Discrete Structures

Beifang Chen

Contents

Chapter 1. Set Theory 5
1.1. Sets and Subsets 5
1.2. Set Operations 6
1.3. Functions 9
1.4. Injection, Surjection, and Bijection 11
1.5. Infinite Sets 13
1.6. Permutations 15

Chapter 2. Number Theory 19
2.1. Divisibility 19
2.2. Greatest Common Divisor 20
2.3. Least Common Multiple 23
2.4. Modulo Integers 24
2.5. RSA Cryptography System 26

Chapter 3. Propositional Logic 29
3.1. Statements 29
3.2. Connectives 29
3.3. Tautology 31
3.4. Methods of Proof 33
3.5. Mathematical Induction 35
3.6. Boolean Functions 35

Chapter 4. Combinatorics 39
4.1. Counting Principle 39
4.2. Permutations 39
4.3. Combination 42
4.4. Combination with Repetition 45
4.5. Combinatorial Proof 46
4.6. Pigeonhole Principle 50
4.7. Relation to Probability 51
4.8. Inclusion-Exclusion Principle 53
4.9. More Examples 57
4.10. Generalized Inclusion-Exclusion Formula 59

Chapter 5. Recurrence Relations 63
5.1. Infinite Sequences 63
5.2. Homogeneous Recurrence Relations 64
5.3. Higher Order Homogeneous Recurrence Relations 67
5.4. Non-homogeneous Equations 68

5.5. Divide-and-Conquer Method 69
5.6. Searching and Sorting 75
5.7. Growth of Functions 75

Chapter 6. Binary Relations 77
6.1. Binary Relations 77
6.2. Representation of Relations 79
6.3. Composition of Relations 80
6.4. Special Relations 82
6.5. Equivalence Relations and Partitions 84

CHAPTER 1

Set Theory

1.1. Sets and Subsets

A **set** is a collection of objects satisfying certain properties; the objects in the collection are called **elements** (or **objects** or **members**). A set is considered to be a whole entity and is different from its elements. Given a set A; we write "$x \in A$" to say that x is an element of A or x belongs to A, and write "$x \notin A$" to say that x is not an element of A or x doesn't belong to A. We usually denote sets by uppercase letters such as $A, B, C, \ldots, X, Y, Z$, and denote the elements of a set by lowercase letters such as $a, b, c, \ldots, x, y, z$, etc.

There are two ways to express a set. One way is to list all elements of the set; the other way is to point out the attributes of the elements of the set. For example, let A be the set of integers whose absolute values are less than or equal to 3. The set A can be described in two ways:

$$A = \{-3, -2, -1, 0, 1, 2, 3\} \quad \text{or} \quad A = \{a \mid a \text{ is an integer, } |a| \leq 3\}.$$

A set X whose elements satisfying Property P is denoted by

$$X = \{x \mid x \text{ satisfies } P\} \quad \text{or} \quad X = \{x : x \text{ satisfies } P\}.$$

In this note, most of time we use the first notation $X = \{x \mid x \text{ satisfies } P\}$, and occasionally use the second notation when there is confusion to use the symbol "$|$".

There are two important things to be noticed about the concept of sets. The first one is that any set, when it is considered as an object, can not be an element of itself, but can be an element of another set. The second one is that for a particular object, it is possible to decide in principle whether or not the object is an element of a given set.

Let A and B be sets of real numbers satisfying the equations $x^2 - 1 = 0$ and $x^4 - 1 = 0$, respectively. In set notation,

$$A = \{x \mid x \in \mathbb{R},\, x^2 - 1 = 0\} \quad \text{and} \quad B = \{x \mid x \in \mathbb{R},\, x^4 - 1 = 0\}.$$

Apparently, the equation to define the elements of A and B are different. However, the sets A and B consist of exactly the same elements, namely, 1 and -1. For this reason we say that A and B are equal to each other, written $A = B$; it does not matter whether or not the sets A and B were defined in different ways.

The sets we will constantly use in our course are the following sets:

$$\begin{array}{lcl}
\mathbb{Z}: & = & \text{the set of integers;} \\
\mathbb{Q}: & = & \text{the set of rational numbers;} \\
\mathbb{R}: & = & \text{the set of real numbers;} \\
\mathbb{C}: & = & \text{the set of complex numbers;} \\
\mathbb{P}: & = & \text{the set of positive integers;} \\
\mathbb{N}: & = & \text{the set of nonnegative integers.}
\end{array}$$

Two sets A and B are called **equal**, written $A = B$, if every element of X is an element of B and every element of B is also an element of A. As usual, we write "$A \neq B$" to say that the sets A and B are not equal. In other words, there is at least one element of A which is not an element of B, or, there is at least one element of B which is not an element of A.

A set A is called a **subset** of a set B, written $A \subset B$, if every element of A is an element of B. Thus, for two sets A and B, $A = B$ if and only if $A \subset B$ and $B \subset A$. A set is called **finite** if it has only finite number of elements; otherwise, it is called **infinite**. For a finite set A, we denote by $|A|$ the number of elements of A; we call $|A|$ the **cardinality** of A.

Consider the set A of real numbers satisfying the equation $x^2 + 1 = 0$. We will see that the set contains no elements at all; we call it empty. The set without any element is called the **empty set**. There is one and only one empty set, and is denoted by $\emptyset$. The empty set $\emptyset$ is a subset of any set and $|\emptyset| = 0$.

EXERCISE 1. Let $A = \{1, 2, 3, 4, a, b, c, d\}$. Identify each of the following as true or false.

(a) $2 \in A$; (b) $3 \notin A$; (c) $c \in A$; (d) $d \notin A$; (e) $6 \in A$; (f) $e \in A$;
(g) $8 \notin A$; (h) $f \notin A$; (i) $\emptyset \in A$; (j) $A \in A$; (k) $\} \in A$; (l) $, \in A$.

EXERCISE 2. List all subsets of a set A, where (a) $A = \{1\}$; (b) $A = \{1, 2\}$; (c) $A = \{1, 2, 3\}$; (d) $A = \{1, 2, 3, 4\}$.

EXERCISE 3. Draw the Venn diagram that represents the following relationships.

(1) $A \subset B$, $A \subset C$, $B \not\subset C$, and $C \not\subset B$.
(2) $x \in A$, $x \in B$, $x \notin C$, $y \in B$, $y \in C$, and $y \notin A$.
(3) $A \subset B$, $x \notin A$, $x \in B$, $A \not\subset C$, $y \in B$, $y \in C$.

1.2. Set Operations

Let A and B be two sets. The **intersection** of A and B, written $A \cap B$, is the set of all elements common to the both sets A and B. In set notation,

$$A \cap B = \{x \mid x \in A \text{ and } x \in B\}.$$

The **union** of A and B, written $A \cup B$, is the set consisting of the elements belonging to either the set A or the set B, that is,

$$A \cup B = \{x \mid x \in A \text{ or } x \in B\}.$$

The **symmetric difference** $A \Delta B$ of A and B is the set

$$A \Delta B = \{x \mid x \in A \text{ or } x \in B, \text{ but } x \notin A \cap B\}.$$

The **relative complement** of A in B is the set consisting of the elements of B that is not in A, that is,

$$B - A = \{x \mid x \in B, x \notin A\}.$$

When we only consider subsets of a fixed set U, this fixed set U is sometimes called a **universal set**. It should be noticed that a universal set is *not* universal; it does not mean that it contains everything. For a universal set U and a subset $A \subset U$, the relative complement $U - A$ is just called the **complement** of A, written

$$\bar{A} = U - A.$$

Since we always consider the elements in U, so, when $x \in \bar{A}$, it is equivalent to $x \notin A$. Similarly, $x \in A$ is equivalent to $x \notin \bar{A}$. To save writing space, we sometimes use the symbol "$\Longleftrightarrow$" instead of writing "is (are) equivalent to." For instance, we may write "$x \in \bar{A}$ is equivalent to $x \notin A$" as "$x \in \bar{A} \Longleftrightarrow x \notin A$."

A convenient way to visualize sets in a universal set U is the Venn diagram. We usually use a rectangle to represent the universal set U and circles or ovals to represent its subsets as follows:

Figure

The **intersection** of a finite number of sets $A_1, A_2, \ldots, A_n$ is the set consisting of elements common to all $A_1, A_2, \ldots, A_n$, that is,

$$\bigcap_{i=1}^{n} A_i = A_1 \cap A_2 \cap \cdots \cap A_n = \Big\{x \mid x \in A_1, x \in A_2, \ldots, x \in A_n\Big\}.$$

Similarly, the **union** of $A_1, A_2, \ldots, A_n$ is the set, each of its element is contained in some A_i, that is,

$$\begin{aligned}\bigcup_{i=1}^{n} A_i &= A_1 \cup A_2 \cup \cdots \cup A_n \\ &= \Big\{x \mid \text{there exists at least one } A_i \text{ such that } x \in A_i\Big\}.\end{aligned}$$

Let $A_1, A_2, \ldots$ be infinitely many sets. We define the **intersection**

$$\bigcap_{i=1}^{\infty} A_i = A_1 \cap A_2 \cap \cdots = \Big\{x \mid x \in A_1, x \in A_2, \ldots\Big\}$$

and the **union**

$$\bigcup_{i=1}^{\infty} A_i = A_1 \cup A_2 \cup \cdots = \Big\{x \mid \text{there exists one } i \text{ such that } x \in A_i\Big\}$$

Let A_i, $i \in I$, be a family of sets, indexed by a set I. We can also define the **intersection**

$$\bigcap_{i \in I} A_i = \Big\{x \mid x \in A_i \text{ for all } i \in I\Big\}$$

and the **union**

$$\bigcup_{i \in I} A_i = \Big\{x \mid x \in A_i \text{ for at least one } i \in I\Big\}.$$

THEOREM 1.1 (**DeMorgan's Law**). *Let A and B be subsets of a universal set U. Then*

(1) $\bar{\bar{A}} = A$,
(2) $\overline{A \cap B} = \bar{A} \cup \bar{B}$,
(3) $\overline{A \cup B} = \bar{A} \cap \bar{B}$.

PROOF. (1) $x \in \bar{\bar{A}} \Longleftrightarrow x \notin \bar{A} \Longleftrightarrow x \in A$.

(2) $x \in \overline{A \cap B} \Longleftrightarrow x \notin A \cap B \Longleftrightarrow x \notin A$ or $x \notin B \Longleftrightarrow x \in \bar{A}$ or $x \in \bar{B} \Longleftrightarrow x \in \bar{A} \cup \bar{B}$.

(3) $x \in \overline{A \cup B} \Longleftrightarrow x \notin A \cup B \Longleftrightarrow x \notin A$ and $x \notin B \Longleftrightarrow x \in \bar{A}$ and $x \in \bar{B} \Longleftrightarrow x \in \bar{A} \cap \bar{B}$. $\square$

The **power set** of a set A, written $\mathcal{P}(A)$, is the set of all subsets of A.

EXAMPLE 1.1. The power set of the set $A = \{a, b, c\}$ is the set

$$\mathcal{P}(A) \quad = \quad \Big\{\emptyset, \{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}, \{a, b, c\}\Big\}.$$

The **Cartesian product** (or just **product**) of two sets A and B, written $A \times B$, is the set consisting of all ordered pairs (a, b), where $a \in A$ and $b \in B$, that is,

$$A \times B \quad = \quad \{(a, b) \mid a \in A \text{ and } b \in B\}.$$

The **product** of a finite family of sets $A_1, \ldots, A_n$ is the set

$$\prod_{i=1}^{n} A_i = A_1 \times \cdots \times A_n = \Big\{(a_1, \ldots, a_n) \mid a_1 \in A_1, \ldots, a_n \in A_n\Big\},$$

the element $(a_1, \ldots, a_n)$ is called an **ordered n-tuple**. The product of an infinite family $A_1, A_2, \ldots$ of sets is the set

$$\prod_{i=1}^{\infty} A_i = A_1 \times A_2 \times \cdots = \Big\{(a_1, a_2, \ldots) \mid a_1 \in A_1, a_2 \in A_2, \ldots\Big\}.$$

Each element of $\prod_{i=1}^{\infty} A_i$ can be considered as an infinite sequence. If $A_1 = A_2 = \cdots = A$, we write

$$\begin{aligned} A^n &= \underbrace{A \times \cdots \times A}_{n}, \\ A^\infty &= \underbrace{A \times A \times \cdots}_{\infty}. \end{aligned}$$

EXAMPLE 1.2. For sets $A = \{0, 1\}$, $B = \{a, b, c\}$, the product A and B is the set

$$A \times B = \{(0, a), (0, b), (0, c), (1, a), (1, b), (1, c)\};$$

and the product and $A^3 = A \times A \times A$ is the set

$$A^3 = \{(0, 0, 0), (0, 0, 1), (0, 1, 1), (1, 0, 0), (1, 0, 1), (1, 1, 0), (1, 1, 1)\}.$$

For the set $\mathbb{R}$ of real numbers, the product $\mathbb{R}^2$ is the 2-dimensional coordinate plane and $\mathbb{R}^3$ is the 3-dimensional coordinate space.

A **sequence** of a nonempty set A is a list of finite or infinite number of objects of A in order:

$$\begin{array}{ll} a_1, a_2, \ldots, a_n & \text{(finite sequence)} \\ a_1, a_2, \ldots & \text{(infinite sequence)} \end{array}$$

where $a_1, a_2, \ldots \in A$. The sequence is called **finite** in the former case and **infinite** in the latter case. A **word of length** n over a nonempty set A is a string

$$a_1 a_2 \cdots a_n,$$

where $a_1, a_2, \ldots, a_n \in A$. The string

$$a_1 a_2 \cdots$$

with $a_1, a_2, \ldots \in A$ is called a **word of infinite length** over A. There is a unique word of length zero, called the **empty word**, and is denoted by λ. The sets of all words of length n, of finite length, and of infinite length over A are denoted by

$$A^{(n)}, \quad A^*, \quad \text{and} \quad A^{(\infty)},$$

respectively. Note that

$$A^* = \bigcup_{n=0}^{\infty} A^{(n)}.$$

EXERCISE 4. Let A be a set, and let A_i, $i \in I$, be a family of sets. Show that

$$\begin{aligned}
\overline{\bigcup_{i\in I} A_i} &= \bigcap_{i\in I} \overline{A_i};\\
\overline{\bigcap_{i\in I} A_i} &= \bigcup_{i\in I} \overline{A_i};\\
A \cap \bigcup_{i\in I} A_i &= \bigcup_{i\in I} (A \cap A_i);\\
A \cup \bigcap_{i\in I} A_i &= \bigcap_{i\in I} (A \cup A_i).
\end{aligned}$$

EXERCISE 5. Let A, B, C be finite sets. Use Venn diagram to show that

$$\begin{aligned}
|A \cup B \cup C| &= |A| + |B| + |C| \\
&\quad -|A \cap B| - |A \cap C| - |B \cap C| + |A \cap B \cap C|.
\end{aligned}$$

1.3. Functions

DEFINITION 1.2. Let X and Y be nonempty sets. A **function** f from X to Y, written

$$f : X \to Y,$$

is a rule that associates each element $x \in X$ with a unique element $y \in Y$, denoted

$$y = f(x).$$

When this rule f is given, the sets X, Y, and $f(X) = \{f(x) \mid x \in X\}$ are called the **domain**, the **codomain**, and the **range** of f, respectively; the element $y(= f(x))$ is called the **image** (or **value**) of x, and x is called the **inverse image** (or **preimage**) of y under f. Functions are also called **mappings** or just **maps**.

Note that for a function $y = f(x)$ in calculus, the variable x is called the **independent variable** and y is called the **dependent variable**.

Let $f : X \to Y$ be a function from a set X to a set Y. It can be viewed as a black-box device

$$x \longrightarrow \boxed{f} \longrightarrow f(x),$$

where the input x is in X and the output $f(x)$ is to be in Y. For a subset $A \subset X$, the **image** of A under f is the set

$$f(A) := \{y \in Y \mid \text{there is } a \in A \text{ such that } y = f(a)\};$$

and for a subset $B \subset Y$, the **inverse image** of B under f is the set

$$f^{-1}(B) := \{x \in X \mid \text{there is } b \in B \text{ such that } b = f(x)\}.$$

The **graph** of f is the set

$$\Gamma(f) = \{(x, y) \in X \times Y \mid y = f(x)\}.$$

The set of all functions from a set A to a set B is sometimes denoted by B^A; that is,

$$B^A := \{f \mid f : A \to B\}.$$

EXAMPLE 1.3. Some ordinary functions.

(1) The usual function $y = x^2$ can be considered as a function $f : \mathbb{R} \to \mathbb{R}$, defined by $f(x) = x^2$ for $x \in \mathbb{R}$. Its domain and codomain are the set $\mathbb{R}$ of real numbers; its range is the set $\mathbb{R}_{\geq 0}$ of nonnegative real numbers.
(2) The usual function $y = e^x$ can be considered as a function $f : \mathbb{R} \to \mathbb{R}_+$, defined by $f(x) = e^x$. Its domain is $\mathbb{R}$; its codomain and range are the set $\mathbb{R}_+$ of positive real numbers.
(3) $y = \log x$ is a function from $\mathbb{R}_+$ to $\mathbb{R}$; its domain is $\mathbb{R}_+$ and codomain is $\mathbb{R}$.
(4) $y = |x|$ is a function from $\mathbb{R}$ to the set $\mathbb{R}_{\geq 0}$.
(5) $y = \sin x$ is a function from $\mathbb{R}$ to the closed interval $[-1, 1]$ of real numbers.

EXAMPLE 1.4. Some functions to be appeared in future lectures.

(1) A finite sequence $s_1, s_2, \ldots, s_n$ of a set A can be viewed as a function $s : \{1, 2, \ldots, n\} \to A$, defined by

$$s(k) = s_k, \quad k = 1, 2, \ldots, n.$$

An infinite sequence $s_1, s_2, \ldots$ of A can be viewed as a function $s : \mathbb{P} \to A$, defined by $s(k) = s_k$, $k \in \mathbb{P}$.
(2) The **factorial** is a function $f : \mathbb{N} \to \mathbb{P}$ defined by

$$f(n) = n! = 1 \cdot 2 \cdot 3 \cdots (n-1)n.$$

(3) The **floor function** is the function $\lfloor \cdot \rfloor : \mathbb{R} \to \mathbb{Z}$, defined by

$$\lfloor x \rfloor = \text{the greatest integer less than or equal to } x.$$

(4) The **ceiling function** is the function $\lceil \cdot \rceil : \mathbb{R} \to \mathbb{Z}$, defined by

$$\lceil x \rceil = \text{the smallest integer greater than or equal to } x.$$

(5) For a universal set U, the **characteristic function** of a subset $A \subset U$ is the function $\chi_A : U \to \{0, 1\}$, defined by

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

If U is finite and its elements are listed as $\{a_1, a_2, \ldots, a_n\}$ or simply identify U as the set $\{1, 2, \ldots, n\}$. Then the subsets can be identified as sequences of 0 and 1 of length n. For instance, let $U = \{1, 2, 3, 4, 5, 6, 7, 8\}$, then the subset $A = \{2, 4, 5, 7, 8\}$ corresponds to the sequence

0	1	0	1	1	0	1	1

(6) Let a be a positive integer. Then for any integer b there exist unique integers q and r such that

$$b = qa + r, \quad 0 \leq r < a.$$

We have a function $\mathsf{Quo}_a : \mathbb{Z} \to \mathbb{Z}$, defined by

$$\mathsf{Quo}_a(b) = q, \quad b \in \mathbb{Z};$$

and a function $\mathsf{Rem}_a : \mathbb{Z} \to \{0, 1, 2, \ldots, a-1\}$, defined by

$$\mathsf{Rem}_a(b) = r, \quad b \in \mathbb{Z}.$$

Let $f : X \to \mathbb{R}$ and $g : X \to \mathbb{R}$ be functions. The **addition** of f and g is the function $f + g : X \to \mathbb{R}$ defined by

$$(f+g)(x) = f(x) + g(x), \quad x \in X;$$

the **scalar multiplication** of f by a constant c is the function $cf : X \to \mathbb{R}$ defined by

$$(cf)(x) = cf(x), \quad x \in X.$$

EXERCISE 6. Let χ_A and χ_B be the characteristic functions of subsets A and B of a universal set X, respectively. Express the characteristic functions of $A \cap B$, $A \cup B$, $A \Delta B$, and $B - A$ in terms of χ_A and χ_B, respectively. (Hint: Since $\{0,1\} \subset \mathbb{R}$, the functions χ_A and χ_B can be viewed as functions from A and B to $\mathbb{R}$, respectively. So the linear combination of χ_A and χ_B is meaningful.)

1.4. Injection, Surjection, and Bijection

DEFINITION 1.3. Let X and Y be nonempty sets.

(1) A function $f : X \to Y$ is called **injective** (or **one-to-one**) if distinct elements of X are mapped to distinct elements in Y; that is, for $x_1, x_2 \in X$, if $x_1 \neq x_2$, then $f(x_1) \neq f(x_2)$. An injective function is also called an **injection** (or **one-to-one mapping**).
(2) A function $f : X \to Y$ is called **surjective** (or **onto**) if every element in Y is an image of some elements of X; that is, for any $y \in Y$, there exist $x \in X$ such that $f(x) = y$. A surjective function is also called a **surjection** (or **onto mapping**).
(3) A function $f : X \to Y$ is called **bijective** if it is both injective and surjective. A bijective function is also called a **bijection** (or **one-to-one correspondence**).

EXAMPLE 1.5. (1) The function $f : \mathbb{R} \to \mathbb{R}$, $f(x) = e^x$, is injective, but not surjective.
(2) The function $f : \mathbb{R} \to \mathbb{R}_{\geq 0}$, $f(x) = x^2$, is surjective, but not injective.
(3) The function $f : \mathbb{R} \to \mathbb{R}$, $f(x) = x^3$, is bijective.
(4) The function $f : \mathbb{R}_+ \to \mathbb{R}$, $f(x) = \log x$, is bijective.

DEFINITION 1.4. The **composition** of a function $f : X \to Y$ and a function $g : Y \to Z$ is a function $g \circ f : X \to Z$, defined by

$$(g \circ f)(x) = g(f(x)), \quad x \in X.$$

Whenever composition $g \circ f$ is concerned, we assume that the codomain of f is the same as the domain of g.

THEOREM 1.5. *Let $f : X \to Y$, $g : Y \to Z$, $h : Z \to W$ be functions. Then*

$$h \circ (g \circ f) = (h \circ g) \circ f,$$

as functions from X to W. We write $h \circ g \circ f = h \circ (g \circ f) = (h \circ g) \circ f$.

PROOF. For any $x \in X$, we have

$$\begin{aligned} (h \circ (g \circ f))(x) &= h((g \circ f)(x)) \\ &= h((g(f(x))) \\ &= (h \circ g)(f(x)) \\ &= ((h \circ g) \circ f)(x). \end{aligned}$$

Fig.

$\square$

The **identity function** of a set X is the function $\mathrm{id}_X : X \to X$ such that $\mathrm{id}_X(x) = x$ for all $x \in X$.

DEFINITION 1.6. A function $f : X \to Y$ is called **invertible** if there exists a function $g : Y \to X$ such that for any $x \in X$ and $y \in Y$,

$$g(f(x)) = x \quad \text{and} \quad f(g(y)) = y.$$

In other words, $g \circ f = \mathrm{id}_X$ and $f \circ g = \mathrm{id}_Y$; the function g is called an **inverse** of f, and is denoted by f^{-1}.

EXAMPLE 1.6. Some invertible and non-invertible functions.

(1) The function $f : \mathbb{R} \to \mathbb{R}$, $f(x) = x^3$, is invertible; its inverse is the function $g : \mathbb{R} \to \mathbb{R}$, $g(x) = \sqrt[3]{x}$.
(2) The function $f : \mathbb{R} \to \mathbb{R}_+$, $f(x) = e^x$, is invertible; its inverse is the function $g : \mathbb{R}_+ \to \mathbb{R}$, $g(x) = \log x$.
(3) The function $f : \mathbb{R} \to \mathbb{R}$, $f(x) = x^2$, is not invertible. However, the function $f_1 : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$, $f_1(x) = x^2$, is invertible, and its inverse is the function $g_1 : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$, $g_1(x) = \sqrt{x}$. The function $f_2 : \mathbb{R}_{\leq 0} \to \mathbb{R}_{\geq 0}$, $f_2(x) = x^2$, is invertible; its inverse is $g_2 : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\leq 0}$, $g_2(x) = -\sqrt{x}$.
(4) The function $f : \mathbb{R} \to [-1, 1]$, $f(x) = \sin x$, is not invertible. However, $f_1 : [-\frac{\pi}{2}, \frac{\pi}{2}] \to [-1, 1]$, $f_1(x) = \sin x$, is invertible, and has the inverse $g_1 : [-1, 1] \to [-\frac{\pi}{2}, \frac{\pi}{2}]$, $g_1(x) = \arcsin x$.

THEOREM 1.7. *A function $f : X \to Y$ is invertible if and only if f is one-to-one and onto.*

PROOF. "$\Rightarrow$": Since f is invertible, there is a function $g : Y \to X$ such that $g \circ f = \mathrm{id}_X$ and $f \circ g = \mathrm{id}_Y$. Suppose f is not one-to-one. Then there are $x_1, x_2 \in X$ such that $f(x_1) = f(x_2)$. Thus $x_1 = g(f(x_1)) = g(f(x_2)) = x_2$, a contradiction. So f is one-to-one. On the other hand, for any $y \in Y$, we have an element $g(y) \in X$ and $f(g(y)) = y$. This means that f is onto.

"$\Leftarrow$": Since f is one-to-one and onto, then for any $y \in Y$ there is one and only one element $x \in X$ such that $f(x) = y$. We define a function $g : Y \to X$ by $g(y) = x$, where x is the unique element of X such that $f(x) = y$. Then $(g \circ f)(x) = g(f(x)) = g(y) = x$ and $(f \circ g)(y) = f(g(y)) = f(x) = y$. Thus $g \circ f = \mathrm{id}_X$ and $f \circ g = \mathrm{id}_Y$. $\square$

EXERCISE 7. Show that if $f : X \to Y$ is invertible, then the inverse function of f is unique.

EXERCISE 8. Let $f : X \to Y$ be a function. Let A_i, $i \in I$, be a family of subsets of X. Show that

(1) $f\left(\bigcap_{i \in I} A_i\right) \subset \bigcap_{i \in I} f(A_i)$;
(2) $f\left(\bigcup_{i \in I} A_i\right) = \bigcup_{i \in I} f(A_i)$;
(3) $f^{-1}\left(\bigcap_{i \in I} A_i\right) = \bigcap_{i \in I} f^{-1}(A_i)$;
(4) $f^{-1}\left(\bigcup_{i \in I} A_i\right) = \bigcup_{i \in I} f^{-1}(A_i)$.

EXERCISE 9. Let $f : X \to Y$ and $g : Y \to Z$ be functions. Show that

(1) If f and g are one-to-one, then $g \circ f$ is one-to-one;

(2) If f and g are onto, then $g \circ f$ is onto;
(3) If $g \circ f$ is one-to-one, then f is one-to-one;
(4) If $g \circ f$ is onto, then g is onto.

EXERCISE 10. Let X be a finite set and let $f : X \to X$ be a function. Then f is bijective $\iff$ f is one-to-one $\iff$ f is onto.

EXERCISE 11. Let $f : X \to X$ be an invertible function. Let $f^0 = \mathrm{id}_X$. For any positive integer k, let

$$\begin{aligned} f^k &= f^{k-1} \circ f = \underbrace{f \circ f \circ \cdots \circ f}_{k}, \\ f^{-k} &= f^{-(k-1)} \circ f^{-1} = \underbrace{f^{-1} \circ f^{-1} \circ \cdots \circ f^{-1}}_{k}. \end{aligned}$$

For each $x \in X$, the **orbit** of x under the map f is the set

$$O_f(x) = \{f^k(x) \mid k \in \mathbb{Z}\}.$$

Show that for $x_1, x_2 \in X$, if $O_f(x_1) \cap O_f(x_2) \neq \emptyset$, then $O_f(x_1) = O_f(x_2)$.

EXERCISE 12. Verify that the function $f : \mathbb{R} - \{0,1\} \to \mathbb{R} - \{0,1\}$, defined by $f(x) = \frac{1}{1-x}$, is a bijection; then list all elements of the set $\{f^k \mid k \in \mathbb{Z}\}$.

1.5. Infinite Sets

DEFINITION 1.8. A set A is called **equivalent** to a set B, written $A \sim B$, if there is a one-to-one correspondence from A to B. The quantity to measure the number of elements of a set A is the **cardinality** of A, denoted $|A|$. Two sets have the same cardinality if they are equivalent, that is, they are in one-to-one correspondent.

The cardinality of a finite set is just the number of elements of the set. The empty set has the cardinality 0. For any set A,

$$|A^{(n)}| = |A^n| \quad \text{and} \quad |A^{(\infty)}| = |A^\infty|.$$

If A has exactly m elements, then there is a one-to-one correspondence from A to the set $[m] = \{1, 2, \ldots, m\}$.

DEFINITION 1.9. A set A is called **countable** if it is finite or there is a one-to-one correspondence from A to the set $\mathbb{P}$ of positive integers. Sets that are not countable are called **uncountable**.

PROPOSITION 1.10. *Every infinite set contains an infinite countable subset.*

PROOF. Let A be an infinite set. Select an element from A, say a_1. Since A is infinite, one can select an element from A other than a_1, say a_2. Similarly, one can select an element a_3 from A other than both a_1 and a_2. Since the infinity of A, one can continue this procedure by selecting a sequence of elements one after the other to get an infinite countable subset $\{a_1, a_2, a_3, \ldots\}$. $\square$

THEOREM 1.11. *If A and B are countable subsets, then $A \cup B$ is countable.*

PROOF. It is obviously true when one of A and B is a finite set. Let $A = \{a_1, a_2, \ldots\}$ and $B = \{b_1, b_2, \ldots\}$ be countable infinite sets. If $A \cap B = \emptyset$, then $A \cup B = \{a_1, b_1, a_2, b_2, \ldots\}$ is countable as demonstrated. If $A \cap B \neq \emptyset$, we just need to delete the elements that appeared more than once in the sequence $a_1, b_1, a_2, b_2, \ldots$. Then the leftover is the set $A \cup B$. $\square$

THEOREM 1.12. *Let A_i $(i = 1, 2, \cdots)$ be countable sets and $A_i \cap A_j = \emptyset\,(i \neq j)$. Then $\bigcup_{i=1}^{\infty} A_i$ is countable.*

PROOF. We assume that $A_i = \{a_{i1}, a_{i2}, a_{i3}, \cdots\}$ $(i = 1, 2, \ldots)$. Then the countability of $\bigcup_{i=1}^{\infty} A_i$ can be demonstrated as

$$
\begin{array}{cccccccc}
a_{11} & \rightarrow & a_{12} & & a_{13} & \rightarrow & a_{14} & \cdots \\
 & \swarrow & & \nearrow & & \swarrow & & \\
a_{21} & & a_{22} & & a_{23} & & a_{24} & \cdots \\
\downarrow & \nearrow & & \swarrow & & \nearrow & & \\
a_{31} & & a_{32} & & a_{23} & & a_{34} & \cdots \\
 & \swarrow & & \nearrow & & \swarrow & & \\
a_{41} & & a_{42} & & a_{23} & & a_{44} & \cdots \\
\vdots & & \vdots & & \vdots & & \vdots &
\end{array}
$$

$\square$

The condition of disjointness in Theorem 1.12 can be omitted.

THEOREM 1.13. *The interval $[0, 1]$ of real numbers is uncountable.*

PROOF. Suppose the set $[0, 1]$ is countable; that is, the numbers in $[0, 1]$ can be listed as an infinite sequence $\{\alpha_i\}_{i=1}^{\infty}$. Write all real numbers α_i in infinite decimal forms, say in base 10, as follows:

$$
\begin{aligned}
\alpha_1 &= 0.a_1a_2a_3a_4a_5\cdots, \\
\alpha_2 &= 0.b_1b_2b_3b_4b_5\cdots, \\
\alpha_3 &= 0.c_1c_2c_3c_4c_5\cdots, \\
&\cdots
\end{aligned}
$$

Then we can construct a number $x = 0.x_1x_2x_3\cdots$, defined by

$$
\begin{aligned}
x_1 &= \begin{cases} 1 & \text{if } a_1 = 2 \\ 2 & \text{if } a_1 \neq 2, \end{cases} \\
x_2 &= \begin{cases} 1 & \text{if } b_2 = 2 \\ 2 & \text{if } b_2 \neq 2, \end{cases} \\
x_3 &= \begin{cases} 1 & \text{if } c_3 = 2 \\ 2 & \text{if } c_3 \neq 2, \end{cases} \\
&\cdots
\end{aligned}
$$

The number x is an infinite decimal of $1s$ and $2s$ and is a real number between 0 and 1. Since $x_1 \neq a_1$, $x_2 \neq b_2$, $x_3 \neq c_3$, and so on, it follows that $x \neq \alpha_1$, $x \neq \alpha_2$, $x \neq \alpha_3$, etc. Thus x is not in the list $\alpha_1, \alpha_2, \ldots$; that is, x is not a real number between 0 and 1, a contradiction. $\square$

THEOREM 1.14 (**Bernstein**). *For sets A and B, if there are subsets $A_1 \subset A$ and $B_1 \subset B$ such that $A_1 \sim B$ and $A \sim B_1$, then $A \sim B$.*

Exercise 13. For real numbers $a, b \in \mathbb{R}$ such that $a < b$, find a one-to-one correspondence between the open interval (a, b) and the closed interval $[a, b]$, where $(a, b) = \{x \in \mathbb{R} \mid a < x < b\}$ and $[a, b] = \{x \in \mathbb{R} \mid a \leq x \leq b\}$.

Exercise 14. Show that the union of countably many countable sets is still countable. That is, if A_i are countable sets, $i = 1, 2, \ldots$, not necessarily disjoint, then $\bigcup_{i=1}^{\infty} A_i$ is countable.

Exercise 15. Let A be a countable set. Show that A^* is countable.

Exercise 16. Let A be a set with at least two elements. Show that A^* is in one-to-one correspondent with a subset of $A^{(\infty)}$.

Exercise 17. Let $B = \{0, 1\}$ and let $B^\infty = \{a_1 a_2 \cdots \mid a_1, a_2, \ldots \in B\}$ be the set of words with infinite length. Show that B^∞ is uncountable.

1.6. Permutations

A bijective function from a finite set A to itself is called a **permutation of** A. If $A = \{a_1, a_2, \ldots, a_n\}$ is a set of n objects and $f : A \to A$ is a permutation, then

$$f(a_1), f(a_2), \ldots, f(a_n)$$

is the same collection of $a_1, a_2, \ldots, a_n$, and may be in different order. Let

$$f(a_1) = a_{i_1},\ f(a_2) = a_{i_2},\ \ldots,\ f(a_n) = a_{i_n}.$$

The permutation f is usually expressed by the array

$$f = \begin{pmatrix} a_1 & a_2 & \ldots & a_n \\ a_{i_1} & a_{i_2} & \ldots & a_{i_n} \end{pmatrix}$$

and sometimes by the word

$$a_{i_1} a_{i_2} \cdots a_{i_n}.$$

Let a be an element of A. The sequence

$$a, f(a), f^2(a), f^3(a), \ldots$$

must eventually return to a and then repeat the pattern. Let k be the smallest positive integer such that $f^k(a) = a$. The sequence $a, f(a), f^2(a), \ldots, f^{k-1}(a)$ is called a **cycle of length** k of the permutation f, and is denoted by

$$\big(a\, f(a)\, f^2(a)\ \cdots\ f^{k-1}(a)\big)\,.$$

Of course,

$$\big(f(a)\, f^2(a)\ \cdots\ f^k(a)\big)\,,\ \big(f^2(a)\, f^3(a)\ \cdots\ f^{k+1}(a)\big)\,,\ \ldots$$

are also cycles of f. However, they are viewed as the same cycle. For instance, for the set $A = \{1, 2, 3, 4, 5, 6, 7, 8\}$, the permutation

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 6 & 7 & 5 & 4 & 3 & 8 & 2 & 1 \end{pmatrix}$$

has four cycles (168), (27), (53), (4). The permutation can be written as (681)(27)(53)(4). However, the way of writing σ in this fashion is not unique. We may make this kind of writing unique by requiring that the leading element of each cycle to be the largest element inside the cycle and requiring that all leading elements of cycles to be increasing. Hence the permutation σ can be uniquely written in the cycle

$$\sigma = (4)(53)(72)(816).$$

THEOREM 1.15. *Every permutation of* $\{1, 2, \ldots, n\}$ *can be written as disjoint cycles in a unique way that the leading element of each cycle is the largest element in the cycle and all leading elements of cycles are in increasing order.*

Let n be fixed. For a permutation σ of $\{1, 2, \ldots, n\}$, let σ can be written as disjoint cycles. Note that each cycle can be viewed as a permutation of $\{1, 2, \ldots, n\}$. For instance, when $n = 8$ and $\sigma = (4)(53)(72)(816)$, the cycle (53) can be viewed as the permutation $(1)(2)(4)(53)(6)(7)(8)$, the cycle (816) can be viewed as the permutation $(2)(3)(4)(5)(816)(7)$, and the cycle (4) can be viewed as the identity permutation. We then have

$$(4)(53)(72)(816) = (53) \circ (72) \circ (816).$$

A permutation σ of $\{1, 2, \ldots, n\}$ is called a **transposition** if it has one cycle of length 2 and all other cycles are of length 1. For instance, if $\sigma = (a_i a_j)$ is a transposition, then $a_i \neq a_j$, $\sigma(a_i) = a_j$, $\sigma(a_j) = a_i$, and $\sigma(a) = a$ for all $a \neq a_i$, $a \neq a_j$. Note that each cycle can be written as composition of transpositions. If $(a_1 a_2 \ldots a_k)$ is a cycle, then

$$(a_1 a_2 \cdots a_k) = (a_1 a_k) \circ (a_1 a_{k-1}) \circ \cdots \circ (a_1 a_3) \circ (a_1 a_2)$$

For instance, the cycle (8164) can be written as

$$(8164) = (84) \circ (86) \circ (81).$$

COROLLARY 1.16. *Every permutation of* $\{1, 2, \ldots, n\}$ *can be written as composition of finite number of transpositions.*

A permutation is called **even** if it can be written as composition of even number of transpositions, and it is called **odd** if it can be written as composition of odd number of transpositions.

PROPOSITION 1.17.
(1) *The product of two even permutations is even.*
(2) *The product of two odd permutations is even.*
(3) *The product of an even permutation and an odd permutation is odd.*
(4) *There are* $\frac{n!}{2}$ *even permutations and* $\frac{n!}{2}$ *odd permutations of* $\{1, 2, \ldots, n\}$.

Let $\sigma = a_1 a_2 \cdots a_n$ be a permutation of $\{1, 2, \ldots, n\}$. A pair (a_i, a_j) with $i < j$ is called an **inversion** of σ if $a_i > a_j$. The number of inversions of σ is denoted by $\mathrm{inv}(\sigma)$.

EXERCISE 18. Consider the permutation 387169425 of the set $\{1, 2, \ldots, 9\}$, that is,

$$\left(\begin{array}{ccccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 3 & 8 & 7 & 1 & 6 & 9 & 4 & 2 & 5 \end{array} \right).$$

(1) Find the permutation σ^{-1}.
(2) Write the permutation in disjoint cycles.
(3) Write the permutation as the composition of cycles.
(4) Write it as the composition of transpositions.
(5) Determine its parity.
(6) Find the total number of inversions of the permutation.

EXERCISE 19. Consider the permutation

$$\sigma = \left(\begin{array}{ccccc} 1 & 2 & \ldots & n-1 & n \\ n & n-1 & \ldots & 2 & 1 \end{array} \right).$$

(1) Write σ as a product of minimal number of transpositions;
(2) Determine the parity of σ;
(3) Find $\operatorname{inv}(\sigma)$.

Exercise 20. Let $a_1a_2\cdots a_n$ be a permutation of $\{1,2,\ldots,n\}$.

(1) Find the parity of the permutation $a_na_{n-1}\cdots a_2a_1$ in terms of the parity of $a_1a_2\cdots a_n$;
(2) Find $\operatorname{inv}(a_na_{n-1}\cdots a_2a_1)$ in terms of $\operatorname{inv}(a_1a_2\cdots a_n)$.

CHAPTER 2

Number Theory

2.1. Divisibility

Leopold Kronecker said: "God created integers, all else are the work of man." We assume that the set of integers are well defined and we are familiar with the properties of integers such as addition, subtraction, multiplication, and division. In particular, we assume the following axiom for subsets of integers bounded below.

Axiom. For every nonempty subset of integers, if it is bounded below, then it has a unique minimum integer.

It follows easily from the axiom that for every subset of integers, if it is bounded above, then it has a unique maximum integer. Given two integers a and b with $a \neq 0$, we say that a **divides** b, written $a|b$, if there exists an integer q such that

$$b = qa.$$

When this is true we say that a is a **factor** (or **divisor**) of b, and say that b is a **multiple** of a. Obviously, any integer n has divisors, ± 1 and $\pm n$, called the **trivial divisors** of n. The divisors of n other than the trivial divisors are called **nontrivial divisors**. Note that every integer is a divisor of 0. A positive integer p ($\neq 1$) is called a **prime** if its positive divisors are only the trivial divisors 1 and p. A positive integer is called **composite** if it is not a prime. The first few primes are listed as follows:

$$2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, 53, 57.$$

PROPOSITION 2.1. *Let a, b, c be nonzero integers.*

(a) *If $a|b$ and $b|a$, then $a = \pm b$.*
(b) *If $a|b$ and $b|c$, then $a|c$.*
(c) *If $a|b$ and $a|c$, then $a|(bx + cy)$ for $x, y \in \mathbb{Z}$.*

PROOF. (a) Let $b = q_1 a$ and $a = q_2 b$ for some integers q_1 and q_2. Then

$$b = q_1 q_2 b.$$

Dividing both sides by b, we have $q_1 q_2 = 1$. It follows that $q_1 = q_2 = \pm 1$. Thus $b = \pm a$.

(b) Let $b = q_1 a$ and $c = q_2 b$ for some integers q_1 and q_2. Then $c = q_1 q_2 a$, that is, $a|c$.

(c) Let $b = q_1 a$ and $c = q_2 a$ for $q_1, q_2 \in \mathbb{Z}$. Then for $x, y \in \mathbb{Z}$,

$$bx + cy = (q_1 x + q_2 y)a,$$

that is, $a|(bx + cy)$. □

THEOREM 2.2. *There are infinitely many prime numbers.*

PROOF. Suppose there are finitely many primes, say, $p_1, p_2, \ldots, p_k$. Then the integer

$$a = p_1 p_2 \cdots p_k + 1$$

is not divisible by any of the primes $p_1, p_2, \ldots, p_k$ because the remainders of a dividing by $p_1, p_2, \ldots, p_k$ respectively are always 1. This means that a has no prime factors. By definition of primes, the integer a must be a prime, and this prime is larger than all primes $p_1, p_2, \ldots, p_k$, a contradiction. $\square$

THEOREM 2.3 (**Division Algorithm**). *For any integers a and b, where $a > 0$, there are unique integers q and r such that*

$$b = qa + r, \qquad 0 \leq r < a.$$

PROOF. Consider the set $S = \{b - ta \geq 0 \mid t \in \mathbb{Z}\}$. Obviously, S is nonempty and is bounded below. Then S has the unique minimum element r, that is, there is an unique integer q such that $b - qa = r$. We claim that $r < a$. Suppose $r \geq a$, then $b - (q+1)a = r - a \geq 0$ shows that $r - a$ is an element of S. This is contrary to that r is the minimum element of S. $\square$

2.2. Greatest Common Divisor

For integers a and b, not simultaneously 0, a **common divisor** of a and b is an integer c such that $c|a$ and $c|b$. Clearly, there are finitely many common divisors for a and b; the very greatest one is called the **greatest common divisor** of a and b, and is denoted by $\gcd(a, b)$. For convenience, we assume $\gcd(0, 0) = 0$. Two integers a and b are called **coprime** (or **relatively prime**) if $\gcd(a, b) = 1$.

THEOREM 2.4. *Let d be the greatest common divisor of integers a and b, that is, $d = \gcd(a, b)$. Then there exist integers x and y such that*

$$d = ax + by.$$

PROOF. It is obviously true when $a = b = 0$. Assume that a and b are not simultaneously zero. We consider the set $S = \{au + bv \mid u, v \in \mathbb{Z}\}$ and the set $S_+ = S \cap \mathbb{Z}_+$. Note that S_+ is bounded below and $S_+ \neq \emptyset$ because $a^2 + b^2 > 0$. Let s be the smallest integer in S_+ and write

$$s = au_0 + bv_0$$

for some $u_0, v_0 \in \mathbb{Z}$. We claim that $d = s$.

Clearly, d divides every integer in S because $d|a$ and $d|b$. In particular, $d|s$. We then have $d \leq s$. To show that $s \leq d$, we claim that d divides every integer in S. In fact, for any $au + bv \in S$ with $u, v \in \mathbb{Z}$, let

$$au + bv = qs + r, \qquad 0 \leq r < s.$$

Then $r = a(u - qv_0) + b(v - qu_0) \in S$; so $d|r$. If r was positive, then $r \in S_+$ and s could not be the smallest integer in S_+. Thus $r = 0$; so $s|(au + bv)$. In particular, taking $(u, v) = (0, 1)$ and $(u, v) = (1, 0)$, we see that s is a common divisor of a and b. Hence by definition of gcd, $s \leq d$. $\square$

THEOREM 2.5. *For integers a, b, q, and r, if*

$$b = qa + r,$$

then

$$\gcd(a, b) = \gcd(a, r).$$

PROOF. Let $d_1 = \gcd(a,b)$ and let $d_2 = \gcd(a,r)$. Obviously, $d_1|a$; $d_1|r$ because $r = b - qa$ and $d_1|b$. This means that d_1 is a common divisor of a and r. Thus $d_1 \leq d_2$. On the other hand, $d_2|a$; $d_2|b$ because $b = qa + r$ and $d_2|r$. This means that d_2 is a common divisor of a and b. Hence, $d_2 \leq d_1$. Therefore $d_1 = d_2$. $\square$

The above proposition gives rise a simple constructive method to calculate gcd by repeating the Division Algorithm. For example, $\gcd(297, 3627)$ can be calculated as follows:

$$\begin{aligned}
3627 &= 12 \cdot 297 + 63, \\
297 &= 4 \cdot 63 + 45, \\
63 &= 1 \cdot 45 + 18, \\
45 &= 2 \cdot 18 + 9, \\
18 &= 2 \cdot 9;
\end{aligned}$$

$$\begin{aligned}
\gcd(297, 3627) &= \gcd(63, 297) \\
&= \gcd(45, 63) \\
&= \gcd(18, 45) \\
&= \gcd(9, 18) \\
&= 9.
\end{aligned}$$

The procedure to calculate $\gcd(297, 3627)$ applies to any pair of nonnegative integers. Let a be a positive integer and b a nonnegative integer. Repeating the Division Algorithm will produce finite sequences of nonnegative integers q_i and r_i such that

$$\begin{array}{lcll}
b & = & q_0 a + r_0, & 0 \leq r_0 < a, \\
a & = & q_1 r_0 + r_1, & 0 \leq r_1 < r_0, \\
r_0 & = & q_2 r_1 + r_2, & 0 \leq r_2 < r_1, \\
r_1 & = & q_3 r_2 + r_3, & 0 \leq r_3 < r_2, \\
 & \vdots & & \\
r_{k-2} & = & q_k r_{k-1} + r_k, & 0 \leq r_k < r_{k-1}, \\
r_{k-1} & = & q_{k+1} r_k + r_{k+1}, & r_{k+1} = 0.
\end{array}$$

Notice that the sequence $\{r_i\}$ is strictly decreasing; it ends eventually to 0 at some step, say, the remainder r_{k+1} becomes zero in the very first time, i.e., $r_{k+1} = 0$ and $r_i \neq 0$ for all $0 \leq i \leq k$. Reverse the sequence $\{r_i\}_{i=0}^{k}$ and make substitutions as follows:

$$\begin{array}{lcl}
\gcd(a,b) & = & r_k, \\
r_k & = & r_{k-2} - q_k r_{k-1}, \\
r_{k-1} & = & r_{k-3} - q_{k-1} r_{k-2}, \\
 & \vdots & \\
r_1 & = & a - q_1 r_0, \\
r_0 & = & b - q_0 a.
\end{array}$$

We see that $\gcd(a,b)$ can be expressed as an integral linear combination of a and b. This procedure is known as the **Euclidean Algorithm**.

EXAMPLE 2.1. The greatest common divisor of 297 and 3627, written as an integral linear combination of 297 and 3627, can be obtained as follows:

$$\begin{aligned} \gcd(297, 3627) &= 45 - 2 \cdot 18 \\ &= 45 - 2(63 - 45) \\ &= 3 \cdot 45 - 2 \cdot 63 \\ &= 3(297 - 4 \cdot 63) - 2 \cdot 63 \\ &= 3 \cdot 297 - 14 \cdot 63 \\ &= 3 \cdot 297 - 14(3627 - 12 \cdot 297) \\ &= 171 \cdot 297 - 14 \cdot 3627. \end{aligned}$$

PROPOSITION 2.6. *For integers a, b, c, if $a|bc$ and $\gcd(a, b) = 1$, then $a|c$.*

PROOF. By the Euclidean Algorithm, there are integers x and y such that $ax + by = 1$. Then

$$c = 1 \cdot c = (ax + by)c = acx + bcy.$$

It is clear that $a|c$ because $a|ac$ and $a|bc$. $\square$

THEOREM 2.7 (**Unique Factorization**). *Every integer $a \geq 2$ can be uniquely factorized into the form*

$$a = p_1^{e_1} p_2^{e_2} \cdots p_m^{e_m},$$

where $p_1, p_2, \ldots, p_m$ are distinct primes, $e_1, e_2, \ldots, e_m$ are positive integers, and $p_1 < p_2 < \cdots < p_s$.

PROOF. If a has only the trivial divisors, then a itself is a prime, and it obviously has unique factorization. If a has some nontrivial divisors, then

$$a = bc$$

for some positive integers b and c other than 1 and a. Obviously, $b < a$ and $c < a$. By induction, the positive integers b and c have factorizations into primes. Consequently, a has a factorization into primes.

Let $a = q_1^{f_1} q_2^{f_2} \cdots a_n^{f_n}$ be another factorization, where $q_1, q_2, \ldots, q_n$ are distinct primes, $f_1, f_2, \ldots, f_n$ are positive integers, and $q_1 < q_2 < \cdots < q_n$. We claim that $m = n$, $p_i = q_i$, $e_i = f_i$ for all $1 \leq i \leq m$.

Suppose $p_1 < q_1$. Then p_1 is distinct from the primes $q_1, q_2, \ldots, q_n$. It is clear that $\gcd(p_1, q_i) = 1$ and so $\gcd(p_1, q_i^{f_i}) = 1$ for all $1 \leq i \leq n$. Note that $p_1 | q_1^{f_1} q_2^{f_2} \cdots a_n^{f_n}$. Since $\gcd(p_1, q_1^{f_1}) = 1$, by Proposition 2.6, we have $p_1 | q_2^{f_2} \cdots a_n^{f_n}$. Since $\gcd(p_1, q_2^{f_2}) = 1$, again by Proposition 2.6, we have $p_1 | q_3^{f_2} \cdots a_n^{f_n}$. Repeating the argument, we finally obtain that $p_1 | q_n^{f_n}$, which is contrary to $\gcd(p_1, q_n^{f_n}) = 1$. We thus conclude that $p_1 \geq q_1$. Similarly, $p_1 \leq q_1$. Hence $p_1 = q_1$. Next we claim that $e_1 = f_1$.

Suppose $e_1 < f_1$. Then

$$p_2^{e_2} \cdots p_m^{e_m} = p_1^{f_1 - e_1} q_2^{f_2} \cdots q_n^{f_n}.$$

This implies that $p_1 | p_2^{e_2} \cdots p_m^{e_m}$. If $m = 1$, it would imply that p_1 divides 1, which is impossible because p_1 is a prime. If $m \geq 2$, note that $\gcd(p_1, p_i) = 1$ and so $\gcd(p_1, p_i^{e_i}) = 1$ for all $2 \leq i \leq m$; by the same token of applying Proposition 2.6 repeatedly, we have $p_1 | p_m^{e_m}$, which is contrary to $\gcd(p_1, p_m^{e_m}) = 1$. This means that we must have $e_1 \geq f_1$. Similarly, $e_1 \leq f_1$. Hence $e_1 = f_1$.

Now we have obtained $p_2^{e_2} \cdots p_m^{e_m} = q_2^{f_2} \cdots q_n^{f_n}$. If $m < n$, then by induction we have $p_1 = q_1, \ldots, p_m = q_m$ and $e_1 = f_1, \ldots, e_m = f_m$. Thus $1 = q_{m+1}^{f_{m+1}} \cdots q_n^{f_n}$; this is impossible because $q_{m+1}, \ldots, q_n$ are primes. So $m \geq n$. Similarly, $m \leq n$. Hence we have $m = n$. By induction on $m = n$, we obtain that $e_2 = f_2, \ldots, e_m = f_m$. $\square$

PROPOSITION 2.8. *A positive integer d is the* gcd *of integers a and b if and only if*

(a) *$d|a$ and $d|b$,*
(b) *if $c|a$ and $c|b$, then $c|d$.*

PROOF. The conditions are obviously sufficient. For necessity, it is clear that (a) is necessary. If $d = \gcd(a, b)$, then there exist integers x and y such that $ax + by = d$. Thus for any common divisor c of a and b, c obviously divides the linear combination $ax + by$; consequently, $c|d$. $\square$

2.3. Least Common Multiple

For two integers a and b, a positive integer m is called a **common multiple** of a and b if $a|m$ and $b|m$. The smallest integer among the common multiples of a and b is called the **least common multiple** of a and b, and is denoted by $\operatorname{lcm}(a, b)$.

PROPOSITION 2.9. *For any nonnegative integers a and b,*

$$ab = \gcd(a, b) \operatorname{lcm}(a, b).$$

PROOF. Let $a = p_1^{e_1} p_2^{e_2} \cdots p_n^{e_n}$ and $b = p_1^{f_1} p_2^{f_2} \cdots p_n^{f_n}$, where $p_1 < p_2 < \cdots < p_n$, e_i and f_i are nonnegative integers, $1 \leq i \leq n$. Then by the Unique Factorization Theorem,

$$\begin{aligned} \gcd(a, b) &= p_1^{g_1} p_2^{g_2} \cdots p_n^{g_n}, \\ \operatorname{lcm}(a, b) &= p_1^{h_1} p_2^{h_2} \cdots p_n^{h_n}, \end{aligned}$$

where $g_i = \min(e_i, f_i)$ and $h_i = \max(e_i, f_i)$ for all $1 \leq i \leq n$. It is clear that

$$ab = p_1^{g_1+h_1} p_2^{g_2+h_2} \cdots p_n^{g_n+h_n},$$

and $g_i + h_i = e_i + f_i$ for all $1 \leq i \leq n$. $\square$

PROPOSITION 2.10. *An integer m is the least common multiple of integers a and b if and only if*

(1) *$a|m$, $b|m$;*
(2) *if $a|n$ and $b|n$, then $m|n$.*

PROOF. By the Unique Factorization Theorem. $\square$

EXAMPLE 2.2. Find all integer solutions for the linear equation $25x + 65y = 10$.

Solution: By the Division Algorithm,

$$\begin{aligned} 65 &= 2 \cdot 25 + 15, \\ 25 &= 15 + 10, \\ 15 &= 10 + 5. \end{aligned}$$

Then by the Euclidean Algorithm,

$$\begin{aligned} \gcd(25, 65) &= 15 - 10 \\ &= 15 - (25 - 15) \\ &= -25 + 2 \cdot 15 \\ &= -25 + 2 \cdot (65 - 2 \cdot 25) \\ &= -5 \cdot 25 + 2 \cdot 65. \end{aligned}$$

Thus the integer solutions are given by

$$\begin{cases} x &= 2(-5)+13k &= -10+13k \\ y &= 2\cdot 2-5k &= 4-5k, \end{cases} \quad k\in\mathbb{Z}.$$

THEOREM 2.11. *Let a and b be integers, not simultaneously zero, and $d = \gcd(a,b)$. Then the linear Diophantine equation*

$$ax+by=c$$

has an integer solution if and only if $d|c$. Moreover, if $(x,y)=(u,v)$ is a special solution, then all solutions are given by

$$\begin{cases} x &= u+\frac{bk}{\gcd(a,b)} \\ y &= v-\frac{ak}{\gcd(a,b)}, \end{cases} \quad k\in\mathbb{Z}.$$

PROOF. It is clear that $(x,y)=(u,v)+\frac{1}{d}(kb,-ka)$ are integer solutions. We only need to show that all integer solutions of $ax+by=0$ are of the form

$$(x,y)=\frac{k}{d}(b,a).$$

Since $ax=-by$, we have $a|by$ and $b|ax$. This means that ax and by are both common multiples of a and b. Thus $\operatorname{lcm}(a,b)$ is a common divisor of ax and by, that is, $ax=k\operatorname{lcm}(a,b)$ and $by=-k\operatorname{lcm}(a,b)$ for some $k\in\mathbb{Z}$. Thus

$$\begin{cases} x &= \frac{k\operatorname{lcm}(a,b)}{a} &= \frac{bk}{\gcd(a,b)} \\ y &= -\frac{k\operatorname{lcm}(a,b)}{b} &= -\frac{ak}{\gcd(a,b)}. \end{cases}$$

□

2.4. Modulo Integers

Given a positive integer n. We consider the equivalence relation $\sim_n$ defined by

$$x\sim_n y \quad \text{if and only if} \quad y-x\equiv 0 \pmod{n}.$$

It is clear that there are n equivalence classes; the quotient set $\mathbb{Z}/\sim_n$ is denoted by $\mathbb{Z}_n$, that is,

$$\mathbb{Z}_n=\Big\{[0],[1],[2],\ldots,[n-1]\Big\},$$

where $[a]=\{kn+a\,|\,k\in\mathbb{Z}\}$ for any integer a. A standard notation for $\mathbb{Z}_n$ is $\mathbb{Z}/n\mathbb{Z}$.

We can define an addition and a multiplication on $\mathbb{Z}_n$ in a natural way:

$$\begin{aligned} [a]+[b] &= [a+b], \\ [a]\,[b] &= [ab]. \end{aligned}$$

The element $[0]$ is the **zero** in $\mathbb{Z}_n$ and $[1]$ is the **identity** in $\mathbb{Z}_n$. Sometimes, we just write $[a]$ as a. However, for $[a]=[b]$, we write it as

$$a\equiv b \pmod{n}.$$

The addition and multiplication are well defined for integers modulo n. In fact, let a,a',b,b' be integers such that $[a]=[a']$ and $[b]=[b']$, that is,

$$a\equiv a' \pmod{n} \quad \text{and} \quad b\equiv b' \pmod{n}.$$

Then $a'-a=kn$ and $b'-b=ln$ for some $k,l\in\mathbb{Z}$. Thus

$$(a'+b')-(a+b)=(a'-a)+(b'-b)=(k+l)n,$$

$$a'b' - ab = a'b' - ab' + ab' - ab = (a' - a)b' + a(b' - b) = (kb' + al)n.$$

This means that $[a + b] = [a' + b']$ and $[ab] = [a'b']$, that is,

$$a + b \equiv a' + b' \pmod{n} \qquad \text{and} \qquad ab \equiv a'b' \pmod{n}.$$

PROPOSITION 2.12. *The addition and multiplication in $\mathbb{Z}_n$ satisfy the following properties:*

(1) $\big([a] + [b]\big) + [c] = [a] + \big([b] + [c]\big)$,
(2) $[a] + [b] = [b] + [a]$,
(3) $[a] + [0] = [a]$,
(4) $[a]([b] + [c]) = [a][b] + [a][c]$,
(5) $\big([a][b]\big)\, [c] = [a]\, \big([b][c]\big)$,
(6) $[a][b] = [b][a]$,
(7) $[a][1] = [a]$,
(8) $[a][0] = [0]$.

For any $[a] \in \mathbb{Z}_n$, there is unique element $[b]$ such that $[a] + [b] = [0]$; we write $[b] = -[a]$, called the **negative element** of $[a]$. An element $[a] \in \mathbb{Z}_n$ is called **invertible** if there is an element $[b]$ in $\mathbb{Z}_n$ such that

$$[a][b] = [1];$$

the element $[b]$ is called the **inverse** of $[a]$, and is denoted by $[a]^{-1}$. If $[a]$ is invertible, its inverse is unique. In fact, if $[c]$ is also an inverse of $[a]$, that is, $[a][c] = [1]$, then

$$[b] = [1][b] = ([a][c])[b] = ([c][a])[b] = [c]([a][b]) = [c][1] = [c].$$

We denote by $\mathbb{Z}_n^*$ the set of all invertible elements of $\mathbb{Z}_n$.

THEOREM 2.13. *An element $[a]$ in $\mathbb{Z}_n$ is invertible if and only if a is coprime with n. Thus*

$$\mathbb{Z}_n^* = \left\{ [a] \;\middle|\; \gcd(a, n) = 1 \right\}.$$

PROOF. By definition of invertibility, $[a]$ is invertible $\iff$ there is an element $[b]$ such that $[a][b] = [1] \iff ab \equiv 1 \pmod{n} \iff ab = kn + 1$ for some $k \in \mathbb{Z} \iff ab - nk = 1$ for some $k \in \mathbb{Z} \iff \gcd(a, n) = 1$. $\square$

COROLLARY 2.14. *If p is a prime, then every nonzero element in $\mathbb{Z}_p$ is invertible.*

For any positive integer n, let $\phi(n)$ denote the number of integers a in $[1, n]$ such that $\gcd(a, n) = 1$. The function $\phi : \mathbb{P} \to \mathbb{P}$ is known as the **Euler function**. If p is a prime, then $\phi(p) = p - 1$.

THEOREM 2.15 (**Euler's Theorem**). *If $[a]$ is invertible in $\mathbb{Z}_n$, that is,* $\gcd(a, n) = 1$, *then*

$$a^{\phi(n)} \equiv 1 \pmod{n}.$$

PROOF. Let $m = \phi(n)$. Then $\mathbb{Z}_n^*$ has m elements. Let

$$\mathbb{Z}_n^* = \{[a_1], [a_2], \ldots, [a_m]\} \quad \text{and} \quad [u] = [a_1][a_2] \cdots [a_m].$$

Since $[a]$ is invertible, there is a one-to-one correspondence $f_{[a]} : \mathbb{Z}_n^* \to \mathbb{Z}_n^*$ defined by

$$f_{[a]}([x]) = [a][x] = [ax].$$

The inverse of $f_{[a]}$ is the map $f_{[a]^{-1}} = f_{[b]}$, where $[a][b] = [1]$. Then

$$[u] = [a_1][a_2]\cdots[a_m] = [aa_1][aa_2]\cdots[aa_m] = [a^m][a_1][a_2]\cdots[a_m]$$

Thus $[a^m] = [1]$. This means that $a^{\phi(n)} \equiv 1 \pmod{n}$. $\square$

COROLLARY 2.16 (**Fermat's Little Theorem**). *Let p be a prime. If $p \nmid a$, then*

$$a^{p-1} \equiv 1 \pmod{p}.$$

PROOF. Apply the Euler Theorem and $\phi(p) = p - 1$. $\square$

2.5. RSA Cryptography System

Common sense tells us that if we know how a secret message was encoded then we can easily decode it. For example, suppose we are given the message *EJTDSFUF NBUIFNBUJDT*, and are told that it was encoded by replacing each letter in the original message by the one that immediately follows it in the alphabet. To decode the message, all we have to do is to replace each letter in the message by the one that proceeds it, yielding *DISCRETE MATHEMATICS*. Since the equivalence of coding and decoding, the encryption process must be kept secret for security reason.

In modern communication networks, the network security leads to two situations: The first was that only highly trusted individuals were allowed to access to the encoding key. This means that the code could not be used by many people. The second was just opposite, that many people were given the secret. So the decoding process was known in principle. The problem with this is that the process is hardly unknown in practice due to computational complexity.

Imagine a network where a number of individuals (corporations, banks, governments) send messages to each other over public wires, so that eavesdropping is possible. The problem is to ensure the privacy of such communications, as to guarantee against fake messages (forgeries).

The usual solution to the problem is to send the messages in a cryptography know only to the two parties involved (the person-in-the-street calls it a "code", but this word already has at least two other meanings for computer scientists, and it therefore not used here). The difficulty with this solution is that it requires pre-arrangement, and therefore communication is limited. Also, it requires $\binom{n}{2}$ different cryptography, if n individuals are involved.

THEOREM 2.17 (**RSA**). *Let $n = pq$ be a positive number, where p and q are primes. Let e and d be positive integers such that $\gcd(e, \phi(n)) = 1$ and $ed \equiv 1 \pmod{\phi(n)}$. Let $E : \mathbb{Z}_n \to \mathbb{Z}_n$ be defined by*

$$E(x) = x^e \pmod{n},$$

and let $D : \mathbb{Z}_n \to \mathbb{Z}_n$ be defined by

$$D(y) = y^d \pmod{n}.$$

Then E and D are inverses of each other.

PROOF. For any $x \in \mathbb{Z}_n = \{0, 1, 2, \ldots, n-1\}$,

$$(D \circ E)(x) = (E \circ D)(x) = x^{ed} \pmod{n}.$$

Our aim is to show that $x^{ed} \equiv x \pmod{n}$. We divide it into three cases.

Case 1: $x = 0$. It is trivial that $x^{ed} \equiv x \pmod{n}$.

Case 2: $\gcd(x, n) = 1$. Since $ed \equiv 1 \pmod{\phi(n)}$, there exists an integer k such that $ed = k\phi(n) + 1$. Then

$$x^{ed} = x^{k\phi(n)+1} = (x^{\phi(n)})^k x.$$

Since $\gcd(x, n) = 1$, by the Euler Theorem, $x^{\phi(n)} \equiv 1 \pmod{n}$. Obviously,

$$x^{ed} \equiv x \pmod{n}.$$

Case 3: $\gcd(x, n) \neq 1$. Since $n = pq$ and p, q are primes, we have either $x = kp$ for some $1 \leq k < q$ or $x = lq$ for some $1 \leq l < p$. In the former case, we have

$$x^{ed} = (kp)^{k\phi(n)+1} = (kp)^{k(p-1)(q-1)+1} = ((kp)^{q-1})^{k(p-1)}(kp)$$

because $\phi(n) = \phi(p)\phi(q) = (p-1)(q-1)$. Since $q \nmid kp$, by the Fermat Little Theorem, $(kp)^{q-1} \equiv 1 \pmod{q}$. Thus

$$x^{ed} \equiv kp \equiv x \pmod{q}.$$

Note that $x^{ed} \equiv (kp)^{ed} \equiv 0 \equiv x \pmod{p}$. Since $\gcd(p, q) = 1$, it follows that $x^{ed} \equiv x \pmod{n}$. $\square$

A **RSA public key cryptography system** is a tuple (S, n, e, d, E, D), where $S = \{1, 2, \ldots, n-1\}$, $n = pq$, p and q are distinct primes, e and d are positive integers such that $\gcd(e, \phi(n)) = 1$ and $ed \equiv 1 \pmod{\phi(n)}$, E and D are bijective functions $S \to S$ defined by $E(x) = x^e \pmod{n}$ and by $D(x) = x^d \pmod{n}$, respectively. The integer e is called the **encryption number** and d the **decryption number**; the function E is called the **encryption map** and D the **decryption map**.

In a RSA public key cryptography system (S, n, e, d, E, D), the numbers n and e are public, while the numbers p, q, d are secret. Mathematically speaking, knowing the numbers n and e in (S, n, e, d, E, D), the numbers p, q, d are known in principle. However, if the primes p and q are selected to large enough, say having 100 digits, then it is impossible in practice to find out the numbers p, q, d. The difficulty to find p, q, d is based on the difficulty to factor integers.

Example 2.3. Let $p = 7, q = 11$. Then $n = pq = 77$, $\phi(n) = \phi(p)\phi(q) = 6 \cdot 10 = 60$. If $e = 7$, then $d = 43$. If $e = 11$, then $d = 11$. If $e = 13$, then $d = 37$. If $e = 17$, then $d = 53$.

Example 2.4. Let $p = 11$, $q = 13$. Then $n = pq = 143$, $\phi(n) = (p-1)(q-1) = 120$. Then there are RSA systems $e = 7, d = 43$; $e = 11, d = 11$; and $e = 13, d = 37$. For the RSA system $e = 13$, $d = 37$, we have

$$E(2) = 2^{13} \equiv 41 \pmod{143}$$

($2^2 = 4$, $2^4 = 16$, $2^8 = 16^2 \equiv 113$, $2^{13} = 2^8 \cdot 2^4 \cdot 2 \equiv 113 \cdot 16 \cdot 2 \equiv 41$); and

$$D(41) = 41^{37} \equiv 2 \pmod{143}$$

($41^2 \equiv 108$, $41^4 \equiv 108^2 \equiv 81$, $41^8 \equiv 81^2 \equiv -17$, $41^{16} \equiv 17^2 \equiv 3$, $41^{32} \equiv 9$, $41^{37} = 41^{32} \cdot 41^4 \cdot 41 \equiv 2$).

CHAPTER 3

Propositional Logic

3.1. Statements

By a **mathematical statement** (or just **statement**) we mean a declarative sentence that is either true or false, but not both. The **truth value** (true or false) for any statement can be determined and is not ambiguous in any sense. For example, the following sentences are statements.

(1) Today is 1st of July 1997.
(2) The course number of Discrete Structure in HKUST is Math132.
(3) The equation $x^2 + y^2 = z^2$ has no positive integer solutions.
(4) There are 7,523,804 people in Hong Kong.

However, many sentences in daily life languages are not mathematical statements. For instance, the following sentences are not mathematical statements.

(1) How are you?
(2) Hong Kong is a big city.
(3) What a beautiful campus!
(4) This sentence is false.

For the last sentence above, if we say that the sentence is true, then it is false. If, on the other hand, we claim that the sentence is false, then it is true. Such sentences will not be considered as mathematical statements. Statements are usually denoted by lowercase letters such as p, q, r, ..., etc.

3.2. Connectives

Given several statements, we wish to set up rules by which we can decide the truth of various combinations of the given statements. New statements can be formed by using connectives "not", "and", and "or".

The **Negation** of a statement p is the statement "not p", denoted $\neg p$. The truth values of $\neg p$ are given by the table

p	$\neg p$
T	F
F	T

The **conjunction** of two statements p and q is the statement "p and q", denoted $p \wedge q$. Its truth values are given by the table

p	q	$p \wedge q$
T	T	T
T	F	F
F	T	F
F	F	F

The **disjunction** of statements p and q is the statement "p or q", denoted $p \vee q$. Its truth table is

p	q	$p \vee q$
T	T	T
T	F	T
F	T	T
F	F	F

The **conditional implication** from a statement p to a statement q is the statement "if p, then q". The statement p is called the **hypothesis** of this implication and q the **conclusion**. This logical connector is symbolized by $p \to q$, and its truth table is defined by

p	q	$p \to q$
T	T	T
T	F	F
F	T	T
F	F	T

Whenever p is false, the implication is irrelevant and the argument is valid for any conclusion, thus it was assigned the true value T.

Example 3.1. Let

p : It is a week day.
q : I go to school.

Then the statement $p \to q$ is the sentence

If it is a week day, then I go to school.

This example may help the read to understand why the truth table for $p \to q$ is given above. Let us say that, suppose it is really a week day, and I did go to school; then the statement is logically "right"; so the statement $p \to q$ receives a true value T. Suppose it is really a week day and I did not go to school when it is a week day; then there is something wrong; so the statement $p \to q$ receives a false value F. However, suppose it is not a week day (say weekend or holiday); then I don't need go to school, so it is all right either I go to school or not go to school; the statement $p \to q$ always receive a true value T.

The **Biconditional Implication** of statements p and q is the statement $(p \to q) \wedge (q \to p)$. Its truth table is given by

p	q	$p \leftrightarrow q$
T	T	T
T	F	F
F	T	F
F	F	T

Let p and q be statements. The **converse** of the statement $p \to q$ is the statement $q \to p$. The **inverse** of $p \to q$ is the statement $\neg p \to \neg q$. The **contrapositive** form of $p \to q$ is the statement $\neg q \to \neg p$.

Example 3.2. Let

p : I got SARS.
q : I stayed in hospital.

The converse, inverse, and contrapositive forms of $p \to q$ are given, respectively, as follows:

$q \to p$:	If I stayed in hospital, then I got SARS.
$\neg p \to \neg q$:	If I didn't get SARS, then I didn't stay in hospital.
$\neg q \to \neg p$:	If I didn't stay in hospital, then I didn't get SARS.

Sometimes we need to consider a family of statements $P(x)$ indexed by a variable x (such statement form indexed by a variable is called a **predicate**). We need **universal quantifier** to express whether all of the statements are simultaneously true or one of them is false. We also need **existential quantifier** to express whether at least one of the statements is true or all of them are false. The universal quantification of a predicate $P(x)$ is the statement "for all values of x $P(x)$ is true," denoted $\forall x\, P(x)$. This means that the statement "$\forall x\, P(x)$" has true value when all $P(x)$ have true value and "$\forall x\, P(x)$" has false value when one of $P(x)$ has false value. For example, let $P(x)$ denote $x+1 < 4$, where x are real numbers. Then $\forall x\, P(x)$ is a false statement because $P(4)$ is not a true statement. The existential quantification of a predicate $P(x)$ is the statement "there exists a value of x for which $P(x)$ is true," denoted $\exists x\, P(x)$. This means that $\exists x\, P(x)$ has true value when there is at least one x such that $P(x)$ has true value and $\exists x\, P(x)$ has false value when all statements $P(x)$ have false value. For example, let $Q(x,y,z)$ denote $x^2 + y^2 = z^2$. Then $\exists x \exists y \exists z\, Q(x,y,z)$ is a true statement, because $Q(3,4,5)$ is a true statement.

Note that here the index x in a predicate is not a propositional variable and its values are sometimes specified by its domain X. So we may have sentences "$\forall x \in X, P(x)$" and "$\exists x \in X, P(x)$". For instance, let $\sqrt{x}$ be the square root of real numbers x and let $P(x)$ denote the statement that $\sqrt{x}$ is irrational. Then the statement

"for all primes x the number $\sqrt{x}$ is irrational"

can be expressed as "$\forall$ primes $x, P(x)$".

3.3. Tautology

A statement is called a **tautology** if it is always true for all possible values of its propositional variables; a **contradiction** if it is always false; and a **contingency** if it can be either true or false, depending on the truth values of its propositional variables. For instance, $(p \to q) \vee \neg q$ is a tautology; $(p \to q) \wedge p \wedge \neg q$ is a contradiction; and $(p \to q) \vee \neg p$ is a contingency.

Two statements p and q are said to be **logically equivalent** or simply **equivalent**, written $p \equiv q$ or even $p = q$, if $p \leftrightarrow q$ is a tautology; that is, p and q have the same truth values.

Proposition 3.1. *Let p, q, r be arbitrary statements. Then*

(1) $p \wedge q \equiv q \wedge p$
(2) $p \vee q \equiv q \vee q$
(3) $p \wedge (q \wedge r) \equiv (p \wedge q) \wedge r$
(d) $p \vee (q \vee r) \equiv (p \vee q) \vee r$
(4) $p \wedge (q \vee r) \equiv (p \wedge q) \vee (p \wedge r)$
(5) $p \vee (q \wedge q) \equiv (p \vee q) \wedge (p \vee r)$
(6) $p \wedge p \equiv p$
(7) $p \vee p \equiv p$

(8) $\neg(\neg p) \equiv p$
(9) $\neg(p \wedge q) \equiv \neg p \vee \neg q$
(10) $\neg(p \vee q) \equiv \neg p \wedge \neg q$

EXAMPLE 3.3. $(p \to q) \leftrightarrow (\neg p) \vee q$ is a tautology.

p	q	$p \to q$	$\neg p$	$\neg p \vee q$	$(p \to q) \leftrightarrow (\neg p \vee q)$
T	T	T	F	T	T
T	F	F	F	F	T
F	T	T	T	T	T
F	F	T	T	T	T

EXAMPLE 3.4. $(p \to q) \leftrightarrow (\neg q \to \neg p)$ is a tautology.

p	q	$p \to q$	$\neg q$	$\neg p$	$\neg q \to \neg p$	$(p \to q) \leftrightarrow (\neg q \to \neg p)$
T	T	T	F	F	T	T
T	F	F	T	F	F	T
F	T	T	F	T	T	T
F	F	T	T	T	T	T

For instance, let us consider the statement

"If I got SARS, then I stayed in hospital."

Let p denote "I got SARS" and let q denote "I stayed in hospital." One feel that statement "If I got SARS, then I stayed in hospital" is logically equivalent to the statement

"If I didn't stay in hospital, then I didn't get SARS."

It is also logically equivalent to

"I didn't get SARS or I stayed in hospital."

THEOREM 3.2. (1) $(p \to q) \equiv (\neg p) \vee q$
(2) $(p \to q) \equiv (\neg q \to \neg p)$
(3) $(p \leftrightarrow q) \equiv (p \to q) \wedge (q \to p)$

THEOREM 3.3. (1) $\neg(\forall x\ P(x)) \equiv \exists x\ \neg P(x)$
(2) $\neg(\exists x\ P(x)) \equiv \forall x\ \neg P(x)$
(3) $\forall x\ (P(x) \wedge Q(x)) \equiv (\forall x\ P(x)) \wedge (\forall x\ Q(x))$
(4) $\exists x\ (P(x) \vee Q(x)) \equiv (\exists x\ P(x)) \vee (\exists x\ Q(x))$
(5) $(\forall x\ P(x)) \vee (\forall x\ Q(x)) \to \forall x\ (P(x) \vee Q(x))$ *is a tautology.*
(6) $\exists x\ (P(x) \wedge Q(x)) \to (\exists x\ P(x)) \wedge (\exists x\ Q(x))$ *is a tautology.*
(7) $((\exists x\ P(x)) \to (\forall x\ Q(x))) \to \forall x\ (P(x) \to Q(x))$ *is a tautology.*
(8) $\exists x\ (P(x) \to Q(x)) \equiv (\forall x\ P(x)) \to (\exists x\ Q(x))$

PROOF. (1)-(4) are trivial.

(5) If the statement $(\forall x\, P(x)) \vee (\forall x\, Q(x))$ has T value, then $(\forall x\, P(x)) = T$ or $(\forall x\, Q(x)) = T$, say $(\forall x\, P(x)) = T$. Obviously, $(\forall x\, P(x) \vee Q(x)) = T$. Note that $(\forall x\, P(x)) \vee (\forall x\, Q(x))$ and $(\forall x\, P(x) \vee Q(x))$ are not equivalent.

(6) It is an equivalent form of (5).

(7) $(\exists x\, P(x)) \to (\forall x\, Q(x)) \equiv \neg(\exists x\, P(x)) \vee (\forall x\, Q(x)) \equiv (\forall x\, \neg P(x)) \vee (\forall x\, Q(x))$; $\forall x\, (P(x) \to Q(x)) \equiv \forall x\, (\neg P(x) \vee Q(x))$. The tautology follows from (5).

(8) $(\exists x\, P(x) \to Q(x)) = (\exists x\, \neg P(x) \vee Q(x))$. It follows from (4) that $(\exists x\, \neg P(x) \vee Q(x))$ is equivalent to

$$(\exists x\, \neg P(x)) \vee (\exists x\, Q(x)) = \neg(\forall x\, P(x)) \vee (\exists x\, Q(x)) = (\forall x\, (x)) \to (\exists x\, Q(x)).$$

$\square$

Definition 3.4. A subset of connectives is called **adequate** if every statement can be represented by a statement form containing only connectives from that subset.

Theorem 3.5. *The subset $\{\neg, \vee, \forall\}$ is adequate. In this adequate subset, $\vee$ can be replaced by either $\wedge$ or $\to$; and $\forall$ can be replaced by $\exists$.*

3.4. Methods of Proof

Let p and q be statements. If $p \to q$ is a tautology, then we say that q **follows logically** from p, and write $p \Rightarrow q$. The statement $p \to q$ is also called a **theorem**. For statements $p_1, p_2, \cdots, p_n$, if

$$(p_1 \wedge p_2 \wedge \cdots \wedge p_n) \Rightarrow q,$$

that is, $(p_1 \wedge p_2 \wedge \cdots \wedge p_n) \to q$ is a tautology, we say that q **follows logically** from $p_1, p_2, \cdots, p_n$, and write

$$\begin{array}{c} p_1 \\ p_2 \\ \vdots \\ \underline{p_n} \\ q \end{array}$$

The statements $p_1, p_2, \cdots, p_n$ are called the **hypothesis** (or **premises**) and q the **conclusion**. To prove the theorem $p \Rightarrow q$, it means to show that the implication $p \to q$ is a tautology. Arguments based on tautology are called **rules of inference**. The true of rules of inference is universal, and is independent of the context of the truth values of the simple statements involved.

Modus Ponens, also called **Rule of Detachment** (method of affirming), is the inference

$$\begin{array}{l} p \\ \underline{p \to q} \\ q \end{array}$$

This means that the statement $(p \wedge (p \to q)) \to q$ is a tautology.

p	q	$p \to q$	$p \wedge (p \to q)$	$(p \wedge (p \to q)) \to q$
T	T	T	T	T
T	F	F	F	T
F	T	T	F	T
F	F	T	F	T

Law of Syllogism, also known as **Chain Rule**, is the inference

$$\begin{array}{l} p \to q \\ \underline{q \to r} \\ p \to r \end{array}$$

This means that the statement $((p \to q) \wedge (q \to r)) \to (p \to r)$ is a tautology. In fact, the statement has false value if and only if $(p \to q) \wedge (q \to r)$ has true value and $p \to r$ has false value. Then both $p \to q$ and $q \to r$ have true value, but p must have true value and r must have false value. Thus q must have true value by the true value of $p \to q$. Since $q \to r$ has true value, then r has true value, a contradiction.

EXAMPLE 3.5. If two integers a and b are even, then their sum $a + b$ is even.

PROOF.

STATEMENT	REASON
1. $a = 2a'$, $b = 2b'$.	Hypothesis and the definition of even
2. $a + b = 2a' + 2b'$.	Step 1 and the meaning of = and +
3. $a + b = 2(a' + b') = 2c$.	Factoring
4. $a + b$ is even.	Step 3 and the definition of even

Note that we have not yet proved that $a + b$ is even; we have simply proved *"If a and b are even, then $a+b$ is even."* The above informal argument can be made into the following formal argument.

SYMBOL	STATEMENT	REASON
1. p	a and b are even.	Hypothesis
2. $p \to q$	If a and b are even, then $a = 2a'$ and $b = 2b'$.	Definition of even
3. $q \to r$	If $a = 2a'$ and $b = 2b'$, then $a + b = 2a' + 2b'$.	Meaning of = and +
4. $p \to r$	If a and b are even, then $a + b = 2a' + 2b'$	Steps 2 and 3, and the Law of the Syllogism
5. $r \to s$	If $a + b = 2a' + 2b'$, then $a + b = 2(a' + b') = 2c$.	Factoring
6. $p \to s$	If a and b are even, then $a + b = 2c$.	Steps 4 and 5, and the Law of the Syllogism
7. $s \to t$	If $a + b = 2c$, then $a + b$ is even.	Definition of even
8. $p \to t$	If a and b are even, then $a + b$ is even	Steps 6 and 7, and the Law of Syllogism
9. t	$a + b$ is even.	Steps 1 and 8, and the Rule of Detachment

□

Proof by Contradiction, also called **Modus Tollens** (method of denying), is the inference

$$\begin{array}{c} p \to q \\ \underline{\neg q} \\ \neg p \end{array}$$

This means that the statement $((p \to q) \wedge \neg q) \to \neg p$ is a tautology.

EXAMPLE 3.6. If n^2 is even, then n is even. $(p \to q)$

PROOF.

STATEMENT	REASON
1. $n = 2m + 1$	Definitiomn of odd
2. $n^2 = (2m+1)^2 = 4m^2 + 4m + 1$	Meaning of $=$, $+$; and
$= 2(2m^2 + 2m) + 1 = 2M + 1$	using of algebra
3. n^2 is odd	Step 2 and definition of odd

The informal argument can be made into the following formal argument.

SYMBOL	STATEMENT	REASON
1. $\neg q$	n is not even.	Denying
2. $\neg q \to r$	If n is not even, then $n = 2m + 1$.	Meaning of not even
3. $r \to s$	If $n = 2m + 1$, then $n^2 = 2(2m^2 + 2m) + 1 = 2M + 1$.	Using of algebra
4. $\neg q \to s$	If n is not even, then $n^2 = 2M + 1$.	Law of Syllogism
5. $s \to t$	If $n^2 = 2M + 1$, then n is odd.	Definition of odd
6. $\neg q \to t$	If n is not even, then $n^2 = 2M + 1$.	Law of Syllogism
7. t	$n^2 = 2M + 1$.	Rule of Detachment
8. $\neg p$	n^2 is not even.	Meaning of not even

□

3.5. Mathematical Induction

Mathematical Induction (MI) is the following inference about a family of statements $P(k)$, indexed by positive integers k

$$\begin{array}{l} P(1) \\ \forall k\ P(k) \to P(k+1) \\ \hline \forall k\ P(k) \end{array}$$

Mathematical Induction is a consequence of applying the Modus Ponens and the Law of Syllogism again and again.

EXAMPLE 3.7. For any positive integer n,

$$\sum_{i=1}^{n} i^2 = 1^2 + 2^2 + \cdots + n^2 = \frac{n(n+1)(2n+1)}{6}.$$

3.6. Boolean Functions

The truth values T and F are sometimes denoted by 1 and 0, and write $B = \{0, 1\}$. The nth product B^n is called the n-dimensional binary space. Any function $f : B^n \to B$ is called a **Boolean function** of n variables; the value of f for Boolean variables $(x_1, \ldots, x_n) \in B^n$ is denoted by $f(x_1, \ldots, x_n)$; the variables $x_1, \ldots, x_n$ and their negations $\bar{x}_1, \ldots, \bar{x}_n$ are called **literals**. The variables $x_1, \ldots, x_n$ care called **Boolean variables** and can be viewed as variables of statements. A **Boolean formula** is a sentence consisting of some literals and connectives $\wedge$ and $\vee$.

THEOREM 3.6. *Any Boolean function can be written as a Boolean formula with literals and connectives of disjunction and conjunction.*

Proof. We proceed by induction on the number of variables. For one variable x, there are four Boolean functions: $f_1(1) = f_1(0) = 1$; $f_2(1) = f_2(0) = 0$; $f_3(1) = 1, f_3(0) = 0$; $f_4(1) = 0, f_4(0) = 1$. It is clear that

$$f_1(x) = x \vee \bar{x}; \quad f_2(x) = x \wedge \bar{x}; \quad f_3(x) = x; \quad f_4(x) = \bar{x}.$$

Assume it is true for $n-1$ variables; consider a Boolean function f of n variables $(x_1, \ldots, x_n)$. Define two Boolean functions of $n-1$ variables as follows:

$$g_1(x_2, \ldots, x_n) = f(1, x_2, \ldots, x_n), \tag{3.1}$$

$$g_0(x_2, \ldots, x_n) = f(0, x_2, \ldots, x_n). \tag{3.2}$$

Then it is not hard to verify

$$f(x_1, \ldots, x_n) = (x_1 \wedge g_1(x_2, \ldots, x_n)) \vee (\bar{x}_1 \wedge g_0(x_2, \ldots, x_n)).$$

By induction hypothesis, g_1 and g_0 can be written as Boolean formulas of literals and connectives of disjunction and conjunction; so does f. □

Example Given the Boolean function

(x_1, x_2, x_3)	$f(x_1, x_2, x_3)$
(0,0,0)	1
(0,0,1)	1
(0,1,0)	0
(0,1,1)	0
(1,0,0)	0
(1,0,1)	1
(1,1,0)	0
(1,1,1)	1

It can be written as

$$f(x_1, x_2, x_3) = (x_1 \wedge g_1(x_2, x_3)) \vee (\bar{x} \wedge g_0(x_2, x_3)),$$

where

$$\begin{aligned} g_1(x_2, x_3) &= f(1, x_2, x_3) \\ &= (x_2 \wedge g_{11}(x_3)) \vee (\bar{x}_2 \wedge g_{10}(x_3)), \\ g_{11}(x_3) &= f(1, 1, x_3) = x_3, \\ g_{10}(x_3) &= f(1, 0, x_3) = x_3 \end{aligned}$$

and

$$\begin{aligned} g_0(x_2, x_3) &= (x_2 \wedge g_{01}(x_3)) \vee (\bar{x}_2 \wedge g_{00}(x_3)) \\ g_{01}(x_3) &= f(0, 1, x_3) = x_3 \wedge \bar{x}_3, \\ g_{00}(x_3) &= f(1, 1, x_3) = x_3 \wedge \bar{x}_3. \end{aligned}$$

Then

$$\begin{aligned} g_1(x_2, x_3) &= (x_2 \wedge x_3) \vee (\bar{x}_2 \wedge x_3) = x_3, \\ g_0(x_2, x_3) &= (x_2 \wedge x_3 \wedge \bar{x}_3) \vee (\bar{x}_2 \wedge (x_3 \vee \bar{x}_3)) = \bar{x}_2. \end{aligned}$$

Hence

$$f(x_1, x_2, x_3) = (x_1 \wedge x_3) \vee (\bar{x}_1 \wedge \bar{x}_2).$$

EXERCISES

(1) Consider the statement

$$\textit{If } 1 = 4, \textit{ then } 1 = 2.$$

Proof. Since $1 + 3 = 4$ and $1 = 4$, we have $0 = 3$. Dividing both sides of $0 = 3$ by 3, we further have $0 = 1$. Hence $1 = 0 + 1 = 1 + 1 = 2$. Is the proof a true argument? What can you conclude from the statement and proof?

(2) Define the connectives "$\downarrow$" and "Δ" by

p	q	$p \downarrow q$
T	T	F
T	F	F
F	T	F
F	F	T

and

p	q	$p\Delta q$
T	T	F
T	F	T
F	T	T
F	F	F

respectively. Find the truth tables for

$$(p \downarrow q) \downarrow r,\ (p \downarrow q) \wedge (p \downarrow r),\ (p \downarrow q) \downarrow (p \downarrow r),\ (p \wedge q)\Delta p,\ (p\Delta q)\Delta(q\Delta r).$$

(3) Let

$p:$ John is a student of Computer Science Department in HKUST.
$q:$ John takes course Math132

Write the English sentences of the converse, inverse and contrapositive forms for the statement $p \to q$. Write the English sentence for the statements

$$\neg p \vee q,\ \neg q \vee p$$

(4) Show that the set $\{\neg, \to, \exists\}$ is adequate. Are the sets $\{\neg, \downarrow, \forall\}$ and $\{\neg, \Delta, \exists\}$ adequate?

(5) Show that the statement

$$(\forall x\, P(x)) \vee (\forall x\, Q(x)) \to \forall x\, (P(x) \vee Q(x))$$

is a tautology. Is the converse of the statement a tautology? If yes, prove it. If no, find a counterexample.

(6) Show that if statements p and $p \to q$ are tautologies then q is a tautology. Give a daily life example of the argument.

(7) Express $p \downarrow q$ and $p\Delta q$ in terms of p, q, and other connectives without $\downarrow$ and Δ.

(8) If $p \to q$ and $q \to r$ are tautologies, then $p \to r$ is a tautology.

(9) If $p \to q$ and $\neg q$ are tautologies, then $\neg p$ is a tautology.

CHAPTER 4

Combinatorics

4.1. Counting Principle

For two tasks T_1 and T_2 to be performed in sequence, if the task T_1 can be performed in m ways, and for each of these m ways the task T_2 can be performed in n ways, then the task sequence T_1T_2 can be performed in mn ways. Using the set language notation, let X be the set of plans to have task T_1 performed and Y the set of plans to have the task T_2 performed, then the product

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

is the set of plans to have the task sequence T_1T_2 to be performed, and

$$|X \times Y| = |X||Y|.$$

Example Suppose a lady has three hats, seven shirts, five skirts, and four pairs of shoes. Assume all hats, shirts, skirts, and shoes are distinct. In how many ways can the lady dress herself by selecting each from the hats, the shirts, the skirts, and the shoes?

$$\text{answer} = 3 \cdot 7 \cdot 5 \cdot 4 = 420.$$

Example Courses Calculus, Linear Algebra, and Discrete mathematics are taught by twenty, fifteen, and ten different instructors respectively in a meg university. In how many ways can a student take two of the three courses by selecting instructors?

$$\text{answer} = 20 \cdot 15 + 20 \cdot 10 + 15 \cdot 10 = 650.$$

Let X and Y be arbitrary finite sets and let $f : X \to Y$ be a function from X to Y. If the inverse image $f^{-1}(y) = \{x \in X \mid f(x) = y\}$ has equal number of elements for all $y \in Y$, that is, $|f^{-1}(y)| = k$, then

$$|Y| = \frac{|X|}{k}.$$

4.2. Permutations

Let A be a set of n objects. An arrangement of r elements from A in linear order is called an **r-permutation of n objects**. The number of r-permutations of n objects is denoted by $P(n, r)$. In forming an r-permutation of n objects, the 1st element can be selected in n choices, the 2nd element in $n - 1$ choices, and so on. Thus

$$P(n, r) = n(n-1)\cdots(n-r+1).$$

In particular, when $r = n$, an n-permutation of n objects is simply called a **permutation of n objects**. The number of permutations of n objects is given by

$$n! := n(n-1)(n-2)\cdots 3 \cdot 2 \cdot 1,$$

where $n!$ is read 'n factorial'.

Example Find the number of possible seating plans for four persons to be seated at a round table.
Solution. Let $A = \{a, b, c, d\}$ be the set of four persons. We denote by X the set of all permutations of A and by Y the set of all round permutations of A. We define a map $f : X \to Y$ such that for any permutation $x_1x_2x_3x_4$, $f(x_1x_2x_3x_4)$ is the round permutation by letting x_1 sit next to x_4 and keeping all other persons in same order. It is clear that the map f is onto; and for each round permutation there are 4 permutations sending to the same round permutation. For instance, the four permutations

$$x_1x_2x_3x_4, \quad x_2x_3x_4x_1, \quad x_3x_4x_1x_2, \quad x_4x_3x_2x_1$$

are sent to a same round permutation. We then have

$$4! = |X| = 4|Y|.$$

Therefore

$$|Y| = \frac{4!}{4} = 3! = 6.$$

PROPOSITION 4.1. *The number of round permutations of n objects is*

$$\frac{n!}{n} = (n-1)!.$$

COROLLARY 4.2. *The number of necklaces with n (≥ 3) distinct marbles is*

$$\frac{(n-1)!}{2}.$$

Elements in a set are always considered distinct. When considering indistinguishable elements we need the concept of multisets. By a **multiset** we mean a collection of objects such that some of them may be identically the same, called **indistinguishable**. Let A be a multiset of n objects such that there are k distinguishable types of objects. If there are n_i indistinguishable objects for the ith type, where $1 \leq i \leq k$, the multiset is called a **multiset of type** $(n_1, n_2, \ldots, n_k)$.

Example How many ways to arrange 6 color balls of the same size, of which two are white, 3 are black, and 2 is red, in linear order?

Solution. Let the 6 balls be represented by w, w, b, b, b, r, r. Let us label the balls of the same color by numbers to get 6 distinct balls w_1, w_2, b_1, b_2, b_3, r_1, r_2. Then each permutation of the 6 balls without labels produces $12 = 2!3!2!$ permutations of the 6 distinct balls (6 color balls with labels). More precisely, let X be the set of permutations of $b_1, b_2, b_3, w_1, w_2, r_1, r_2$ and Y the set of permutations of b, b, b, w, w, r, r. Then there is a map $f : X \to Y$, sending each permutation of $b_1, b_2, b_3, w_1, w_2, r_1, r_2$ to a permutation of b, b, b, w, w, r, r by erasing the labels. For instance,

$$2!3!2!\left\{\begin{array}{ll} r_1b_1w_1b_2b_3r_2w_2 & r_2b_1w_1b_2b_3r_1w_2 \\ r_1b_1w_1b_3b_2r_2w_2 & r_2b_1w_1b_3b_2r_1w_2 \\ r_1b_2w_1b_1b_3r_2w_2 & r_2b_2w_1b_1b_3r_1w_2 \\ r_1b_2w_1b_3b_1r_2w_2 & r_2b_2w_1b_3b_1r_1w_2 \\ r_1b_3w_1b_1b_2r_2w_2 & r_2b_3w_1b_1b_2r_1w_2 \\ r_1b_3w_1b_2b_1r_2w_2 & r_2b_3w_1b_2b_1r_1w_2 \\ r_1b_1w_2b_2b_3r_2w_1 & r_2b_1w_2b_2b_3r_1w_1 \\ r_1b_1w_2b_3b_2r_2w_1 & r_2b_1w_2b_3b_2r_1w_1 \\ r_1b_2w_2b_1b_3r_2w_1 & r_2b_2w_2b_1b_3r_1w_1 \\ r_1b_2w_2b_3b_1r_2w_1 & r_2b_2w_2b_3b_1r_1w_1 \\ r_1b_3w_2b_1b_2r_2w_1 & r_2b_3w_2b_1b_2r_1w_1 \\ r_1b_3w_2b_2b_1r_2w_1 & r_2b_3w_2b_2b_1r_1w_1 \end{array}\right\} \overset{f}{\mapsto} rbwbbrw.$$

Clearly, f is onto. For each permutation P of b, b, b, w, w, r, r, its inverse image $f^{-1}(P)$ consists of 2!3!2! permutations of $b_1, b_2, b_3, w_1, w_2, r_1, r_2$. Thus $|X| = 24|Y|$, that is,

$$|Y| = \frac{|X|}{2!3!2!} = \frac{7!}{2!3!2!} = 210.$$

THEOREM 4.3. *The number of permutations of n objects of type $(n_1, n_2, \ldots, n_k)$ is*

$$\frac{n!}{n_1!n_2!\cdots n_k!}.$$

Example How many ways to put five same calculus books, three same physics books, and two same chemistry book in a bookshelf?

$$\text{answer} = \frac{10!}{5!3!2!} = 2520.$$

COROLLARY 4.4. *The number of sequences of* 0 *and* 1 *of length n with exact r* 1*s and* $(n-r)$ 0*s is given by*

$$\frac{n!}{r!(n-r)}.$$

Example Counting the number of nondecreasing coordinate paths from the origin (0,0) to the point (6,4).

Solution. Note that each such path can be viewed as a walk by moving to the right and up. If we denote the moving of one unit to the right by R and the moving of one unit up by U, then each such path can be viewed as a sequence of R and U of length 10 with 6 Rs and 4 Us. Thus the answer is

$$\frac{10!}{6!4!} = 210.$$

PROPOSITION 4.5. *The number of non-decreasing coordinate paths from* $(0,0)$ *to* (a, b)*, where a and b are both non-negative integers, is given by*

$$\frac{(a+b)!}{a!b!}.$$

Thinking Problem Find a formula for the number of round permutations of n objects of type $(n_1, n_2, \ldots, n_k)$.

Example How many possible seating plans can be made for r people to be seated at a round table of n seats, leaving $n-r$ seats empty? (Assume $n \geq r \geq 1$)

Solution. Each seating plan produces n permutations of n objects of type $(\underbrace{1,\ldots,1}_{r}, n-r)$. Then the answer is given by

$$\frac{\binom{n}{1,\ldots,1,n-r}}{n} = \frac{n!}{(n-r)!n}.$$

4.3. Combination

A **combination** is a collection of objects (order is immaterial) from a given source of objects. An **r-combination of n objects** is a collection of r objects from a source of n objects, that is, an r-subset of an n-set. The number of r-combinations n objects is denoted by

$$\binom{n}{r},$$

read 'n choose r'.

Example Find the number of 3-subsets of a 5-set $A = \{a,b,c,d,e\}$

Solution. First Method: Let X be the set of all permutations of the five elements a,b,c,d,e and let Y be the set of all 3-subsets of A. Then there is a map $f : X \to Y$, sending each permutation $x_1x_2x_3x_4x_5$ to the 3-subset $\{x_1,x_2,x_3\}$, that is, $f(x_1x_2x_3x_4x_5) = \{x_1,x_2,x_3\}$. It is clear that f is onto, and the inverse image of every 3-subset from Y has $3!2!$ permutations in X. For instance,

$$12 = 3!2! \left\{ \begin{array}{c} acebd \\ aecbd \\ caebd \\ ceabd \\ eacbd \\ ecabd \\ acedb \\ aecdb \\ caedb \\ ceadb \\ eacdb \\ ecadb \end{array} \right\} \overset{f}{\mapsto} \{a,c,e\}$$

Thus $|X| = 3!2!|Y|$ and the answer is

$$\binom{5}{3} = \frac{5!}{3!2!} = 10.$$

Second Method: Let X be the set of 3-permutations of A and Y the set of 3-subsets of A. Define $f : X \to Y$ by $f(x_1x_2x_3) = \{x_1,x_2,x_3\}$ for each 3-permutation $x_1x_2x_3 \in X$. Obviously, there are $3!$ permutations of $\{x_1,x_2,x_3\}$ sent to $\{x_1,x_2,x_3\}$. We then have

$$|Y| = \frac{|X|}{3!} = \frac{P(6,3)}{3!}.$$

THEOREM 4.6. *The number of r-combinations of n objects is*

$$\binom{n}{r} = \frac{n!}{r!(n-r)!} = \frac{P(n,r)}{r!}.$$

THEOREM 4.7. (Binomial Theorem or Binomial Expansion)

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}.$$

Proof.

$$\begin{aligned}
(x+y)^n &= \underbrace{(x+y)(x+y)\cdots(x+y)}_{n} \\
&= \sum u_1 u_2 \cdots u_n \ (u_i = x \text{ or } y,\ 1 \le i \le n) \\
&= \sum_{k=0}^{n} \left\{ \begin{array}{l} \text{number of sequences of } x \text{ and } y \text{ of} \\ \text{length } n \text{ with } k\ x\text{'s and } (n-k)\ y\text{'s} \end{array} \right\} \\
&= \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}.
\end{aligned}$$

□

A collection of k disjoint subsets of an n-set is called a **combination of n objects of type** $(n_1, n_2, \ldots, n_k)$ if the k subsets have the cardinalities $n_1, n_2, \ldots, n_k$ and $n = n_1 + n_2 + \cdots + n_k$. A combination of n objects of type $(n_1, n_2, \ldots, n_k)$ can be viewed as a placement of n objects into k boxes so that the 1st box contains n_1 objects, the 2nd box contains n_2 objects, and so on. The number of combinations of n objects of type $(n_1, n_2, \ldots, n_k)$ is denoted by

$$\binom{n}{n_1, n_2, \ldots, n_k},$$

read 'n choose n_1, n_2, dot dot dot, and n_k'.

Example How many ways can six distinct objects be placed into three boxes so that the 1st box contains two objects, the 2nd box contains three objects, and the 3rd box contains one object?

Solution. Given a set $A = \{a, b, c, d, e, f\}$ of six objects. Let X be the set of permutations of A, and let Y be the set of placements of elements of A into three boxes so that the 1st box receives two elements, the 2nd box receives thee elements, and the 3rd box receives one element. Then there is map $f : X \to Y$, sending each permutation $x_1x_2x_3x_4x_5x_6$ of A to the placement $\{x_1, x_2\}\{x_3, x_4, x_5\}\{x_6\}$. It is clear that f is onto. For each placement $\{x_1, x_2\}\{x_3, x_4, x_5\}\{x_6\}$, its inverse image

$f^{-1}(\{x_1, x_2\}\{x_3, x_4, x_5\}\{x_6\})$ has $2!3!1!$ permutations of A. For instance,

$$
2!3!1! \left\{ \begin{array}{c} acbdfe \\ acbfde \\ acdbfe \\ acdfbe \\ acfbde \\ acfdbe \\ cabdfe \\ cabfde \\ cadbfe \\ cadfbe \\ cafbde \\ cafdbe \end{array} \right\} \leftrightarrow \begin{array}{|c|c|c|} \hline ac & abf & e \\ \hline ac & bfd & e \\ \hline ac & dbf & e \\ \hline ac & dfb & e \\ \hline ac & fbd & e \\ \hline ac & fdb & e \\ \hline ca & bdf & e \\ \hline ca & bfd & e \\ \hline ca & dbf & e \\ \hline ca & dfb & e \\ \hline ca & fbd & e \\ \hline ca & fdb & e \\ \hline \end{array} \overset{f}{\mapsto} \begin{array}{|c|c|c|} \hline a,\ c & b,\ d,\ f & e \\ \hline \end{array} \leftrightarrow \{a, c\}\{b, d, f\}\{e\}
$$

Thus $|X| = 2!3!1!|Y|$. Therefore the answer is

$$|Y| = \frac{|X|}{2!3!1!} = \frac{6!}{2!3!1!} = \binom{6}{2, 3, 1} = 60.$$

THEOREM 4.8. *The number of ways to place n distinct objects into k distinct boxes so that the 1st box contains n_1 objects, the 2nd box contains n_2 objects, ..., the kth box contains n_k objects is given by*

$$\binom{n}{n_1, n_2, \ldots, n_k} = \frac{n!}{n_1!n_2!\cdots n_k!}.$$

Note: When considering placement of n objects to two boxes of type $(r, n-r)$, we write

$$\binom{n}{r} := \binom{n}{r, n-r} = \frac{n!}{r!(n-r)!}.$$

THEOREM 4.9. (Multinomial Theorem and Multinomial Expansion)

$$(x_1 + x_2 + \cdots + x_k)^n = \sum_{\substack{n_1+n_2+\cdots+n_k=n \\ n_1\geq 0, n_2\geq 0, \ldots, n_k\geq 0}} \binom{n}{n_1, n_2, \ldots, n_k} x_1^{n_1} x_2^{n_2} \cdots x_k^{n_k}.$$

Proof.

$$
\begin{aligned}
(x_1 + x_2 + \cdots + x_k)^n &= \underbrace{(x_1 + x_2 + \cdots + x_k)\cdots(x_1 + x_2 + \cdots + x_k)}_{n} \\
&= \sum u_1 u_2 \cdots u_n \ \ (\text{where } u_i = x_1, x_2, \ldots, x_k,\ 1 \leq i \leq n) \\
&= \sum \left\{ \begin{array}{c} \text{number of sequences of } x_1,\ x_2,\ \ldots,\ x_k \text{ of} \\ \text{length } n \text{ with } n_1\ x_1\text{'s, } n_2\ x_2\text{'s, } \ldots,\ n_k\ x_k\text{'s} \end{array} \right\} \\
&= \sum_{\substack{n_1+n_2+\cdots+n_k=n \\ n_1\geq 0, n_2\geq 0, \ldots, n_k\geq 0}} \binom{n}{n_1, n_2, \ldots, n_k} x_1^{n_1} x_2^{n_2} \cdots x_k^{n_k}.
\end{aligned}
$$

□

4.4. Combination with Repetition

We now consider combinations with repetition allowed. The number of **r-combinations of n objects with repetition allowed** is denoted by

$$\left\langle {n \atop r} \right\rangle .$$

Example How many ways to take seven objects with repetition allowed from a set $A = \{a, b, c, d\}$ of four objects?

Solution. Take a collection of seven objects from A with repetition allowed, say, a, a, b, b, b, c, d, we insert bars (denoted by the symbol 1) between the a's and b's, the b's and the c's, the c's and d's; and denote the objects a, b, c, d by the same symbol 0. Then each collection of seven objects from A is encoded into a sequence of 0 and 1 of length 10 with seven 0's and three 1's. For instance,

$$\begin{array}{rcccccclcc} aa & 1 & bbb & 1 & c & 1 & d & & \mapsto & 0010001010 \\ aaa & 1 & bbbb & 1 & & 1 & & & \mapsto & 0001000011 \\ & 1 & bb & 1 & ccccc & 1 & & & \mapsto & 1001000001 \\ a & 1 & & 1 & ccc & 1 & ddd & & \mapsto & 0110001000 \end{array}$$

Note that different collections of seven objects from A with repetition allowed are encoded into different sequences, and every sequence of 0 and 1 of length 10 with exact seven 0's and three 1's can be obtained in this way. Thus

$$\left\langle {4 \atop 7} \right\rangle = \binom{4+7-1}{7} = \binom{10}{7}.$$

THEOREM 4.10. *The number of r-combinations of n objects with repetition allowed is*

$$\left\langle {n \atop r} \right\rangle = \binom{n+r-1}{r}.$$

Example Eight students plan to have dinner together in a restaurant where the menu shows 20 different dishes. Now each student decides to order one dish. How many possible combinations of dishes can be ordered by the students for their dinner?

Solution.

$$\text{Answer } = \left\langle {20 \atop 8} \right\rangle = \binom{20+8-1}{8} = \binom{27}{8}.$$

THEOREM 4.11. *The number of nonnegative integer solutions for the equation*

$$x_1 + x_2 + \cdots + x_n = r$$

is given by

$$\left\langle {n \atop r} \right\rangle = \binom{n+r-1}{r}.$$

Example There are five types of color T-shirts on sale, black, blue, green, orange, and white. John is going to buy ten T-shirts; he has to buy at least two blues and two oranges, and at least one for all other colors. Find the number of ways that John can select ten T-shirts.

Solution. We label the colors black, blue, green, orange and white by numbers $1, 2, 3, 4, 5$. Let x_i be the number of T-shirts that John would select for the ith color T-shirt. Then the problem is to find the number of integer solutions for the equation

$$x_1 + x_2 + \cdots + x_5 = 10,$$

where $x_1 \geq 1, x_2 \geq 2, x_3 \geq 1, x_4 \geq 2, x_5 \geq 1$.

$$\text{Answer } = \left\langle {5 \atop 3} \right\rangle = \binom{5+3-1}{3} = \binom{7}{3} = 35.$$

Example In how many ways can a student order eight dumplings from three different kinds? In how many ways can a student eat eight dumplings from the three kinds, assuming that there are infinitely many supply of dumplings for each kind and that the dumplings of the same kind should be eaten consecutively one by one?

$$\left\langle {3 \atop 8} \right\rangle = \binom{10}{8} = 45.$$

$$\text{answer} = \sum_{k=1}^{3} k! \left\langle {k \atop 8-k} \right\rangle$$

4.5. Combinatorial Proof

A proof for a result by using bijection is sometimes called a **combinatorial proof** of the result. Here are a few examples.

Example

$$\binom{n+1}{r} = \binom{n}{r-1} + \binom{n}{r}.$$

Proof. Let A_n be a set $\{a_1, a_2, \ldots, a_n\}$ of n elements and let A_{n+1} be the $\{a_1, \ldots, a_n, a_{n+1}\}$ of $n+1$ elements. The r-subsets of A_{n+1} can be divided into two kinds: the r-subsets of A_n and the r-subsets of A_{n+1} that are not subsets of A_n. There are $\binom{n}{r}$ r-subsets of the first kind. Each r-subset of the second kind must contain the element a_{n+1}. Thus the r-subsets of the second kind can be obtained by taking all $(r-1)$-subsets of A_n and adding the element a_{n+1} to each of them; so there are $\binom{n}{r-1}$ r-subsets of the second kind. Thus we have $\binom{n+1}{r} = \binom{n}{r} + \binom{n}{r-1}$.

$$\begin{array}{c} 1 \\ 1 \;\; 2 \;\; 1 \\ 1 \;\; 3 \;\; 3 \;\; 1 \\ 1 \;\; 4 \;\; 6 \;\; 4 \;\; 1 \\ 1 \;\; 5 \;\; 10 \;\; 10 \;\; 5 \;\; 1 \end{array}$$

EXAMPLE 4.1.

$$\binom{2n}{n} = \binom{n}{0}^2 + \binom{n}{1}^2 + \cdots + \binom{n}{n}^2$$

Proof. Let us consider n-combinations of $2n$ balls of the same size, of which n balls are white and the other n balls are black. Each such combination can be obtained by taking k balls from the n white balls and taking $(n-k)$ balls from the n black balls, where k ranges from 0 to n. Then we have

$$\begin{aligned}\binom{2n}{n} &= \binom{n}{0}\binom{n}{n}+\binom{n}{1}\binom{n}{n-1}+\cdots+\binom{n}{n}\binom{n}{0}\\ &= \binom{n}{0}^2+\binom{n}{1}^2+\cdots+\binom{n}{n}^2.\end{aligned}$$

Exercises

(1) A computer user name consists of three English letters followed by five digits. How many different user names can be made?

(2) A set lunch is to include a soup, a main course, and a drink. Suppose a customer can select from three soups, five main courses, and four drinks. How many different set lunches can be selected.

(3) Give a procedure for determining the number of zeros at the end of $n!$. Justify your procedure and make examples for 12! and 26!.

(4) Find the number of different permutations of the letters in HONGKONG.

(5) A bookshelf is to be used to display 10 math books. Suppose there are 8 kinds of calculus books, 6 kinds of linear algebra books, and 5 kinds of discrete math books. It is required that books for the same subject should be displayed together.
   (a) Find the number of ways to display 10 distinct books so that there are 5 calculus books, 3 linear algebra books, and 2 discrete math books.
   (b) Find the number of ways to display 10 books (not necessarily distinct) so that there are 5 calculus books, 3 linear algebra books, and 2 discrete math books.

(6) There are n men and n women to form a circle or a line, $n \geq 2$. Find the number of patterns of
   (a) circles could be formed so that each man is next to at least one woman;
   (b) straight lines could be formed so that each man is next to at least one woman.

(7) Four identical six-sided dice are tossed simultaneously and numbers showing on the top faces are recorded as a multiset of four elements. How many different multisets are possible?

(8) Find the number of non-decreasing coordinate paths from the origin $(0,0,0)$ to the lattice point (a,b,c).

(9) In how many ways can a six-card hand be dealt from a deck of 52 cards.

(10) How many different eight-card hands with five red cards and three black cards can be dealt from a deck of 52 cards?

(11) Fortune draws are arranged to select six ping pang balls simultaneously from a box in which 20 are orange and 30 are white. A draw is lucky if it consists of three orange and thee white balls. What is the chance of a lucky draw?

(12) Determine the number of integer solutions for $x_1+x_2+x_3+x_4 \leq 38$, where
   (a) $x_i \geq 0$, $1 \leq i \leq 5$.

(b) $x_1 \geq 0$, $x_2 \geq 2$, $x_3 \geq -2$, $3 \leq x_4 \leq 8$.

(13) Determne the number of nonnegative integer solutions to the pair of equations

$$x_1 + x_2 + x_3 = 8, \quad x_1 + x_2 + x_3 + x_4 + x_5 + x_6 = 18.$$

(14) Let M be a multiset of type $(n_1, n_2, \ldots, n_k)$ such that $n_i \geq 1$ for $1 \leq i \leq k$. If the numbers $n_1, n_2, \ldots, n_k$ are all coprime with $n = n_1 + n_2 + \cdots + n_k$, then the number of round permutations of M is

$$\frac{\binom{n}{n_1, n_2, \ldots, n_k}}{n}.$$

The formula is actually valid when $\gcd(n_1, \ldots, n_k) = 1$, but we didn't define the gcd yet for more than two integers. Find a counterexample if the conditions are not satisfied.

(15) Find the number of non-decreasing lattice paths from the origin $(0, 0)$ to a non-negative lattice point (a, b), allowing only horizontal, vertical, and diagonal unit moves; that is, allowing moves $(x, y) \to (x + 1, y)$, $(x, y) \to (x, y + 1)$ and $(x, y) \to (x + 1, y + 1)$.

(16) *Find the number of non-decreasing lattice paths from the origin $(0, 0)$ to a non-negative lattice point (a, b), allowing arbitrary straight moves from one lattice point to another lattice point; that is, allowing moves $(x, y) \to (x + k, y + h)$, where k and h are non-negative integers such that $(k, h) \neq (0, 0)$.

Hint for Ex. 14: Let M be a multiset of type $(n_1, \ldots, n_k)$ and $S(M)$ the set of all permutations of M. Define $\sigma : S(M) \to S(M)$ by

$$\sigma(x_1 x_2 x_3 \cdots x_n) = x_2 x_3 \cdots x_n x_1.$$

A permutation $w = x_1 x_2 \cdots x_n$ is called **primitive** if the permutations

$$w, \sigma(w), \sigma^2(w), \ldots, \sigma^{n-1}(w)$$

are distinct.

For a non-primitive permutation $x_1 x_2 \cdots x_n$ of M, there are integers a and b, $0 \leq a < b < n$, such that

$$\sigma^a(x_1 x_2 \cdots x_n) = \sigma^b(x_1 x_2 \cdots x_n).$$

Let $l = b - a$. Then

$$\sigma^l(x_1 x_2 \cdots x_n) = x_1 x_2 \cdots x_n.$$

That is,

$$x_{l+1} x_{l+2} \cdots x_n x_1 x_2 \cdots x_l = x_1 x_2 \cdots x_{n-l} x_{n-l+1} x_{n-l+2} \cdots x_n.$$

This is equivalent to saying that

$$x_{l+1} = x_1,\ x_{l+2} = x_2,\ \ldots,\ x_n = x_{n-l};\ \ x_1 = x_{n-l+1},\ x_2 = x_{n-l+2},\ \ldots,\ x_l = x_n.$$

We claim that $l | n$. In fact, suppose l doesn't divide n, then $n = ql + r$ with $0 < r < l$. Then

$$x_{l+1} = x_1,\ x_{l+2} = x_2,\ \ldots,\ x_{2l} = x_l,\ x_{2l+1} = x_1,\ \ldots,\ x_n = x_{ql+r} = x_r.$$

Thus

$$x_{l+1}x_{l+2}\cdots x_n x_1 x_2\cdots x_l = \underbrace{\underbrace{x_1x_2\cdots x_l}\cdots\underbrace{x_1x_2\cdots x_l}}_{q-1} x_1x_2\cdots x_r x_1x_2\cdots x_l.$$

On the other hand,

$$x_{l+1}x_{l+2}\cdots x_n x_1 x_2\cdots x_l = x_1x_2\cdots x_n = \underbrace{\underbrace{x_1x_2\cdots x_l}\cdots\underbrace{x_1x_2\cdots x_l}}_{q} x_1x_2\cdots x_r.$$

This means that

$$x_1x_2\cdots x_r x_1x_2\cdots x_l = x_1x_2\cdots x_l x_1x_2\cdots x_r.$$

Continuing this procedure, one can find a positive integer s such that $n = sd$ and

$$x_1x_2\cdots x_n = \underbrace{\underbrace{x_1x_2\cdots x_s}\cdots\underbrace{x_1x_2\cdots x_s}}_{d}.$$

Let s_i be the number of elements of the ith type in the segment $x_1x_2\cdots x_s$. Then $s_i d = n_i$. So $d|n_i$, $1 \le i \le k$. Hence d divides $\gcd(n_1,\ldots,n_k)$. We have shown that if $\gcd(n_1,\ldots,n_k) = 1$ then there is no positive integer $l < n$ such that $\sigma^\ell(x_1x_2\cdots x_n) = x_1x_2\cdots x_n$, that is, every permutation is primitive. Thus the number of round permutations is $\binom{n}{n_1,\ldots,n_k}/n$.

Now let $m = \gcd(n_1,\ldots,n_k)$ and let D_m be the set of divisors of m. Then D_m is a finite partially ordered set by its divisibility. For each $d \in D$, let

$$\begin{aligned} f(d) :&= \text{number of permutations of type } (dn_1/m,\ldots,dn_k/m),\\ &= \binom{dn/m}{dn_1/m,\ldots,dn_k/m}\\ g(d) :&= \text{number of primitive permutations of type } (dn_1/m,\ldots,dn_k/m). \end{aligned}$$

Note that each permutation $x_1x_2\cdots x_{dn/m}$ of type $(dn_1/m,\ldots,dn_k/m)$ can be classified as a unique primitive permutation $x_1x_2\cdots x_{d'n/m}$ for some d' such that $d'|d$ and

$$x_1x_2\cdots x_{dn/m} = \underbrace{\underbrace{x_1x_2\cdots x_{d'n/m}}\cdots\underbrace{x_1x_2\cdots x_{d'n/m}}}_{d/d'}.$$

Then

$$f(d) = \sum_{d'|d} g(d').$$

By the Möbius inversion formula, we have

$$g(d) = \sum_{d'|d} \mu(d') f\left(\frac{d}{d'}\right) = \sum_{d'|d} \mu\left(\frac{d}{d'}\right) f(d').$$

Thus the number of round permutations of type $(n_1, \ldots, n_k)$ is given by

$$\begin{aligned}
RP(n_1, \ldots, n_k) &= \sum_{d|m} g(d) \cdot \frac{1}{dn/m} \\
&= \frac{1}{n} \sum_{d|m} \sum_{d'|d} \mu\left(\frac{d}{d'}\right) f(d') \cdot \frac{m}{d} \\
&= \frac{1}{n} \sum_{d'|m} f(d') \sum_{d|m, d'|d} \mu\left(\frac{d}{d'}\right) \cdot \frac{m}{d} \\
&= \frac{1}{n} \sum_{d'|m} f(d') \sum_{k|(m/d')} \mu(k) \cdot \frac{m/d'}{k} \\
&= \frac{1}{n} \sum_{d'|m} f(d') \phi\left(\frac{m}{d'}\right) \\
&= \frac{1}{n} \sum_{d|m} f\left(\frac{m}{d}\right) \phi(d) \\
&= \frac{1}{n} \sum_{d|m} \binom{n/d}{n_1/d, \ldots, n_k/d} \phi(d).
\end{aligned}$$

Here μ is the **Möbius function**, defined for positive integers by

$$\mu(d) = \begin{cases} 1 & \text{if } d = 1, \\ (-1)^k & \text{if } d \text{ is a product of } k \text{ distinct primes,} \\ 0 & \text{if } d \text{ has a repeated prime factor;} \end{cases}$$

and ϕ is **Euler's function**, defined for positive integers by

$$\phi(d) = \text{number of integers coprime to } d \text{ in } [1, d].$$

$$\text{answer:} \sum_{k=0}^{\min\{a,b\}} \binom{a+b-k}{a-k, b-k, k}.$$

For any such path with k diagonal moves ($0 \leq k \leq \min\{a, b\}$), the number of horizontal moves should be $a-k$ and the number of vertical moves should be $b-k$.

4.6. Pigeonhole Principle

THEOREM 4.12. (Pigeonhole Principle) *If n objects are placed into m boxes and $n > m$, then there is at least one box which contains two or more objects.*

The pigeonhole principle is a common Chinese saying: When pigeons are put into pigeonholes with more pigeons than pigeonholes, there are at least two pigeons must be put in a same pigeonhole.

Example Among any five integers between 1 and 8 inclusive, there are at least two of them adding up to 9.

Solution. We can divide the set $\{1, 2, \cdots, 8\}$ into four disjoint subsets where each has two elements adding up to 9: $\{1, 8\}$, $\{2, 7\}$, $\{3, 6\}$, and $\{4, 5\}$. When selecting five numbers from these four subsets, at least two of the five selected numbers must come from a same subset of the four subsets. Thus their addition is 9.

Example Show that in any group of two or more persons there are at least two having the same number of friends (It is assumed that if a person x is a friend of a person y then y is also a friend of x).

Solution. Assume that there are n persons in the group. The number of friends of a person x should be between 0 and $n-1$. If there is one person x^* who has $n-1$ friends, then everyone is a friend of x^*. So both 0 and $n-1$ can not be numbers of friends of some people in the group. Thus the pigeonhole principle tells us that there are at least two people who have the same number of friends.

Example Show that if $a_1, a_2, \ldots, a_k$ are integers (not necessarily distinct), then some of them can be added up to a multiple of k.

Solution. Consider the integers of the following $k+1$ objects:

$$0,\ a_1,\ a_1+a_2,\ a_1+a_2+a_3,\ \cdots,\ a_1+a_2+\cdots+a_k. \tag{4.1}$$

Note that integers modulo k can only be $0, 1, 2, \ldots, k-1$. By the Pigeonhole Principle there are at least two integers in (4.1), say

$$a_1 + \cdots + a_i \quad \text{and} \quad a_1 + \cdots + a_j,$$

whose remainders modulo k are the same. The number $a_1 + \cdots + a_i$ could be the 0, the very first element in (4.1). Thus

$$a_{i+1} + a_{i+2} + \cdots + a_j$$

is a multiple of k.

Example Given 10 distinct integers $a_1, a_2, \ldots, a_{10}$ such that $0 \le a_i < 100$, can we find a subset of $\{a_1, \ldots, a_{10}\}$ such that the sum of numbers in the subset with sign is zero?

Solution. Consider all possible partial sums of the selected numbers $a_1, a_2, \ldots, a_{10}$. The values of these sums should be between 0 and 1000. Note that the number of subsets of 10 objects is $2^{10} = 1024$. By the Pigeonhole Principle there are at least two subsets A and B of $\{a_1, a_2, \ldots, a_{10}\}$ such that the sum of the elements of A and the sum of the elements of B are the same, that is,

$$\sum_{a_i \in A} a_i = \sum_{a_j \in B} a_j.$$

Now we move all elements from the right side to the left; the elements in both A and B will be canceled. Thus sum of the elements of $A \Delta B$ with positive sign for the elements in $A - B$ and negative sign for the elements in $B - A$ is equal to 0.

THEOREM 4.13. *If n objects are placed in m boxes, then one of the boxes must contain at least $\lceil \frac{n}{m} \rceil$ objects, where $\lceil \frac{n}{m} \rceil$ denotes the smallest integer greater or equal to $\frac{n}{m}$.*

4.7. Relation to Probability

There are lot problems in our daily life about chance, the possibility or probability. When we flip a coin, we have two possible **outcomes**, Head and Tail. If the coin is fair, the chance to have the outcome – Head – is one-half or 50%. When we roll a pair of dice we may have outcomes – a collection of pairs of numbers between 1 and 6. The chance of the event of the outcomes that the sum of the pair is even is one-half. For instance, we may be interested in computing the probability of the event of the outcomes that the sum of the pair is 8.

DEFINITION 4.14. Any collection of outcomes in a probabilistic experiment is called an **event**. If each outcome is equally likely to be happened, we define

$$\textbf{Probability of even } A = P(A) = \frac{\textit{Total number of favorite outcomes}}{\textit{Total number of possible outcomes}}.$$

Example What is the probability of selecting three distinct numbers from $1, 2, \ldots, 11$ so that two are less that 5, one is equal to 5, and four are larger than 5?

Solution. The total number of possible outcomes is $\binom{11}{7}$, and the total number of favorite outcomes is $\binom{4}{2}\binom{1}{1}\binom{6}{4}$. Then the probability is

$$\frac{\binom{4}{2}\binom{1}{1}\binom{6}{4}}{\binom{11}{7}}.$$

Example Find the probability that no two persons have the same birthday in a party of 40 people.

Solution. The total number of possible outcomes is 365^{40} and the total number of favorite outcomes is $\binom{365}{40}40!$. The probability is

$$\frac{\binom{365}{40}40!}{365^{40}} \approx 0.109.$$

Example What is the probability of rolling a pair of dice so that the sum of numbers on the top faces is 8?

Solution. Since there is no order between the two dice, there are twenty-one possible outcomes

$$\{i, j\},\ 1 \le i \le j \le 6$$

and three favorite outcomes $\{2, 6\}$, $\{3, 5\}$, $\{4, 4\}$. So the answer might be $\frac{3}{21} = \frac{1}{7}$.

One may color the two dice as black and white so that the two dice are ordered. There are $36 (= 6 \times 6)$ possible outcomes and five favorite outcomes $(2, 6)$, $(3, 5)$, $(4, 4)$, $(5, 3)$, $(6, 2)$. The answer should be $\frac{5}{36}$. Which one is correct and why?

Example Find the probability of rolling four dice simultaneously so that the sum of points is exactly 9.

Solution. The total number of possible outcomes is 6^4. The total number of favorite outcomes is the number of positive integer solutions of the equation

$$x_1 + x_2 + x_3 + x_4 = 9$$

which is equivalent to the number of non-negative integer solutions of the equation

$$y_1 + y_2 + y_3 + y_4 = 5.$$

Thus the answer is given by

$$\frac{\left\langle {4 \atop 5} \right\rangle}{6^4} = \frac{7}{162} \approx \frac{1}{23}.$$

Exercises

(1) Show that there must be 90 ways to choose six numbers from 1 to 15 so that all the choices have the same sum.
(2) Show that if five points are selected in a square whose sides have length 2, then there are at leat two points whose distance is at most $\sqrt{2}$.
(3) Prove that if any 14 numbers from 1 to 25 are chosen, then one of them is a multiple of another.
(4) Twenty disks numbered 1 through 20 are placed face down on a table. Disks are selected one at a time and turned over until 10 disks have been chosen. If two of the disks add up to 21, the play loses. Is it possible to win this game?
(5) Show that it is impossible to arrange the numbers $1, 2, \ldots, 10$ in a circle so that every triple of consecutively placed numbers has a sum less than 15.

4.8. Inclusion-Exclusion Principle

Let U be a finite set. For two subsets A_1 and A_2 of U, we have

$$|A_1 \cup A_2| = |A_1| + |A_2| - |A_1 \cap A_2|,$$

or equivalently,

$$|\bar{A}_1 \cap \bar{A}_2| = |U| - |A_1| - |A_2| + |A_1 \cap A_2|.$$

For three subsets A_1, A_2, A_3 of U, we have

$$\begin{aligned} |A_1 \cup A_2 \cup A_3| &= |A_1| + |A_2| + |A_3| \\ &\quad -|A_1 \cap A_2| - |A_1 \cap A_3| - |A_2 \cap A_3| \\ &\quad +|A_1 \cap A_2 \cap A_3|, \end{aligned}$$

or equivalently,

$$\begin{aligned} |\bar{A}_1 \cap \bar{A}_2 \cap \bar{A}_3| &= |U| - |A_1| - |A_2| - |A_3| \\ &\quad +|A_1 \cap A_2| + |A_1 \cap A_3| + |A_2 \cap A_3| \\ &\quad -|A_1 \cap A_2 \cap A_3|. \end{aligned}$$

These formulas and similar kinds for more subsets are called **Inclusion-Exclusion Principle**.

Example The pin numbers Hang Seng Bank card are nonnegative integers of six digits. How many pin numbers can be made so that the triple 444 doesn't appear?

Solution. Let U be the set of all possible secrete codes. Then $|U| = 10^6$. Let

$$\begin{aligned} A_1 &= \text{set of codes of the form } 444xxx, \\ A_2 &= \text{set of codes of the form } x444xx, \\ A_3 &= \text{set of codes of the form } xx444x, \\ A_4 &= \text{set of codes of the form } xxx444, \end{aligned}$$

where x varies from 0 to 9. Then $|A_1| = |A_2| = |A_3| = |A_4| = 10^3$; $|A_1 \cap A_2| = |A_2 \cap A_3| = |A_3 \cap A_4| = 10^2$, $|A_1 \cap A_3| = |A_2 \cap A_4| = 10$, $|A_1 \cap A_4| = 1$; $|A_1 \cap A_2 \cap A_3| = |A_2 \cap A_3 \cap A_4| = 10$, $|A_1 \cap A_2 \cap A_4| = |A_1 \cap A_3 \cap A_4| = 1$; $|A_1 \cap A_2 \cap A_3 \cap A_4| = 1$. Thus

$$\text{answer} = 10^6 - 4 \cdot 10^3 + (3 \cdot 10^2 + 2 \cdot 10 + 1) - (2 \cdot 10 + 2) + 1 = 996310.$$

Example Find the number of positive integer solutions for the linear equation $x_1 + x_2 + x_3 = 8$.

Solution. Let A_i be the set of non-negative integer solutions of the above equation such that $x_i = 0$, $1 \le i \le 3$. Then

$$|U| = \left\langle {3 \atop 8} \right\rangle = \binom{3+8-1}{8} = \binom{10}{8} = 45;$$

$$|A_1| = |A_2| = |A_3| = 9;$$

$$|A_1 \cap A_2| = |A_1 \cap A_3| = |A_2 \cap A_3| = 1;$$

$$|A_1 \cap A_2 \cap A_3| = 0.$$

By the Inclusion-Exclusion Principle,

$$\begin{aligned} |\bar{A}_1 \cap \bar{A}_2 \cap \bar{A}_3| &= |U| - |A_1| - |A_2| - |A_3| \\ &\quad + |A_1 \cap A_2| + |A_1 \cap A_3| + |A_2 \cap A_3| \\ &\quad - |A_1 \cap A_2 \cap A_3| \\ &= 45 - 3 \cdot 9 + 3 \cdot 1 - 0 = 21. \end{aligned}$$

Of course, one can easily get the same answer by setting $x_i - 1 = y_i$, $1 \le i \le 3$. Then the answer is the number of non-negative integer solutions of the equation $y_1 + y_2 + y_3 = 5$. That is,

$$\left\langle {3 \atop 5} \right\rangle = \binom{3+5-1}{5} = \binom{7}{5} = 21.$$

THEOREM 4.15. *Let U be a finite set. Let A_1, A_2, $\cdots$, A_n be subsets of U. Then*

$$\begin{aligned} |A_1 \cup A_2 \cup \cdots \cup A_n| &= |A_1| + |A_2| + \cdots |A_n| \\ &\quad - \Big[|A_1 \cap A_2| + |A_1 \cap A_3| + \cdots + |A_1 \cap A_n| \\ &\quad + |A_2 \cap A_3| + |A_2 \cap A_4| + \cdots + |A_2 \cap A_n| \\ &\quad + \cdots + |A_{n-1} \cap A_n| \Big] + \\ &\quad + \Big[|A_1 \cap A_2 \cap A_3| + \cdots + |A_{n-2} \cap A_{n-1} \cap A_n| \Big] \\ &\quad - \cdots \\ &\quad + (-1)^{n-1} |A_1 \cap A_2 \cap \cdots \cap A_n| \\ &= \sum_{k=1}^{n} (-1)^{k-1} \sum_{1 \le i_1 < \cdots < i_k \le n} |A_{i_1} \cap \cdots \cap A_{i_k}|; \end{aligned}$$

or equivalently,

$$
\begin{aligned}
|\bar{A}_1 \cap \bar{A}_2 \cap \cdots \cap \bar{A}_n| &= |U| - \sum_i |A_i| + \sum_{i<j} |A_i \cap A_j| \\
&\quad - \sum_{i<j<k} |A_i \cap A_j \cap A_k| \\
&\quad + \cdots \\
&\quad + (-1)^n |A_1 \cap A_2 \cap \cdots \cap A_n| \\
&= |U| + \sum_{k=1}^{n} (-1)^k \sum_{1 \le i_1 < \cdots < i_k \le n} |A_{i_1} \cap \cdots \cap A_{i_k}| . \qquad (4.2)
\end{aligned}
$$

Proof. For each element $x \in U$, we show that x contributes the same count to both sides of (4.2).

Case I: $x \notin A_1 \cup A_2 \cup \cdots \cup A_n$. Note that the element x is counted once in

$$\overline{A_1 \cup A_2 \cup \cdots \cup A_n} = \bar{A}_1 \cap \bar{A}_2 \cap \cdots \cap \bar{A}_n,$$

once in U, and 0 times in all

$$A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}, \quad i_1 < i_2 < \cdots < i_k.$$

Thus x is counted once on both sides of (4.2).

Case II: $x \in A_1 \cup A_2 \cup \cdots \cup A_n$. We assume that x belongs to exactly r subsets of $A_1, A_2, \ldots, A_n$, say $A_{t_1}, A_{t_2}, \ldots, A_{t_r}$. Then x is counted $\binom{r}{0}$, $\binom{r}{1}$, $\binom{r}{2}$, $\binom{r}{3}$, $\ldots$, $\binom{r}{r}$ times in

$$U, \quad \sum_i |A_{t_i}|, \quad \sum_{i<j} |A_{t_i} \cap A_{t_j}|, \quad \sum_{i<j<k} |A_{t_i} \cap A_{t_j} \cap A_{t_k}|, \quad \ldots, \quad |A_{t_1} \cap A_{t_2} \cap \cdots \cap A_{t_r}|$$

respectively. Consequently, the contribution of x on the right side of (4.2) is

$$\binom{r}{0} - \binom{r}{1} + \binom{r}{2} - \binom{r}{3} + \cdots + (-1)^r \binom{r}{r} = [1 + (-1)]^r = 0.$$

Therefore x is counted zero times on both sides of (4.2). $\square$

The subsets $A_1, \ldots, A_n$ of U may be given by conditions or properties $c_1, \ldots, c_n$ on the elements of U respectively. Let N be the number of elements of U, $\bar{N}$ the number of elements of U satisfying none of the conditions $c_1, \ldots, c_n$. Let $N(c_{i_1} c_{i_2} \cdots c_{i_k})$ be the number of elements of U satisfying the conditions $c_{i_1}, c_{i_2}, \cdots, c_{i_k}$, $1 \le i_1 < i_2 < \cdots < i_k \le n$. Then the Inclusion-Exclusion Principle can be stated as

$$
\begin{aligned}
\bar{N} &= N - \Big[N(c_1) + N(c_2) + \cdots + N(c_n) \Big] \\
&\quad + \Big[N(c_1 c_2) + N(c_1 c_3) + \cdots + N(c_1 c_n) \\
&\qquad + N(c_2 c_3) + N(c_2 c_4) + \cdots + N(c_2 c_n) \\
&\qquad + \cdots + N(c_{n-1} c_n) \Big] \\
&\quad + \Big[N(c_1 c_2 c_3) + \cdots + N(c_{n-2} c_{n-1} c_n) \Big] \\
&\quad + \cdots + (-1)^n N(c_1 c_2 \cdots c_n). \qquad (4.3)
\end{aligned}
$$

Let N_k be the number of elements of U satisfying at least k conditions, $0 \leq k \leq n$, that is,

$$N_k = \sum_{i_1 < i_2 < \cdots < i_k} |A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}|.$$

Then the Inclusion-Exclusion Formula can be further stated as

$$|\bar{A}_1 \cap \bar{A}_2 \cap \cdots \cap \bar{A}_n| = N_0 - N_1 + N_2 - N_3 + \cdots + (-1)^n N_n. \tag{4.4}$$

Example Find the number of possible HK telephone numbers having no 14?

Proof. Let U be the set of HK phone numbers. Obviously, $|U| = 10^8$. Let A_i be the set of phone numbers such that the ith digit is 1 and the $(i+1)$th digit is 4, $1 \leq i \leq 7$. Then

$$\begin{aligned}
N_1 &= \sum_{i=1}^{7} |A_i| = \binom{7}{1} 10^6, \\
N_2 &= \sum_{1 \leq i < j \leq 7} |A_i \cap A_j| = \binom{6}{2} 10^4, \\
N_3 &= \sum_{1 \leq i < j < k \leq 7} |A_i \cap A_j \cap A_k| = \binom{5}{3} 10^2, \\
N_4 &= \sum_{1 \leq i < j < k < l \leq 7} |A_i \cap A_j \cap A_k \cap A_l| = \binom{4}{4} 10^0.
\end{aligned}$$

Thus the answer is

$$10^8 - \binom{7}{1} 10^6 + \binom{6}{2} 10^4 - \binom{5}{3} 10^2 + \binom{4}{4} 10^0 = 93149001.$$

There is an algebraic proof for the Inclusion-Exclusion formula. For subset A of U, the **characteristic function** of A is defined by

$$1_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A. \end{cases}$$

For functions $f : U \to \mathbf{R}$ and $g : U \to \mathbf{R}$ and for any real number a, we define functions

$$\begin{array}{lll}
f + g : U \to \mathbf{R} & \text{by} & (f+g)(x) = f(x) + g(x), \\
f - g : U \to \mathbf{R} & \text{by} & (f-g)(x) = f(x) - g(x), \\
fg : U \to \mathbf{R} & \text{by} & fg(x) = f(x)g(x), \\
af : U \to \mathbf{R} & \text{by} & (af)(x) = af(x).
\end{array}$$

PROPOSITION 4.16. *For subsets A and B of U,*

(a) $1_{A \cap B} = 1_A 1_B$,
(b) $1_{\bar{A}} = 1_U - 1_A$, *where* $\bar{A} = U - A$,
(c) $1_U f = f$ *for any function* $f : U \to \mathbf{R}$,
(d) $1_{A \cup B} = 1_A + 1_B - 1_{A \cap B}$.

For a function $f : U \to \mathbf{R}$, the **weight** of f is the number

$$|f| = \sum_{x \in U} f(x).$$

It is clear that

$$|af + bg| = a|f| + b|g|.$$

Note that $\bar{A}_1 \cap \cdots \cap \bar{A}_n$ is the set of elements of U satisfying none of the conditions $c_1, \ldots, c_n$. The set $A_{i_1} \cap \cdots \cap A_{i_k}$ consists of the elements of U satisfying the conditions $c_{i_1}, \ldots, c_{i_k}$. On the one hand by Proposition 4.16, we have

$$\begin{aligned}
1_{\bar{A}_1 \cap \cdots \cap \bar{A}_n} &= 1_{\bar{A}} \cdots 1_{\bar{A}_n} \\
&= (1_U - 1_{A_1}) \cdots (1_U - 1_{A_1}) \\
&= \sum f_1 f_2 \cdots f_n \text{ (each } f_i \text{ is either } 1_U \text{ or } -1_{A_i}) \\
&= \underbrace{1_U \cdots 1_U}_{n} + \sum_{\substack{1 \le i_1 < \cdots < i_k \le n \\ 1 \le k \le n}} \underbrace{1_U \cdots 1_U}_{n-k} (-1_{A_{i_1}}) \cdots (-1_{A_{i_k}}) \\
&= 1_U + \sum_{k=1}^{n} (-1)^k \sum_{1 \le i_1 < \cdots < i_k \le n} 1_{A_{i_1}} \cdots 1_{A_{i_k}} \\
&= 1_U + \sum_{k=1}^{n} (-1)^k \sum_{1 \le i_1 < \cdots < i_k \le n} 1_{A_{i_1} \cap \cdots \cap A_{i_k}}.
\end{aligned}$$

On the other hand,

$$1_{\bar{A}_1 \cap \cdots \cap \bar{A}_n} = 1_{\overline{A_1 \cup \cdots \cup A_n}} = 1_U - 1_{A_\cup \cdots \cup A_n}.$$

Then

$$1_{A_\cup \cdots \cup A_n} = \sum_{k=1}^{n} (-1)^{k-1} \sum_{1 \le i_1 < \cdots < i_k \le n} 1_{A_{i_1} \cap \cdots \cap A_{i_k}}.$$

Thus

$$\begin{aligned}
|A_1 \cup \cdots \cup A_n| &= |1_{A_1 \cup \cdots \cup A_n}| \\
&= \sum_{k=1}^{n} (-1)^{k-1} \sum_{1 \le i_1 < \cdots < i_k \le n} |1_{A_{i_1} \cap \cdots \cap A_{i_k}}| \\
&= \sum_{k=1}^{n} (-1)^{k-1} \sum_{1 \le i_1 < \cdots < i_k \le n} |A_{i_1} \cap \cdots \cap A_{i_k}|.
\end{aligned}$$

4.9. More Examples

Example Let A and B be finite sets, $|A| = m$ and $|B| = n$. Let $C(m, n)$ be the number of surjective functions from A to B. What is $C(m, n)$?

Solution. Let U be the set of all functions from A to B, and let $B = \{b_1, \ldots, b_n\}$. We call a function $f : A \to B$ to satisfy condition c_i if b_i is not contained in $f(A)$. Then $N_0 (= n^m)$ is the number of functions from A to B; $\bar{N}$ is the number of surjective functions from A to B; and $N(c_{i_1} \cdots c_{i_k})$ is the number of functions f from A to B such that $\{b_{i_1}, \ldots, b_{i_k}\}$ is not contained in $f(A)$. Note that each function $f : A \to B$ such that $\{b_1, \dots, b_{i_k}\} \not\subset f(A)$ can be identified as a function from A to $B \backslash \{b_{i_1}, \dots, b_{i_k}\}$. Thus

$$N(c_{i_1} \cdots c_{i_k}) = (n-k)^m$$

and

$$N_k = \binom{n}{k} (n-k)^m, \quad 0 \le k \le n.$$

Therefore

$$C(m,n) = \bar{N} = \sum_{k=0}^{n} (-1)^k \binom{n}{k} (n-k)^m.$$

When $m < n$, obviously $C(m,n) = 0$. We then have

$$\sum_{k=0}^{n} (-1)^k \binom{n}{k} (n-k)^m = 0.$$

Note that $C(m,n)$ can be interpreted as the number of ways to place the elements of A into n distinct boxes so that no one box is empty. We have

$$C(m,n) = \sum_{\substack{i_1+\cdots+in=m \\ i_1,\dots,in \ge 1}} \binom{m}{i_1, \dots, i_n}.$$

We obtain the identity

$$\sum_{\substack{i_1+\cdots+in=m \\ i_1,\dots,in \ge 1}} \binom{m}{i_1, \dots, i_n} = \sum_{k=0}^{n} (-1)^k \binom{n}{k} (n-k)^m.$$

THEOREM 4.17. *Let $\phi(n)$ be the number of integers x in $[1,n]$ that are coprime with n. Then for any positive integer $n = p_1^{e_1} \cdots p_r^{e_r}$, where $p_1, \dots, p_r$ are distinct primes and $e_1, \dots, e_r$ are positive integers, the Euler function $\phi(n)$ is given by*

$$\phi(n) = n \prod_{k=1}^{r} \left(1 - \frac{1}{p_k}\right).$$

Proof. Let $U = \{1, 2, \dots, n\}$. For the distinct primes $p_1, \dots, p_r$, we define conditions $c_1, \dots, c_r$. An element $x \in U$ is called to satisfy condition c_i if p_i divides x, $1 \le i \le k$. Let A_i be the set of elements of U satisfying the condition c_i. Then

$$A_i = \left\{p_i, 2p_i, 3p_i, \dots, \left(\frac{n}{p_i}\right) p_i\right\}, \quad 1 \le i \le r.$$

For $1 \le i_1 < \cdots < i_k \le r$, let $q = p_{i_1} \cdots p_{i_k}$, then

$$A_{i_1} \cap \cdots \cap A_{i_k} = \left\{q, 2q, 3q \dots, \left(\frac{n}{q}\right) q\right\}.$$

Thus $|A_{i_1} \cap \cdots \cap A_{i_k}| = \frac{n}{p_{i_1} \cdots p_{i_k}}$ for all $1 \leq i_1 < \cdots < i_k \leq r$. Therefore

$$\begin{aligned}
\phi(n) &= n + \sum_{k=1}^{r} (-1)^k \sum_{1 \leq i_1 < \cdots < i_k \leq n} |A_{i_1} \cap \cdots \cap A_{i_k}| \\
&= n + \sum_{k=1}^{r} (-1)^k \sum_{1 \leq i_1 < \cdots < i_k \leq n} \frac{n}{p_{i_1} \cdots p_{i_k}} \\
&= n \left[1 - \left(\frac{1}{p_1} + \cdots + \frac{1}{p_r} \right) \right. \\
&\quad + \left(\frac{1}{p_1 p_2} + \frac{1}{p_1 p_3} + \cdots + \frac{1}{p_1 p_r} + \cdots + \frac{1}{p_{r-1} p_r} \right) \\
&\quad - \left(\frac{1}{p_1 p_2 p_3} + \frac{1}{p_1 p_2 p_4} + \cdots + \frac{1}{p_1 p_2 p_r} + \cdots + \frac{1}{p_{r-2} p_{r-1} p_r} \right) \\
&\quad \left. + \cdots + (-1)^r \frac{1}{p_1 p_2 \cdots p_r} \right] \\
&= n \prod_{k=1}^{r} \left(1 - \frac{1}{p_k} \right).
\end{aligned}$$

Example For $n = 36 = 2^2 3^2$,

$$\phi(36) = 36 \left(1 - \frac{1}{2} \right) \left(1 - \frac{1}{3} \right) = 12.$$

In fact, the integers from 1 to 36 that are coprime with 36 can be listed as

$$1,\ 5,\ 7,\ 11,\ 13,\ 17,\ 19,\ 23,\ 25,\ 29,\ 31,\ 35.$$

Example A permutation of $\{1, 2, \ldots, n\}$ is called a **dearrangement** if every i $(1 \leq i \leq n)$ is not placed at the ith position. We call an arrangement of $\{1, 2, \ldots, n\}$ to satisfy condition c_i if i is placed at the ith position. Let d_n be the number of dearrangements of $\{1, 2, \ldots, n\}$. Then

$$\begin{aligned}
d_n &= N(\bar{c}_1 \bar{c}_2 \cdots \bar{c}_n) \\
&= n! + \sum_{k=1}^{n} (-1)^k \sum_{1 \leq i_1 < \cdots < i_k \leq n} (n-k)! \\
&= \sum_{k=0}^{n} (-1)^k \binom{n}{k} (n-k)! \\
&= n! \sum_{k=1}^{n} (-1)^k \frac{1}{k!} \simeq n! e^{-1}.
\end{aligned}$$

4.10. Generalized Inclusion-Exclusion Formula

THEOREM 4.18. *Let U be a finite set and $|U| = N$. Let $c_1, \ldots, c_n$ be some properties about the elements of U. Let E_k denote the number of elements of U satisfying exactly m properties of $c_1, \ldots, c_n$, then*

(4.5)

$$E_m = \binom{m}{0} N_m - \binom{m+1}{1} N_{m+1} + \binom{m+2}{2} N_{m+2} - \cdots + (-1)^{n-m} \binom{n}{n-m} N_n.$$

Proof. For each $x \in U$ we count the contribution of x on both sides of (4.5) and divide the situation into three cases.

Case I: the element x satisfies fewer than m conditions. In this case the contributions of x on both sides are 0.

Case II: the element x satisfies exactly m conditions, say, $c_{i_1}, \cdots, c_{i_m}$. In this case the contribution of x on the left side is 1; and the contribution of x on the right side is also 1 becasue x is counted once in N_m and 0 times in N_k for all $k > m$.

Case III: the element x satisfies exactly r $(> m)$ conditions, say, $c_{i_1}, \cdots, c_{i_r}$. Then the contribution of x on the left side is 0. On the right side the contributions of x for N_m, N_{m+1}, …, N_r are $\binom{r}{m}$, $\binom{r}{m+1}$, …, $\binom{r}{r}$ respectively; and the contributions of x for N_k with $k > r$ are all 0. So the contribution of x on the right side is

$$\binom{m}{0}\binom{r}{m} - \binom{m+1}{1}\binom{r}{m+1} + \cdots + (-1)^{r-m}\binom{r}{r-m}\binom{r}{r}.$$

Now it is easy to see that

$$\begin{aligned}
\sum_{i=0}^{r-m}(-1)^i\binom{m+i}{i}\binom{r}{m+i} &= \sum_{i=0}^{r-m}(-1)^i\frac{(m+i)!}{i!m!}\cdot\frac{r!}{(m+i)!(r-m-i)!} \\
&= \sum_{i=0}^{r-m}(-1)^i\frac{r!}{m!(r-m)!}\cdot\frac{(r-m)!}{i!(r-m-i)!} \\
&= \sum_{i=0}^{r-m}(-1)^i\binom{r}{m}\binom{r-m}{i} \\
&= \binom{r}{m}\sum_{i=0}^{r-m}(-1)^i\binom{r-m}{i} \\
&= \binom{r}{m}[(-1)+1]^{r-m} = 0.
\end{aligned}$$

Thus the contributions of x on both sides are the same. □

THEOREM 4.19. *For any positive integer n,*

$$\sum_{d|n}\phi(d) = n.$$

Proof. Let S be the set of pairs (d, k) of integers such that

$$d|n, \quad 1 \le k \le d, \quad \gcd(k, d) = 1.$$

Then S is a subset of the grid $U = \{1, 2, \ldots, n\} \times \{1, 2, \ldots, n\}$. The characteristic function of S can be viewed as an n by n matrix whose (d, k) entry is 1 if $(d, k) \in S$ and 0 if $(d, k) \notin S$. For each fixed d such that $d|n$, the number of 1s in the dth row is $\phi(d)$. Thus

$$|S| = \sum_{d|n}\phi(d).$$

Now it suffices to show that $|S| = n$. To this end we construct a bijection between S and $\{1, 2, \ldots, n\}$. Let $f : S \to \{1, 2, \ldots, n\}$ be defined by

$$f(d, k) = kn/d.$$

Since $1 \le k \le d$ and $d|n$, we have $f(d, k) = kn/d \le n$; that is, f is well-defined.

For $(d', k') \neq (d, k)$, if $f(d, k) = f(d', k')$, that is, $kn/d = k'n/d'$, then $kd' = k'd$. Since $\gcd(d, k) = 1$ and $\gcd(d', k') = 1$, we have $d|d'$, $d'|d$, $k|k'$, and $k'|k$. Then $d = d'$ and $k = k'$, that is, f is injective.

For any $1 \leq m \leq n$, let $g_m = \gcd(m, n)$, $d_m = n/g_m$, $k_m = m/g_m$. Then $d_m|n$, $k_m \leq d_m$, and $\gcd(d_m, k_m) = 1$, that is, $(d_m, k_m) \in S$. Now we have $f(d_m, k_m) = k_m n/d_m = m$, that is, f is surjective. □

Exercises

(1) In how many ways to arrange the letters E, I, M, O, T, U, Y so that YOU, ME and IT would not occur?

(2) Six passengers on a msall airplane are randomly assigned to the six seats on the plane. On the return trip they are again randomly assigned seats.
  (a) What is the chance that every passenger has the same seats on both trips?
  (b) What is the probability that exactly five passengers have the same seats ob both trips?
  (c) What is the probability that at least one passenger has the seat on both trips?

(3) Show that for any positive integer n,
$$\phi(n) = \sum_{d|n} \mu(d) \left(\frac{n}{d}\right),$$
where μ is the **Möbius function** defined by
$$\mu(d) = \begin{cases} 1 & \text{if } d = 1, \\ (-1)^k & \text{if } d \text{ is a product of } k \text{ distinct primes,} \\ 0 & \text{if } d \text{ has a repeated prime factor.} \end{cases}$$

(4) There are n couples to be arranged in a straightline. Find the number of ways to arrange them so that
  (a) there is no couple to be next to each other.
  (b) men and women alternate and there is no couple next to each other.

(5) There are n couples to be seated at a round table. Find the number of seating plans so that
  (a) there is no couple to be next to each other.
  (b) men and women alternate and there is no couple next to each other.

(6) There are n persons seated in a circle. Find the numbers of ways to be re-seated so that no k consecutive persons in the original seating order appear.

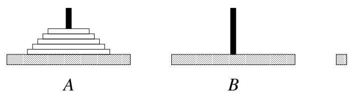
CHAPTER 5

Recurrence Relations

5.1. Infinite Sequences

DEFINITION 5.1. An infinite sequence is a function from the set of positive integers to the set of real or complex numbers.

Example The game of **Hanoi Tower** is to play with a set of disks of graduated size with holes in their centers and a playing board having three spokes for holding the disks; see Figure **??**. The object of the game is to transfer all the disks from spoke A to spoke C by moving one disk at a time without placing a larger disk on top of a smaller one. What is the minimal number of moves required when there are n disks?



Solution. Let a_n be the minimum number of moves to transfer n disks from one spoke to another. Then $\{a_n \mid n \geq 1\}$ defines an **infinite sequence**. The first few terms of the sequence $\{a_n\}$ can be listed as

$$1,\ 3,\ 7,\ 15,\ \ldots$$

We are interested in finding a closed formula to compute a_n for arbitrary n.

In order to to move n disks from spoke A to spoke C, one must move the first $n-1$ disks from spoke A to spoke B by a_{n-1} moves, then move the last (also the largest) disk from spoke A to spoke C by one move, and then remove the $n-1$ disks again from spoke B to spoke C by a_{n-1} moves. Thus the total number of moves should be

$$a_n = a_{n-1} + 1 + a_{n-1} = 2a_{n-1} + 1.$$

This means that the sequence $\{a_n \mid n \geq 1\}$ satisfies the recurrence relation

$$\left\{ \begin{array}{lcl} a_n & = & 2a_{n-1} + 1, \quad n \geq 1 \\ a_1 & = & 1. \end{array} \right. \tag{5.1}$$

Given a recurrence relation for a sequence with initial conditions, solving the recurrence relation means to find a formula to express the general term a_n of the sequence.

For the sequence $\{a_n \mid n \geq 0\}$ defined by the recurrence relation (5.1), if we apply the recurrence relation again and again, we have

$$\begin{aligned}
a_1 &= 2a_0 + 1 \\
a_2 &= 2a_1 + 1 = 2(2a_0 + 1) + 1 = 2^2 a_0 + 2 + 1 \\
a_3 &= 2a_2 + 1 = 2(2^2 a_0 + 2 + 1) = 2^3 a_0 + 2^2 + 2 + 1 \\
a_4 &= 2a_3 + 1 = 2(2^3 a_0 + 2^2 + 2 + 1) = 2^4 a_0 + 2^3 + 2^2 + 2 + 1 \\
&\vdots \\
a_n &= 2^n a_0 + 2^{n-1} + 2^{n-2} + \cdots + 2 + 1 = 2^n a_0 + 2^n - 1.
\end{aligned}$$

Let $a_0 = 0$. The general term is given by

$$a_n = 2^n - 1, \ \ n \geq 1.$$

5.2. Homogeneous Recurrence Relations

Any recurrence relation of the form

$$x_n = a x_{n-1} + b x_{n-2} \tag{5.2}$$

is called a **second order homogeneous linear recurrence relation**.

Let $x_n = s_n$ and $x_n = t_n$ be two solutions, i.e.,

$$s_n = a s_{n-1} + b s_{n-2} \quad \text{and} \quad t_n = a t_{n-1} + b t_{n-2}.$$

Then for constants c_1 and c_2

$$\begin{aligned}
c_1 s_n + c_2 t_n &= c_1(a s_{n-1} + b s_{n-2}) + c_2(a t_{n-1} + b t_{n-2}) \\
&= a(c_1 s_{n-1} + c_2 t_{n-1}) + b(c_1 s_{n-2} + c_2 t_{n-2}).
\end{aligned}$$

This means that $x_n = c_1 s_n + c_2 t_n$ is a solution of (5.2).

THEOREM 5.2. *Any linear combination of solutions of a homogeneous recurrence linear relation is also a solution.*

In solving the first order homogeneous recurrence linear relation $x_n = a x_{n-1}$, it is clear that the general solution is $x_n = a^n a_0$. This means that $x_n = a^n$ is a solution. This suggests that, for the second order homogeneous recurrence linear relation (5.2), we may have the solutions of the form

$$x_n = r^n.$$

Indeed, put $x_n = r^n$ into (5.2); we have

$$r^n = a r^{n-1} + b r^{n-2} \quad \text{or} \quad r^{n-2}(r^2 - ar - b) = 0.$$

Thus either $r = 0$ or

$$r^2 - ar - b = 0. \tag{5.3}$$

The equation (5.3) is called the **characteristic equation** of (5.2).

THEOREM 5.3. *If the characteristic equation* (5.3) *has two distinct roots r_1 and r_2, then the general solution for* (5.2) *is given by*

$$x_n = c_1 r_1^n + c_2 r_2^n.$$

If the characteristic equation (5.3) *has only one root r, then the general solution for* (5.2) *is given by*

$$x_n = c_1 r^n + c_2 n r^n.$$

Proof. When the characteristic equation (5.3) has two distinct roots r_1 and r_2 it is clear that both $x_n = r_1^n$ and $x_n = r_2^n$ are solutions of (5.2).

Now assume that (5.2) has only one root r. Then $a^2 + 4b = 0$ and $r = a/2$, i.e.,

$$b = -\frac{a^2}{4} \quad \text{and} \quad r = \frac{a}{2}.$$

We verify that $x_n = nr^n$ is a solution of (5.2). In fact,

$$ax_{n-1} + bx_{n-2} = a(n-1)\left(\frac{a}{2}\right)^{n-1} + \left(-\frac{a^2}{4}\right)(n-2)\left(\frac{a}{2}\right)^{n-2} = n\left(\frac{a}{2}\right)^n = x_n.$$

□

Remark There is heuristic method to explain why $x_n = nr^n$ is a solution when the two roots are the same. If two roots r_1 and r_2 are distinct but very close to each other, then $r_1^n - r_2^n$ is a soltuion. So is $(r_1^n - r_2^n)/(r_1 - r_2)$. It follows that the limit

$$\lim_{r_2 \to r_1} \frac{r_1^n - r_2^n}{r_1 - r_2} = nr_1^{n-1}$$

would be a solution. Thus its multiple $x_n = r_1(nr_1^{n-1}) = nr_1^n$ by the constant r_1 is also a solution. Please note that this is *not* a mathematical proof, but a mathematical idea.

Example Find a general formula for the **Fibonacci sequence**

$$\begin{cases} f_n &= f_{n-1} + f_{n-2} \\ f_0 &= 0 \\ f_1 &= 1 \end{cases}$$

Solution. The characteristic equation $r^2 = r + 1$ has two distinct roots

$$r_1 = \frac{1+\sqrt{5}}{2} \quad \text{and} \quad r_2 = \frac{1-\sqrt{5}}{2}.$$

Then the general solutions is

$$f_n = c_1\left(\frac{1+\sqrt{5}}{2}\right)^n + c_2\left(\frac{1-\sqrt{5}}{2}\right)^n.$$

Set

$$\begin{cases} 0 = c_1 + c_2 \\ 1 = c_1\left(\frac{1+\sqrt{5}}{2}\right) + c_2\left(\frac{1-\sqrt{5}}{2}\right). \end{cases}$$

We have $c_1 = -c_2 = \frac{1}{\sqrt{5}}$. Thus

$$f_n = \frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^n - \frac{1}{\sqrt{5}}\left(\frac{1-\sqrt{5}}{2}\right)^n, \quad n \geq 0.$$

Remark The Fibonacci sequence f_n is an integer sequence, but it "looks like" a sequence of irrational numbers from its general formula above.

Example Find the solution for the recurrence relation

$$\begin{cases} x_n &= 6x_{n-1} - 9x_{n-2} \\ x_0 &= 2 \\ x_1 &= 3 \end{cases}$$

Solution. The characteristic equation

$$r^2 - 6r + 9 = 0 \Leftrightarrow (r-3)^2 = 0$$

has only one root $r = 3$. Then the general solution is

$$x_n = c_1 3^n + c_2 n 3^n.$$

The initial conditions $x_0 = 2$ and $x_1 = 3$ imply that $c_1 = 2$ and $c_2 = -1$. Thus the solution is

$$x_n = 2 \cdot 3^n - n \cdot 3^n = (2-n)3^n, \ \ n \geq 0.$$

Example Find the solution for the recurrence relation

$$\begin{cases} x_n &= 2x_{n-1} - 5x_{n-2}, \quad n \geq 2 \\ x_0 &= 1 \\ x_1 &= 5 \end{cases}$$

Solution. The characteristic equation

$$r^2 - 2r + 5 = 0 \Leftrightarrow (x - 1 - 2i)(x - 1 + 2i) = 0$$

has two distinct complex roots $r_1 = 1 + 2i$ and $r_2 = 1 - 2i$. The initial conditions imply that

$$c_1 + c_2 = 1 \quad c_1(1+2i) + c_2(1-2i) = 5.$$

So $c_1 = \frac{1-2i}{2}$ and $c_2 = \frac{1+2i}{2}$. Thus the solutions is

$$\begin{aligned} x_n &= \frac{1-2i}{2} \cdot (1+2i)^n + \frac{1+2i}{2} \cdot (1-2i)^n \\ &= \frac{5}{2}(1+2i)^{n+1} + \frac{5}{2}(1-2i)^{n+1}, \ \ n \geq 0. \end{aligned}$$

Remark The sequence is obviously a real sequence. However, its general formula involves complex numbers.

Example Two persons A and B gamble dollars on the toss of a fair coin; A has \$70 and B has \$30. In each play either A wins \$1 from B or loss \$1 to B; and the game is played without stop until one wins all the money of the other or goes forever. Find the following probabilities.

(a) A wins all the money of B.
(b) A loss all his moeny to B.
(c) The game continues forever.

Solution. Either A or B can keep track of the game simply by counting their own money; their position (money) can be one of the numbers $0, 1, 2, \ldots, 100$. Let p_n be the probability that A reaches 100 at position n. After one toss, A enters into either position $n+1$ or position $n-1$. The new probability that A reaches 100 is either p_{n+1} or p_{n-1}. Since the probability of A moving to position $n+1$ or $n-1$ from n is $\frac{1}{2}$. We thus have the recurrence relation

$$\begin{cases} p_n &= \frac{1}{2}p_{n+1} + \frac{1}{2}p_{n-1} \\ p_0 &= 0 \\ p_{100} &= 1 \end{cases}$$

The characteristic equation is $r^2 - 2r + 1 = 0$; it has only one root $r = 1$. The general solutions is

$$p_n = c_1 + c_2 n.$$

Apply the boundary conditions $p_0 = 0$ and $p_{100} = 1$; we have $c_1 = 0$ and $c_2 = \frac{1}{100}$. Thus

$$p_n = \frac{n}{100}, \quad 0 \le n \le 100.$$

Of course, $p_n = \frac{n}{100}$ for $n > 100$ is nonsense to the original problem. The probabilities for (a), (b), and (c) are 70%, 30%, and 0, respectively.

The recurrence relation $p_n = \frac{1}{2}p_{n+1} + \frac{1}{2}p_{n-1}$ can be directly solved. In fact, the recurrence relation can be simplied to

$$p_{n+1} - p_n = p_n - p_{n-1}.$$

Then $p_{n+1} - p_n = p_n - p_{n-1} = \cdots = p_1 - p_0$. Since $p_0 = 0$, we have $p_n = p_{n-1} + p_1$. Apply the recurrence relation again and again, we obtain

$$p_n = p_0 + np_1.$$

Applying the conditions $p_0 = 0$ and $p_{100} = 1$, we have $p_n = \frac{n}{100}$.

5.3. Higher Order Homogeneous Recurrence Relations

For the **higher order homogeneous recurrence relation**

$$x_{n+k} = a_1 x_{n+k-1} + a_2 x_{n+k-2} + \cdots + a_{n-k} x_n, \quad n \ge 0 \tag{5.4}$$

the **characteristic equation** is

$$r^k = a_1 r^{k-1} + a_2 r^{k-1} + \cdots + a_{n-k+1} r + a_{n-k} \tag{5.5}$$

or

$$r^k - a_1 r^{k-1} - a_2 r^{k-1} - \cdots - a_{n-k+1} r - a_{n-k} = 0.$$

Theorem 5.4. *For the recurrence relation* (5.4)*, if its characteristic equation* (5.5) *has distinct roots* $r_1, r_2, \ldots, r_k$*, then the general solution for* (5.4) *is*

$$x_n = c_1 r_1^n + c_2 r_2^n + \cdots + c_k r_k^n$$

where $c_1, c_2, \ldots, c_k$ *are arbitrary constants. If the characteristic equation has repeated roots* $r_1, r_2, \ldots, r_s$ *of multiplicities* $m_1, m_2, \ldots, m_s$ *respectively, then the general solution of* (5.4) *is a linear combination of the solutions*

$$\begin{array}{llll} r_1^n, & nr_1^n, & \ldots, & n^{m_1-1} r_1^n; \\ r_2^n, & nr_2^n, & \ldots, & n^{m_2-1} r_2^n; \\ \ldots; & & & \\ r_s^n, & nr_s^n, & \ldots, & n^{m_s-1} r_s^n. \end{array}$$

Example Find an explicit formula for the sequence given by the recurrence relation

$$\begin{cases} x_n = 15x_{n-2} - 10x_{n-3} - 60x_{n-4} + 72x_{n-5} \\ x_0 = 1, \ x_1 = 6, \ x_2 = 9, \ x_3 = -110, \ x_4 = -45 \end{cases}$$

Solution. The characteristc equation $r^5 = 15r^3 - 10r^2 - 60r + 72$ can be simpliefied as

$$(r-2)^3(r+3)^2 = 0.$$

So there are two roots $r_1 = 2$ with multiplicity 3 and $r_2 = -3$ with multiplicity 2. Then the general solution is given by

$$x_n = c_1 2^n + c_2 n 2^n + c_3 n^2 2^n + c_4(-3)^n + c_5 n(-3)^n.$$

The initial condition means that

$$\begin{cases} c_1 & & & +c_4 & & = & 1 \\ 2c_1 & +2c_2 & +2c_3 & -3c_4 & -3c_5 & = & 1 \\ 4c_1 & +8c_2 & +16c_3 & +9c_4 & +18c_5 & = & 1 \\ 8c_1 & +24c_2 & +72c_3 & -27c_4 & -81c_5 & = & 1 \\ 16c_1 & +64c_2 & +256c_3 & +81c_4 & +324c_5 & = & 1 \end{cases}$$

Solving the linear system we have

$$c_1 = 2,\ c_2 = 3,\ c_3 = -2,\ c_4 = -1,\ c_5 = 1.$$

5.4. Non-homogeneous Equations

A recurrence relation of the form

$$x_n = ax_{n-1} + bx_{n-2} + f(n) \tag{5.6}$$

is called a **non-homogeneous recurrence relation**. Let $x_n^{(s)}$ be a solution of (5.6), called a **special solution**, then the general solution for (5.6) is

$$x_n = x_n^{(s)} + x_n^{(h)}, \tag{5.7}$$

where $x_n^{(h)}$ is the general solution for the corresponding homogeneous equation

$$x_n = ax_{n-1} + bx_{n-2}. \tag{5.8}$$

THEOREM 5.5. *Let $f(n) = cr^n$ for the non-homogeneous recurrence relation (5.6). Let r_1 and r_2 be the roots of the characteristic equation of the corresponding homogeneous equation (5.8).*

(a) If $r \neq r_1$, $r \neq r_2$, then $x_n^{(s)}$ can be of the form $x_n^{(s)} = Ar^n$.
(b) If $r = r_1$, $r_1 \neq r_2$, then $x_n^{(s)}$ can be of the form $x_n^{(s)} = Anr^n$.
(c) If $r = r_1 = r_2$, then $x_n^{(s)}$ can be of the form $x_n^{(s)} = An^2r^n$.

where A is a constant to be determined in all cases.

Proof. (a) Put $x_n = Ar^n$ into the recurrence relation (5.6), we have

$$Ar^n = aAr^{n-1} + bAr^{n-2} + cr^n.$$

We may assume $r \neq 0$; otherwise the recurrence relation is homogeneous. Then

$$A(r^2 - ar - b) = cr^2.$$

Note that $r^2 - ar - b \neq 0$ because r is not a root of the characteristc equation. Thus when

$$A = \frac{cr^2}{r^2 - ar - b}$$

the sequence $x_n = Ar^n$ is a special solution.

(b) Since $r = r_1 \neq r_2$, it is clear that $x_n = nr^n$ is not a solution for its corresponding homogeneous equation, i.e.,

$$nr^2 - a(n-1)r - b(n-2) = n(r - ar - b) + ar + 2b = ar + 2b \neq 0.$$

Put $x_n = Anr^n$ into (5.6); we have

$$Anr^n = aA(n-1)r^{n-1} + bA(n-2)r^{n-2} + cr^n,$$

If $r \neq 0$, we then have $A(nr^2 - a(n-1)r - b(n-2)) = cr^2$. So

$$A = \frac{cr^2}{ar + 2b}$$

the sequence $x_n = Anr^n$ is a special solution.

(c) Since $r = r_1 = r_2$, it follows that $x_n = n^2r^n$ is not a solution of the corresponding homogeneous equation, i.e.,

$$n^2r^2 - a(n-1)^2r - b(n-2)^2 = n^2(r^2 - ar - b) + 2n(ar + 2b) - a - 4b = -a - 4b \neq 0.$$

Put $x_n = An^2r^n$ into (5.6); we have

$$Ar^{n-2}\left(n^2r^2 - a(n-1)^2r - b(n-2)^2\right) = cr^n.$$

If $r \neq 0$, we then have

$$A = -\frac{cr^2}{a + 4b}$$

and the sequence $x_n = An^2r^n$ is a non-homogeneous solution. □

Example Consider the non-homogeneous equation

$$\begin{cases} x_n &= 3x_{n-1} + 10x_{n-2} + 7 \cdot 5^n \\ x_0 &= 4 \\ x_1 &= 3 \end{cases}$$

Solution. The characteristic equation $r^2 - 3r - 10 = 0$ has roots $r_1 = 5$ and $r_2 = -2$. A special solution can be given by

$$x_n = \frac{cr^2}{3a + 2b} n5^n = n5^{n+1}$$

and the general solution is

$$x_n = n5^{n+1} + c_1 5^n + c_2(-2)^n.$$

The initial condition implies $c_1 = -2$ and $c_2 = 6$. Thus

$$x_n = n5^{n+1} - 2 \cdot 5^n + 6(-2)^n.$$

5.5. Divide-and-Conquer Method

Suppose we have a job of size n to be done. If the size n is large and the job is complicated, we may divide the job into smaller jobs of the same type and of the same size, then conquer the smaller problems and use the results to construct a solution for the original problem of size n. This is the nature of the so-called **Divide-and-Conquer** method.

Example Suppose there are n $(= 2^k)$ student files, indexed by the student id numbers as

$$A = \{a_1, a_2, \ldots, a_n\}.$$

Given a particular file $a \in A$, what is the number of comparisons needed in worst case to find the position of the file a?

Solution. Let x_n be the number of comparisons needed to find the position of the file a in worst case. Then the answer depends on whether or not the files are sorted.

Case I: The files in A are not sorted. Then the answer is at most n comparisons.

Case II: The files in A are sorted in the order $a_1 < a_2 < \cdots < a_n$.

a_1	a_2	$\cdots$	$a_{\frac{n}{2}-1}$	$a_{\frac{n}{2}}$	$a_{\frac{n}{2}+1}$	$\cdots$	a_{n-1}	a_n

We may compare the file a with $a_{\frac{n}{2}}$. If $a = a_{\frac{n}{2}}$, the job is done by one comparison. If $a < a_{\frac{n}{2}}$, consider the subset $\{a_1, a_2, \ldots, a_{\frac{n}{2}}\}$. If $a > a_{\frac{n}{2}}$, consider the subset

$\{a_{\frac{n}{2}+1}, a_{\frac{n}{2}+2}, \ldots, a_n\}$. Then the number of comparisons is at most $x_{\frac{n}{2}} + 1$. We thus obtain a recurrence relation

$$\begin{cases} x_n &= x_{\frac{n}{2}} + 1 \\ x_1 &= 1 \end{cases}$$

Applying the recurrence relation again and again, we obtain

$$x_n = x_{\frac{n}{2}} + 1 = x_{\frac{n}{2^2}} + 2 = x_{\frac{n}{2^3}} + 3 = \cdots = x_{\frac{n}{2^k}} + k = x_1 + k = k + 1.$$

Since $n = 2^k$, we have $k = \log_2 n$. Therefore

$$x_n = \log_2 n + 1.$$

Example Let $S = \{a_1, a_2, \ldots, a_n\} \subset \mathbf{Z}$, where $n = 2^k$ and $k \geq 1$. How many number of comparisons are needed in worst case to find the minimum in S? We assume that the numbers in S are not sorted.

Solution. The number of comparisons depends on the method we employed. If all possible pairs of elements in S are compared, then the minimum element will be found, and the number of comparisons in worst case is $\binom{n}{2} = \frac{n(n-1)}{2} = O(n^2)$. Of course this is not best possible.

There is another method to find a better solution. Let x_n be the minimal number of comparisons needed in worst case to find the minimum in S. Obviously, $x_n = 1$. For $n = 2^k$ and $k \geq 1$, we may divide S into two subsets

$$\begin{array}{ll} S_1 = \{a_1, a_2, \ldots, a_{\frac{n}{2}}\}, & |S_1| = \frac{n}{2}, \\ S_2 = \{a_{\frac{n}{2}+1}, a_{\frac{n}{2}+2}, \ldots, a_n\}, & |S_2| = \frac{n}{2}. \end{array}$$

It takes $x_{\frac{n}{2}}$ comparisons to find the minimum m_1 for S_1 and the minimum m_2 for S_2. Then compare m_1 with m_2 to determine the minimum element in S. In this way the total number of comparisons for S in worst case is $2x_{\frac{n}{2}} + 1$. We thus obtain a recurrence relation

$$\begin{cases} x_n &= 2x_{\frac{n}{2}} + 1 \\ x_2 &= 1 \end{cases}$$

Apply the recurrence relation again and again; we have

$$\begin{aligned} x_n &= 2\left(2x_{\frac{n}{2^2}} + 1\right) = 2^2 x_{\frac{n}{2^2}} + 2 + 1 \\ &= 2^2\left(2x_{\frac{n}{2^3}} + 1\right) + 2 + 1 = 2^3 x_{\frac{n}{2^3}} + 2^2 + 2 + 1 \\ &= \cdots = 2^k x_{\frac{n}{2^k}} + 2^{k-1} + \cdots + 2 + 1 \\ &= 2^{k-1} + \cdots + 2 + 1 = \frac{2^k - 1}{2 - 1} \\ &= n - 1 = O(n). \end{aligned}$$

□

We hope that we understand the nature of divide-and-conquer method by the above examples. In order to solve a problem of size n, if the size n is large and the problem is complicated, we divide the problem into a smaller subproblems of the same type and of the same size $\lceil \frac{n}{b} \rceil$, where $a, b \in \mathbf{Z}_+$, $1 \leq a < n$ and $1 < b < n$. Then we solve the a smaller subproblems and use the results to construct a solution for the original problem of size n. We are especially interested in the case where $n = b^k$ and $b = 2$.

THEOREM 5.6. (Divide-and-Conquer Algorithm)

(a) *The time to solve the initial problem of size $n = 1$ is a constant $c \geq 0$.*
(b) *The time to break the given problem of size n into a smaller same type subproblems, together with the time to construct a solution for the original problem by using the solutions for the a subproblems, is a function $h(n)$.*

Our concern here is to figure out the time complexity function $f(n)$ which is given by the recurrence relation

$$\begin{cases} f(1) = c \\ f(n) = af(\frac{n}{b}) + h(n), \quad n = b^k, k \geq 1 \end{cases}$$

THEOREM 5.7. *Let $f : \mathbf{Z}_+ \to \mathbf{R}$ be a function defined by the recurrence relation*

$$\begin{cases} f(1) = c \\ f(n) = af(\frac{n}{b}) + c, \quad n = b^k, k \geq 1 \end{cases}$$

where a, b, c are positive integers and $b \geq 2$. Then

$$\begin{cases} f(n) = c(\log_b n + 1) & \text{if} \quad a = 1 \\ f(n) = \frac{c(an^{\log_b a} - 1)}{a-1} & \text{if} \quad a \neq 1 \end{cases}$$

Proof. Applying the recurrence relation, we obtain

$$\begin{array}{lcl} f(n) & = & af(\frac{n}{b}) + c \\ af(\frac{n}{b}) & = & a^2 f(\frac{n}{b^2}) + ac \\ a^2 f(\frac{n}{b^2}) & = & a^3 f(\frac{n}{b^3}) + a^2 c \\ & \vdots & \\ a^{k-2} f(\frac{n}{b^{k-2}}) & = & a^{k-1} f(\frac{n}{b^{k-1}}) + a^{k-2} c \\ a^{k-1} f(\frac{n}{b^{k-1}}) & = & a^k f(\frac{n}{b^k}) + a^{k-1} c \end{array}$$

Adding both sides of the k equations and canceling the common summands, we have

$$f(n) = a^k f\left(\frac{n}{b^k}\right) + \left(c + ac + a^2 c + \cdots + a^{k-1} c\right).$$

Since $n = b^k$ and $f(1) = c$, we further have

$$f(n) = c\left(1 + a + a^2 + \cdots + a^k\right).$$

If $a = 1$, then $f(n) = c(k+1)$. Note that $n = b^k$ implies $k = \log_b n$. Hence

$$f(n) = c\left(\log_b n + 1\right).$$

If $a \neq 1$, then $f(n) = \frac{c(a^{k+1} - 1)}{a-1}$. Since $k = \log_b n$, it follows that

$$a^k = a^{\log_b n} = \left(b^{\log_b a}\right)^{\log_b n} = \left(b^{\log_b n}\right)^{\log_b a} = n^{\log_b a}.$$

Therefore

$$f(n) = \frac{c\left(an^{\log_b a} - 1\right)}{a-1}.$$

□

Exercises

(1) Find an explicit formua for each of the sequences defined by the recurrence relations with initial conditions.
   (a) $x_n = 5x_{n-1} + 3$, $x_1 = 3$.

(b) $x_n = 3x_{n-1} + 5n$, $x_1 = 5$.
(c) $x_n = 2x_{n-1} + 15x_{n-2}$, $x_1 = 2$, $x_2 = 4$.
(d) $x_n = 4x_{n-1} + 5x_{n-2}$, $x_1 = 3$, $x_2 = 5$.
(e) $x_n = 3x_{n-1} - 2x_{n-2}$, $x_0 = 2$, $x_1 = 4$.
(f) $x_n = 6x_{n-1} - 9x_{n-2}$, $x_0 = 3$, $x_1 = 9$.

(2) Show that if s_n and t_n are solutions for the non-homogeneous linear recurrence relation

$$x_n = ax_{n-1} + bx_{n-2} + f(n),\ n \geq 2,$$

then $x_n = s_n - t_n$ is a solution for the homogeneous linear recurrence relation

$$x_n = ax_{n-1} + bx_{n-2},\ n \geq 2.$$

(3) Let the characteristic equation for the homogeneous linear recurrence relation

$$x_n = ax_{n-1} + bx_{n-2},\ n \geq 2$$

have two distinct roots r_1 and r_2. Show that every solution can be written in the form

$$x_n = c_1 r_1^n + c_2 r_2^n$$

for some constants c_1 and c_2.

(4) * Let $A_1, A_2, \ldots, A_{n+1}$ be $k \times k$ matrices. Let C_n be the number of ways to evaluate the product $A_1 A_2 \cdots A_{n+1}$ by choosing different orders in which to do the n multiplications.
(a) Find a recurrence relation with an initial condition for the sequence C_n.
(b) Verify that the sequence $\frac{1}{n+1}\binom{2n}{n}$ satisfies your recurrence relation and conclude that $C_n = \frac{1}{n+1}\binom{2n}{n}$. (The numbers C_n are called **Catalan numbers**.)

(5) Find a general formula for the recurrence relation

$$x_n = ax_{n-1} + b + cn$$

in terms of x_0, where a, b, c are real constants.

(6) Find an explicit formua for each of the sequences defined by the recurrence relations with initial conditions.
(a) $x_n = 5x_{\frac{n}{3}} + 5$, $x_1 = 5$, $n = 2^k$, $k \geq 0$.
(b) $x_n = x_{\lfloor \frac{n}{2} \rfloor} + 3$, $x_1 = 4$, $n \geq 1$.
(c) $x_{2n} = 2x_n + 5 - 7n$, $x_1 = 0$.

(7) Let $f(n)$ be a real sequence defined for $n = 1, b, b^2, \ldots$, and satisfy the recurrence relation

$$f(n) = af\left(\frac{n}{b}\right) + h(n),$$

where $b \geq 2$ is an integer. Show that

$$f(n) = f(1)n^{\log_b a} + \sum_{i=0}^{-1+\log_b n - 1} a^i h\left(\frac{n}{b^i}\right).$$

(8) Let $f(n)$ be a real sequence defined for $n = 1, b, b^2, b^3, \ldots$, and satisfy the recurrence relation

$$f(n) = af\left(\frac{n}{2}\right) + a_0 + a_1 n + \cdots + a_k n^k,$$

where $a, b, a_0, a_1, \ldots, a_k$ are real constatnts, $a > 0$ and $b > 1$. Show that

(a) If $a = b^i$ for some $0 \le i \le k$, then

$$f(n) = f(1)n^i + a_i n^i \log_b n + \sum_{j=1, j\ne i}^{k} \frac{b^j a_j}{b^j - b^i}\left(n^j - n^i\right).$$

(b) If $a \ne b^i$ for all $0 \le i \le k$, then

$$f(n) = f(1)n^{\log_b a} + \sum_{j=0}^{k} \frac{b^j a_j}{b^j - a}\left(n^j - n^{\log_b a}\right).$$

Solution: For the recurrence relation

$$x_n = ax_{n-1} + b + cn$$

we have

$$\begin{aligned}
x_n &= ax_{n-1} + b + cn \\
&= a(ax_{n-2} + b + c(n-1)) + b + cn = a^2 x_{n-2} + b(a+1) + c(a(n-1) + n) \\
&= a^2(ax_{n-3} + b + c(n-2)) + b(a+1) + c(a(n-1) + n) \\
&= a^3 x_{n-3} + b(a^2 + a + 1) + c(a^2(n-2) + a(n-1) + n) \\
&\vdots \\
&= a^n x_0 + b(a^{n-1} + \cdots + a + 1) + c(a^{n-1} + 2a^{n-2} + \cdots + (n-1)a + n) \\
&= a^n x_0 + b(a^{n-1} + \cdots + a + 1) + c(a^{n-1} + \cdots + a + 1) \\
&\quad + c(a^{n-2} + \cdots + a + 1) + \cdots + c(a^2 + a + 1) + c(a+1) + c
\end{aligned}$$

If $a = 1$, then

$$x_n = x_0 + bn + \frac{cn(n+1)}{2}.$$

If $a \ne 1$, then

$$\begin{aligned}
x_n &= a^n x_0 + \frac{b(a^n - 1)}{a-1} + c\left(\frac{a^n - 1}{a-1} + \frac{a^{n-1} - 1}{a-1} + \cdots + \frac{a^2 - 1}{a-1} + \frac{a-1}{a-1}\right) \\
&= a^n x_0 + \frac{b(a^n - 1)}{a-1} + \frac{c}{a-1}\left(a^n + a^{n-1} \cdots + a - n\right) \\
&= a^n x_0 + \frac{b(a^n - 1)}{a-1} + \frac{c(a^{n+1} - 1)}{(a-1)^2} - \frac{c(n-1)}{a-1}.
\end{aligned}$$

Catland numbers We have $C_0 = 1$, $C_1 = 1$, $C_2 = 2$. The product $A_1 A_2 \cdots A_{n+1}$ can be obtained by the multiplication of two matrices in n ways, i.e.,

$$A_1 A_2 \cdots A_{n+1} = (A_1 \cdots A_k)(A_{k+1} \cdots A_{n+1}),\ 1 \le k \le n.$$

This yields the recurrence relation

$$C_n = \sum_{k=1}^{n} C_{k-1} C_{n-k} = \sum_{i+j=n-1} C_i C_j.$$

Consider the generating function

$$F(x) = \sum_{n=0}^{\infty} C_n x^n.$$

Note that

$$F(x)^2 = \sum_{n=0}^{\infty}\left(\sum_{i+j=n} C_iC_j\right)x^n = \sum_{n=0}^{\infty} C_{n+1}x^n = \frac{F(x)}{x} - \frac{1}{x}.$$

Then

$$xF(x)^2 - F(x) + 1 = 0.$$

We thus have

$$F(x) = \frac{1 \pm \sqrt{1-4x}}{2x}.$$

Since

$$\begin{aligned}
\sqrt{1-4x} &= \sum_{n=0}^{\infty}\binom{\frac{1}{2}}{n}(-4x)^n \\
&= 1 + \sum_{n=1}^{\infty}\frac{\frac{1}{2}\cdot(\frac{1}{2}-1)\cdots(\frac{1}{2}-n+1)}{n!}2^{2n}(-1)^n x^n \\
&= \sum_{n=0}^{\infty}\frac{(-1)(-3)(-5)\cdots(-2(n-1)+1)}{n!}2^n(-1)^n x^n \\
&= -\sum_{n=0}^{\infty}\frac{1\cdot 3\cdot 5\cdots(2(n-1)-1)}{n!}2^n x^n \\
&= 1 - 2\sum_{n=1}^{\infty}\frac{(2(n-1))!}{n!(n-1)!}x^n \\
&= 1 - 2\sum_{n=0}^{\infty}\frac{(2n)!}{n!(n+1)!}x^{n+1},
\end{aligned}$$

we conclude that

$$F(x) = \frac{1-\sqrt{1-4x}}{2x} = \sum_{n=0}^{\infty}\frac{(2n)!}{n!(n+1)!}x^n = \sum_{n=0}^{\infty}\frac{1}{n+1}\binom{2n}{n}x^n.$$

Therefore $C_n = \frac{1}{n+1}\binom{2n}{n}$.

Euler's Problem In how many different ways can a labeled convex n-gon be divided into triangles by non-intersecting diagonals?

Solution. Let c_n be number of ways for an $(n+2)$-gon in the problem. Then $c_1 = 1$, $c_2 = 2$, and $c_3 = 5$. Consider a convex $(n+3)$-gon $V_1V_2\cdots V_{n+3}$.

In each decomposition of the $(n+3)$-gon, the segment V_1V_{n+3} is a side of some triangle in the decomposition; and the third vertex of such a triangle is one of the vertices $V_2, V_3, \ldots, V_{n+2}$. Let the thrid vertex be V_{k+2}, $0 \le k \le n$. Then we have one $(k+2)$-gon $V_1V_2\cdots V_{k+2}$ and one $(n-k+2)$-gon $V_{k+2}V_{k+3}\cdots V_{n+3}$. There are c_k ways to divide $V_1V_2\cdots V_{k+2}$ into triangles and c_{n-k} ways to divide $V_{k+2}V_{k+3}\cdots V_{n+3}$ into triangles. We thus have the recurrence relation

$$c_{n+1} = \sum_{k=0}^{n} c_kc_{n-k},$$

where $c_0 = 1$. So $c_n = \frac{(2n)!}{n!(n+1)!} = \frac{\binom{2n}{n}}{n+1}$.

5.6. Searching and Sorting

Given an array $A = \langle a_1, a_2, \cdots, a_n \rangle$ and an object S, determine the position of S in A, that is, find an index i such that $a_i = S$ (if such an i exists).

Algorithm SEQSEARCH
Step 1 Input A and S.
Step 2 For $i = 1$ to n do
Step 3 If $a_i = S$, then output i and stop.
Step 4 Output "S not in A" and stop.

5.7. Growth of Functions

For functions f and g defined on the set $\mathbf{Z}^+$ of positive integers, if there exist positive constant C and K such that

$$|f(n)| \leq C|g(n)| \quad \text{for all} \quad n \geq K,$$

then f is said to be **big-Oh** of g, write f is $O(g)$. This means that f grows no faster than g. We say that f and g have the **same order** if f is $O(g)$ and g is $O(f)$. If f is $O(g)$, but g is not $O(f)$, then we say that f is of **lower order** than g or g **grows faster** than f.

EXAMPLE 5.1. *For Example 6 in Lectures 12 and 13, the number of comparisons $f(n)$ is a function of integers, and in Case I, f is $O(n)$; in Case II, f is $O(\log n)$.*

For Example 7, the number of comparisons $f(n)$ is also function of positive integers, and for Solution I, f is $O(n^2)$, but for Solution II, f is $O(n)$.

For two functions f and g defined on positive integers, f is $O(g)$ if and only if there exists a constant C such that

$$\lim_{n \to \infty} \frac{f(n)}{g(n)} \leq C.$$

CHAPTER 6

Binary Relations

6.1. Binary Relations

The notion of a relation between two sets of objects is quite common and intuitively clear. Let X be the set of all living human females and Y the set of all living human males. The wife-husband relation R can be defined from X to Y. Thus, for $x \in X$ and $y \in Y$, we say that x is related to y by the relation R if x is a wife of y, and write xRy. To describe the relation R, we may take the collection of all ordered pairs (x, y) such that x is related to y by R; the collection of related ordered pairs is simply a subset of the product set $X \times Y$. This motivates the definition of relations.

Definition 6.1. Let X and Y be nonempty sets. By a **binary relation** (or just **relation**) from X to Y we mean a subset $R \subset X \times Y$. If $(x, y) \in R$, we say that x is related to y by R, written xRy. If $(x, y) \notin R$, we say that x is not related to y, and written $x\bar{R}y$. For $x \in X$, we define

$$R(x) = \{y \in Y \mid xRy\} = \{y \in Y \mid (x, y) \in R\};$$

and for a subset $A \subset X$, define

$$R(A) = \{y \in Y \mid \textit{there exists } x \in A \textit{ such that } xRy\}.$$

If $X = Y$, we say that R is a binary relation on X.

Since binary relations from X to Y are subsets of $X \times Y$, one can define intersection, union, and complement for binary relations. For a relation $R \subset X \times Y$, the **complementary relation** of R is the binary relation $\bar{R} \subset X \times Y$, defined by

$$x\bar{R}y \Leftrightarrow (x, y) \notin R;$$

and the **inverse relation** of R is the binary relation $R^{-1} \subset Y \times X$, defined by

$$yR^{-1}x \Leftrightarrow (x, y) \in R.$$

Example Consider a family A with five children, Amy, Bob, Charlie, Debbie, and Eric. We abbreviate the names to their first letters so that $A = \{a, b, c, d, e\}$.

(a) The "brother-sister" relation R_{bs} is the set

$$R_{bs} = \{(b, a), (b, d), (c, a), (c, d), (e, a), (e, d)\}.$$

(b) The "sister-brother" relation R_{sb} is the set

$$R_{sb} = \{(a, b), (a, c), (a, e), (d, b), (d, c), (d, e)\}.$$

(c) The "brother" relation R_b is the set

$$\{(b, b), (b, c), (b, e), (c, b), (c, c), (c, e), (e, b), (e, c), (e, e)\}.$$

(d) The "sister" relation R_s is the set

$$\{(a, a), (a, d), (d, a), (d, d)\}.$$

The "brother-sister" relation R_{bs} is the inverse of the "sister-brother" relation R_{sb}; that is, $R_{bs} = R_{sb}^{-1}$. The "brother or sister" relation is the union of the "brother" relation and the "sister" relation; that is, $R_b \cup R_s$. The complementary relation of "brother or sister" relation is the "brother-sister or sister-brother" relation; that is, $\overline{R_b \cup R_s} = R_{bs} \cup R_{sb}$.

Example

(a) The graph of the equation, $\frac{x^2}{3^2} + \frac{y^2}{2^2} = 1$, defines a binary relation on the set $\mathbb{R}$ of real numbers; its graph is an ellipse.
(b) The relation "less than", denoted $<$, is a binary relation on $\mathbb{R}$, defined by $a < b$ if and only if a is less than b. As a subset of $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$, the relation is also given by the set $\{(a, b) \in \mathbb{R}^2 \mid a \text{ is less than } b\}$.
(c) The relation "greater than or equal to" is a binary relation $\geq$ on $\mathbb{R}$, defined by $a \geq b$ if and only if a is greater than or equal to b. As a subset of $\mathbb{R}^2$, the relation is given by the set $\{(a, b) \in \mathbb{R}^2 \mid a \text{ is greater than or equal to } b\}$.
(d) The divisibility relation $\mid$ about integers, defined by $a|b$ if and only if a divides b, is a binary relation on the set $\mathbf{Z}$ of integers.

Example A function $f : X \to Y$ can be viewed as a relation from X to Y; it is a relation $f \subset X \times Y$ such that $|f(x)| = 1$ for all $x \in X$.

Proposition 6.2. *Let R be a binary relation from X to Y. Let A and B be subsets of X.*

(a) If $A \subset B$, then $R(A) \subset R(B)$.
(b) $R(A \cup B) = R(A) \cup R(B)$.
(c) $R(A \cap B) \subset R(A) \cap R(B)$.

Proof. (a) For any $y \in R(A)$, there is an $x \in A$ such that xRy. Since $A \subset B$, we have $y \in R(B)$. Thus $R(A) \subset R(B)$.

(b) For any $y \in R(A \cup B)$, there is an $x \in A \cup B$ such that xRy. If $x \in A$, then $y \in R(A)$. If $x \in B$, then $y \in R(B)$. In either case, $y \in R(A) \cup R(B)$. Thus $R(A \cup B) \subset R(A) \cup R(B)$. On the other hand, it follows from (a) that $R(A) \subset R(A \cup B)$ and $R(B) \subset R(A \cup B)$. Therefore $R(A) \cup R(B) \subset R(A \cup B)$.

(c) It follows from (1) that $R(A \cap B) \subset R(A)$ and $R(A \cap B) \subset R(B)$. Thus $R(A \cap B) \subset R(A) \cap R(B)$. □

Proposition 6.3. *Let R_1 and R_2 be relations from X to Y. If $R_1(x) = R_2(x)$ for all $x \in X$, then $R_1 = R_2$.*

Proof. If xR_1y, then $y \in R_1(x)$. Since $R_1(x) = R_2(x)$, we have $y \in R_2(x)$. Then xR_2y. A similar argument shows that if xR_2y then xR_1y. Thus $R_1 = R_2$. □

Let R be a relation on a set X. A **path of length** k in R from x to y is a finite sequence $x = x_0, x_1, \ldots, x_k = y$, beginning with x and ending with y, such that

$$x_0Rx_1,\ x_1Rx_2,\ \ldots,\ x_{k-1}Rx_k.$$

A path that begins and ends at the same vertex is called a **cycle**. For a fixed positive integer k, we define a relation R^k on X as follows:

$$xR^ky \Leftrightarrow \text{ there is a path of length } k \text{ from } x \text{ to } y.$$

We may also define a relation R^∞ on X by letting

$$xR^\infty y \Leftrightarrow \text{ there is some path from } x \text{ to } y.$$

The relation R^∞ is sometimes called the **connectivity relation** for R. It is clear that

$$R^\infty = R \cup R^2 \cup R^3 \cup \cdots = \bigcup_{k=1}^{\infty} R^k.$$

The **reachability relation** of R is the relation R^* on X defined by

$$xR^*y \Leftrightarrow \text{ either } x = y \text{ or } xR^\infty y;$$

that is,

$$R^* = I \cup R \cup R^2 \cup R^3 \cup \cdots = \bigcup_{k=0}^{\infty} R^k,$$

where I is the identity relation on X, defined by xIy if and only if $x = y$. We always assume that $R^0 = I$ for any relation R.

6.2. Representation of Relations

Binary relations are the most important relations among all relations. Ternary relations, quaternary relations, and multi-relations can be studied by binary relations. We introduce two methods to represent a binary relation, one by a matrix and the other one by a directed graph.

Definition 6.4. Let R be a binary relation from $X = \{x_1, x_2, \ldots, x_m\}$ to $Y = \{y_1, y_2, \ldots, y_n\}$. The **matrix** of R is an $m \times n$ matrix $M_R = [a_{ij}]$ whose (i, j)-entry is

$$a_{ij} = \begin{cases} 1 & \textit{if} \quad (x_i, y_j) \in R \\ 0 & \textit{if} \quad (x_i, y_j) \notin R. \end{cases}$$

The matrix M_R is called a **Boolean matrix**. If $X = Y$, M_R is a square matrix.

For $m \times n$ Boolean matrices $M_1 = [a_{ij}]$ and $M_2 = [b_{ij}]$, if $a_{ij} \le b_{ij}$ for all (i, j)-entries, we write $M_1 \le M_2$. The matrix of the "brother-sister" relation R_{bs} on the set $A = \{a, b, c, d, e\}$ is the square matrix

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \end{bmatrix}$$

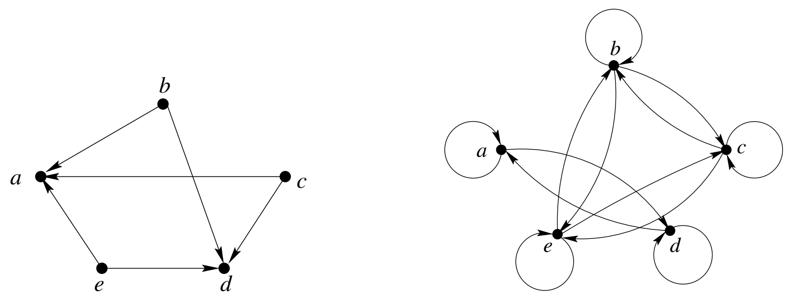
and the matrix of the "brother or sister" relation is the square matrix

$$\begin{bmatrix} 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 \end{bmatrix}$$

Another way to describe a binary relation is to draw a directed graph. Let R be a binary relation on a finite set $V = \{v_1, v_2, \ldots, v_n\}$. For each element $v_i \in V$, we draw a solid dot and name it by v_i, called a **vertex**. For two vertices v_i and v_j, if $v_i R v_j$, we draw an arrow from v_i to v_j, called a **directed edge**; if $v_i = v_j$, the arrow becomes a **directed loop**. The resulted graph is a directed graph, called the **digraph** of R, denoted $D(R)$. Note that the directed edges of a digraph may have to cross each other when drawing the digraph on a plane. However, the intersection points of directed edges are not considered to be vertices of the digraph. The **in-degree** of a vertex $v \in V$ is the number of vertices u such that $(u, v) \in R$, denoted $\operatorname{ideg}(v)$; the **out-degree** of v is the number of vertices w such that $(v, w) \in R$, denoted $\operatorname{odeg}(v)$. If R is a relation from X to Y, we define

$$\begin{array}{rcll} \operatorname{odeg}(x) & = & |R(x)| & \text{for all } x \in X, \\ \operatorname{ideg}(y) & = & |R^{-1}(y)| & \text{for all } y \in Y. \end{array}$$

The digraphs of the "brother-sister" relation R_{bs} and the "brother or sister" relation $R_b \cup R_s$ are demonstrated in the following.



(a) R_{bs} (b) $R_b \cup R_s$

FIGURE 1. The digraphs of relations R_{bs} and $R_b \cup R_s$.

PROPOSITION 6.5. *For the digraph $D(R)$ of a binary relation R on V,*

$$\sum_{v \in V} \operatorname{ideg}(v) = \sum_{v \in V} \operatorname{odeg}(v) = |R|.$$

If R is a relation from X to Y, then

$$\sum_{x \in X} \operatorname{odeg}(x) = \sum_{y \in Y} \operatorname{ideg}(y) = |R|.$$

Proof. Trivial. $\square$

6.3. Composition of Relations

DEFINITION 6.6. Let R be a relation from X to Y, and S a relation from Y to Z. The **composition** of R and S is a relation $S \circ R$ from X to Z, defined by

$$x(S \circ R)z \Leftrightarrow \text{ there is an element } y \in Y \text{ such that } xRy \text{ and } ySz.$$

If $X = Y$, then R is a relation on X; we have $R^2 = R \circ R$ and $R^k = R^{k-1} \circ R$ for $k \geq 2$.

Note that in the composition of R and S we consider R as the first relation and S the second, and the notation $S \circ R$ is backward. However, many people use $R \circ S$ as a name for what we have called $S \circ R$. Such usage is inconsistent with the notation for functional composition and causes some confusion. To avoid misunderstanding and for aesthetic reason, we often write RS for $S \circ R$.

Example For the "brother-sister relation," "sister-brother relation," "brother relation," and "sister relation" on $A = \{a, b, c, d, e\}$, we have

$$\begin{array}{lll} R_{bs}R_{sb} = R_b, & R_{sb}R_{bs} = R_s, & R_{bs}R_s = R_{bs}, \\ R_{bs}R_{bs} = \emptyset, & R_bR_b = R_b, & R_bR_s = \emptyset. \end{array}$$

Let X_1, X_2, ..., X_n, X_{n+1} be nonempty sets. Let R_i be a relation from X_i to X_{i+1}, $1 \le i \le n$. We define a relation $R_1R_2 \cdots R_n$ from X_1 to X_{n+1} by

$$xR_1R_2 \cdots R_n y$$

if and only if there is a sequence $x = x_1, x_2, \ldots, x_n, x_{n+1} = y$ such that

$$x_1R_1x_2,\ x_2R_2x_3,\ \ldots,\ x_nR_nx_{n+1}.$$

THEOREM 6.7. *Let R_1 be a relation from X_1 to X_2, R_2 a relation from X_2 to X_3, and R_3 a relation from X_3 to X_4. Then $R_1(R_2R_3)$ and $(R_1R_2)R_3$ are relations from X_1 to X_4, and*

$$R_1R_2R_3 = R_1(R_1R_3) = (R_1R_2)R_3.$$

Proof. For any $x \in X_1$ and $y \in X_4$, we have

$$\begin{array}{rcl} xR_1(R_2R_3)y & \Leftrightarrow & \exists x_2 \in X_2 \text{ s.t. } xR_1x_2,\ x_2R_2R_3y \\ & \Leftrightarrow & \exists x_2 \in X_2 \text{ s.t. } xR_1x_2; \exists x_3 \in X_3 \text{ s.t. } x_2R_2x_3,\ x_3R_3y \\ & \Leftrightarrow & \exists x_2 \in X_2, x_3 \in X_3 \text{ s.t. } xR_1x_2,\ x_2R_2x_3,\ x_3R_3y \\ & \Leftrightarrow & xR_1R_2R_3y. \end{array}$$

Similarly, $x(R_1R_2)R_3y \Leftrightarrow xR_1R_2R_3y$. □

PROPOSITION 6.8. *Let R be a relation from X to Y, R_i ($i \in I$) a family of relations from Y to Z, and S a relation from Z to W. Then*

(a) $R\left(\bigcup_{i\in I} R_i\right) = \bigcup_{i\in I} RR_i$;
(b) $\left(\bigcup_{i\in I} R_i\right) S = \bigcup_{i\in I} R_iS$.

Proof. (a) For any $x(R \cup_i R_i)z$, there exists $y \in Y$ such that xRy and $y(\cup_i R_i)z$. Then there is one j such that yR_iz. Thus $x(RR_i)z$, and so $x(\bigcup_i RR_i)z$. Conversely, for any $x(\cup_i RR_i)z$, there is one i such that $x(RR_i)z$. Then there exists $y \in Y$ such that xRy and yR_iz. Of course, $y(\cup_i R_i)z$. Thus $x(R \cup_i R_i)z$. The proof for (b) is similar. □

For the convenience of representing composition of relations, we introduce two operations $\wedge$ and $\vee$ on real numbers. For $a, b \in \mathbb{R}$, define

$$a \wedge b = \min\{a, b\},$$

$$a \vee b = \max\{a, b\}.$$

LEMMA 6.9. *For $a, b, c \in \mathbb{R}$,*

$$a \wedge (b \vee c) = (a \wedge b) \vee (a \wedge c),$$

$$a \vee (b \wedge c) = (a \vee b) \wedge (a \vee c).$$

Proof. We only prove the first formula. The second one is similar.

Case 1: $b \leq c$. If $a \geq c$, then the left side is $a \wedge (b \vee c) = a \wedge c = c$. The right side is $(a \wedge b) \vee (a \wedge c) = b \vee c = c$. If $b \leq a \leq c$, then the left side is $a \wedge (b \vee c) = a \wedge c = a$. The right side is $(a \wedge b) \vee (a \wedge c) = b \vee a = a$. If $a \leq b \leq c$, then the left side is $a \wedge (b \vee c) = a \wedge c = a$. The right side is $(a \wedge b) \vee (a \wedge c) = a \vee a = a$.

Case 2: $b \geq c$. If $a \leq c$, then $a \wedge (b \vee c) = a \wedge c = a$ and $(a \wedge b) \vee (a \wedge c) = a \vee a = a$. If $b \geq a \geq c$, then $a \wedge (b \vee c) = a \wedge c = a$ and $(a \wedge b) \vee (a \wedge c) = a \vee c = a$. If $a \geq b$, then $a \wedge (b \vee c) = a \wedge b = b$ and $(a \wedge b) \vee (a \wedge c) = b \vee c = b$. □

For real numbers $a_1, a_2, \ldots, a_n$, we define

$$\bigwedge_{i=1}^{n} a_i = \min\{a_1, a_2, \ldots, a_n\},$$

$$\bigvee_{i=1}^{n} a_i = \max\{a_1, a_2, \ldots, a_n\}.$$

Let $A = [a_{ij}]$ be an $m \times n$ matrix and $B = [b_{jk}]$ an $n \times p$ matrix. The **Boolean multiplication** of A and B is an $m \times p$ matrix $A * B = [c_{ik}]$, whose (i, k)-entry is

$$c_{ik} = \bigvee_{j=1}^{n} (a_{ij} \wedge b_{jk}).$$

THEOREM 6.10. *Let R be a relation from $X = \{x_1, \ldots, x_m\}$ to $Y = \{y_1, \ldots, y_n\}$ and let S be a relation from Y to $Z = \{z_1, \ldots, z_p\}$. If M_R, M_S, and M_{RS} are matrices of the relations R, S, and RS respectively, then*

$$M_{RS} = M_R * M_S.$$

Proof. Write $M_R = [a_{ij}]$, $M_S = [b_{jk}]$, $M_R * M_S = [c_{ik}]$, and $M_{RS} = [d_{ik}]$. It suffices to show that $c_{ik} = d_{ik}$ for each (i, k)-entry of the matrices. In fact, if $c_{ik} = 1$, it forces that $a_{ij} \wedge b_{jk} = 1$ for at least one j, say j_0. Then $a_{ij_0} = b_{j_0k} = 1$. This means that $x_i R y_{j_0}$ and $y_{j_0} S z_k$. Thus $x_i RS z_k$ by definition; so $d_{ik} = 1$. If $c_{ik} = 0$, then $a_{ij} \wedge b_{jk} = 0$ for all j's; that is, there is no $y_j \in Y$ such that both $x_i R y_j$ and $y_j S z_k$. Thus by definition x_i is not related to z_k by RS. Therefore $d_{ik} = 0$. This completes the proof of $c_{ik} = d_{ik}$. □

6.4. Special Relations

Most of time we are interested in some special relations satisfying certain properties. For instance, the "less than" relation on the set of real numbers satisfies the so-called transitive property: if $a < b$ and $b < c$, then $a < c$.

DEFINITION 6.11. A binary relation R on a set X is called

(a) **reflexive** if xRx for all x in X;
(b) **symmetric** if xRy implies yRx;
(c) **transitive** if xRy and yRz imply xRz.

The relation R is called an **equivalence relation** if it is reflexive, symmetric, and transitive; and in this case, if xRy, we say that x and y are **equivalent**.

The relation $I = I_X = \{(x, x) \mid x \in X\}$ is called the **identity relation**, and X^2 is called the **complete relation** on X.

Example Many family relations are binary relations on the set of human beings.

(a) The brother relation R_b: $xR_by \Leftrightarrow x$ and y are both males and have the same parents. (symmetric and transitive)
(b) The sister relation R_s: $xR_sy \Leftrightarrow x$ and y are both females and have the same parents. (symmetric and transitive)
(c) The brother-sister relation R_{bs}: $xR_{bs}y \Leftrightarrow x$ is male, y is female, x and y have the same parents.
(d) The sister-brother relation R_{sb}: $xR_{sb} \Leftrightarrow x$ is female, y is male, and x and y have the same parents.
(e) The generalized brother relation R'_b: $xR'_by \Leftrightarrow x$ and y are both males and have the same father or the same mother. (symmetric)
(f) The generalized sister relation R'_s: $xR'_sy \Leftrightarrow x$ and y are both females and have the same father or mother. (symmetric)
(g) The relation R: $xRy \Leftrightarrow x$ and y have the same parents. (reflexive, symmetric, and transitive; equivalence relation)
(h) The relation R': $xR'y \Leftrightarrow x$ and y have the same father or the same mother. (reflexive and symmetric)

Example

(a) The "less than" relation $<$ on the set of real numbers is a transitive relation.
(b) The "less than or equal to" relation $\leq$ on the set of real numbers is a reflexive and transitive relation.
(c) The divisibility relation on the set of positive integers is a reflexive and transitive relation.
(d) Given a positive integer n; the **congruence of modulo** n is a relation $\equiv_n$ on $\mathbb{Z}$, defined by $a \equiv_n b$ if and only if $b - a$ is a multiple of n. The standard notation for $a \equiv_n b$ is $a \equiv b \pmod{n}$. The relation $\equiv_n$ is an equivalence relation on $\mathbb{Z}$.

THEOREM 6.12. *Let R be a relation on a set X. Then*

(a) R is reflexive $\Leftrightarrow I \subset R \Leftrightarrow$ all diagonal entries of M_R are 1.
(b) R is symmetric $\Leftrightarrow R = R^{-1} \Leftrightarrow M_R$ is a symmetric matrix.
(c) R is transitive $\Leftrightarrow R^2 \subset R \Leftrightarrow M_R^2 \leq M_R$.

Proof. (a) and (b) are trivial.

(c) "R is transitive $\Rightarrow R^2 \subset R$." For any $(x, y) \in R^2$, there exists $z \in X$ such that $(x, z) \in R$ and $(z, y) \in R$. Since R is transitive, we have $(x, y) \in R$. So $R^2 \subset R$.

"$R^2 \subset R \Rightarrow R$ is transitive." For $(x, z) \in R$ and $(z, y) \in R$, we have $(x, y) \in R^2 \subset R$. Then $(x, y) \in R$. So R is transitive.

Note that for any relations R and S on X, $R \subset S$ if and only if $M_R \leq M_S$. Also note that the matrix M_{R^2} of R^2 is M_R^2. We thus have that $R^2 \subset R$ if and only if $M_R^2 \leq M_R$. $\square$

6.5. Equivalence Relations and Partitions

The most important relations among binary relations are equivalence relations. We will see that an equivalence relation on a set X will partition X into disjoint equivalence classes.

Example Consider the congruence relation $\equiv_3$ on $\mathbb{Z}$. For each $a \in \mathbb{Z}$, define

$$[a] = \{b \in \mathbb{Z} \mid a \equiv_3 b\} = \{b \in \mathbb{Z} \mid a \equiv b \pmod 3\}.$$

It is clear that $\mathbf{Z}$ is partitioned into three disjoint subsets

$$\begin{aligned} [0] &= \{0, 3, \pm 6, \pm 9, \ldots\} = \{3k \mid k \in \mathbb{Z}\}, \\ [1] &= \{1, 1 \pm 3, 1 \pm 6, 1 \pm 9, \ldots\} = \{3k+1 \mid k \in \mathbb{Z}\}, \\ [2] &= \{2, 2 \pm 3, 2 \pm 6, 2 \pm 9, \ldots\} = \{3k+2 \mid k \in \mathbb{Z}\}. \end{aligned}$$

Moreover, $[3k] = [0]$, $[3k+1] = [1]$, and $[3k+2] = [2]$ for all $k \in \mathbb{Z}$.

Theorem 6.13. *Let $\sim$ be an equivalence relation on a set X. For each x of X, let $[x] = \{y \in X \mid x \sim y\}$. Then*

(a) $[x] \neq \emptyset$ *for any* $x \in X$,
(b) $[x] = [y]$ *if* $x \sim y$,
(c) $[x] \cap [y] = \emptyset$ *if* $x \not\sim y$,
(d) $X = \bigcup_{x \in X} [x]$.

Each subset $[x]$ is called an **equivalence class**, *and x is called a* **representative** *of $[x]$.*

Proof. (a) Each $[x]$ is obviously nonempty because $\sim$ is reflexive.

(b) For any $z \in [x]$, we have $x \sim z$ by definition of $[x]$. Since $x \sim y$, we have $y \sim x$ by the symmetric property of $\sim$. Then $y \sim x$ and $x \sim z$ imply that $y \sim z$ by transitivity of $\sim$. Thus $z \in [y]$ by definition of $[y]$; that is, $[x] \subset [y]$. Since $\sim$ is symmetric, we have $[y] \subset [x]$. Therefore $[x] = [y]$.

(c) Suppose $[x] \cap [y]$ is not empty, say $z \in [x] \cap [y]$. Then $x \sim z$ and $y \sim z$. By symmetry of $\sim$, we have $z \sim y$. Thus $x \sim y$ by transitivity of $\sim$, a contradiction.

(d) This is obvious because $x \in [x]$ for any $x \in X$. □

Definition 6.14. A **partition** of a nonempty set X is a collection $\mathcal{P} = \{A_j | j \in J\}$ of nonempty subsets of X such that

(a) $A_i \cap A_j = \emptyset$ if $i \neq j$,
(b) $X = \bigcup_{j \in J} A_j$.

Each subset A_j is called a **block** of the partition $\mathcal{P}$.

Theorem 6.15. *Let $\mathcal{P}$ be a partition of a set X. Let $R_{\mathcal{P}}$ be the relation on X, defined by*

$$xR_{\mathcal{P}}y \Leftrightarrow \textit{there exists a block } A_j \in \mathcal{P} \textit{ such that } x, y \in A_j.$$

Then $R_{\mathcal{P}}$ is an equivalence relation on X, called the **equivalence relation determined by $\mathcal{P}$**.

Proof. (a) For each $x \in X$, there exits one A_j such that $x \in A_j$. Then by definition of $R_{\mathcal{P}}$, $xR_{\mathcal{P}}x$. So $R_{\mathcal{P}}$ is reflexive. (b) If $xR_{\mathcal{P}}y$, then there is one A_j such that $x, y \in A_j$. Again by definition of $R_{\mathcal{P}}$, $yR_{\mathcal{P}}x$. Thus $R_{\mathcal{P}}$ is symmetric. (c) If $xR_{\mathcal{P}}y$ and $yR_{\mathcal{P}}z$, then there exist A_i and A_j such that $x, y \in A_i$ and $y, z \in A_j$. Obviously, $z \in A_i \cap A_j$. Since $\mathcal{P}$ is a partition. It forces that $A_i = A_j$. Thus $xR_{\mathcal{P}}z$.

This shows that $R_{\mathcal{P}}$ is transitive. $\square$

For an equivalence relation R on a set X, the collection $\mathcal{P}_R = \{[x] : x \in X\}$ of equivalence classes of R forms a partition of X, called the **quotient set** of X under R. Let $E(X)$ denote the set of all equivalence relations on X and $P(X)$ the set of all partitions of X. Define functions

$$f : E(X) \to P(X) \quad \text{by} \quad f(R) = \mathcal{P}_R,$$
$$g : P(X) \to E(X) \quad \text{by} \quad g(\mathcal{P}) = R_{\mathcal{P}}.$$

Then f and g satisfy the following properties.

THEOREM 6.16. *For any equivalence relation R on X and any partition $\mathcal{P}$ of X,*

$$(g \circ f)(R) = R \quad \textit{and} \quad (f \circ g)(\mathcal{P}) = \mathcal{P}.$$

In other words, f and g are inverse of each other.

Proof. Note that $(g \circ f)(R) = g(f(R))$ and $(f \circ g)(\mathcal{P}) = f(g(\mathcal{P}))$. We have

$$\begin{array}{lcllcl} x[g(f(R))(R)]y & \Leftrightarrow & \exists A \in f(R) \text{ s.t. } x, y \in A & \Leftrightarrow & xRy; \\ A \in f(g(\mathcal{P})) & \Leftrightarrow & \exists x \in X \text{ s.t. } A = g(\mathcal{P})(x) & \Leftrightarrow & A \in \mathcal{P}. \end{array}$$

Thus $g(f(R)) = R$ and $f(g(\mathcal{P})) = \mathcal{P}$. $\square$

EXAMPLE 6.1. *Let $\mathbb{Z}_+$ be the set of positive integers. Define the relation $\sim$ on $\mathbb{Z} \times \mathbb{Z}_+$ by*

$$(a, b) \sim (c, d) \;\Leftrightarrow\; ad = bc.$$

Is $\sim$ an equivalence relation? If Yes, what are the equivalence classes?

Let R be a relation on a set X. The **reflexive closure** is a reflexive relation $r(R)$ on X such that

(a) $R \subset r(R)$;
(b) if R' is a reflexive relation on X and $R \subset R'$, then $r(R) \subset R'$.

The **symmetric closure** of R is a symmetric relation $s(R)$ on X such that

(a) $R \subset s(R)$;
(b) if R' is a symmetric relation on X and $R \subset R'$, then $s(R) \subset R'$.

The **transitive closure** of R is a transitive relation $t(R)$ on X such that

(a) $R \subset t(R)$;
(b) if R' is a transitive relation on X and $R \subset R'$, then $t(R) \subset R'$.

Obviously, the reflexive, symmetric, and transitive closures of R need to be unique respectively.

THEOREM 6.17. *For any relation R on a set X,*

(a) $r(R) = R \cup I$;
(b) $s(R) = R \cup R^{-1}$;
(c) $t(R) = R^{\infty} = \bigcup_{k=1}^{\infty} R^k$.

Proof. (a) and (b) are obvious. (c) Note that $R \subset \bigcup_{k=1}^{\infty} R^k$ and

$$\left(\bigcup_{i=1}^{\infty} R^i\right)\left(\bigcup_{j=1}^{\infty} R_j\right) = \bigcup_{i,j=1}^{\infty} R^i R^j = \bigcup_{i,j=1}^{\infty} R^{i+j} = \bigcup_{k=2}^{\infty} R^k \subset \bigcup_{k=1}^{\infty} R^k.$$

This shows that $\bigcup_{k=1}^{\infty} R^k$ is transitive and $R \subset \bigcup_{k=1}^{\infty} R^k$. Since any transitive relation which contains R must contain R^k for all positive integers k. Then $\bigcup_{k=1}^{\infty} R^k$ is the transitive closure of R. $\square$

THEOREM 6.18. *Let R be a relation on a set X of n elements. Then*

$$t(R) = R \cup R^2 \cup \cdots \cup R^{n-1}.$$

In particular, if R is reflexive, then $t(R) = R^{n-1}$.

Proof. It suffices to show that $R^l \subset \bigcup_{k=1}^{n-1} R^k$ for all $l \geq n$; and this is equivalent to showing $R^l \subset \bigcup_{k=1}^{l-1} R^k$ for all $l \geq n$. Let $(x, y) \in R^l$. There exist $x_1, \ldots, x_{l-1} \in X$ such that all (x, x_1), (x_1, x_2), $\ldots$, (x_{l-1}, y) belong to R. Since $l \geq n$, two elements of $x = x_0$, x_1, x_2, $\ldots$, x_{l-1} ,$x_l = y$ must be the same, say, $x_i = x_j$ with $i < j$. Then (x_0, x_1), $\ldots$, (x_{i-1}, x_i), (x_j, x_{j+1}), $\ldots$, $(x_{l-1}, x_l) \in R$ imply that $(x, y) = (x_0, x_l) \in R^{l+i-j} \subset \bigcup_{k=1}^{l-1} R^k$. Thus $R^l \subset \bigcup_{k=1}^{l-1} R^k$.

If R is reflexive, we have $R^k \subset R^{k+1}$ for all $k \geq 1$. So $t(R) = R^{n-1}$. $\square$

PROPOSITION 6.19. *Let R be a relation on a set X. Then $I \cup t(R \cup R^{-1})$ is an equivalence relation. In particular, if R is reflexive and symmetric, then $t(R)$ is an equivalence relation.*

Proof. Since $I \cup t(R \cup R^{-1})$ is reflexive and transitive, we only need to show that $I \cup t(R \cup R^{-1})$ is symmetric. For $(x, y) \in I \cup t(R \cup R^{-1})$, if $x = y$, obviously $(y, x) \in I \cup t(R \cup R^{-1})$. If $x \neq y$, then $(x, y) \in t(R \cup R^{-1})$. Thus $(x, y) \in (R \cup R^{-1})^k$ for some $k \geq 1$; that is, there is a sequence $x = x_0, x_1, \ldots, x_k = y$ such that $(x_i, x_{i+1}) \in R \cup R^{-1}$, $0 \leq i \leq k - 1$. Since $R \cup R^{-1}$ is symmetric, we have $(x_{i+1}, x_i) \in R \cup R^{-1}$ for all $0 \leq i \leq k - 1$. This means that $(y, x) \in (R \cup R^{-1})^k$. So $(y, x) \in I \cup t(R \cup R^{-1})$. We have proved that $I \cup t(R \cup R^{-1})$ is symmetric.

Now if R is reflexive and symmetric, then $t(R) = I \vee t(R \cup R^{-1}) = t(R)$. So $t(R)$ is an equivalence relation. $\square$

Let R be a relation on a set X. The **reachability relation** of R is a relation R^* on X, defined by xR^*y if and only if either $x = y$ or there exist $x_1, x_2, \ldots, x_k$ such that $(x, x_1), (x_1, x_2), \ldots, (x_k, y) \in R$; that is, $R^* = I \cup t(R)$.

THEOREM 6.20. *Let R be a relation on a set X. Let M and M^* be the matrices of R and R^* respectively. If $|X| = n$, then*

$$M^* = I \vee M \vee M^2 \vee \cdots \vee M^{n-1}.$$

Moreover, if R is reflexive, then $R^k \subset R^{k+1}$ for all $k \geq 1$ and $M^* = M^{n-1}$.

Proof. It follows from Theorem 6.18. $\square$

Let R be a relation on $X = \{x_1, \ldots, x_n\}$. If $y_0, y_1, \ldots, y_m$ is a path in R, the vertices $y_1, \ldots, y_{m-1}$ are called **interior vertices** of the path. For each $1 \leq k \leq n$, we define a Boolean matrix $W_k = [w_{ij}]$, where $w_{ij} = 1$ if there is a path in R from x_i to x_j whose interior vertices are contained in $\{x_1, \ldots, x_k\}$. Since the interior vertices of any path in R is contained in $X = \{x_1, \ldots, x_n\}$, the (i, j)-entry of W_n is equal to 1 if there is a path in R from x_i to x_j; that is, $W_n = M_{R^\infty}$. We set $W_0 = M_R$. Then we have a sequence of Boolean matrices

$$W_0 = M_R,\ W_1,\ W_2,\ \ldots, W_n.$$

We will give an algorithm, called **Warshall's algorithm**, to compute W_k from W_{k-1}.

Let $W_{k-1} = [s_{ij}]$ and $W_k = [t_{ij}]$. If $t_{ij} = 1$, then there must be a path

$$x_i = y_0, y_1, \ldots, y_m = x_j$$

from x_i to x_j whose interior vertices are contained in $\{x_1, \ldots, x_k\}$. We may assume that the interior vertices $y_1, \ldots, y_{m-1}$ are distinct. If x_k is not an interior vertex of this path, then all interior vertices must be actually contained in $\{x_1, \ldots, x_{k-1}\}$, so $s_{ij} = 1$. If x_k is an interior of the path, say $x_k = y_l$, we then have two subpaths

$$x_i = y_0, y_1, \ldots, y_l = x_k \quad \text{and} \quad x_k = y_l, y_{l+1}, \ldots, y_m = x_j$$

whose interior vertices are both contained in $\{x_1, \ldots, x_{k-1}\}$, so $s_{ik} = 1$ and $s_{kj} = 1$. Thus

$$t_{ij} = 1 \Leftrightarrow \begin{cases} s_{ij} = 1 \text{ or} \\ s_{ik} = 1 \text{ and } s_{kj} = 1. \end{cases}$$

Warshall's Algorithm Working on the Boolean matrix W_{k-1} to produce W_k.

(a) If W_{k-1} has 1 in (i,j)-entry, so is W_k; keep 1 there.
(b) If W_{k-1} has 0 in (i,j)-entry, then check the entries (i,k) and (k,j) in W_{k-1}. If both entries are 1, then change the (i,j)-entry in W_{k-1} to 1. Otherwise, keep 0 there.

Example Consider the relation R on $A = \{1,2,3,4,5\}$, defined by the Boolean matrix

$$M_R = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 \end{bmatrix}.$$

By Warshall's algorithm, we have

$$W_0 = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 \end{bmatrix} \to W_1 = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1* \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 \end{bmatrix} \to W_2 = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 \end{bmatrix} \to$$

$$W_3 = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 \\ 1* & 1 & 1 & 0 & 1* \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 1* & 0 & 1 & 1 & 1* \end{bmatrix} \to W_4 = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 1 \end{bmatrix} \to W_5 = \begin{bmatrix} 1* & 0 & 1* & 1* & 1 \\ 1 & 1 & 1 & 1* & 1 \\ 1 & 0 & 1* & 1* & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 1 \end{bmatrix}.$$

DEFINITION 6.21. A binary relation R on a set X is called

(a) **asymmetric** if xRy implies $y\bar{R}x$;
(b) **antisymmetric** if xRy and yRx imply $x = y$.

EXERCISES

(1) Let R be a binary relation from X to Y. Let A and B be subsets of X.
(a) If $A \subset B$, then $R(A) \subset R(B)$.

(b) $R(A \cup B) = R(A) \cup R(B)$.
(c) $R(A \cap B) \subset R(A) \cap R(B)$.

(2) Let R_1 and R_2 be relations from X to Y. If $R_1(x) = R_2(x)$ for all $x \in X$, then $R_1 = R_2$.

(3) Let $a, b, c \in \mathbb{R}$. Then

$$a \wedge (b \vee c) = (a \wedge b) \vee (a \wedge c),$$

$$a \vee (b \wedge c) = (a \vee b) \wedge (a \vee c).$$

(4) Let R be a relation from X to Y, R_i $(i \in I)$ a family of relations from Y to Z, and S a relation from Z to W. Then
(a) $R\left(\bigcup_{i\in I} R_i\right) = \bigcup_{i\in I} RR_i$;
(b) $\left(\bigcup_{i\in I} R_i\right) S = \bigcup_{i\in I} R_i S$.

(5) Let R_i $(1 \le i \le 3)$ be relations on $A = \{a, b, c, d, e\}$, whose Boolean matrices are given by

$$M_1 = \begin{bmatrix} 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \; M_2 = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \; M_3 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 \end{bmatrix}$$

respectively.
(a) Draw the digraphs of the relations R_1, R_2, R_3.
(b) Find the Boolean matrices for the relations R_1^{-1}, $R_2 \cup R_3$, R_1R_1, $R_1R_1^{-1}$, $R_1^{-1}R_1$, $R_1 \cup R_1^{-1}$; and verify that $R_1R_1^{-1} = R_2$, $R_1^{-1}R_1 = R_3$.
(c) Verify that $R_2 \cup R_3$ is an equivalence relation and find the quotient set $A/(R_2 \cup R_3)$.

(6) Let R be a relation on $\mathbb{Z}$ defined by xRy if $x + y$ is an even integer.
(a) Show that R is an equivalence relation on $\mathbb{Z}$.
(b) Find all equivalence classes of the relation R.

(7) Let $X = \{1, 2, \ldots, 10\}$ and let R be a relation on X such that aRb if and only if $|a - b| \le 2$. Determine whether R is an equivalence relation. Let M_R be the matrix of R; compute M_R^8.

(8) A relation R on a set X is called a **preference relation** if R is reflexive and transitive. Show that $R \cup R^{-1}$ is an equivalence relation.

(9) Let n be a positive integer. The congruence relation $\sim$ of modulo n is an equivalence relation on $\mathbb{Z}$. Let $\mathbb{Z}_n$ denote the quotient set $\mathbb{Z}/\!\sim$. For any integer $a \in \mathbb{Z}$, we define $f_a : \mathbb{Z}_n \to \mathbb{Z}_n$ by $f_a([x]) = [ax]$. Find the cardinality of the set $f_a(\mathbb{Z}_n)$.

(10) For a positive integer n, let $\phi(n)$ be the number of positive integers $x \le n$ such that $\gcd(x, n) = 1$, called **Euler's function**. Let R be the relation on $X = \{1, 2, \ldots, n\}$, defined by

$$xRy \text{ if and only if } x \le y,\ y|n, \text{ and } \gcd(x, y) = 1.$$

(a) For each $y \in X$, find the cardinality $|R^{-1}(y)|$.
(b) Show that

$$|R| = \sum_{x|n} \phi(x).$$

(c) Show that $|R| = n$ by proving that the function $f : R \to X$, defined by $f(x, y) = \frac{xn}{y}$, is a bijection.

(11) Let X be a set of n elements. Show that the number of equivalence relations on X is

$$\sum_{k=0}^{n}(-1)^k \sum_{l=k}^{n} \frac{(l-k)^n}{k!(l-k)!}.$$

(Hint: equivalence relations are in one-to-one correspondence with partitions.)