$

On the other hand, we have

$$\deg(r_2(x) - r_1(x)) \leq \max\{\deg r_1(x), \deg r_2(x)\} < \deg b(x).$$

The contradiction shows that $q_1(x) = q_2(x)$, which further implies $r_1(x) = r_2(x)$. $\square$

Exercise 4.3. Find the quotient and the remainder of the division of polynomials.

$$(x^4 + 2x^3 - 3x + 1) \div (x + 2), \quad (3x^5 + 4x^3 - 2x + 5) \div (x^2 + x + 2).$$

Exercise 4.4. Prove that the remainder of the division of a polynomial $a(x)$ by $b(x) = x - x_0$ is $a(x_0)$.

4.2 Decimal Expansion

In decimal expression, the number 24681 actually means

$$24681 = 20000 + 4000 + 600 + 80 + 1 = 2 \cdot 10^4 + 4 \cdot 10^3 + 6 \cdot 10^2 + 8 \cdot 10 + 1.$$

In general, the *decimal expansion* of a natural number n is

$$n = r_k \cdot 10^k + r_{k-1} \cdot 10^{k-1} + \cdots + r_2 \cdot 10^2 + r_1 \cdot 10 + r_0.$$

Here r_i is an integer satisfying $0 \leq r_i < 10$, and $r_k \neq 0$. Moreover, the decimal expression of the number is $n = r_k r_{k-1} \cdots r_2 r_1 r_0$.

The number 10 is the *base* of the decimal expression. Any natural number $b > 1$ can be used as the base.

Example 4.2.1. To get the expression of 24681 based on $b = 13$, we repeatedly divide the quotient by 13:

$$\begin{aligned} 24681 &= 1898 \cdot 13 + 7, \\ 1898 &= 146 \cdot 13 + 0, \\ 146 &= 11 \cdot 13 + 3. \end{aligned}$$

Then we combine these and get

$$\begin{aligned} 24681 &= 1898 \cdot 13 + 7 \\ &= (146 \cdot 13 + 0) \cdot 13 + 7 \\ &= ((11 \cdot 13 + 3) \cdot 13 + 0) \cdot 13 + 7 \\ &= 11 \cdot 13^3 + 3 \cdot 13^2 + 0 \cdot 13^1 + 7. \end{aligned}$$

This gives the base 13 expansion

$$24681 = 11, 3, 0, 7_{[13]}.$$

The subscript [13] indicates the base of the expansion.

The digits are the remainders obtained in the repeated division. Therefore the *digits* r_i in a base b expansion are integers r_i satisfying $0 \le r_i < b$. In other words, $r_i = 0, 1, 2, \dots, n-1$. For example, the base 13 digits are $0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12$. The general base b expression is $r_k, r_{k-1}, \dots, r_2, r_1, r_{0[b]}$.

The commas are needed to divide the digits, in case $b > 10$. If we do not use commas, then $11117_{[13]}$ may also have the following meaning

$$1, 1, 3, 0, 7_{[13]} = 3 \cdot 13^4 + 1 \cdot 13^3 + 3 \cdot 13^2 + 0 \cdot 13 + 7 = 31272.$$

On the other hand, if $b \le 10$, then there is no ambiguity, and we may omit the commas

$$11307_{[8]} = 1, 1, 3, 0, 7_{[8]} = 1 \cdot 8^4 + 1 \cdot 8^3 + 3 \cdot 8^2 + 0 \cdot 8 + 7 = 4807.$$

Proposition 4.2.1. *For any $n, b \in \mathbb{N}$ satisfying $b > 1$, we have the unique expansion*

$$n = r_k b^k + r_{k-1} b^{k-1} + \cdots + r_2 b^2 + r_1 b + r_0,$$

where the integers $r_0, r_1, r_2, \dots, r_{k-1}, r_k$ satisfy

$$0 \le r_i < b, \quad r_k \ne 0.$$

Proof. We fix b and induct on n.

For $n = 1$, we have the expansion for $k = 0$ and $r_0 = 1$.

Next assume all natural numbers $< n$ has the expansion. By Lemma 4.1.1, we have $n = qb+r$ for some integers q and r satisfying $0 \le r < b$. By $n \ge qb$ and $b > 1$, we know $q < n$.

If $q = 0$, then $n = r < b$, and we have the expansion of n for $k = 0$ and $r_0 = n$.

If $q > 0$, then by $q < n$, we may apply the inductive assumption to q and get

$$q = r_k b^k + r_{k-1} b^{k-1} + \cdots + r_2 b^2 + r_1 b + r_0.$$

This implies the expansion of n

$$n = qb + r = r_l b^{l+1} + r_{l-1} b^l + \cdots + r_2 b^3 + r_1 b^2 + r_0 b + r,$$

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 $r_0 = 1$ because if $k > 0$, then

$$r_k b^k + r_{k-1} b^{k-1} + \cdots + r_2 b^2 + r_1 b + r_0 \ge r_k b^k > r_k \ge 1 = n.$$

Now assume the uniqueness of the expansion for all natural numbers $< n$. Then an expansion of n can be rewritten as

$$n = qb + r, \quad q = r_k b^{k-1} + r_{k-1} b^{k-2} + \cdots + r_2 b + r_1, \quad r = r_0.$$

This satisfies the conditions in Lemma 4.1.1. Therefore q and r_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 $r_1, r_2, \ldots, r_{k-1}, r_k$. □

Example 4.2.2. In computer, the *binary* (base 2) and the hexadecimal (base 16) are the most commonly used expressions. For example, we have

$$4807 = 2^{12} + 2^9 + 2^7 + 2^6 + 2^2 + 2^1 + 1 = 1001011000111_{[2]}.$$

We also have

$$\begin{aligned} 4807 &= 300 \cdot 16 + 7 \\ &= (18 \cdot 16 + 12) \cdot 16 + 7 \\ &= (1 \cdot 16 + 2) \cdot 16^2 + 12 \cdot 16 + 7 \\ &= 1 \cdot 16^3 + 2 \cdot 16^2 + 12 \cdot 16 + 7 \\ &= 1, 2, 12, 7_{[16]}. \end{aligned}$$

The hexadecimal expression usually uses the following notations for the digits between 10 and 15

$$A = 10, \quad B = 11, \quad C = 12, \quad D = 13, \quad E = 14, \quad F = 15.$$

Then we may omit commas and get

$$\begin{aligned} 4807 &= 1, 2, C, 7_{[16]} = 12C7_{[16]}, \\ 2C3A_{[16]} &= 2, 13, 3, 11_{[16]} = 2 \cdot 16^3 + 13 \cdot 16^2 + 3 \cdot 16 + 11 = 11579, \\ ABCD_{[16]} &= 10, 11, 12, 13_{[16]} = 10 \cdot 16^3 + 11 \cdot 16^2 + 12 \cdot 16 + 13 = 28621. \end{aligned}$$

Exercise 4.5. Express the number in binary and hexadecimal forms.

$$6, \quad 20, \quad 200, \quad 1024, \quad 12345.$$

Exercise 4.6. Express in decimal form (commas are omitted).

1. $1000_{[2]}$.
2. $11011_{[2]}$.
3. $101010_{[2]}$.
4. $101010_{[3]}$.
5. $37_{[16]}$.
6. $3C_{[16]}$.
7. $1000_{[16]}$.
8. $3AB_{[16]}$.

Exercise 4.7. Carry out the long operations (commas are omitted).

1. $1101_{[2]} + 110_{[2]}$.
2. $1101_{[2]} - 110_{[2]}$.
3. $1101_{[2]} \times 110_{[2]}$.
4. $1101_{[2]} \div 110_{[2]}$.
5. $120_{[3]} + 21_{[3]}$.
6. $120_{[3]} - 21_{[3]}$.
7. $120_{[3]} \times 21_{[3]}$.
8. $120_{[3]} \div 21_{[3]}$.

Example 4.2.3. The following shows how to convert decimal expansion into base $100 = 10^2$ expansion

$$\begin{aligned} 4807 = 4, 8, 0, 7_{[10]} &= 4 \cdot 10^3 + 8 \cdot 10^2 + 20 \cdot 10^1 + 7 \cdot 10^0 \\ &= (4 \cdot 10 + 8) \cdot 10^2 + (0 \cdot 10 + 7) \\ &= 48 \cdot 100^1 + 23 \cdot 100^0 = 48, 23_{[100]}. \end{aligned}$$

We find the conversion from base 10 to base 100 means dropping every other comma. Of course the reverse is adding comma. Here are more examples

$$\begin{aligned} 12345 &= 1, 23, 45_{[100]} = 12, 345_{[1000]}, \\ 1, 2, 3_{[100]} &= 1, 02, 03_{[100]} = 10203, \\ 1, 23, 4, 56_{[100]} &= 1230456 = 1, 230, 456_{[1000]}. \end{aligned}$$

We note that, in converting from base 100 to base 10, we need to first fill in necessary 0 to get length 2 digits before removing comma.

Example 4.2.4. The conversion from binary to octal (base $8 = 2^3$) means combining every three digits

$$\begin{aligned} 4807 &= 1001011000111_{[2]} \\ &= (1_{[2]}, 001_{[2]}, 011_{[2]}, 000_{[2]}, 111_{[2]})_{[8]} \\ &= 1, 1, 3, 0, 7_{[8]} = 11307_{[8]}. \end{aligned}$$

The conversion from binary to hexadecimal (base $16 = 2^4$) means combining every four digits

$$\begin{aligned} 4807 &= 1001011000111_{[2]} \\ &= (1_{[2]}, 0010_{[2]}, 1100_{[2]}, 0111_{[2]})_{[16]} \\ &= 1, 2, 12, 7_{[16]}. \end{aligned}$$

For the reverse conversions from octal or hexadecimal to binary, we should not forget to add necessary 0

$$\begin{aligned} 11579 = 2, 13, 3, 11_{[16]} &= (10_{[2]}, 1101_{[2]}, 11_{[2]}, 1011_{[2]})_{[16]} \\ &= (10_{[2]}, 1101_{[2]}, 0011_{[2]}, 1011_{[2]})_{[16]} \\ &= 10110100111011_{[2]}. \\ 12345_{[8]} = 1, 2, 3, 4, 5_{[8]} &= (1_{[2]}, 10_{[2]}, 11_{[2]}, 100_{[2]}, 101_{[2]})_{[8]} \\ &= (1_{[2]}, 010_{[2]}, 011_{[2]}, 100_{[2]}, 101_{[2]})_{[8]} \\ &= 1010011100101_{[2]}. \end{aligned}$$

Exercise 4.8. Express number in different base.

1. $1000_{[10]} =?_{[100]}$.
2. $1000_{[100]} =?_{[10]}$.
3. $1000_{[2]} =?_{[4]}$.
4. $110001111_{[2]} =?_{[8]}$.
5. $1D5B_{[16]} =?_{[2]}$.
6. $53030_{[8]} =?_{[2]}$.
7. $53030_{[9]} =?_{[3]}$.
8. $5, 30, 30_{[36]} =?_{[6]}$.
9. $5, 30, 30_{[100]} =?_{[10]}$.

Finally, we mention the polynomial version of Proposition 4.2.1.

Proposition 4.2.2. *For any polynomials $a(x)$ and $b(x)$ satisfying $\deg b(x) \geq 1$, we have the unique expansion*

$$a(x) = r_k(x)b(x)^k + r_{k-1}(x)b(x)^{k-1} + \cdots + r_2(x)b(x)^2 + r_1(x)b(x) + r_0(x),$$

where the polynomials $r_0(x), r_1(x), r_2(x), \ldots, r_{k-1}(x), r_k(x)$ satisfy

$$\deg r_i(x) < \deg b(x), \quad r_k(x) \neq 0.$$

For example, we have

$$\begin{aligned} x^5 - 2x^4 + x^2 + 3 &= (x^3 - 3x + 2)(x^2 - 3x + 2) + (4x - 1) \\ &= [(x + 3)(x^2 - 3x + 2) + (4x - 4)](x^2 - 3x + 2) + (4x - 1) \\ &= (x + 3)(x^2 - 3x + 2)^2 + (4x - 4)(x^2 - 3x + 2) + (4x - 1). \end{aligned}$$

Exercise 4.9. Expand $x^4 + 2x^3 - 3x + 1$ in base $x + 2$. Note that this means rewrite the polynomial in the variable $y = x + 2$.

Exercise 4.10. Expand $3x^5 + 4x^3 - 2x + 5$ in base $x^2 + x + 2$.

Exercise 4.11. Prove Proposition 4.2.2.

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. This means the remainder of division is 0. In this case, we denote $b \mid a$. If a is not divisible by b, then we denote $b \nmid a$. It is easy to verify the following properties.

Proposition 4.3.1. *The divisibility has the following properties:*

1. $a \mid 0$ *and* $\pm 1 \mid a$ *for any* a.
2. $a \mid b$ *and* $b \mid c \implies a \mid c$.
3. $a \mid b$ *and* $a \mid c \implies a \mid b + c$.
4. $a \mid b$ *and* $c \mid d \implies ac \mid bd$.
5. $a \mid b$ *and* $b \mid a \iff a = \pm b$.

If $b \neq 0$ divides all the numbers, then in particular, b divides 1. This means $1 = qb$ for some integer q. In other words, b is invertible in the integers $\mathbb{Z}$, with q as the inverse. Of course, this means $b = \pm 1$.

Therefore ± 1 are the *invertibles* in $\mathbb{Z}$. As far as divisibility is concerned, there is no difference if we multiply by an invertible. This means that, to discuss divisibility among integers, by multiplying -1 if necessary, we may restrict the discussion to natural numbers only.

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 $a_1, a_2, \ldots, a_k$, we may introduce the set of all positive common divisors

$$D(a_1, a_2, \ldots, a_k) = \{b \in \mathbb{N} \colon b \mid a_1,\ b \mid a_2,\ \ldots,\ b \mid a_k\}.$$

The common divisors can be calculated by the so called *Euclidean*[1] *algorithm*. Take 204 and 90 as an example:

$$\begin{aligned} 204 &= 2 \cdot 90 + 24, & D(204, 90) &= D(90, 24) \\ 90 &= 3 \cdot 24 + 18, & &= D(24, 18) \\ 24 &= 1 \cdot 18 + 6, & &= D(18, 6) \\ 18 &= 3 \cdot 6, & &= D(6, 0) = D(6). \end{aligned}$$

The algorithm is based on the following fact: If $a = qb + r$, then by Proposition 4.3.1, we know $c \mid a$ and $c \mid b$ implies $c \mid r = a - qb$. Conversely, if $c \mid b$ and $c \mid r$, then we know $c \mid a = qb + r$. Therefore we have

$$D(a, b) = D(b, r).$$

The algorithm gives

$$D(204, 90) = D(6) = \{b \in \mathbb{N} \colon b \mid 6\} = \{1, 2, 3, 6\}.$$

This means

$$b \mid 204 \text{ and } b \mid 90 \iff b \mid 6.$$

Moreover, by tracing back the calculations, 6 can be expressed in terms of 204 and 90:

$$\begin{aligned} 6 &= 24 - 1 \cdot 18 \\ &= 24 - 1 \cdot (90 - 3 \cdot 24) = 4 \cdot 24 - 1 \cdot 90 \\ &= 4 \cdot (204 - 2 \cdot 90) - 1 \cdot 90 = 4 \cdot 204 - 9 \cdot 90. \end{aligned}$$

In general, we have the following.

Theorem 4.3.2. *If $a_1, a_2, \ldots, a_k \in \mathbb{Z}$ are not all zero, then there is a unique natural number a, such that $D(a_1, a_2, \ldots, a_k) = D(a)$. Moreover,*

$$a = u_1 a_1 + u_2 a_2 + \cdots + u_k a_k$$

for some integrs $u_1, u_2, \ldots, u_k$.

The property $D(a_1, a_2, \ldots, a_k) = D(a)$ means exactly

$$b \mid a_1,\ b \mid a_2,\ \ldots,\ b \mid a_k \iff b \mid a.$$

The number a is the *greatest common divisor* of $a_1, a_2, \ldots, a_k$, and is denoted

$$a = \gcd(a_1, a_2, \ldots, a_k).$$

[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.

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 $u_1 = 1$ and $u_2 = \cdots = u_k = 0$.

Assume the theorem holds for the cases $\max\{a_1, a_2, \dots, a_k\} < n$. Now consider the case $\max\{a_1, a_2, \dots, a_k\} = n$. After dropping zeros among $a_1, a_2, \dots, a_k$, without loss of generality, we may assume that $a_1, a_2, \dots, a_k$ are nonzero and a_1 is the smallest among $a_1, a_2, \dots, a_k$. We consider two cases $a_1 = a$ and $a_1 < n$.

If $a_1 = n$, then $a_2 = \cdots = a_k = n$, and $D(a_1, a_2, \dots, a_k) = D(n)$. We may also choose $u_1 = 1$ and $u_2 = \cdots = u_k = 0$.

If $a_1 < n$, 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

$$D(a_1, a_2, \dots, a_k) = D(a_1, q_2 a_1 + a_2', \dots, q_k a_1 + a_k') = D(a_1, a_2', \dots, a_k').$$

Since $\max\{a_1, a_2', \dots, a_k'\} = a_1 < n$, we may apply the inductive assumption to $a_1, a_2', \dots, a_k'$ and find unique $a \in \mathbb{N}$, such that $D(a_1, a_2', \dots, a_k') = D(a)$. Moreover, we can find integers $u_1, u_2', \dots, u_k'$, such that

$$a = u_1 a_1 + u_2' a_2' + \cdots + u_k' a_k'.$$

Then we conclude $D(a_1, a_2, \dots, a_k) = D(a)$. Moreover, we have

$$a = (u_1 - q_2 u_2' - \cdots - q_k u_k') a_1 + u_2' a_2 + \cdots + u_k' a_k. \qquad \square$$

The Euclidean algorithm can be applied to several numbers. For example, to find the greatest common divisor of $-36, 204, 90, -114$, we divide by 36

$$204 = 5 \cdot 36 + 24, \quad 90 = 2 \cdot 36 + 18, \quad 114 = 3 \cdot 36 + 12,$$

and get

$$\gcd(-36, 204, 90, -114) = \gcd(36, 24, 18, 12).$$

Then we further divide by 12

$$36 = 3 \cdot 12, \quad 24 = 2 \cdot 12, \quad 18 = 1 \cdot 12 + 6,$$

and get

$$\gcd(36, 24, 18, 12) = \gcd(0, 0, 6, 12) = 6.$$

Moreover, we have

$$\begin{aligned}
6 &= 18 - 1 \cdot 12 \\
&= (90 - 2 \cdot 36) - 1 \cdot (114 - 3 \cdot 36) \\
&= 90 - 1 \cdot 114 + 1 \cdot 36 \\
&= (-1) \cdot (-36) + 0 \cdot 204 + 1 \cdot 90 + 1 \cdot (-114).
\end{aligned}$$

Exercise 4.12. 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 4.13. Explain the common divisors have the following properties

$$D(a, b) = D(-a, b) = D(b, a) = D(a, b, ac) = D(qb + a, b).$$

Exercise 4.14. Prove $\gcd(ac, bc) = \gcd(a, b)c$, $\gcd(a, b, c) = \gcd(a, \gcd(b, c))$.

Since polynomials have the similar division property as integers, the discussion about divisors and greatest common divisors also applies to polynomials.

Let $a(x)$ and $b(x) \neq 0$ be polynomials. If $a(x) = q(x)b(x)$ for some polynomial $q(x)$, then we say $a(x)$ is *divisible* by $b(x)$, or $b(x)$ is a *divisor* of $a(x)$. In this case, we denote $b(x) \mid a(x)$. If $a(x)$ is not divisible by $b(x)$, then we denote $b(x) \nmid a(x)$. Proposition 4.3.1 still holds for polynomials, except ± 1 should be replaced by invertible polynomials, which are nonzero constants.

A polynomial $a(x)$ is invertible, if there is another polynomial $b(x)$ satisfying $a(x)b(x) = 1$. This means exactly that $a(x)$ is a nonzero constant. Therefore invertible polynomials are nonzero constants. The divisibility among polynomials is not changed by multiplying nonzero constants.

Up to multiplying the invertibles ± 1 in $\mathbb{Z}$, we may consider only natural numbers. Up to multiplying invertibles in polynomials, we only need to consider *monic polynomials*

$$a(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_2x^2 + a_1x + a_0, \quad a_n \neq 0.$$

For example, polynomials $2x^3 - 3x + 1$ and $\frac{1}{3}x^2 + x - \frac{2}{3}$ are changed to monic polynomials by multiplying $\frac{1}{2}$ and 3

$$\tfrac{1}{2}(2x^3 - 3x + 1) = x^3 - \tfrac{3}{2}x + \tfrac{1}{2}, \quad 3(\tfrac{1}{3}x^2 + x - \tfrac{2}{3}) = x^2 + 3x - 2.$$

Similar to integers, define the set of all common divisors of given polynomials

$$D(a_1(x), a_2(x), \dots, a_k(x)) = \{b(x) \colon p(x) \mid a_1(x),\ b(x) \mid a_2(x),\ \dots,\ b(x) \mid a_k(x)\}.$$

The set is not changed under the modifications similar to the common divisors for integers. We may apply the Euclidean algorithm to calculate the common divisors.

Theorem 4.3.3. *If $a_1(x), a_2(x), \dots, a_k(x)$ are polynomials, not all zero, then there is a polynomial $a(x)$, such that*

$$D(a_1(x), a_2(x), \dots, a_k(x)) = D(a(x)).$$

Moreover, $a(x)$ is unique up to multiplying nonzero number, and

$$a(x) = u_1(x)a_1(x) + u_2(x)a_2(x) + \cdots + u_k(x)a_k(x)$$

for some polynomials $u_1(x), u_2(x), \dots, u_k(x)$.

The equality $D(a_1(x), a_2(x), \dots, a_k(x)) = D(a(x))$ means

$$b(x) \mid a_1(x),\ b(x) \mid a_2(x),\ \dots,\ b(x) \mid a_k(x) \iff b(x) \mid a(x).$$

Moreover, $a(x)$ is the *greatest common divisor*, and is denoted

$$a(x) = \gcd(a_1(x), a_2(x), \dots, a_k(x)).$$

To find the greatest common divisor of $x^4-x^3+x^2-1$, x^3-1, $-4x^4+4x^3+x^2+1$, we apply the Euclidean algorithm. First, we note that x^3-1 has the lowest degree. Then we divide the other polynomials by x^3-1 to get

$$\begin{aligned} x^4 - x^3 + x^2 - 1 &= (x-1)(x^3-1) + (x^2+x-2), \\ -4x^4 + 4x^3 + x^2 - 1 &= (-4x+4)(x^3-1) + (x^2-4x+3). \end{aligned}$$

This implies

$$D(x^4 - x^3 + x^2 - 1, x^3 - 1, -4x^4 + 4x^3 + x^2 - 1) = D(x^2 + x - 2, x^3 - 1, x^2 - 4x + 3).$$

Among x^2+x-2, x^3-1, x^2-4x+3, we pick the lowest degree polynomial x^2+x-2 (you can also pick x^2-4x+3) and get

$$\begin{aligned} x^3 - 1 &= (x-1)(x^2+x-2) + (3x-3), \\ x^2 - 4x + 3 &= 1(x^2+x-2) + (-5x+5). \end{aligned}$$

This implies

$$\gcd(x^2 + x - 2, x^3 - 1, x^2 - 4x + 3) = \gcd(-5x + 5, 3x - 3, x^2 - 4x + 3).$$

Finally, we get

$$\begin{aligned} x^2 - 4x + 3 &= \tfrac{1}{3}(x-3)(3x-3) + 0, \\ -5x + 5 &= -\tfrac{5}{3}(3x-3) + 0. \end{aligned}$$

Then we get

$$D(x^4 - x^3 + x^2 - 1, x^3 - 1, -4x^4 + 4x^3 + x^2 - 1) = D(3x-3) = D(x-1).$$

Therefore we conclude

$$\gcd(x^4 - x^3 + x^2 - 1, x^3 - 1, -4x^4 + 4x^3 + x^2 - 1) = x - 1.$$

Moreover, the greatest common divisor can be expressed in terms of the original polynomials by tracing back the series of divisions:

$$\begin{aligned} x - 1 &= \tfrac{1}{3}(x^2 - 4x + 3) - \tfrac{1}{3}(x^2 + x - 2) \\ &= \tfrac{1}{3}[(-4x^4 + 4x^3 + x^2 - 1) - (-4x + 4)(x^3 - 1)] \\ &\quad - \tfrac{1}{3}[(x^4 - x^3 + x^2 - 1) - (x-1)(x^3 - 1)] \\ &= -\tfrac{1}{3}(x^4 - x^3 + x^2 - 1) + (x-1)(x^3-1) + \tfrac{1}{3}(-4x^4 + 4x^3 + x^2 - 1). \end{aligned}$$

Exercise 4.15. Find the greatest common divisors and express the results in terms of the original polynomials.

1. $x^5 - x^3 + x^2 - 1$, $x^7 - x^3$.
2. $x^5 - x^3 + x^2 - 1$, $x^7 - x^3$, $x^4 - 2x + 1$.
3. $x^5 - x^3 + x^2 - 1$, $x^7 - x^3$, $x^4 - 2x + 1$, $x^4 + 1$.

Exercise 4.16. Prove Theorem 4.3.3 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

$$b \in \mathbb{N} \text{ and } b \mid p \implies b = 1 \text{ or } b = p,$$

i.e., p is divisible *only* by 1 and itself, then p is a *prime* number. We do not consider $p = 1$ because it is invertible. Invertible numbers are not prime numbers.

A natural number is not prime, if $n = m_1 m_2$ for some natural numbers $m_1, m_2 > 1$. We call n is a *composite* number. Finding all prime numbers is the same as removing all composite numbers. For $n_1 = 2$, this means removing all $2k$, for $k \geq 2$. Then we remove all $3k$, for $k \geq 2$. We do not need to remove $4k$, because they are already removed when we remove all $2k$. Next we remove all $5k$, for $k \geq 2$, and so on.

The following table consists of natural numbers between 2 and 100. We use m_n to indicate $m = nk$ for some $k \geq 2$. We indicate 12_2 instead of 12_3, because 12 is already removed as multiple of 2, and multiples of 3 are removed only after multiples of 2 are removed.

	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

The 25 remaining natural numbers, i.e., the ones without subscripts, are all the prime numbers between 2 and 100.

Theorem 4.4.1. *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,

$$\begin{aligned} 6 &= 2 \cdot 3 = 2^1 \cdot 3^1, \\ 60 &= 2 \cdot 2 \cdot 3 \cdot 5 = 2^2 \cdot 3^1 \cdot 5^1, \\ 600 &= 2 \cdot 2 \cdot 2 \cdot 3 \cdot 5 \cdot 5 = 2^3 \cdot 3^1 \cdot 5^2. \end{aligned}$$

To compare the prime numbers in 6 and 60, we may also write $6 = 2^1 \cdot 3^1 = 2^1 \cdot 3^1 \cdot 5^0$.

Proof. We induct on the natural number $n \geq 2$. The induction starts with $n = 2$. The natural number 2 is the product of one prime number 2.

Next, we assume the theorem holds for all natural numbers m satisfying $1 < m < n$. If n is already a prime number, then n is the product of one prime number n. If n is not a prime number, then we have $n = m_1 m_2$ for some natural numbers $m_1, m_2 \geq 2$. Then we have $2 \leq m_1 < n$ and $2 < m_2 < n$. By the inductive assumption, we know m_1 and m_2 are product of prime numbers. Then $n = m_1 m_2$ is a product of prime numbers. $\square$

We have not yet proved the uniqueness. For this purpose, we establish the following result.

Lemma 4.4.2. *Suppose natural numbers p, m, n satisfy $p \mid mn$. If p is a prime number, then $p \mid mn$ implies $p \mid m$ or $p \mid n$.*

Proof. Let $a = \gcd(p, m)$. Then $a \mid p$ and $a \mid m$. Since p is a prime number, we get $a = p$ or $a = 1$.

If $a = p$, then $p = a \mid m$.

If $a = 1$, then by Proposition 4.3.2, we get $up + vm = 1$ for some integers u, v. Then $n = unp + vmn$. We have $p \mid unp$. By $p \mid m$, we also have $p \mid vmn$. Then we get $p \mid n$. □

Continuation of the Proof of Theorem 4.4.1. Let $n = p_1p_2 \cdots p_k$ and $n = p'_1p'_2 \cdots p'_{k'}$ be two multiplications of prime numbers. We prove that $p_1, p_2, \dots, p_k$ is a permutation of $p'_1, p'_2, \dots, p'_k$, by inducting on n.

For $n = 2$, the only way of expressing 2 as a product of primes is $k = 1$ and $p_1 = 2$.

Suppose the uniqueness property is proved for all natural number m satisfying $2 \le m < n$. Consider $n = p_1p_2 \cdots p_k = p'_1p'_2 \cdots p'_{k'}$. We have $p_1 \mid n = p'_1p'_2 \cdots p'_{k'}$. By Lemma 4.4.2, this implies $p_1 \mid p'_i$. Without loss of generality, we may assume $p_1 \mid p'_1$. Since p'_1 is a prime number, and $p_1 \neq 1$, we get $p_1 = p'_1$. Dividing p_1, we get $m = p_2 \cdots p_k = p'_2 \cdots p'_{k'}$. Moreover, we have $m = \frac{n}{p_1} < n$. We may apply the inductive assumption to m and conclude $p_2, \dots, p_k$ is a permutation of $p'_2, \dots, p'_{k'}$. Then $p_1, p_2, \dots, p_k$ is a permutation of $p'_1, p'_2, \dots, p'_{k'}$. □

An important consequence of Theorem 4.4.1 is the following important result. The proof first appeared in Euclid's The Elements.

Theorem 4.4.3. *There are infinitely many prime numbers.*

Proof. Suppose there are only finitely many prime numbers $p_1, p_2, \dots, p_k$. Then consider the number $p = p_1p_2 \cdots p_k + 1$. Since $p_1, p_2, \dots, p_k$ is the list of *all* prime numbers, by Proposition 4.4.1, one of them, say p_i, must be a divisor of p. Since $p_1p_2 \cdots p_k$ is also divisible by p_i, we conclude that p divides 1, a contradiction. □

By combining the prime factors that appear repeatedly, the product of primes can be written as

$$n = p_1^{e_1}p_2^{e_2} \cdots p_m^{e_m},$$

where $p_1, p_2, \dots, p_m$ are distinct prime numbers and $e_i \in \mathbb{N}$. If a prime number p is none of p_i, we still have $n = p_1^{e_1}p_2^{e_2} \cdots p_m^{e_m}p^0$. In general, any natural number

$$n = 2^{e_2(n)}3^{e_3(n)}5^{e_5(n)} \cdots p^{e_p(n)} \cdots = \prod_{\text{prime } p} p^{e_p(n)}.$$

Here $e_p(n) \in \mathbb{N} \cup \{0\}$ is the *exponent* of p in n. Moreover, for each n, we have $e_p(n) > 0$ for only finitely many prime numbers p. For example, By $60 = 2^2 \cdot 3^1 \cdot 5^1 \cdot 7^0 \cdot 11^0$, we have

$$e_2(60) = 2,\ e_3(60) = 1,\ e_5(60) = 1,\ e_p(60) = 0 \text{ for prime number } p \geq 7.$$

We have $e_p(1) = 0$ for all p, and we define $e_p(0) = -\infty$ for all p. The exponent for zero is consistent with the expectation that $p^{-\infty} = 0$.

By $\dfrac{p^m}{p^n} = p^{m-n}$, we get

$$\frac{204}{90} = \frac{2^2 3^1 17^1}{2^1 3^2 5^1} = 2^{2-1}3^{1-2}5^{0-1}17^{1-0} = 2^1 3^{-1} 5^{-1} 17^1.$$

Therefore the division of prime factorisations corresponds to the subtraction of the exponents. This gives the exponents of rational numbers

$$e_2(\tfrac{204}{90}) = 1,\ e_3(\tfrac{204}{90}) = -1,\ e_5(\tfrac{204}{90}) = -1,\ e_{17}(\tfrac{204}{90}) = 1,$$

and

$$e_p(\tfrac{204}{90}) = 0 \text{ for prime } p \geq 7 \text{ and } p \neq 17.$$

The multiplication by invertibles does not change the exponent. We have $e_p(-r) = e_p(r)$ for $r \in \mathbb{Q}$. By $p^a p^b = p^{a+b}$, we get

$$e_p(rs) = e_p(r) + e_p(s), \quad r, s \in \mathbb{Q}.$$

We know a nonzero rational number $r \in \mathbb{Z}$ if and only if all $e_p(r) \geq 0$. Therefore for $a, b \in \mathbb{N}$, we have $a \mid b$ if and only if $e_p(a) \leq e_p(b)$ for all prime p. For example, $b \mid 204$ and $b \mid 90$ means

$$\begin{aligned}
e_2(b) &\leq \min\{e_2(204), e_2(90)\} = \min\{2, 1\} = 1,\\
e_3(b) &\leq \min\{e_3(204), e_3(90)\} = \min\{1, 2\} = 1,\\
e_5(b) &\leq \min\{e_5(204), e_5(90)\} = \min\{0, 1\} = 0,\\
&\ \ \vdots
\end{aligned}$$

This implies the greatest common divisor $a = \gcd(204, 90)$ satisfies

$$e_2(a) = 1,\ e_3(a) = 1,\ e_p(a) = 0 \text{ for all } p \geq 5.$$

In other words, we have $a = 2^1 3^1 = 6$. In general, we have the following result.

Proposition 4.4.4. *For nonzero integers $a_1, a_2, \dots, a_k$, we have*

$$e_p(\gcd(a_1, a_2, \dots, a_k)) = \min\{e_p(a_1), e_p(a_2), \dots, e_p(a_k)\}.$$

For example, we have

$$\begin{gathered}
\gcd(-36, 204, 90, -114) = \gcd(2^2 3^2,\ 2^2 3^1 17^1,\ 2^1 3^2 5^1,\ 2^1 3^1 19^1) = 2^1 3^1 = 6,\\
\gcd(22275, 38115) = \gcd(3^4 5^2 11^1,\ 3^2 5^1 7^1 11^2) = 3^2 5^1 11^1 = 495.
\end{gathered}$$

Exercise 4.17. Suppose $n > 1$ is a natural number. Prove that, if $p \nmid n$ for all prime $p \le \sqrt{n}$, then n is a prime number.

Exercise 4.18. Factor 1053, -390, 247, -500 into products of primes. Then use the result to find the greatest common divisor.

Exercise 4.19. Let n be a natural number.

1. Prove that $3 \mid n$ and $5 \mid n \implies 15 \mid n$.
2. Is it true that $3 \mid n$ and $20 \mid n \implies 60 \mid n$? Explain.
3. Is it true that $6 \mid n$ and $10 \mid n \implies 60 \mid n$? Explain.

Exercise 4.20. 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 4.21. Prove that the exponent of a prime p in an integer $a \neq 0$ is the biggest non-negative number e satisfying $p^e \mid a$.

Exercise 4.22. The following introduces a concept complementary to the greatest common divisor.

1. For nonzero integers $a_1, a_2, \dots, a_k$, let $c \in \mathbb{N}$ be determined by
$$e_p(c) = \max\{e_p(a_1), e_p(a_2), \dots, e_p(a_k)\}.$$
Prove that for any integer b,
$$a_1 \mid b,\ a_2 \mid b,\ \dots,\ a_k \mid b \iff c \mid b.$$
The number c is the *least common multiple*, and is denoted $\mathrm{lcm}(a_1, a_2, \dots, a_k)$.
2. Find the least common multiple of 1053, -390, 247, -500.
3. Prove that for any two nonzero natural numbers m, n, we have
$$\gcd(m, n) \cdot \mathrm{lcm}(m, n) = mn.$$

4. We know the greatest common divisor is not changed by certain modifications. Is the least common multiple not changed by the same modifications?

Next we turn to polynomials. If a *non-constant* polynomial $p(x)$ satisfies

$$a(x) \mid p(x) \implies a(x) = r \text{ or } a(x) = rp(x) \text{ for some constant } r \neq 0,$$

i.e., $p(x)$ is divisible *only* by constants and constant multiples of itself, then $p(x)$ is called an *irreducible polynomial*.

Recall that nonzero constants are the invertible polynomials, and multiplying nonzero constants do not change divisibility. Up to multiplying nonzero constants, we only need to consider monic polynomials.

All the results about prime numbers and factorisations are based on division $a = qb + r$. Since polynomials also have division $a(x) = q(x)b(x) + r(x)$, all the discussions and results about prime factorisations remain valid for polynomials. In particular, we may use Euclidean algorithm to calculate the greatest common divisor, and any polynomial is the unique (up to multiplying nonzero constants, and up to permutation) product of irreducible polynomials.

Consider $x^4 - 1 = (x^2+1)(x^2-1) = (x^2+1)(x+1)(x-1)$. By $\deg(a(x)b(x)) = \deg a(x) + \deg b(x)$, it is easy to deduce that the degree 1 polynomials $p(x) = x - x_0$ are irreducible. On the other hand, if $x^2 + 1$ were *reducible* (i.e., not irreducible), then $x^2 + 1 = (x - x_1)(x - x_2)$ must be a product of two degree one polynomials. In particular, we have $x_1^2 + 1 = (x_1 - x_1)(x_1 - x_2) = 0$ and similarly $x_2^2 + 1 = 0$. However, there is no *real* number x_1 satisfying $x_1^2 + 1 = 0$. Therefore $x^2 + 1$ is also irreducible, and $x^4 - 1 = (x^2+1)(x^2-1) = (x^2+1)(x+1)(x-1)$ is a factorization into a product of irreducible polynomials.

Note that the impossibility of $x_1^2 + 1 = 0$ is due to the fact that we are restricted to real numbers only. If we are allowed to use complex numbers, then $x^2 + 1 = (x+i)(x-i)$, $i = \sqrt{-1}$, is a further factorization. Therefore x^2+1 is not irreducible *as a complex polynomial*, although it is irreducible as a real polynomial.

Thus the irreducibility of polynomials depend on the numbers allowed to be the coefficients. For another example, let us consider $x^2 + 2x - 2$. It the polynomial were reducible, then we must have $x^2 + 2x - 2 = (x - x_1)(x - x_2)$. This implies x_1 and x_2 are two roots of $x^2 + 2x - 2$. The roots are $-1 \pm \sqrt{3}$, which are real but not rational. Therefore $x^2 + 2x - 2$ is irreducible as a rational polynomial and is reducible as a real polynomial.

Exercise 4.23. Consider the irreducibility of rational polynomials.

1. Factor $x^6 - 1$, $x^4 - 2x + 1$, $x^4 + 4$ into products of irreducible rational polynomials.

2. Use the result above to find $\gcd(x^6 - 1, x^4 - 2x + 1)$ and $\gcd(x^6 - 1, x^4 - 2x + 1, x^4 + 4)$.

3. What is the condition for a degree two rational polynomial to be irreducible?

Exercise 4.24. Consider the irreducibility of degree one polynomials.

1. Prove that any degree one polynomial $x - x_0$ 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 4.25. Consider the irreducibility of degree two polynomials.

1. Show that $x^2 + 2x - 3$ and $2x^2 + 4x - 3$ 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 if $a - b$ is divisible by n. We denote

$$a \equiv b \pmod{n}.$$

In Example 2.5.5, we explained that congruent mod n is an equivalence. In Example 2.5.12, we denote the quotient set by ($[a]$ is changed to more commonly used $\bar{a}$ here)

$$\mathbb{Z}_n = \mathbb{Z}/n\mathbb{Z} = \{\bar{0}, \bar{1}, \dots, \overline{n-1}\}.$$

Moreover, the quotient map $\mathbb{Z} \to \mathbb{Z}_n$ essentially means taking the remainder:

$$24681 = 1898 \cdot 13 + 7 \implies \overline{24681} = \bar{7} \text{ in } \mathbb{Z}_{13}.$$

We may add and multiply congruence classes by

$$\bar{a} + \bar{b} = \overline{a+b}, \quad \bar{a}\bar{b} = \overline{ab}.$$

Strictly speaking, we need to verify the operations are 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 addition and multiplication in $\mathbb{Z}_{16}$:

$$\begin{aligned}
\overline{13} + \overline{10} &= \overline{23} = \overline{1 \cdot 16 + 7} = \bar{7}.\\
\overline{13} - \overline{10} &= \bar{3}.\\
\overline{10} - \overline{13} &= \overline{-3} = \overline{(-1) \cdot 16 + 13} = \overline{13}.\\
\overline{13} \times \overline{10} &= \overline{130} = \overline{8 \cdot 16 + 2} = \bar{2}.\\
\overline{13} \times \bar{5} &= \overline{65} = \overline{5 \cdot 16 + 1} = \bar{1}.
\end{aligned}$$

By the the associativity, commutativity, and distributivity of the operations in $\mathbb{Z}$, we have the same properties in $\mathbb{Z}_n$. It is also easy to see that $\bar{0}$ and $\bar{1}$ behave like 0 and 1 for integers:

$$\bar{a} + \bar{0} = \bar{a} = \bar{0} + \bar{a}, \quad \bar{a}\bar{1} = \bar{a} = \bar{1}\bar{a}.$$

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 $-\bar{a} = \overline{-a}$.

Although the addition and multiplication in $\mathbb{Z}_n$ are very much like the same operations in $\mathbb{Z}_n$, there are some major differences. First, there is no order for $\mathbb{Z}_n$. Note that instead of $n > 0$ in $\mathbb{Z}$, we have $\bar{n} = \bar{0}$ in $\mathbb{Z}_n$.

The second difference is the reciprocal, i.e., invertible numbers. The only invertibles in $\mathbb{Z}$ are ± 1. However, we have $\bar{2}\bar{3} = \bar{6} = \bar{1}$ in $\mathbb{Z}_5$. Therefore it is possible to divide $\bar{2}$ or $\bar{3}$ in $\mathbb{Z}_5$:

$$\frac{\bar{4}}{\bar{3}} = \bar{4} \cdot \bar{2} = \bar{8} = \bar{3}, \quad \frac{\bar{3}}{\bar{2}} = \bar{3} \cdot \bar{3} = \bar{9} = \bar{4}, \quad \text{in } \mathbb{Z}_5.$$

Proposition 4.5.1. *$\bar{a} \in \mathbb{Z}_n$ is invertible if and only if* $\gcd(a, n) = 1$.

We denote all the invertibles by

$$\mathbb{Z}_n^* = \{\bar{a} \in \mathbb{Z}_n \colon \gcd(a, n) = 1\}.$$

Proof. The invertibility means there is a reciprocal $\bar{b} \in \mathbb{Z}_n$ satisfying $\bar{a}\bar{b} = \bar{1}$. If there is such b, then we have $ab - 1 = qn$ for some $q \in \mathbb{Z}$. This implies

$$c \mid a \text{ and } c \mid n \implies c \mid 1 = ab - qn \implies c = \pm 1.$$

This proves $\gcd(a, n) = 1$.

Conversely, suppose $\gcd(a, n) = 1$. By Theorem 4.3.2, we have $ua + vn = 1$ for some $u, v \in \mathbb{Z}$. This implies $\bar{a}\bar{u} = \overline{au} = \overline{1 - vn} = \bar{1}$. Therefore $\bar{u}$ is the reciprocal of a. ∎

Example 4.5.1. For a primes number p, we have

$$\gcd(a,p) = \begin{cases} p, & \text{if } p \mid a \\ 1, & \text{if } p \nmid a \end{cases}.$$

Then Proposition 4.5.1, we get

$$(\mathbb{Z}_p)^* = \mathbb{Z}_p - \{\bar{0}\} = \{\bar{1}, \bar{2}, \dots, \overline{p-1}\}.$$

In other words, all nonzero numbers are invertible.

The reciprocals of nonzero numbers in $\mathbb{Z}_5$ are

$$\bar{1}^{-1} = \bar{1},\ \bar{2}^{-1} = \bar{3},\ \bar{3}^{-1} = \bar{2},\ \bar{4}^{-1} = \bar{4}, \quad \text{in } \mathbb{Z}_5.$$

The reciprocals of nonzero numbers in $\mathbb{Z}_{13}$ are (if $\bar{a}^{-1} = \bar{b}$, then $\bar{b}^{-1} = \bar{a}$)

$$\bar{1}^{-1} = \bar{1},\ \bar{2}^{-1} = \bar{7},\ \bar{3}^{-1} = \bar{9},\ \bar{4}^{-1} = \overline{10},\ \bar{5}^{-1} = \bar{8},\ \bar{6}^{-1} = \overline{11}, \quad \text{in } \mathbb{Z}_{13}.$$

Example 4.5.2. By $12 = 2^2 \cdot 3$, we know $\gcd(a, 12) = 1$ means $2 \nmid a$ and $3 \nmid a$. We list all natural numbers up to 11, and indicate all multiples 2 and 3 (including 2 and 3, unlike the way to get prime numbers):

$$1,\ 2_2,\ 3_3,\ 4_2,\ 5,\ 6_2,\ 7,\ 8_2,\ 9_3,\ 10_2,\ 11.$$

After deleting these multiples, we get the invertibles in $\mathbb{Z}_{12}$:

$$\mathbb{Z}_{12}^* = \{\bar{1}, \bar{5}, \bar{7}, \overline{11}\}.$$

Then by $\bar{5} \cdot \bar{5} = \overline{25} = \bar{1}$, and $\bar{7} \cdot \bar{7} = \overline{49} = \bar{1}$, and $\overline{11} \cdot \overline{11} = \overline{121} = \bar{1}$, we get

$$\bar{1}^{-1} = \bar{1},\ \bar{5}^{-1} = \bar{5},\ \bar{7}^{-1} = \bar{7},\ \overline{11}^{-1} = \overline{11}, \quad \text{in } \mathbb{Z}_{12}.$$

Exercise 4.26. Find invertible elements in $\mathbb{Z}_8$ and their reciprocals.

Exercise 4.27. List all the invertibles in $\mathbb{Z}_{30}$.

Exercise 4.28. Let p be a prime number, and let e be a natural number. How many invertibles are in $\mathbb{Z}_{p^e}$? How about invertible are in $\mathbb{Z}_{p^e q^f}$, where p, q are distinct prime numbers?

Exercise 4.29. What $\bar{a} \in \mathbb{Z}_n$ has the property that $\bar{a}\bar{b} = \bar{0} \implies \bar{b} = \bar{0}$?

The property in Proposition 4.5.1 is called coprime. In general, a collection of integers $a_1, a_2, \dots, a_k$ are *coprime* if

$$\gcd(a_1, a_2, \dots, a_k) = 1.$$

In other words, the only natural number that divides all $a_1, a_2, \ldots, a_k$ is 1. By Proposition 4.3.2, for coprime integers $a_1, a_2, \ldots, a_k$, we can find integrs $u_1, u_2, \ldots, u_k$ satisfying

$$u_1 a_1 + u_2 a_2 + \cdots + u_k a_k = 1.$$

Conversely, given the equality, any natural number dividing all $a_1, a_2, \ldots, a_k$ must also divide the combination 1. Therefore the only common divisor is 1, and $a_1, a_2, \ldots, a_k$ are coprime.

Suppose $a_1, a_2, \ldots, a_k$ are not coprime. Then some integer $b > 1$ divides all a_i. Then any prime factor p of b also divides all a_i. Therefore $a_1, a_2, \ldots, a_k$ are not coprime, if and only if they have common prime factor. In other words, $a_1, a_2, \ldots, a_k$ are coprime, if and only if they do not have common prime factor.

Example 4.5.3. Both $204, 90$ have prime factor 2 (i.e., divisible by 2). Therefore $204, 90$ are not coprime.

We wish to add more number to get a coprime set. To avoid prime factor 2, we add an odd number. If we add 99, then $204, 90, 99$ has common prime factor 3. Therefore $204, 90, 99$ are still not coprime.

Therefore the number we add should not have 2 or 3 as prime factors. If we add $175 = 5^2 \cdot 7$, then 5 is not a factor of 204, and 7 is not a factor of90. Therefore $204, 90, 175$ are coprime.

Example 4.5.4. We argue that, if a, c are coprime, and b, c are coprime, then ab, c are coprime. Our argument is by contrapositive.

Suppose ab, c are not coprime. Then there is a prime p satisfying $p \mid ab$ and $p \mid c$. By Lemma 4.4.2, $p \mid ab$ implies $p \mid a$ or $p \mid b$. If $p \mid a$ and $p \mid c$, then a, c are not coprime. If $p \mid b$ and $p \mid c$, then b, c are not coprime. We conclude either a, c are coprime, or b, c are coprime.

By repeatedly using the property, we know that, if a_i and c are coprime for all i, then $a_1 a_2 \cdots a_k$ and c are coprime.

Exercise 4.30. Suppose $a_1, a_2, \ldots, a_k$ are coprime. Prove that $a_1, a_2, \ldots, a_k, a_{k+1}$ are coprime.

Exercise 4.31. Suppose $a, a_1, a_2, \ldots, a_k$ are coprime, and $b, a_1, a_2, \ldots, a_k$ are also coprime. Prove that $ab, a_1, a_2, \ldots, a_k$ are coprime.

Exercise 4.32. Prove that a, b, c are coprime if and only if $\gcd(a, b), c$ are coprime. Extend the property to more integers.

Chapter 5

Counting

5.1 Finite Counting

We count Bill, Bob, and Mary as three people by the following process: First point to Bill and call "one". Then point to Bob and say "two". Finally point to Mary and say "three". Mathematically, we did the counting by establishing the following one-to-one correspondence from the set {Bill, Bob, and Mary} to the set $\{1, 2, 3\}$:

$$\text{Bill} \to 1, \quad \text{Bob} \to 2, \quad \text{Mary} \to 3.$$

Definition 5.1.1. A set X has n elements, and denoted $|X| = n$, if there is a one-to-one correspondence between X and the set $\{1, 2, \dots, n\}$ of natural numbers. The number of element in the empty set is zero.

By the definition, and the composition of invertible maps is invertible, if there is a one-to-one correspondence between X and Y, then $|X| = |Y|$.

We need to verify that the concept is well defined. In other words, if there is one one-to-one correspondence from X to $\{1, 2, \dots, n\}$ and another one-to-one correspondence from X to $\{1, 2, \dots, m\}$, then $m = n$. By combining the two one-to-one correspondences, we get a one-to-one correspondence from $\{1, 2, \dots, n\}$ to $\{1, 2, \dots, m\}$. Therefore the problem boils down to the following.

Proposition 5.1.2. *If there is a one-to-one correspondence from* $\{1, 2, \dots, n\}$ *to* $\{1, 2, \dots, m\}$*, then* $m = n$*.*

Proof. We induct on n. For $n = 1$, if $f\colon \{1\} \to \{1, 2, \dots, m\}$ is a one-to-one correspondence, then f is onto. This implies $m = f(1) = 1$.

Now assume the proposition holds for $n - 1$. Consider a one-to-one correspondence $f\colon \{1, 2, \dots, n\} \to \{1, 2, \dots, m\}$. Let $f(n) = k$. Then the restriction

$$\tilde{f}\colon \{1, 2, \dots, n-1\} \to \{1, 2, \dots, k-1, k+1, \dots, m\}$$

of f is still a one-to-one correspondence. Combining $\tilde{f}$ with the one-to-one correspondence

$$h(j) = \begin{cases} j, & \text{if } j < k \\ j-1, & \text{if } j > k \end{cases} : \{1, 2, \ldots, k-1, k+1, \ldots, m\} \to \{1, 2, \ldots, m-1\},$$

we get a one-to-one correspondence

$$g(i) = \begin{cases} f(i), & \text{if } f(i) < k \\ f(i)-1, & \text{if } f(i) > k \end{cases} : \{1, 2, \ldots, n-1\} \to \{1, 2, \ldots, m-1\}.$$

Then by the inductive assumption, we get $n - 1 = m - 1$. This implies $m = n$. $\square$

We remark that the following facts are used in the proof:

1. If $f\colon X \to Y$ is a one-to-one correspondence, and $A \subset X$ is a subset, then the restriction $\tilde{f}\colon X - A \to Y - f(A)$ is also a one-to-one correspondence.

2. The composition of one-to-one correspondences is a one-to-one correspondence.

Proposition 5.1.3. *For finite sets X and Y, we have $|X \cup Y| = |X| + |Y| - |X \cap Y|$.*

The formula means the following. When we count the umber of elements in $X \cup Y$, we add the numbers of X and number of Y together. However, the sum counts the elements in $X \cap Y$ twice. Then we we need to subtract $|X \cap Y|$ to balance the overcounting.

Proof. First assume X and Y are disjoint. Let $|X| = m$ and $|Y| = n$. Then we have one-to-one correspondences

$$f\colon X \to \{1, 2, \ldots, m\}, \quad g\colon Y \to \{1, 2, \ldots, n\}.$$

Composing g with the one-to-one correspondence

$$i \mapsto i + m\colon \{1, 2, \ldots, n\} \to \{m+1, m+2, \ldots, m+n\},$$

we get the one-to-one correspondence

$$g(x) + m\colon Y \to \{m+1, m+2, \ldots, m+n\}.$$

Combining f and $g + m$, we get a one-to-one correspondence

$$h(x) = \begin{cases} f(x), & \text{if } x \in X \\ g(x) + m, & \text{if } x \in Y \end{cases} : X \sqcup Y \to \{1, 2, \ldots, m+n\}.$$

This implies $|X \sqcup Y| = m + n$.

In general, we have disjoint unions

$$\begin{aligned} X &= (X - Y) \sqcup (X \cap Y), \\ Y &= (Y - X) \sqcup (X \cap Y), \\ X \cup Y &= (X - Y) \sqcup (Y - X) \sqcup (X \cap Y). \end{aligned}$$

Let $|X - Y| = m$, $|Y - X| = n$, $|X \cap Y| = k$. Then we get

$$|X| = m + k, \quad |Y| = n + k, \quad |X \cup Y| = m + n + k.$$

Then

$$|X| + |Y| - |X \cap Y| = (m + k) + (n + k) - k = m + n + k = |X \cup Y|.$$

This completes the proof of the first equality. □

Example 5.1.1. Suppose among total of 45 people, 18 are younger than thirty, 15 are between twenty and forty years old, and 7 are between twenty and thirty. The question is how many are older than forty. Consider

$$\begin{aligned} X &= \text{all people} & |X| &= 45, \\ A &= \text{younger than thirty} & |A| &= 18, \\ B &= \text{between twenty and forty} & |B| &= 15, \\ A \cap B &= \text{between twenty and thirty} & |A \cap B| &= 7. \end{aligned}$$

Then $X - (A \cup B)$ is all the people older than forty, and

$$\begin{aligned} |X - (A \cap B)| &= |X| - |A \cup B| = |X| - (|A| + |B| - |A \cap B|) \\ &= 45 - (18 + 15 - 7) = 19. \end{aligned}$$

Example 5.1.2. Let X be the set of natural numbers between (and including) 20 and 40. Let X_3, X_5 be the numbers in X divisible by $3, 5$. Then $|X| = 40 - 20 + 1 = 21$, $|X_3| = \frac{1}{3}(39 - 21) + 1 = 7$, and $|X_5| = \frac{1}{5}(40 - 20) + 1 = 5$.

The numbers not divisible by 3 is $X - X_3$, and $|X - X_3| = 21 - 7 = 14$. The numbers not divisible by 5 is $X - X_5$, and $|X - X_5| = 21 - 5 = 16$.

The numbers divisible by 3 and 5 is $X_3 \cap X_5$, and are the numbers divisible by 15. The only such number between 20 and 40 is $2 \cdot 15 = 30$. Therefore $|X_3 \cap X_5| = 1$.

The numbers divisible by 3 or 5 is $X_3 \cup X_5$, and $|X_3 \cup X_5| = |X_3| + |X_5| - |X_3 \cap X_5| = 7 + 5 - 1 = 11$.

The numbers divisible by 3 but not by 5 is $X_3 - X_5$. By the disjoint union $X_3 = (X_3 - X_5) \sqcup (X_3 \cap X_5)$, we get $|X_3 - X_5| = |X_3| - |X_3 \cap X_5| = 7 - 1 = 6$.

What about divisible by 5 but not by 3? What about not divisible by 3 and not divisible by 5? What about not divisible by 3 or not divisible by 5?

The equality in Proposition 5.1.3 can be extended to the union of more subsets

$$
\begin{aligned}
|X_1 \cup X_2 \cup \cdots \cup X_n| &= \sum_{k=1}^{n} (-1)^{k+1} \sum_{1 \le i_1 < i_2 < \cdots < i_k \le n} |X_{i_1} \cup X_{i_2} \cup \cdots \cup X_{i_k}| \\
&= |X_1| + |X_2| + \cdots + |X_n| \\
&\quad - |X_1 \cap X_2| - |X_1 \cap X_3| - \cdots - |X_{n-1} \cap X_n| \\
&\quad + |X_1 \cap X_2 \cap X_3| + |X_1 \cap X_2 \cap X_4| + \cdots + |X_{n-2} \cap X_{n-1} \cap X_n| \\
&\quad \cdots \\
&\quad + (-1)^{n+1} |X_1 \cap X_2 \cap \cdots \cap X_n|.
\end{aligned}
$$

The extended formula is the *inclusion-exclusion principle*.

The following is the proof for the case $n = 3$, by repeatedly using Proposition 5.1.3

$$
\begin{aligned}
|X_1 \cup X_2 \cup X_3| &= |X_1 \cup X_2| + |X_3| - |(X_1 \cup X_2) \cap X_3| \\
&= |X_1| + |X_2| - |X_1 \cap X_2| + |X_3| - |(X_1 \cap X_3) \cup (X_2 \cap X_3)| \\
&= |X_1| + |X_2| + |X_3| - |X_1 \cap X_2| \\
&\quad - (|X_1 \cap X_3| + |X_2 \cap X_3| - |(X_1 \cap X_3) \cap (X_2 \cap X_3)|) \\
&= |X_1| + |X_2| + |X_3| - |X_1 \cap X_2| \\
&\quad - |X_1 \cap X_3| - |X_2 \cap X_3| + |X_1 \cap X_2 \cap X_3|.
\end{aligned}
$$

Exercise 5.1. Consider all natural numbers up to 100.

1. How many are divisible by 2? by 3? by 5?
2. How many are not divisible by 2? by 3? by 5?
3. How many are divisible by 2 and 3? by 2 and 5? by 3 and 5?
4. How many are divisible by at least two of 2, 3 and 5?
5. How many are not divisible by at least two of 2, 3 and 5?
6. How many are divisible by 2, 3, but not by 5?

Exercise 5.2. There are 5 students studying math, 18 students not studying math, 8 girl students not studying math, and 3 boy students studying math.

1. What is the total number of students?
2. How many boy students do not study math?
3. How many boy students are there?

You may let S, M, B, G be students, math students, boy students, girl students. Then express different groups of students in terms of these sets.

Exercise 5.3. Prove the inclusion-exclusion principle.

Exercise 5.4. Let X and Y be finite sets. Prove the following are equivalent.

- $|X| \leq |Y|$.
- There is a one-to-one map $f\colon X \to Y$.
- There is an onto map $g\colon Y \to X$.

Proposition 5.1.4. *For finite sets X and Y, we have $|X \times Y| = |X||Y|$.*

Proof. Let $|X| = m$ and $|Y| = n$. Then we have one-to-one correspondences

$$f\colon X \to \{1, 2, \dots, m\}, \quad g\colon Y \to \{1, 2, \dots, n\}.$$

Then we get a one-to-one correspondence

$$(x, y) \mapsto (f(x), g(y))\colon Z = \{1, 2, \dots, m\} \times \{1, 2, \dots, n\}.$$

We only need to find a one-to-one correspondence between Z and $\{1, 2, \dots, mn\}$. The following is an explicit formula

$$u(i, j) = n(i-1) + j\colon Z \to \{1, 2, \dots, mn\}.$$

The following is the inverse. For any k in the right, we have $0 \leq k - 1 \leq mn - 1$. Then we divide $k - 1$ by n to get $k - 1 = qn + r$, where q, r are integers satisfying $0 \leq q < m$ and $0 \leq r < n$. Then we define

$$v(k) = (q + 1, r + 1)\colon \{1, 2, \dots, mn\} \to Z.$$

Then you may verify $v(u(i, j)) = (i, j)$, and $u(v(k)) = k$. □

Example 5.1.3. How many maps are there from X to Y?

A map $f\colon X \to Y$ means for each $x \in X$, we choose an element $y = f(x) \in Y$. This gives a one-to-one correspondence

$$\mathrm{Map}(X \to Y) \longleftrightarrow \times_X Y.$$

The right side is X copies of Y, each copy labeled by one element of X. For example, for $X = \{1, 2, 3\}$ and $Y = \{a, b\}$, the one-to-one correspondence is

$$\mathrm{Map}(\{1, 2, 3\} \to \{a, b\}) \longleftrightarrow \{a, b\}_1 \times \{a, b\}_2 \times \{a, b\}_3.$$

Then the map $f(1) = a, f(2) = a, f(3) = b$ on the left ccorrespondes to (a, a, b) on the right. Conversely, (b, a, b) on the right corresponds to the function $g(1) = b, g(2) = a, g(3) = b$ on the left. In general, a map $f\colon X \to Y$ on the left corresponds to $f(x)$ in the x-th copy of Y on the right. Conversely, if an element on the right has y in the x-th copy of Y, then the corresponding map has $f(x) = y$.

By the one-to-one correspondence and Proposition 5.1.4, we get

$$|\mathrm{Maps}(X \to Y)| = |\times_X Y| = \underbrace{|Y||Y|\cdots|Y|}_{|X|} = |Y|^{|X|}.$$

This is the reason we also denote $\mathrm{Maps}(X \to Y)$ by Y^X.

The discussion above sticks to the definition of the number of elements. Here is a more intuitive explanation. A map f means that for each $x \in X$, we choose an element $f(x) \in Y$. The number of choices, for each fixed x, is $|Y|$. Now we need to make $|X|$ many such choices, and the choices are all independent. Therefore the total number of choices is $|X|$ copies of $|Y|$ multiplies together, which is $|Y|^{|X|}$.

Example 5.1.4. How many subsets are there in a finite set X. In other words, what is $|\mathcal{P}(X)|$?

For each subset $A \subset X$, we introduce a map

$$\chi_A(x) = \begin{cases} \text{yes}, & \text{if } x \in A \\ \text{no}, & \text{if } x \notin A \end{cases} : X \to \{\text{yes}, \text{no}\}.$$

Conversely, for any map $\chi : X \to \{\text{yes}, \text{no}\}$, we recover a subset $A = \{x\colon \chi(A) = 1\} \in \mathcal{P}(X)$. This gives a one-to-one correspondence

$$A \mapsto \chi_A\colon \mathcal{P}(X) \to \mathrm{Map}(X, \{\text{yes}, \text{no}\}).$$

Then by Example 5.1.3, we get

$$|\mathcal{P}(X)| = |\mathrm{Map}(X, \{\text{yes}, \text{no}\})| = |\{\text{yes}, \text{no}\}|^{|X|} = 2^{|X|}.$$

In mathematics, we actually use 1 and 0 for yes and no. Then χ_A is the *characteristic function* of A. See Exercise 2.44.

Example 5.1.5. Let $E(m, n)$ be the natural number pairs (i, j), such that $1 \le i \le m$, $1 \le j \le n$, and $i - j$ is even. Let $O(m, n)$ be the similar collection, except $i - j$ is odd. Then $|E(m, n)| + |O(m, n)| = mn$.

We have a one-to-one correspondence

$$(i, j) \in E(m, n) \longleftrightarrow (i, j + 1) \in O(m, n + 1) - O(m, 1).$$

Therefore $|E(m, n)| = |O(m, n + 1)| - |O(m, 1)|$. By the similar reason, we get $|O(m, n)| = |E(m, n + 1)| - |E(m, 1)|$. Then we get

$$|E(m, n)| = (|E(m, n + 2)| - |E(m, 1)|) - |O(m, 1)| = |E(m, n + 2)| - m.$$

Then for $n = 2k$, we have

$$|E(m, 2k)| = |E(m, 2(k-1))| + m = |E(m, 2(k-2))| + 2m = \cdots = km = \frac{1}{2}mn.$$

For $n = 2k+1$, we have

$$\begin{aligned} |E(m, 2k+1)| &= |E(m, 2(k-1)+1)| + m = |E(m, 2(k-2)+1)| + 2m = \cdots \\ &= |E(m,1)| + km = \begin{cases} l + km = \frac{1}{2}mn, & \text{if } m = 2l \\ l + 1 + km = \frac{1}{2}(mn+1), & \text{if } m = 2l+1 \end{cases}. \end{aligned}$$

We conclude that, if one of m, n is even, then $|E(m,n)| = |O(m,n)| = \frac{1}{2}mn$. If both m, n are odd, then $|E(m,n)| = \frac{1}{2}(mn+1)$ and $|O(m,n)| = \frac{1}{2}(mn-1)$.

Exercise 5.5. Verify u and v in the proof of Proposition 5.1.4 are inverse of each other.

Exercise 5.6. Use induction and Proposition 5.1.3 to prove Proposition 5.1.4.

Exercise 5.7. Let $E(n)$ and $O(n)$ be the numbers of subsets of $\{1, 2, \ldots, n\}$ with even and odd number of elements. Explain $E(n) = E(n-1) + O(n-1)$ and $O(n) = O(n-1) + E(n-1)$. Then find $|E(n)|$ and $|O(n)|$.

Exercise 5.8. Let

$$X_i(n) = \{A \subset \{1, 2, \ldots, n\} \colon |A| = 3k + i \text{ for some integer } k\}, \quad i = 0, 1, 2.$$

1. Prove $|X_0(n)| + |X_1(n)| + |X_2(n)| = 2^n$.
2. Prove $|X_0(n)| = |X_0(n-1)| + |X_2(n-1)|$, $|X_1(n)| = |X_1(n-1)| + |X_0(n-1)|$, and $|X_2(n)| = |X_2(n-1)| + |X_1(n-1)|$.
3. Prove that the vector $\vec{y}(n) = (|X_0(n)|, |X_1(n)|, |X_2(n)|) - \frac{2^{n+1}}{3}(1,1,1)$ satisfies $\vec{y}(n) = -\vec{y}(n-3)$.
4. Find the formulae for $|X_0(n)|, |X_1(n)|, |X_2(n)|$.

5.2 Cardinality

If there is a one-to-one correspondence between a set X and $\{1, 2, \ldots, n\}$, then X is a *finite set* and $|X| = n$ is the number of elements in X. The empty set is also finite with $|\emptyset| = 0$. If $X \neq \emptyset$, and there is no one-to-one correspondence between a set X and $\{1, 2, \ldots, n\}$, for any $n \in \mathbb{N}$, then X is an *infinite set*.

Intuitively, we feel many sets, such as $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{R}$, $\mathbb{R}^2$, are infinite. Inspired by the definition for finite sets, we may still compare sizes of any two (including infinite) sets.

Definition 5.2.1. Two sets X and Y have the same *cardinality* (or same *size*), and denoted $|X| = |Y|$, if there is a one-to-one correspondence between X and Y.

Exercise 5.9. Prove $\mathbb{N}$ is infinite by showing that a map $f\colon \{1, 2, \dots, n\} \to \mathbb{N}$ cannot be onto.

Exercise 5.10. Prove that $|X| = |Y|$ and $|Y| = |Z|$ imply $|X| = |Z|$.

Exercise 5.11. Prove that $|X_1| = |X_2|$ and $|Y_1| = |Y_2|$ imply $|X_1 \times X_2| = |Y_1 \times Y_2|$.

Exercise 5.12. Prove that $|X| = |Y|$ implies $|\mathcal{P}(X)| = |\mathcal{P}(Y)|$.

Example 5.2.1. The following is a one-to-one correspondence

$$f(a) = \begin{cases} 2a, & \text{if } a > 0 \\ 1 - 2a, & \text{if } a \le 0 \end{cases} : \mathbb{Z} \to \mathbb{N}.$$

Therefore natural numbers and integers have the same cardinality: $|\mathbb{N}| = |\mathbb{N}^2|$. We may loosely say that the number of integers is the same as the number of natural numbers.

Example 5.2.2. The following one-to-one correspondence shows that $\mathbb{N}$ and $\mathbb{N}^2$ have the same number of elements

$$f(m, n) = \frac{m(m-1)}{2} + n\colon \mathbb{N}^2 \to \mathbb{N}.$$

Combined with $|\mathbb{Z}| = |\mathbb{N}|$, we get $|\mathbb{Z}| = |\mathbb{N}| = |\mathbb{N}^2| = |\mathbb{Z}^2|$.

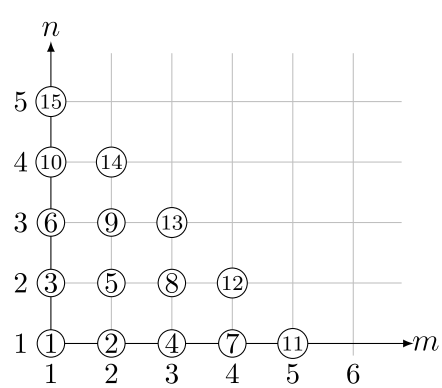

Figure 5.1: One-to-one correspondence between $\mathbb{N}^2$ and $\mathbb{N}$.

Example 5.2.3. The one-to-one correspondence $\frac{t}{1+|t|}\colon \mathbb{R} \to (-1,1)$ shows that $|\mathbb{R}| = |(-1,1)|$. Similarly, we know $\mathbb{R} = (a,b)$ for any open interval, including the possibility that $a = -\infty$ or $b = \infty$.

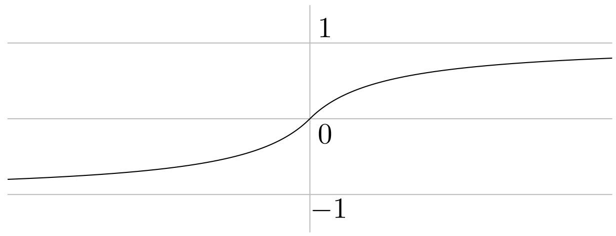

Figure 5.2: $\frac{t}{1+|t|}$ is a one-to-one correspondence between $\mathbb{R}$ and $(-1,1)$.

In general, it is quite difficult to actually construct a suitable one-to-one correspondence. The following important result simplifies the task quite a lot.

Theorem 5.2.2 (Cantor-Schröder-Bernstein). *If there is a one-to-one map from X to Y and another one-to-one map from Y to X, then $|X| = |Y|$.*

We may define $|X| \le |Y|$, and loosely say Y has bigger size than X, if there is a one-to-one map $X \to Y$. In particular, by the inclusion map, a subset has smaller size. The theorem says that $|X| \le |Y|$ and $|X| \ge |Y|$ implies $|X| = |Y|$.

Proof. Let $f\colon X \to Y$ and $g\colon Y \to X$ be one-to-one maps. Then we have a sequence of smaller and smaller subsets in X

$$X \supset g(Y) \supset gf(X) \supset gfg(Y) \supset gfgf(X) \supset gfgfg(Y) \supset \cdots .$$

We have the similar sequence in Y

$$Y \supset f(X) \supset fg(Y) \supset fgf(X) \supset fgfg(Y) \supset fgfgf(X) \supset \cdots .$$

For one-to-one map h and $A \supset B$, we have $h(A - B) = h(A) - h(B)$. Then the differences between adjacent subsets in the sequence are

$$\begin{array}{rr}
X\colon & Y\colon \\
X - g(Y) = U, & Y - f(X) = V, \\
g(Y) - gf(X) = g(V), & f(X) - fg(Y) = f(U), \\
gf(X) - gfg(Y) = gf(U), & fg(Y) - fgf(X) = fg(V), \\
gfg(Y) - gfgf(X) = gfg(V), & fgf(X) - fgfg(Y) = fgf(U), \\
gfgf(X) - gfgfg(Y) = gfgf(U), & fgfg(Y) - fgfgf(X) = fgfg(V), \\
gfgfg(Y) - gfgfgf(X) = gfgfg(V), & fgfgf(X) - fgfgfg(Y) = fgfgf(U), \\
\vdots & \vdots
\end{array}$$

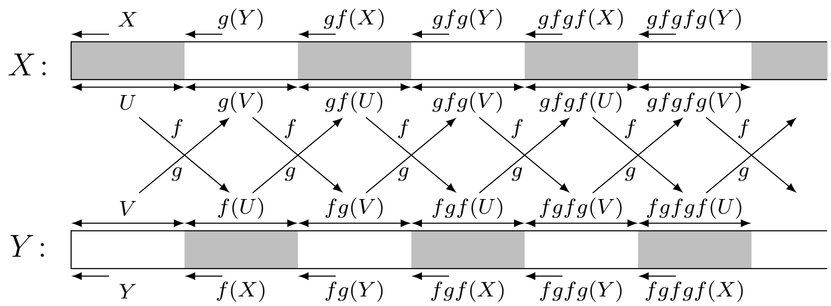

Figure 5.3: Cantor-Schroeder-Bernstein Theorem: $U = X - g(Y)$ and $V = Y - f(X)$.

Let us further denote these differences

$$\begin{aligned}
U &= U_0, & V &= V_0, \\
g(V) &= g(V_0), & f(U) &= f(U_0), \\
gf(U) &= U_1, & fg(V) &= V_1, \\
gfg(V) &= g(V_1), & fgf(U) &= f(U_1), \\
gfgf(U) &= U_2, & fgfg(V) &= V_2, \\
gfgfg(V) &= g(V_2), & fgfgf(U) &= f(U_2), \\
&\vdots & &\vdots
\end{aligned}$$

Both X and Y are the disjoint unions of these differences

$$\begin{aligned}
X &= U_0 \sqcup g(V_0) \sqcup U_1 \sqcup g(V_1) \sqcup U_2 \sqcup g(V_2) \sqcup \cdots \\
&= (U_0 \sqcup U_1 \sqcup U_2 \sqcup \cdots) \sqcup g(V_0 \sqcup V_1 \sqcup V_2 \sqcup \cdots) = A \sqcup g(B), \\
Y &= U_0 \sqcup g(V_0) \sqcup U_1 \sqcup g(V_1) \sqcup U_2 \sqcup g(V_2) \sqcup \cdots \\
&= (V_0 \sqcup V_1 \sqcup V_2 \sqcup \cdots) \sqcup f(U_0 \sqcup U_1 \sqcup U_2 \sqcup \cdots) = B \sqcup f(A),
\end{aligned}$$

In Figure 5.2, A and $f(A)$ are the shaded parts, and B and $g(B)$ are the white parts. The one-to-one maps f and g restrict to one-to-one maps $f_A \colon A \to f(A)$ and $g_B \colon B \to g(B)$. Moreover, f_A and g_B are already onto. Therefore f_A and g_B are invertible. The two invertible maps combine to give invertible maps between X and Y

$$\varphi(x) = \begin{cases} x, & \text{if } x \in A \\ g_B^{-1}(x), & \text{if } x \in g(B) \end{cases} : X \to Y,$$

$$\psi(x) = \begin{cases} y, & \text{if } y \in B \\ f_A^{-1}(y), & \text{if } y \in f(A) \end{cases} : Y \to X.$$

Therefore $|X| = |Y|$. $\square$

Example 5.2.4. Unlike Example 5.2.3, it is hard to find a one-to-one correspondence between $\mathbb{R}$ and a closed interval. In fact, continuous one-to-one correspondence exists only between $\mathbb{R}$ and open intervals. On the other hand, by the inclusion $(0,1) \subset [0,1] \subset \mathbb{R}$, we get $|\mathbb{R}| = |(0,1)| \le |[0,1]| \le |\mathbb{R}|$. Then by Theorem 5.2.2, we get $|[0,1]| = |\mathbb{R}|$. In general, for any interval I (including half open and half closed) consisting of more than two points, we have $|I| = |\mathbb{R}|$.

Example 5.2.5. Since $\mathbb{Z}$ is a subset of $\mathbb{Q}$. We have $|\mathbb{Z}| \le |\mathbb{Q}|$.

On the other hand, any nonzero rational number r can be *uniquely* written as $r = \dfrac{a}{b}$, with $b > 0$ and a, b coprime. We also write $0 = \dfrac{0}{1}$. Then $r \mapsto (a,b)\colon \mathbb{Q} \to \mathbb{Z} \times \mathbb{Z}$ is a one-to-one map. Therefore $|\mathbb{Q}| \le |\mathbb{Z} \times \mathbb{Z}| = |\mathbb{Z}|$, where the equality is from Example 5.2.2.

By Theorem 5.2.2, we conclude $|\mathbb{Z}| = |\mathbb{Q}|$.

Example 5.2.6. We compare $\mathbb{R}$ with $\mathbb{R}^2$. By $\mathbb{R}\times 0 \subset \mathbb{R}^2$, we get $|\mathbb{R}| = |\mathbb{R}\times 0| \le |\mathbb{R}^2|$.

For the reverse, by Example 5.2.5, we know the interval $[0,1)$ has the same size as $\mathbb{R}$. Any $x \in [0,1)$ has binary expansion

$$x = 0.n_1n_2n_3\cdots_{[2]} = \frac{n_1}{2} + \frac{n_2}{2^2} + \frac{n_3}{2^3} + \cdots, \quad n_i = 0 \text{ or } 1.$$

Such expansion is ambiguous when $n_i = 0$ and $n_{i+1} = n_{i+2} = n_{i+3} = \cdots = 1$. In this case, we use

$$\frac{1}{2^{i+1}} + \frac{1}{2^{i+2}} + \frac{1}{2^{i+3}} + \cdots = \frac{1}{2^i},$$

and choose $n_i = 1, n_{i+1} = n_{i+2} = n_{i+3} = \cdots = 0$ instead. For example, we will not use $0.100111\cdots_{[2]}$ (all digits in $\cdots$ are 1). Instead, we will use $0.101_{[2]} = 0.10100\cdots_{[2]}$ (all digits in $\cdots$ are 0). In this way, we fix unique binary expansion for each $x \in [0,1)$.

For any pair $(x,y) \in [0,1) \times [0,1)$, we write down their unique binary expansions

$$x = 0.m_1m_2m_3\cdots_{[2]}, \quad y = 0.n_1n_2n_3\cdots_{[2]}.$$

Then we define

$$f(x,y) = 0.m_1n_1m_2n_2m_3n_3\cdots_{[2]} : [0,1) \times [0,1) \to [0,1).$$

In other words, we put x in the odd binary places and put y in the even binary places. For example, we have $f(0.101_{[2]}, 0.001_{[2]}) = 0.101011_{[2]}$. Of course given $f(x,y)$, it is easy to recover x and y, by respectively picking even or odd binary places only. Therefore f is a one-to-one map, and we get $|[0,1) \times [0,1)| \subset |[0,1)|$. By Example 5.2.5, this implies $|\mathbb{R}^2| \le |\mathbb{R}|$.

By Theorem 5.2.2, we conclude $|\mathbb{R}^2| = |\mathbb{R}|$.

Example 5.2.7. We compare the power set $\mathcal{P}(\mathbb{N})$ with the real numbers $\mathbb{R}$.

An element $A \in \mathcal{P}(\mathbb{N})$ is a subset $A \subset \mathbb{N}$. The subset gives a number via the decimal expansion

$$f(A) = \sum_{k \in A} \frac{1}{10^k} \colon \mathcal{P}(\mathbb{N}) \to [0,1].$$

The idea is that the k-th decimal place is 1 for $k \in A$, and is 0 for $k \notin A$. For example, we have $f(\{1,4,5,6,8,\dots\}) = 0.10011101\cdots$. The map is one-to-one, and we get $|\mathcal{P}(\mathbb{N})| \le |[0,1]| = |\mathbb{R}|$.

For the reverse, we use the same idea, but with binary expansion. For any $x \in [0,1)$, we fix the unique binary expansion of x as in Example 5.2.6, and then define

$$g(x) = \{i \colon n_i = 1\} \subset \mathbb{N} \colon (0,1) \to \mathcal{P}(\mathbb{N}).$$

For example, we have $g(0.10011101\cdots) = \{1,4,5,6,8,\dots\}$. Since the binary expansion is uniquely determined by the subset $g(x)$ of binary places with 1 (the other binary places have 0), the map g is one-to-one, and we get $|\mathbb{R}| = |[0,1)| \le |\mathcal{P}(\mathbb{N})|$.

By Theorem 5.2.2, we conclude $|\mathcal{P}(\mathbb{N})| = |\mathbb{R}|$.

We remark that we used base 10 for f, and base 2 for g. The reason is that the ambiguities in the expansions are 9 for base 10 and 1 for base 2. To get f one-to-one, we need to use base 10, or any base except 2. To get g one-to-one, we need to use base 2.

Exercise 5.13. Prove that the following contain the same number of elements:

1. Real number: $\mathbb{R}$.
2. Unit circle: $\{(x,y) \in \mathbb{R}^2 \colon x^2 + y^2 = 1\}$.
3. Open unit disk: $\{(x,y) \in \mathbb{R}^2 \colon x^2 + y^2 < 1\}$.
4. Closed square: $\{(x,y) \in \mathbb{R}^2 \colon |x| \le 1, |y| \le 1\}$.
5. Unit sphere: $\{(x,y,z) \in \mathbb{R}^3 \colon x^2 + y^2 + z^2 = 1\}$.

Exercise 5.14. Prove $|\mathbb{R}^2| = |\mathbb{R}|$ by proving $|\mathcal{P}(\mathbb{N}) \times \mathcal{P}(\mathbb{N})| = |\mathcal{P}(\mathbb{N})|$, and using Example 5.2.1.

Exercise 5.15. Prove $|X| \le |Y|$ and $|Y| \le |Z|$ imply $|X| \le |Z|$.

Exercise 5.16. In Example 5.1.4, we know $\mathcal{P}(X) = \mathrm{Map}(X, \{0,1\})$. Therefore $|\mathcal{P}(\mathbb{N})| = |\mathbb{R}|$ can be interpreted as $|\mathrm{Map}(\mathbb{N}, Y)| = |\mathbb{R}|$ in case Y has two elements. Prove that the equality holds for any finite Y.

Exercise 5.17. Prove $|\mathrm{Map}(\mathbb{N},\mathbb{N})| = |\mathbb{R}|$ by using the following idea: A map $\mathrm{Map}(\mathbb{N},\mathbb{N})$ is a sequence of natural numbers $n_1, n_2, \dots$. Corresponding to the sequence, you may construct the number

$$f(n_1, n_2, \dots) = 0.\underbrace{0\cdots01}_{n_1}\underbrace{0\cdots01}_{n_2}\cdots.$$

5.3 Countability

Historically, natural numbers were invented for counting. Therefore we have the following definition.

Definition 5.3.1. A set is *countable* if there is a one-to-one map from the set into $\mathbb{N}$.

The countability of X means $|X| \le |\mathbb{N}|$. By Examples 5.2.1, 5.2.2, 5.2.5, we know $\mathbb{N}^n, \mathbb{Z}^n, \mathbb{Q}^n$ are countable. In fact, they are *countably infinite*.

Theorem 5.3.2. *A countable set is either finite, or has the same size as* $\mathbb{N}$.

The proposition means that, the elements in a countable set can be lined up and successively labeled by the natural numbers. Therefore a countable set is either $\{x_1, x_2, \dots, x_n\}$ or $\{x_1, x_2, \dots, x_n, \dots\}$.

Proof. Suppose there is a one-to-one map from X into $\mathbb{N}$. Then we may use the map to identify X with its image in $\mathbb{N}$. As far as the proposition is concerned, we may assume $X \subset \mathbb{N}$.

We inductively define $f\colon \mathbb{N} \to X$ by

$$f(1) = \min X, \quad f(n+1) = \min(X - \{f(1), f(2), \dots, f(n)\}).$$

If the induction stops after n-th step, then we get $X = \{f(1), f(2), \dots, f(n)\}$. The already constructed part of f gives an onto map

$$f\colon \{1, 2, \dots, n\} \to X = \{f(1), f(2), \dots, f(n)\}.$$

Moreover, for any $1 \le k \le n$, by $f(k) = \min(X - \{f(1), f(2), \dots, f(k-1)\})$, we know $f(k) \ne f(l)$ for all $1 \le l < k$. This means f is one-to-one. Therefore f is a one-to-one correspondence, and X is finite.

If the induction does not stop, then we get a map $f\colon \mathbb{N} \to X$. By the same reason as above, we know f is one-to-one. Therefore $|\mathbb{N}| \le |X|$. On the other hand, by $X \subset \mathbb{N}$, we get $|X| \le |\mathbb{N}|$. Then by Theorem 5.2.2, we get $|X| = |\mathbb{N}|$. $\square$

Proposition 5.3.3. *The subset of a countable set is countable. The union of countably many countable sets is countable. The product of finitely many countable sets is countable.*

Proof. A countable set is defined by one-to-one map into $\mathbb{N}$. For a subset, the composition with the inclusion map is still one-to-one. Therefore the subset is still countable.

Let $X_1, X_2, \cdots$ be countable sets. Then we have one-to-one maps $f_i \colon X_i \to \mathbb{N}$. We rewrite the union as a disjoint one

$$X = X_1 \cup X_2 \cup X_3 \cup \cdots = X_1 \sqcup (X_2 - X_1) \sqcup (X_3 - X_1 - X_2) \sqcup \cdots .$$

Then we define $f \colon X \to \mathbb{N} \times \mathbb{N}$ by

$$f(x) = (f_i(x), i) \in \mathbb{N} \times \mathbb{N} \text{ if } x \in X_i - X_1 - X_2 - \cdots - X_{i-1}.$$

Here the disjoint union is used to make sure f is well defined. Then the one-to-one properties of the maps f_i imply f is one-to-one. Then $|X| \leq |\mathbb{N} \times \mathbb{N}| = |\mathbb{N}|$. This proves that X is countable.

Let X and Y be countable. Then we have one-to-one maps $f \colon X \to \mathbb{N}$ and $g \colon Y \to \mathbb{N}$. Then the map

$$(f, g) \colon X \times Y \to \mathbb{N} \times \mathbb{N},\ (x, y) \mapsto (f(x), g(y))$$

is also one-to-one. Therefore $|X \times Y| \leq |\mathbb{N} \times \mathbb{N}| = |\mathbb{N}|$. This proves that $X \times Y$ is countable. The proof for the product of three or more countable sets is similar. □

Example 5.3.1. Let $\mathcal{P}_n(\mathbb{N})$ be the collection of subsets of $\mathbb{N}$ containing n numbers. Then any $A \in \mathcal{P}_n(\mathbb{N})$ is $A = \{k_1, k_2, \ldots, k_n\}$, with $k_1 < k_2 < \ldots < k_n$. We define a map

$$f(\{k_1, k_2, \ldots, k_n\}) = (k_1, k_2, \ldots, k_n) \colon \mathcal{P}_n(\mathbb{N}) \to \mathbb{N}^n.$$

The map is clearly one-to-one. Since $\mathbb{N}^n$ is countable, this implies $\mathcal{P}_n(\mathbb{N})$ is countable.

The collection $\mathcal{P}_{\text{finite}}(\mathbb{N})$ of all finite subsets of $\mathbb{N}$ is $\cup_{n=0}^{\infty} \mathcal{P}_n(\mathbb{N})$. As a union of countably many countable sets, we know $\mathcal{P}_{\text{finite}}(\mathbb{N})$ is also countable.

Finally, since $\mathcal{P}_n(\mathbb{N})$ and $\mathcal{P}_{\text{finite}}(\mathbb{N})$ are not finite, by Theorem 5.3.2, we get $|\mathcal{P}_n(\mathbb{N})| = |\mathcal{P}_{\text{finite}}(\mathbb{N})| = |\mathbb{N}|$.

Exercise 5.18. Prove that the product of two countable sets $X \times Y$ is countable, by writing $X \times Y$ as a union of countably many copies of X.

Exercise 5.19. If X is infinite, prove that $|X| \geq |\mathbb{N}|$.

Exercise 5.20. Suppose X is countably infinite, and $A \subset X$ is a finite subset. Prove that $X - A$ is still countably infinite.

Theorem 5.3.4. *For any set X, $|\mathcal{P}(X)| \neq |X|$.*

The one-to-one map $f(x) = \{x\} \colon X \to \mathcal{P}(X)$ implies $|X| \leq |\mathcal{P}(X)|$. The theorem means $|X| < |\mathcal{P}(X)|$.

Proof. Suppose there is a one-to-one correspondence $f\colon X \to \mathcal{P}(X)$. Consider the subset of X

$$A = \{x \in X : x \notin f(x)\} \in \mathcal{P}(X).$$

Since f is onto, we have $A = f(y)$ for some $y \in X$. There are two possibilities for y.

1. If $y \in A$, then $y \in f(y)$. By the definition of A, this means $y \notin A$.
2. If $y \notin A$, then $y \notin f(y)$. By the definition of A, this means $y \in A$.

We get contradiction in both cases. Therefore there is no one-to-one correspondence between X and $\mathcal{P}(X)$. □

Example 5.3.2. By Example 5.2.7, we know $|\mathcal{P}(\mathbb{N})| = |\mathbb{R}|$. By Theorem 5.3.4, we have $|\mathcal{P}(\mathbb{N})| > |\mathbb{N}|$. Therefore $|\mathbb{R}| > |\mathbb{N}|$, and $\mathbb{R}$ is uncountable.

Exercise 5.21. Determine countability.

1. Polynomials with rational coefficients.
2. Continuous functions on $[0, 1]$.
3. Real numbers, such that the decimal expansion does not have digit 8.
4. Even numbers.
5. Complex numbers with rational real and imaginary parts.
6. Irrational numbers.
7. All intervals.

Exercise 5.22. Prove that the set of continuous functions on $\mathbb{R}$ has the same size as $\mathbb{R}$.

Exercise 5.23. Prove that the set of integrable functions on $[0, 1]$ has the same size as $\mathcal{P}(\mathbb{R})$.