Daoud Siniora

Discrete Mathematics

Lecture Notes

$x \in A$	$p \to q$	$A + B$	$\mathbb{R} \setminus \mathbb{Q}$	$\forall x P(x)$
$a \sim b$	$\lvert\emptyset\rvert = 0$	$p \vee \neg p$	$x \preceq y$	${}^{n}P_k$
$\phi \leftrightarrow \psi$	$\bigcup_{i\in\mathbb{N}} A_i$	$\alpha \equiv \beta$	$f \circ g$	$\mathbb{N} \subseteq \mathbb{Z}$
$AI_n = A$	$[x]_R$	$R \cap S$	$P(x, y)$	$\beta \iff \theta$
$\lvert\mathbb{N}\rvert = \aleph_0$	$A \times B$	$2^n < n!$	$m \mid n$	$f(x)$
$\mathcal{P}(S)$	$\sum_{k=1}^{n} k$	$[a_{11}\ a_{12}]$	$\deg(v)$	$\lvert\mathbb{N}\rvert < \lvert\mathbb{R}\rvert$
$\exists y Q(y)$	$1 \wedge 0$	$n \bmod m$	$A \odot B$	${}^{n}C_k$
$\gcd(a, b)$	K_n	aRb	$\lvert\mathbb{N}\rvert = \lvert\mathbb{Q}\rvert$	$x \neq 0$

Mathematics and Actuarial Science Department
The American University in Cairo

Contents

Mathematics

Mathematics is the study of patterns, order, structure, relation, quantity, space, and rate of change through logical reasoning and abstract thinking.

The origins of mathematics date back to 3000 BC when the Babylonians in the historical region of Mesopotamia (present day Syria and Iraq) and ancient Egyptians used arithmetic, algebra, and geometry for counting, construction, astronomy, taxation, and financial calculations. Later around 500 BC the ancient Greeks started treating mathematics as a subject of its own right, studying the field in a systemic way. Around 300 BC the Greek mathematician Euclid of Alexandria introduced the axiomatic method of studying mathematics which is still used until this present day. Euclid's famous book "*Elements*" is considered the most influential and successful textbook of all time. In the 3^{rd} century BC lived the Greek mathematician Archimedes of Syracuse who is considered to be the greatest mathematician of ancient history. In the 2^{nd} century BC, Hipparchus of Nicaea founded trigonometry. In the 3^{rd} century AD, Diophantus of Alexandria wrote his text "*Arithmetica*" dealing with solving algebraic equations. In Alexandria too, in the 4^{th} century AD lived the famous female mathematician and philosopher Hypatia.

One of the earliest developments of the concept of zero as a number was done by the Mayan civilisation (1800 BC – 900 AD) in Mexico and Central America. The Mayans used a base-20 (vigesimal) numeral system, and represented the number zero by a shell, one by a dot, and five by a bar. On the other side of the earth and during the course of the first millennium AD mathematics evolved in India where Indian mathematicians invented the decimal numeral system which the world uses today and they also developed the concept of zero. Advances in trigonometry in India developed definitions of sine and cosine functions. Indian mathematical works were written in the language of Sanskrit.

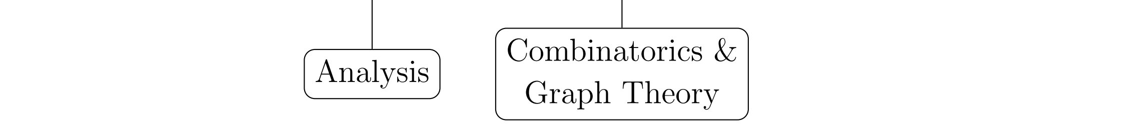
During the Golden Age of Islam, mainly the 9^{th} and 10^{th} centuries, mathematics experienced important developments building on the work of the Greeks. Notable achievements include the development of algebra and the addition of the decimal point to the Arabic numeral system. Distinguished mathematician of this era include the Persians Muhammad ibn Musa Al-Khwarizmi and Omar Khayyam.

Greek and Arabic texts on mathematics were translated into Latin starting from the

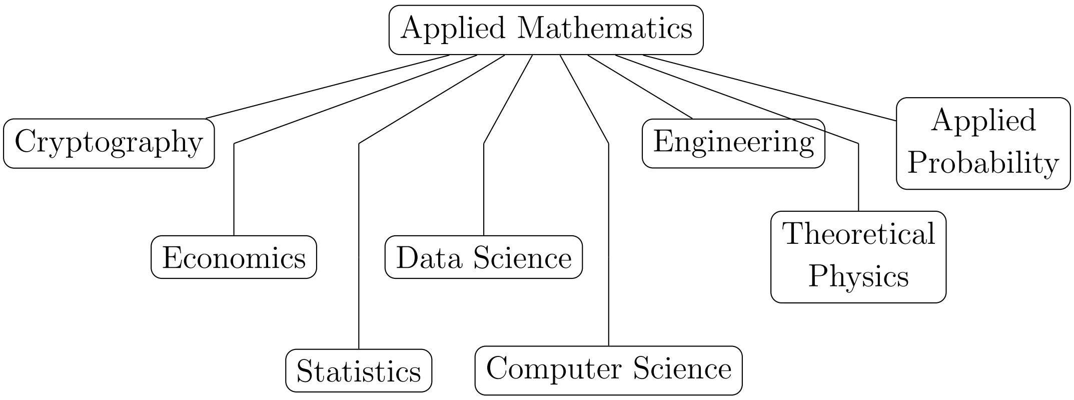
12th century and made available to Medieval Europe. This lead to new mathematical developments at an accelerating pace during the subsequent centuries, most notably, the groundbreaking work of both Isaac Newton and Gottfried Wilhelm Leibniz in the development of infinitesimal calculus during the 17th century. Subsequent eminent mathematicians include Leonhard Euler in the 18th century, Carl Friedrich Gauss in the 19th century, and Kurt Gödel in the 20th century who made a breakthrough with his Incompleteness Theorem. In 2014, the Iranian mathematician Maryam Mirzakhani became the first woman to be awarded the Fields Medal, the most prestigious award in mathematics.

Contemporary mathematics may be divided into two main areas: pure mathematics and applied mathematics.

Pure mathematics is concerned with studying abstract mathematical ideas for their own sake, a process which provides intellectual stimulation and an aesthetic appreciation of the beauty of mathematics. Pure mathematics may be included in the same category as art, music, philosophy, and poetry. Although pure mathematicians are not primarily motivated by real-world applications, results in pure mathematics often turn out to play an important role in practical applications, possibly after hundreds of years. Major areas of pure mathematics include the following.



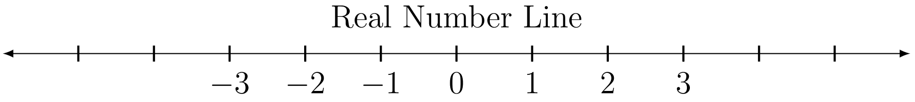
Applied mathematics is the field where mathematical methods and reasoning are applied to real-world problems and phenomena in fields such as science, engineering, social sciences, and industry. Applied mathematics includes the following areas.



Discrete mathematics is the part of pure mathematics studying discrete objects, that is, objects consisting of distinct, separated, or unconnected elements. An example of such an object is the set of positive integers.

$$1,\ 2,\ 3,\ 4,\ 5,\ 6,\ 7,\ 8,\ 9,\ 10,\ 11,\ 12,\ 13,\ 14,\ 15,\ 16,\ 17,\ \ldots$$

The term "discrete mathematics" is used in contrast with"continuous mathematics" which deals with objects based on the real numbers.



Discrete mathematics meets many fields of mathematics such as logic, set theory, number theory, abstract algebra, combinatorics, graph theory, and probability theory. It also provides mathematical foundations for computer science courses such as data structures, algorithms, database theory, formal languages, and operating systems.

This set of notes is based on the textbook "*Discrete Mathematics and Its Applications*" by Kenneth Rosen.

Chapter 1

Logic and Proofs

Logic and proofs are among the foundations of mathematics. The word *logic* in the Greek language means“the word” or “what is spoken”, it also means “thought” or “reason”. Logic consists of two parts: a *precise language* together with a list of *deduction rules.* We use the language of logic to express our thoughts and statements, and we use the deduction rules to build arguments that decide the validity of our statements. Logic is the foundation of mathematical reasoning, and it has important applications in computer science, for instance, in programming languages, verification of the correctness of programs, and artificial intelligence.

1.1 Propositional Logic

Propositional logic was first developed by the Greek philosopher Aristotle more than 2300 years ago. The basic building blocks of this type of logic are called propositions.

Definition. A *proposition* is a declarative statement that is either true or false, but not both.

Example. The following statements are propositions. They all declare an opinion.

(i) Cairo is the capital of Egypt.

(ii) Sydney is the capital of Australia.

(iii) The Rosetta Stone is in Cairo.

(iv) The number $\sqrt{2}$ is a rational number.

(v) $4 \times 487 = 1948$.

The first is a true proposition. The second is false, as Australia’s capital is Canberra. The third is false, as the Rosetta Stone is in the British Museum in London. We will prove in this chapter that $\sqrt{2}$ is not a rational number, and so the fourth statement is a false proposition. The fifth is a true proposition. $\Diamond$

Example. The following statements are not propositions.

(i) What time is it?

(ii) Bring me coffee.

(iii) $x^2 + x = 12$.

(iv) $x^2 + y^2 = 25$.

The first and second statements do not declare any facts, and so they are not propositions. The third and fourth declare an opinion but they are neither true nor false, and so they are not propositions. However they could be turned into propositions when we assign values to the variables x and y. We call such statements predicates and they will be studied in Section 1.4 of Predicate Logic. $\Diamond$

Aristotle (384 BC - 322 BC) was born in Stagirus in Greece. He became an orphan at a young age, and at the age of 17 he joined Plato's Academy in Athens. There he attended Plato's lectures for 20 years, and when Plato died in 347 BC, Aristotle was not chosen to succeed him because he had different views from those of Plato. Instead, Aristotle joined the court of King Hermias of Atarneus, and he married Pythias, the niece of the King. When the Persians defeated Hermias, Aristotle moved to Mytilene where he tutored the son of King Philip of Macedonia, who later became Alexander the Great. After the death of King Philip, Aristotle returned to Athens and set up his own school, the Lyceum. Aristotle's followers were called the peripatetics, which means "to walk about," because Aristotle often walked around as he discussed philosophical questions. Aristotle taught at the Lyceum for 13 years where he gave advanced lectures to his students and popular lectures to a broad audience. Aristotle wrote systematic treatises on logic, philosophy, psychology, physics, and natural history.

We will use symbols such as

$$p,\ q,\ r,\ s,\ t,\ p_1,\ p_2,\ p_3,\ \ldots$$

to denote variables that represent propositions. Such letters are called *propositional variables.* The *truth value* of a proposition is either *true* (denoted by T) or *false* (denoted by F). Our goal now is to build new propositions from those we already have. These methods were developed by George Boole in his famous book *"The Laws of Thought"* that was published in 1854.

George Boole (1815 – 1864) was an English mathematician, philosopher, and logician. His first contribution to logic was in 1848 when he published "*The Mathematical Analysis of Logic*". He was appointed in 1849 professor of mathematics at Queen's College in Cork, Ireland. In his book *"The Laws of Thought"* he introduced Boolean algebras which are named in his honour. In the nineteenth century, Great Britain used Boole's textbooks on differential equations and on difference equations.

♣ Logical Operators

The simplest propositions are the propositional variables. We will now introduce several *logical operators* such as negation, conjunction, disjunction, and implication to be able to construct more complicated propositions called *compound propositions.*

Definition. Let p be a propositional variable. The *negation* of p, denoted by $\neg p$, is the proposition "it is not the case that p", and we define the truth value of $\neg p$ to be false when p is true, and to be true when p is false.

We present the truth values of $\neg p$ in a table called the *truth table.*

Negation Truth Table

p	$\neg p$
T	F
F	T

Example. Let p be the proposition "Karim can swim" and suppose p is true. Then $\neg p$ stands for "Karim cannot swim" and $\neg p$ has false as its truth value. ♢

Definition. The *conjunction* of propositions p and q is the proposition $p \wedge q$ which states "p and q". The truth value of $p \wedge q$ is true only when both p and q are true and is false otherwise.

Conjunction Truth Table

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

Example. Let p be the proposition“George Boole published The Laws of Thought”, and q be “Boole worked in Ireland”.
Then the proposition $p \wedge q$ is the statement “George Boole published The Laws of Thought and he worked in Ireland”. ♢

Definition. The *disjunction* of propositions p and q is the proposition $p \vee q$ which states “p or q”. The proposition $p \vee q$ is true when p, or q, or both are true. Otherwise, when both are false it is false.

Disjunction Truth Table

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

Example. Suppose that p is the proposition “$3 < 7$”, while q is the proposition “$3 = 7$”. Then the disjunction $p \vee q$ is the proposition “$3 \leq 7$”.

Definition. The *exclusive-or* of propositions p and q is the proposition $p \oplus q$ which states “p or q but not both”. The proposition $p \oplus q$ is true only when exactly one of p or q is true.

Exclusive-or Truth Table

p	q	$p \oplus q$
T	T	F
T	F	T
F	T	T
F	F	F

The Conditional Statement

Definition. Let p and q be propositions. The *conditional statement* $p \to q$ is the proposition “if p, then q”. We call p the *hypothesis* or *premise*, and we call q the *conclusion*. The proposition $p \to q$ is false only when the hypothesis p is true and the conclusion q is false, otherwise $p \to q$ is true.

The conditional statement $p \to q$ asserts that q holds on the condition that p holds. In other words, if p holds, then q must hold too.

Conditional Statement Truth Table

p	q	$p \rightarrow q$
T	T	T
T	F	F
F	T	T
F	F	T

Suppose that we know that some conditional statement $p \rightarrow q$ is true. Thus, it is either the first row, third row, or fourth row of the truth table above. Furthermore, if you knew that the hypothesis p is also true, then this corresponds to exactly the first row, and therefore the conclusion q must be true. On the other hand, if you knew that the hypothesis p is false, then this corresponds to the third and fourth rows where the conclusion q could be true or could be false.

In summary, when the conditional statement is true and the hypothesis is true, then the conclusion must be true. However, when the conditional statement is true and the hypothesis is false, then the conclusion could be true or false. Finally, a conditional statement is false precisely when the hypothesis is true and the conclusion is false.

The conditional statement $p \rightarrow q$ is also called an *implication* and it may be read in several ways.

(i) If p, then q.

(ii) p implies q.

(iii) q if p.

(iv) p is a sufficient condition for q.

(v) q is a necessary condition for p.

(vi) q holds whenever p holds.

(vii) q follows from p.

(viii) p only if q.

(ix) whenever p holds, q holds.

(x) q whenever p.

Example. The following are conditional statements.

(i) If water molecules exist, then oxygen atoms exist.

(ii) If Laila learns mathematics, then Laila will find a job.

(iii) If I am healthy, then I will go to class.

(iv) If there is rain, there are clouds.

(v) If there are clouds, there is rain.

(vi) Having processors is necessary for manufacturing computers. ♢

We form new conditional statements starting with the conditional statement $p \to q$.

(i) The *contrapositive* of $p \to q$ is the proposition $\neg q \to \neg p$.

(ii) The *converse* of $p \to q$ is the proposition $q \to p$.

(iii) The *inverse* of $p \to q$ is the proposition $\neg p \to \neg q$.

Example. Consider the following conditional statement.

If there is electricity, then there is light.

- The contrapositive is "If there is no light, then there is no electricity."
- The converse is "If there is light, then there is electricity."
- The inverse is "If there is no electricity, then there is no light." ♢

Definition. Let p and q be propositions. The *biconditional statement* $p \leftrightarrow q$ is the proposition "p if and only if q". The biconditional statement is true only when both p and q have the same truth value, otherwise it is false.

Biconditional Statement Truth Table

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

We may read $p \leftrightarrow q$ in several ways, such as:

(i) p if and only if q.

(ii) p is sufficient and necessary for q.

(iii) p is equivalent to q.

(iv) If p then q, and conversely.

(v) p iff q ("iff" is an abbreviation for "if and only if").

We collect the truth values of the logical operators in the truth table below.

p	q	$p \wedge q$	$p \vee q$	$p \oplus q$	$p \to q$	$p \leftrightarrow q$
T	T	T	T	F	T	T
T	F	F	T	T	F	F
F	T	F	T	T	T	F
F	F	F	F	F	T	T

Using the propositional variables and these logical operators we can create more and more complicated sentences, called compound propositions. Roughly speaking, a *compound proposition* is a proposition that is obtained starting from the propositional variables $p, q, r, s, \ldots$ and repeatedly applying the logical operators $\neg, \wedge, \vee, \rightarrow, \leftrightarrow$. Strictly speaking, we define compound propositions recursively as follows.

Definition. We define a *compound proposition* recursively as follows.

(i) Any propositional variable is a compound proposition.

(ii) If α and β are compound propositions, then $\neg\alpha$, and $\alpha \wedge \beta$, and $\alpha \vee \beta$, and $\alpha \rightarrow \beta$, and $\alpha \leftrightarrow \beta$ are compound propositions as well.

You may think of compound propositions as the well-formed sentences that we are able to write in our language of propositional logic, in the same way we write sentences in natural languages like Arabic, Greek, and English.

Example. There are infinitely many compound propositions. The following are examples of compound propositions in the variables p, q, r, s. Read them out loud.

- $p \wedge q$
- $\neg p \leftrightarrow q$
- $(p \vee s) \leftrightarrow r$
- $(q \rightarrow r) \rightarrow s$
- $(p \wedge q) \rightarrow \neg r$
- $(p \wedge q) \rightarrow (r \vee s)$
- $\big((p \rightarrow q) \leftrightarrow (r \vee \neg q)\big) \wedge \big(r \vee (p \rightarrow s)\big)$ ◊

The Greek Alphabet

Alpha $\mathrm{A}\ \alpha$	Beta $\mathrm{B}\ \beta$	Gamma $\Gamma\ \gamma$	Delta $\Delta\ \delta$	Epsilon $\mathrm{E}\ \epsilon$	Zeta $\mathrm{Z}\ \zeta$	Eta $\mathrm{H}\ \eta$	Theta $\Theta\ \theta$
Iota $\mathrm{I}\ \iota$	Kappa $\mathrm{K}\ \kappa$	Lambda $\Lambda\ \lambda$	Mu $\mathrm{M}\ \mu$	Nu $\mathrm{N}\ \nu$	Xi $\Xi\ \xi$	Omicron $\mathrm{O}\ \mathrm{o}$	Pi $\Pi\ \pi$
Rho $\mathrm{P}\ \rho$	Sigma $\Sigma\ \sigma$	Tau $\mathrm{T}\ \tau$	Upsilon $\Upsilon\ \upsilon$	Phi $\Phi\ \phi$	Chi $\mathrm{X}\ \chi$	Psi $\Psi\ \psi$	Omega $\Omega\ \omega$

A compound proposition is a proposition, and thus it must be true or false, but not both. The truth value of a compound proposition is determined by the truth values of its constituent propositional variables. Given a compound proposition α, any assignment of truth values to the propositional variables in α determines a truth value of α itself. We may compute the truth value for compound propositions using truth tables as illustrated in the example below.

Example. Find the truth table of the compound proposition $(p \vee \neg q) \to (p \wedge q)$.

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

◊

Precedence of logical operators

We usually use parentheses to specify the order in which the logical operators are applied. When parentheses are not used the table below says which logical operator must be applied before the other. For example, the negation operator takes precedence over all other operators.

Precedence	Logical operator
1	$\neg$
2	$\wedge$
3	$\vee$
4	$\to$
5	$\leftrightarrow$

So when you encounter the compound proposition $\neg p \vee q \wedge r \to s$, you understand the following, $((\neg p) \vee (q \wedge r)) \to s$.

1.2 Applications of Logic

♣ Manipulation of Information

Computers represent information using strings of bits (binary digits). A bit is a symbol which takes two possible values 0 or 1. A *bit string* or a *binary string* is a sequence of bits, that is, a sequence of zeros and ones, for example, 1010011001110001.

Let us agree that a 1 bit represents true (T) and a 0 bit represents false (F). In various programming languages the computer bit operations AND, OR, XOR correspond to the logical operators $\wedge, \vee, \oplus$, respectively. These operations are used to manipulate information as illustrated below.

$$\begin{array}{rr} & 0110110110 \\ \text{AND} & 1100011101 \\ \hline & 0100010100 \end{array}$$

♣ Translating Natural Languages

Given an English sentence, we use different propositional variables to represent its basic components and then use logical operators to connect these components producing a compound proposition which reflects the meaning of the original English sentence.

Translate the English sentences below into the language of propositional logic.

- "If you can access the internet from campus, then either you are a computer hacker or a university student."
  Let t be the proposition "You can access the internet from campus", and h be "You are a computer hacker, and s be "You are a university student". Then the English sentence translates to
  $$t \to (h \vee s).$$
- "Dina is tired, but she went to play football."
  Let t represents the proposition "Dina is tired", and f represents "she went to play football". Then the translation is
  $$t \wedge f.$$
- "You cannot ride the roller coaster if you are under one meter tall unless you are older than 16 years old".
  Let r be "you can ride the roller coaster", and u be "you are under one meter tall", and d be "you are older than 16 years old". Then the translation is
  $$(u \wedge \neg d) \to \neg r. \qquad \Diamond$$

♣ System Specifications

System and Software engineers need to guarantee that the specifications of their computer system are consistent, that is, they can all be true at the same time. To check if some system specifications are consistent, we first express each specification using logical expressions, and then find an assignment of truth values that makes all specifications true.

Example. Are the following system specifications consistent?

"The message is stored in the buffer or transmitted." $(s \vee t)$
"The message is not stored in the buffer." $(\neg s)$
"If the message is stored, then it is transmitted". $(s \to t)$

Clearly, these specifications are consistent as the assignment $s = F$ and $t = T$ will make all of them true. ♢

♣ Google Search

Logical connectives are used in searches of information in the internet. They are called Boolean Searches. For example, if you search on Google for

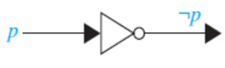
"New AND Cairo AND Universities",

the results will be webpages which include all of the three words New, Cairo, and Universities. One of these pages should be for the American University in Cairo.

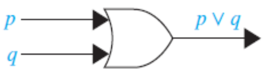
♣ Digital Circuit Design

The idea of applying propositional logic to the design of computer hardware was first observed by the American mathematician and electrical engineer Claude Shannon in his Master's thesis at the Massachusetts Institute of Technology (MIT) in 1938. Shannon is known as "the father of information theory" due to his landmark publication "A Mathematical Theory of Communication", which was published in 1948.

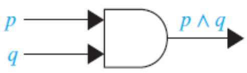
A digital circuit can be constructed from three basic circuits, called *logic gates*: the inverter, the OR gate, and the AND gate. A digital circuit receives input signals. A signal is simply a bit, so 0 for (off), and 1 for (on), and produces output signals.



Inverter

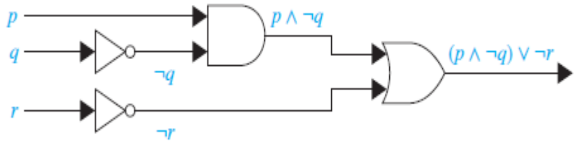


OR gate



AND gate

Example. Construct a digital circuit which receives three input signals p, q, r and produces one output signal given by $(p \wedge \neg q) \vee \neg r$.



◊

♣ Logic Puzzles

We will use the logical reasoning we learned so far to solve some puzzles.

Example. On an island there are two types of people knights and knaves. Knights are honest and always tell the truth. Knaves always lie. Each inhabitant on the island is either a knight or a knave but, of course, not both. You went on a trip to this island and encountered two people A and B. What are the types of A and B if

- Person A said "B is a knight", and
- Person B said "Both of us are of opposite types"?

We will examine the possible cases.
Case 1. Suppose that A is a knight. Then A said the truth, that is, B is a knight. And consequently B said the truth as well, which means that they are of opposite types, however, they are both knights in this case. This means we arrived to a contradiction. Thus, A cannot be a knight.
Case 2. We now know that A must be a knave. This means A lied, and so B is not a knight. This means B must be a knave too. So what B said is a lie, that is, it must be that both are of the same type. There is no problem with this, as they are both knaves.
Therefore, we know that both A and B are knaves. ◊

Exercise. On another island there are three types of people, knights, knaves, and spies. A spy is a person who may or may not lie. When visiting the island, you encounter three people A, B, and C, who are exactly one knight, one knave, and one spy. If possible, determine the types of A, B, and C where

- Person A says "I am the knight",
- Person B says "A is not the knave",
- Person C says "B is not the knave".

Show that A is the knave, B is the spy, and C is the knight. ◊

1.3 Tautologies and Logical Equivalences

In the practice of writing English sentences we often replace a word with its synonym. For example, we may replace the sentence "mathematics is important" with the sentence "mathematics is crucial". Similarly, in the construction of mathematical arguments we replace a statement with another statement which has the same truth value. In this section we study such statements.

Recall that a compound proposition is a statement formed from propositional variables and logical operators. For example $\neg(p \wedge q)$ and $(p \vee q) \rightarrow r$ are compound propositions. Moreover, whenever we assign truth values to the propositional variables within a compound proposition, a resulting truth value is determined for the compound proposition itself. We usually say "proposition" instead of "compound proposition".

Definition.

- A *tautology* is a compound proposition which is always true, no matter what the truth values of its propositional variables.
- A *contradiction* (or *fallacy*) is a compound proposition which is always false, no matter what the truth values of its propositional variables.
- A *contingency* is a compound proposition which is neither a tautology nor a contradiction.

Remark. In a truth table, all the entries in the column of a tautology are true. On the other hand, all the entries in the column of a contradiction are false. Furthermore, the column of a contingency must contain at least one true value and one false value.

Example. The proposition $p \vee \neg p$ is a tautology, while $p \wedge \neg p$ is a contradiction.

p	$\neg p$	$p \vee \neg p$	$p \wedge \neg p$
T	F	T	F
F	T	T	F

Example. The compound proposition $p \rightarrow (p \vee r)$ is a tautology.

p	r	$p \vee r$	$p \rightarrow (p \vee r)$
T	T	T	T
T	F	T	T
F	T	T	T
F	F	F	T

We now give a characterisation of when a conditional statement $\alpha \to \beta$ is a tautology where α and β are compound propositions.

The following two statements are either both true or both false.

(i) The conditional statement $\alpha \to \beta$ is a tautology.

(ii) Whenever the hypothesis α is true, the conclusion β is true as well.

To check this, we will show that if one statement is true, then the other is also true. First, suppose that Statement (i) is true. So $\alpha \to \beta$ is a tautology. We need to show that whenever α is true, then β must be true. So suppose α is true. As β is a proposition, it has only two choices: either β is true or β is false. If β were false, then $\alpha \to \beta$ is $T \to F$ which is false, but this cannot happen because we assumed that $\alpha \to \beta$ is a tautology (it is always true). Thus, β cannot be false, and so β must be true. It follows that Statement (ii) holds true.
Second, let us suppose that Statement (ii) is true. So suppose whenever α is true, β must be true. We need to show that $\alpha \to \beta$ is a tautology. We proceed as follows. Since α is a proposition, it is either true or false. If it were false, then $\alpha \to \beta$ is true regardless of what is the truth value of β. If α is true, then β is true too because we are assuming that Statement (ii) holds, and so $\alpha \to \beta$ is true in this case as well. Thus, $\alpha \to \beta$ is always true, meaning that $\alpha \to \beta$ is a tautology. So Statement (i) holds true.

Remark. The discussion above gives us another way to check when a conditional statement is a tautology. A conditional statement is a tautology if whenever its hypothesis is true, its conclusion is also true.

Example. Show that the compound proposition $p \to (p \vee r)$ is a tautology.
We have already checked in the previous example that this compound proposition is a tautology. We may use the characterisation we just discussed to establish the same result. We will show that whenever the hypothesis is true, the conclusion is also true. So suppose the hypothesis p is true, then the conclusion $p \vee r$ is $T \vee r$ which is true regardless of the truth value of r. We conclude that $p \to (p \vee r)$ is a tautology, because whenever its hypothesis is true, it follows that its conclusion is also true. ♢

Example. Show that the following proposition is a tautology.

$$\big((p \vee q) \wedge (\neg p \vee r)\big) \to (q \vee r).$$

We will show that whenever the hypothesis of this conditional statement is true, its conclusion must be true as well. So suppose that $(p \vee q) \wedge (\neg p \vee r)$ is true. By definition of conjunction, it follows that $(p \vee q)$ is true and $(\neg p \vee r)$ is true as well. As $(p \vee q)$ is true, by definition of disjunction, it must be that either p is true or q is true. On one hand, if p is true, then $\neg p$ is false, and since $(\neg p \vee r)$ is true, it must be that r is true, and so the conclusion $(q \vee r)$ is true regardless of the value of q.

On the other hand, if q is true, then the conclusion $(q \vee r)$ is true regardless of the value of r. Thus, we showed that when the hypothesis is true, the conclusion must be true. Therefore, $\big((p \vee q) \wedge (\neg p \vee r)\big) \to (q \vee r)$ is a tautology. As an exercise, construct its truth table and see that all its values are true. $\Diamond$

♣ Logical Equivalence

It is certainly possible that two different compound propositions have the same truth value in all scenarios, that is, have exactly the same truth table (Why?). Let us study such propositions in greater detail.

Definition. Let α and β be compound propositions. We say that α and β are *logically equivalent* if they have the same truth value in every case of assigning truth values to their propositional variables. We write $\alpha \equiv \beta$ when α and β are logically equivalent.

The compound propositions α and β being logically equivalent means that in every case either α and β are both true or both false. This means that the columns giving their truth values are identical.

Moreover, $\alpha \equiv \beta$ means that the proposition $\alpha \leftrightarrow \beta$ is a tautology. To see this, suppose that $\alpha \equiv \beta$. Then in every case either α and β are both true or both false. This means that in every case $\alpha \leftrightarrow \beta$ is true, that is, $\alpha \leftrightarrow \beta$ is a tautology.

Remark. The symbol $\equiv$ is not a logical connective in propositional logic, and $\alpha \equiv \beta$ is not a compound proposition.

Example. Show that $(p \to q) \equiv (\neg p \vee q)$.

p	q	$p \to q$	$\neg p$	$\neg p \vee q$
T	T	T	F	T
T	F	F	F	F
F	T	T	T	T
F	F	T	T	T

The truth values of propositions $(p \to q)$ and $(\neg p \vee q)$ are the same in all possible cases (the third and fifth columns are identical), which means they are logically equivalent. $\Diamond$

Example. Show that a conditional statement is equivalent to its contrapositive, that is, show that $(p \to q) \equiv (\neg q \to \neg p)$.

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

Notice that the third column and sixth column are identical. ◊

Example. Show that $\neg(p \wedge q) \not\equiv \neg p \wedge \neg q$.
To show that $\neg(p \wedge q)$ is not logically equivalent to $\neg p \wedge \neg q$ we need to show that both propositions don't always have the same truth value. In other words, we need to find at least one scenario where one of them is true while the other false. For instance, consider the case when $p = T$ and $q = F$. Then $\neg(p \wedge q) = \neg(T \wedge F) = \neg F = T$. However, $\neg p \wedge \neg q = (\neg T) \wedge (\neg F) = F \wedge T = F$. Since they have different truth values for the same assignment they are not logically equivalent. Alternatively, construct the corresponding truth tables and check that they are not identical. ◊

De Morgan's Laws

Given any compound propositions p and q, the following hold.

(*i*) The negation of the conjunction is equivalent to the disjunction of the negations.

$$\neg(p \wedge q) \equiv \neg p \vee \neg q$$

(*ii*) The negation of the disjunction is equivalent to the conjunction of the negations.

$$\neg(p \vee q) \equiv \neg p \wedge \neg q$$

Let us prove the second law of De Morgan: $\neg(p \vee q) \equiv \neg p \wedge \neg q$.

p	q	$p \vee q$	$\neg(p \vee q)$	$\neg p$	$\neg q$	$\neg p \wedge \neg q$
T	T	T	F	F	F	F
T	F	T	F	F	T	F
F	T	T	F	T	F	F
F	F	F	T	T	T	T

Compare the fourth and seventh columns. The truth values of both propositions are identical in all cases, showing that $\neg(p \vee q)$ and $\neg p \wedge \neg q$ are logically equivalent.

Example. Use propositional logic to express the negation of "Laila has a mobile phone and she has a computer".
Let p be the proposition "Laila has a mobile phone", and q be "Laila has a computer". The sentence above is $p \wedge q$. The negation of $p \wedge q$ is $\neg(p \wedge q)$ which is equivalent to $(\neg p \vee \neg q)$. So the negation of the statement is: "Either Laila does not have a mobile phone or she does not have a computer". ◊

Famous Logical Equivalences

Let p, q, r be compound propositions, and $\mathbf{T}$ be a compound proposition which is a tautology and $\mathbf{F}$ be a contradiction.

- Identity Laws

  $p \wedge \mathbf{T} \equiv p$
  $p \vee \mathbf{F} \equiv p$

- Domination Laws

  $p \vee \mathbf{T} \equiv \mathbf{T}$
  $p \wedge \mathbf{F} \equiv \mathbf{F}$

- Idempotent Laws

  $p \vee p \equiv p$
  $p \wedge p \equiv p$

- Double-negation Law

  $\neg(\neg p) \equiv p$

- Commutative Laws

  $p \vee q \equiv q \vee p$
  $p \wedge q \equiv q \wedge p$

- Associative Laws

  $(p \vee q) \vee r \equiv p \vee (q \vee r)$
  $(p \wedge q) \wedge r \equiv p \wedge (q \wedge r)$

- Distributive Laws

  $p \vee (q \wedge r) \equiv (p \vee q) \wedge (p \vee r)$
  $p \wedge (q \vee r) \equiv (p \wedge q) \vee (p \wedge r)$

- De Morgan's Laws

  $\neg(p \wedge q) \equiv \neg p \vee \neg q$
  $\neg(p \vee q) \equiv \neg p \wedge \neg q$

- More Logical Equivalences

  $p \rightarrow q \equiv \neg p \vee q$
  $p \rightarrow q \equiv \neg q \rightarrow \neg p$
  $p \vee q \equiv \neg p \rightarrow q$

$p \wedge q \equiv \neg(p \to \neg q)$
$p \leftrightarrow q \equiv (p \to q) \wedge (q \to p)$
$p \leftrightarrow q \equiv \neg p \leftrightarrow \neg q$

- General form of De Morgan's Laws.

  $\neg(p_1 \vee p_2 \vee \cdots \vee p_n) \equiv \neg p_1 \wedge \neg p_2 \wedge \cdots \wedge \neg p_n$
  $\neg(p_1 \wedge p_2 \wedge \cdots \wedge p_n) \equiv \neg p_1 \vee \neg p_2 \vee \cdots \vee \neg p_n$

Example. Prove the following Distributive Law.

$$p \vee (q \wedge r) \equiv (p \vee q) \wedge (p \vee r).$$

p	q	r	$q \wedge r$	$p \vee (q \wedge r)$	$p \vee q$	$p \vee r$	$(p \vee q) \wedge (p \vee r)$
T	T	T	T	T	T	T	T
T	T	F	F	T	T	T	T
T	F	T	F	T	T	T	T
T	F	F	F	T	T	T	T
F	T	T	T	T	T	T	T
F	T	F	F	F	T	F	F
F	F	T	F	F	F	T	F
F	F	F	F	F	F	F	F

The fifth and eighth columns are identical. Thus the truth values of both propositions are the same in all possible cases, which means they are logically equivalent. ♢

Example. Show that $\neg(p \to q) \equiv (p \wedge \neg q)$.

$$\neg(p \to q) \equiv \neg(\neg p \vee q) \equiv \neg(\neg p) \wedge \neg q \equiv p \wedge \neg q. \qquad \diamondsuit$$

Example. Show that $\neg(p \vee (\neg p \wedge q) \equiv \neg p \wedge \neg q$.

$$\begin{aligned} \neg(p \vee (\neg p \wedge q)) &\equiv \neg((p \vee \neg p) \wedge (p \vee q)) \\ &\equiv \neg(\mathbf{T} \wedge (p \vee q)) \\ &\equiv \neg(p \vee q) \\ &\equiv \neg p \wedge \neg q. \qquad \diamondsuit \end{aligned}$$

Example. Show that the proposition $(p \wedge q) \to (p \vee q)$ is a tautology.

$$\begin{aligned} (p \wedge q) \to (p \vee q) &\equiv \neg(p \wedge q) \vee (p \vee q) \\ &\equiv (\neg p \vee \neg q) \vee (p \vee q) \\ &\equiv (\neg p \vee p) \vee (\neg q \vee q) \\ &\equiv \mathbf{T} \vee \mathbf{T} \equiv \mathbf{T}. \qquad \diamondsuit \end{aligned}$$

A compound proposition is called *satisfiable* if there exists an assignment of truth values to its propositional variables which makes it true. Otherwise, it is always false and we say it is unsatisfiable. Clearly, an unsatisfiable proposition is a contradiction.

Example.

- $(p \vee \neg q) \wedge (q \vee \neg r) \wedge (r \vee \neg p)$ is satisfiable. Assign $p = q = r = T$.
- $(p \vee q \vee r) \wedge (\neg p \vee \neg q \vee \neg r)$ is satisfiable. Assign $p = T$ and $q = r = F$. $\Diamond$

Suppose we are given the truth table of some compound proposition α, however, we do not know what α really is. In the next example we will describe a method to construct a compound proposition β such that $\beta \equiv \alpha$.

Example. Use propositional variables p, q, r to construct a compound proposition α realizing the truth table below.

Case	p	q	r	α
1	**T**	**T**	**T**	**T**
2	T	T	F	F
3	T	F	T	F
4	**T**	**F**	**F**	**T**
5	F	T	T	F
6	**F**	**T**	**F**	**T**
7	**F**	**F**	**T**	**T**
8	F	F	F	F

The proposition α can be written as a disjunction of conjunctions of the variables p, q, r and their negations. Such form is called *disjunctive normal form.*

Consider only the rows in which α is true, namely rows $1, 4, 6,$ and 7. For instance in row 4 we have that $\alpha = T$ in the case when $p = T, q = F, r = F$. In this particular case, both α and $p \wedge \neg q \wedge \neg r$ are true. Actually $p \wedge \neg q \wedge \neg r$ is true only in this case, and false otherwise. Repeating this for the other rows where α is true and taking the disjunction of the resulting four propositions we obtain a proposition in disjunctive normal form equivalent to α.

$$\alpha \equiv (p \wedge q \wedge r) \vee (p \wedge \neg q \wedge \neg r) \vee (\neg p \wedge q \wedge \neg r) \vee (\neg p \wedge \neg q \wedge r). \qquad \Diamond$$

Observation.
Any compound proposition is logically equivalent to one in disjunctive normal form.

1.4 Predicate Logic

Propositional logic cannot express the meaning of all statements in mathematics and natural language. For example, we need a precise language and rules to express the following argument.

- All men are mortal.
- Socrates is a man.
- Therefore, Socrates is mortal.

We will introduce a more powerful type of logic called *predicate logic* using which we can express the argument above and a vast number of mathematical statements and arguments.

Let us start by examining the statement "*x is a prime number*". This sentence has two parts:

(i) The subject of the statement which is the variable x.

(ii) The predicate "is a prime number" which states the property of the subject.

We let $P(x)$ denotes the statement "x is a prime number" where P is the predicate or the property of the variable x. Once we assign a value for x, the statement $P(x)$ becomes a proposition and hence has a truth value: true or false. For example, $P(6)$ says "6 is a prime number" which is a false proposition, and $P(17)$ says "17 is a prime number" which is a true proposition.

Example. Consider the predicate $Q(x, y)$ which is the statement "$x = y + 3$". Here the property Q of numbers x and y is that x is 3 more than y. Let us assign the value 2 for x and 5 for y in the predicate $Q(x, y)$. We write $Q(2, 5)$ for such an assignment. Now $Q(2, 5)$ stands for the statement "$2 = 5 + 3$" which is a false proposition. On the other hand, when we assign 5 for x and 2 for y we obtain $Q(5, 2)$ which stands for "$5 = 2 + 3$" and it is a true proposition. ♢

Definition. A *predicate* is a descriptive statement involving some variables which becomes a proposition once values are assigned to these variables. These values come from a fixed set of objects called the *domain* or the *universe of discourse.* We write $P(x_1, x_2, \ldots, x_n)$ for a predicate involving the variables $x_1, x_2, \ldots, x_n$.

Note that a predicate is a statement which is neither true nor false, however, once we specify values for its variables it becomes a proposition which has a truth value: true or false, but not both. Predicates which involve one variable are called unary predicates, those involving two variables are called binary predicates, and those with three variables are called ternary predicates. In general we say an n-ary predicate for a predicate involving n many variables.

Example. Consider the ternary predicate $R(x, y, z)$ which states that "$x + y = z$", where the domain of the variables x, y, z is the set of all real numbers.

Then $R(1,2,3)$ means that we assign $x = 1$, $y = 2$, and $z = 3$ in the predicate $R(x,y,z)$. As a result, we get the true proposition $1+2=3$. On the other hand, $R(0,0,1)$ is a false proposition because $0+0 \neq 1$. $\Diamond$

Remark. Note that all of $x = y$, and $x < y$, and $x \leq y$, and $x > y$, and $x \geq y$ are binary predicates in the variables x and y. Try to substitute number values in these predicates and see how you get propositions. For example, $1 = 1$ and $3 < 5$ are true propositions, while $2 = 7$ and $4 \geq 9$ are false propositions.

Sometimes we have predicates which have different domains for different variables as illustrated in the next example.

Example. Consider the binary predicate $T(x,y)$ which is the statement "x is the capital of y", where the domain of the variable x is all cities and the domain of y is all countries.
Now T(Cairo, Egypt) means that we substitute "Cairo" for x and "Egypt" for y in the predicate $T(x,y)$. Consequently, we get a proposition which states that "Cairo is the capital of Egypt". $\Diamond$

As predicates become propositions when values are assigned to their variables, we can use the logical connectives we developed in propositional logic to obtain new predicates from old ones.

Example. Suppose we are given the predicates $P(x)$ and $Q(x)$. You may think of $P(x)$ as the statement "x is passionate" and $Q(x)$ as "x is quarantined", where the domain of the variables consists of all people. Consider the following new predicates.

- $S(x)$ is the predicate $P(x) \wedge Q(x)$.
  This means that x has property S if x has both properties P and Q. So $S(x)$ means that x is both passionate and quarantined.

- $T(x,y)$ is the predicate $P(x) \wedge \neg Q(y)$.
  This means that a pair (x,y) has property T if x has property P and y does not have property Q.

- $W(x,y,z)$ is the predicate $P(z) \vee (Q(x) \leftrightarrow Q(y))$.
  The triple (x,y,z) has property W if either z has property P, or both x and y have property Q, or both x and y do not have property Q. $\Diamond$

Substituting values for variables in predicates is not the only way to change a predicate into a proposition. We now introduce new logical symbols, called quantifiers, which when applied to predicates, we create propositions.

♣ Quantification

In the English language we use words such as *all, every, many, some, few*, and *none* for quantification. Think of sentences we write when we describe the number or quantity of objects having a certain property. For instance, we may say "*some people are trustworthy*" and "*all birds have wings*". We aim to equip the logic we are building with symbols called quantifiers to do the job of quantification. This type of logic which deals with predicates, logical connectives, and quantifiers is called *predicate logic.*

Let $P(x)$ be a predicate in the variable x with some specified domain. Recall that $P(x)$ says that "x has property P". We will consider two types of quantification: *universal quantification* and *existential quantification.* We use universal quantification to say that every element in the domain has property P. And we use existential quantification to say that there exists at least one element in the domain with property P.

Definition. Let $P(x)$ be a predicate.

- The *universal quantification* of $P(x)$ is the statement "for all values of x in the domain, we have that $P(x)$ is true".
  The universal quantification of $P(x)$ is denoted by $\forall x P(x)$, and the symbol $\forall$ is called the *universal quantifier.*

- The *existential quantification* of $P(x)$ is the statement "there exists an element x in the domain such that $P(x)$ is true".
  The existential quantification of $P(x)$ is denoted by $\exists x P(x)$ and the symbol $\exists$ is called the *existential quantifier.*

The table below lists several ways to read the quantified expressions $\forall x$ and $\exists x$.

$\forall x$	$\exists x$
For all x	There exists an x
For every x	There is an x
For each x	There is at least one x
For any x	For some x
For arbitrary x	For at least one x
Given any x	

We remark that both $\forall x P(x)$ and $\exists x P(x)$ are propositions, that is, each is a statement which is either true or false, but not both. The universal statement $\forall x P(x)$ is declared true if every element in the domain has property P. And $\forall x P(x)$ is false if not every element in the domain has property P, meaning that at least one of the elements in the domain does not have the property P. Such element is called a *counterexample* to the universal statement $\forall x P(x)$. On the other hand, the existential statement $\exists x P(x)$ is true if at least one of the domain elements has property P. And $\exists x P(x)$ is false if no element in the domain has property P.

Note that the truth value of a quantified statement depends on the domain under consideration, and the truth value could change when we change the domain. Without specifying the domain, the truth value of a quantified statement is not defined. The table below summarises when universal and existential statements are true and when they are false.

Statement	True	False
$\forall x\, P(x)$	When for every x in the domain, we have that $P(x)$ is true.	When there is some element c in the domain such that $P(c)$ is false.
$\exists x\, P(x)$	When there is some element a in the domain such that $P(a)$ is true.	When for every x in the domain, we have that $P(x)$ is false.

Example. Let $P(x)$ be the predicate "$x + 1 > x$". What is the truth value of $\forall x\, P(x)$ and $\exists x\, P(x)$ when the domain is the set of real numbers?
The universal statement $\forall x\, P(x)$ states that for every real number x, we have that $x + 1 > x$. This is true because if we choose any number x and add 1 to it, we get a bigger number. So, $P(x)$ is true for every number x. Thus, $\forall x\, P(x)$ is a true proposition. The existential statement $\exists x\, P(x)$ is also true, because we can find some number, say 7, such that $P(7)$ is true since $7 + 1 > 7$. $\lozenge$

Example. Let $Q(x)$ be the predicate "$x < 2$". When the domain is the real numbers, is $\forall x\, Q(x)$ true? Is $\exists x\, Q(x)$ true?
The statement $\forall x\, Q(x)$ is false because it is not for every real number x, we have that $Q(x)$ is true. A counterexample is the element 4 from the domain because $P(4)$ is false since $4 < 2$ is false. The statement $\exists x\, Q(x)$ is true because there is an element in the domain with property Q, for example take the number 0 and see that $Q(0)$ is true since $0 < 2$ is true. $\lozenge$

Example. Let $P(x)$ be the statement "$x^2 \geq x$". So x has property P if the square of x is greater than or equal to itself. Evaluate the truth value of $\forall x\, P(x)$ and $\exists x\, P(x)$ when:

- The domain is the set of real numbers.
  Then $\forall x\, P(x)$ is false since the number $\frac{1}{2}$ serves as a counter example because $P(\frac{1}{2})$ is false since $(\frac{1}{2})^2 \not\geq \frac{1}{2}$. However, $\exists x\, P(x)$ is true because 3 is a real number and $P(3)$ is true.

- The domain is the set of positive integers $\{1, 2, 3, \dots\}$.
  Then $\forall x\, P(x)$ is true because the square of any positive integer is greater than or equal to itself: $1^2 \geq 1$, and $2^2 \geq 2$, and $3^2 \geq 3$, and so on. Also $\exists x\, P(x)$ is true because 3 is a positive integer and $P(3)$ is true. $\Diamond$

Example. Let $S(x)$ be the statement "$x^2 + 1 = 0$" where the domain is the real numbers. Then $\exists x\, S(x)$ is false because when we add 1 to the square of any real number we get a number bigger than or equal to 1, so we never get 0, this is mainly because the square of any real number is nonnegative. In other words, no real number satisfies the property S, and so $\exists x\, S(x)$ is false in the real numbers.
However, if we change our domain to the set of complex numbers, then $\exists x\, S(x)$ becomes a true statement since the complex number i is a witness for the existential statement $\exists x\, S(x)$ because $S(i)$ is true as $i^2 + 1 = (-1) + 1 = 0$. $\Diamond$

In the special case where the domain of variables is finite, say the domain is the set $\{a_1, a_2, \dots, a_n\}$, then we can use conjunction and disjunction to express universal and existential quantification, respectively.

$$\forall x P(x) \text{ means } P(a_1) \wedge P(a_2) \wedge \cdots \wedge P(a_n),$$
$$\exists x P(x) \text{ means } P(a_1) \vee P(a_2) \vee \cdots \vee P(a_n).$$

Example. Let us work in the domain $\{1, 2, 3, 4\}$ and the predicates $P(x)$ which is the statement "$x^2 < 10$" and $Q(x)$ which is the statement "$x^2 > 10$".
The statement $\forall x\, P(x)$ means that every element from $1, 2, 3, 4$ has property P, so it has the same meaning as the compound proposition

$$P(1) \wedge P(2) \wedge P(3) \wedge P(4).$$

Since $P(4)$ is false, we have that $P(1) \wedge P(2) \wedge P(3) \wedge P(4)$ is false, so $\forall x\, P(x)$ is false as well.
The statement $\exists x\, Q(x)$ means that at least one of the elements $1, 2, 3, 4$ has property Q, in other words, it means that

$$Q(1) \vee Q(2) \vee Q(3) \vee Q(4).$$

Since $Q(4)$ is true, we have that $Q(1) \vee Q(2) \vee Q(3) \vee Q(4)$ is true, so $\exists x\, Q(x)$ is also true. $\Diamond$

In the language of predicate logic we create sentences using predicates, logical connectives, and quantifiers. Such statements are called *(first-order) formulas.* Let us explain how we construct formulas in predicate logic.

Definition. We recursively define a *formula* in predicate logic as follows.

(i) Any predicate is a formula.

(ii) If ϕ and ψ are formulas then $\neg\phi$, and $\phi \wedge \psi$, and $\phi \vee \psi$, and $\phi \rightarrow \psi$, and $\phi \leftrightarrow \psi$ and $\forall x\, \phi$, and $\exists x\, \phi$ are formulas as well.

Example. Given unary predicates $P(x)$ and $Q(x)$, below are formulas (or statements) built using these predicates.

- $P(x) \vee \neg Q(y)$
- $P(x) \rightarrow Q(x)$
- $\forall x\, P(x)$
- $\exists y\, Q(y)$
- $(\forall x\, P(x)) \wedge (\exists y\, Q(y))$
- $\forall x\, (P(x) \wedge Q(x))$
- $\forall x \exists y\, (P(x) \vee Q(y))$
- $\forall x\, (P(x) \rightarrow Q(x))$
- $\exists x \exists y \exists z\, (x \neq y \wedge y \neq z \wedge x \neq z \wedge P(x) \wedge P(y) \wedge P(z))$ ♢

We now discuss when two statements (or formulas) in predicate logic are considered logically equivalent.

Definition. Let ϕ and ψ be statements in predicate logic. We say that ϕ is *logically equivalent* to ψ, and write $\phi \equiv \psi$, if in every domain both ϕ and ψ have the same truth value.

We cannot use truth tables to show $\phi \equiv \psi$, but rather we show that when we work in any domain, if ϕ is true in that domain, then so is ψ, and vice versa. To show that $\phi \not\equiv \psi$, all what you need to do is to find one domain where ϕ and ψ have different truth values.

Example. Show that $\forall x(P(x) \wedge Q(x)) \equiv (\forall x\, P(x)) \wedge (\forall x\, Q(x))$.

Choose any domain you like and fix it. First, suppose that $\forall x(P(x) \wedge Q(x))$ is true. Let a be any element in the domain. Then $P(a) \wedge Q(a)$ is true, and so $P(a)$ is true. This means that for any a in the domain, we have that $P(a)$ is true. That is, $\forall x P(x)$ is true. Similarly, we show $\forall x Q(x)$ is also true. Thus, we showed both $\forall x P(x)$ and $\forall x Q(x)$ are both true. Therefore $\forall x P(x) \wedge \forall x Q(x)$ is true.
Second, suppose $\forall x P(x) \wedge \forall x Q(x)$ is true in the domain. Let a be any element in this domain. We know that $\forall x P(x)$ and $\forall x Q(x)$ are both true. So $P(a)$ is true, and so is $Q(a)$. It follows that the proposition $P(a) \wedge Q(a)$ is also true. Thus, $\forall x(P(x) \wedge Q(x))$ is true in the domain.
Therefore, we showed that $\forall x(P(x) \wedge Q(x))$ and $\forall x P(x) \wedge \forall x Q(x)$ are both true or both false in any domain we choose, that is, they are logically equivalent. ♢

The next example shows that we cannot distribute the existential quantifier over a conjunction as we did with the universal quantifier in the previous example.

Example. Show that $\exists x(P(x) \wedge Q(x)) \not\equiv \exists x\, P(x) \wedge \exists x\, Q(x)$.
To show that these two statements are not logically equivalent we need to find at least one domain with interpretations of the predicates such that these two statements have different truth values. For instance, consider the domain of all integer numbers, and interpret $P(x)$ as "x is even", and $Q(x)$ as "x is odd". Consequently, the statement $\exists x(P(x) \wedge Q(x))$ is false because there exists no integer which is both even and odd. However, the statement $\exists x\, P(x) \wedge \exists x\, Q(x)$ is true because there exists an even number in the domain like 12, and also there exists an odd number like 13. ♢

Below are more logical equivalences in predicate logic. We have established the first one, showing that the rest hold is left as an exercise to the reader.

- $\forall x(P(x) \wedge Q(x)) \equiv \forall x\, P(x) \wedge \forall x\, Q(x)$.
- $\exists x(P(x) \vee Q(x)) \equiv \exists x\, P(x) \vee \exists x\, Q(x)$.
- $\forall x(P(x) \rightarrow Q(x)) \equiv \forall x(\neg P(x) \vee Q(x))$.
- De Morgan's Laws for Predicate Logic.
    - $\neg \forall x P(x) \equiv \exists x \neg P(x)$.
    - $\neg \exists x P(x) \equiv \forall x \neg P(x)$.

Example. Consider the predicate "$x^2 > x$" where the domain is the real numbers.

- $\neg \forall x(x^2 > x) \equiv \exists x \neg (x^2 > x) \equiv \exists x(x^2 \leq x)$.
- $\neg \exists x(x^2 = x) \equiv \forall x \neg (x^2 = x) \equiv \forall x(x^2 \neq x)$. ♢

Example. Show that $\neg \forall x(P(x) \rightarrow Q(x)) \equiv \exists x(P(x) \wedge \neg Q(x))$.

$$\begin{aligned} \neg \forall x(P(x) \rightarrow Q(x)) &\equiv \neg \forall x(\neg P(x) \vee Q(x)) \\ &\equiv \exists x \neg(\neg P(x) \vee Q(x)) \\ &\equiv \exists x(P(x) \wedge \neg Q(x). \end{aligned}$$ ♢

Let us use predicate logic to express some English sentences.

Example. Express using predicate logic (that is, using predicates, logical connectives, and quantifiers) the following English sentences. Use the predicate $H(x)$ which stands for "x is a human", and $T(x)$ for "x can think", where the domain of x consists of all mammals.

(i) "All humans can think". Translation: $\forall x(H(x) \rightarrow T(x))$.

(ii) "Some humans can think". Translation: $\exists x(H(x) \wedge T(x))$.

(iii) What is the negation of the first sentence?

$$\begin{aligned}\neg\forall x(H(x) \to T(x)) &\equiv \neg\forall x(\neg H(x) \vee T(x))\\ &\equiv \exists x\neg(\neg H(x) \vee T(x))\\ &\equiv \exists x(H(x) \wedge \neg T(x)).\end{aligned}$$

Thus, the negation of "All humans can think" is $\exists x(H(x) \wedge \neg T(x))$ which says "There is a human which cannot think".

(iv) What is the negation of the second sentence?

$$\begin{aligned}\neg\exists x(H(x) \wedge T(x)) &\equiv \forall x\neg(H(x) \wedge T(x))\\ &\equiv \forall x(\neg H(x) \vee \neg T(x))\\ &\equiv \forall x(H(x) \to \neg T(x)).\end{aligned}$$

Thus, the negation of "Some humans can think" is $\forall x(H(x) \to \neg T(x))$ which says "Every human cannot think". $\Diamond$

1.5 Nested Quantifiers

Given a statement ϕ in predicate logic, we may form the universal quantification of ϕ which is the statement $\forall x\, \phi$. Here we call ϕ the *scope* of the quantifier $\forall x$. In this section, we will study statements where one quantifier is in the scope of another. For example, in the statement $\forall x \exists y\, P(x, y)$, the scope of $\forall x$ is $\exists y\, P(x, y)$, and the scope of $\exists y$ is $P(x, y)$.

Remark. Any occurrence of the variable x in the scope of a quantifier $\exists x$ or $\forall x$ is called *bound*. Otherwise, when an occurrence of x does not appear in the scope of $\exists x$ or $\forall x$ we call it *free*. For example, in the statement $\exists x(x+y=1)$, the occurrence of x in the scope $(x+y=1)$ is bounded by $\exists x$, however, the occurrence of y is free since it is not in the scope of $\exists y$ or $\forall y$.

What is the meaning of $\forall x \exists y\, P(x, y)$?

First, this statement is read from left to right. So it reads "*For all x in the domain, there exists some y in the domain such that $P(x, y)$ is true*". For clarity, let us rewrite this statement using parentheses as follows, $\forall x\big(\exists y\, P(x, y)\big)$. This means that every x in the domain has the property $\big(\exists y\, P(x, y)\big)$. An element x having this property means that there is some element y such that the pair (x, y) has property P. For instance, suppose that the predicate $P(x, y)$ states that "x is the parent of y", and say that the universe is all living people. Then our statement $\forall x \exists y\, P(x, y)$ says that every person x has the property $\exists y\, P(x, y)$, and the property $\exists y\, P(x, y)$ says that there is somebody whose parent is x. So $\forall x \exists y\, P(x, y)$ means that for any person x we choose, we can find somebody y such that x is the parent of y. In other words, any person is the parent of some person. In short, it says, every person has a child. Of course, $\forall x \exists y\, P(x, y)$ is false in our real world because many people do not have children. Now, as an exercise, think of the meaning of $\exists x \forall y\, P(x, y)$.

Example. Let the universe of variable be all positive integers: $1, 2, 3, \ldots$, and let $P(y)$ be "y is prime". Express the following statements in English.

(i) $\forall x\, \exists y\, (x < y)$.
For any positive integer x, there exists a positive integer y such that $x < y$. That is, every positive integer is less than some integer.

(ii) $\exists x\, \forall y\, (y \leq x)$.
There exists a positive integer x such that for any positive integer y we have that $y \leq x$. That is, there is some positive integer which is greater than or equal to any positive integer. In short, there is a maximum positive integer.

(iii) $\forall x\, \exists y\, (P(y) \wedge (x < y))$.
For any positive intger x, there is a positive integer y such that y is a prime number and $x < y$. That is, given any positive integer, we can find a prime number bigger than it. $\Diamond$

Example. Let the domain of the variables be the set $\mathbb{R}$ of all real numbers. Express the following logical statements in English.

- $\forall x\,\forall y\,((0 < x \wedge y < 0) \to (xy < 0))$.
  The statement says: "for any real numbers x and y, if x is positive and y is negative, then their product xy is negative". In short, it says that the product of any positive number with a negative number is a negative number.

- $\forall x \forall y (x + y = y + x)$.
  The statements says: "for all real numbers x and y we have that $x+y = y+x$ is true. In short, it says that addition of real numbers is a commutative operation.

- $\forall x \exists y (x \neq 0 \to xy = 1)$.
  The statement says: "for any real number x, there exists a real number y such that if x is nonzero, then the product of x and y is equal to 1. In short, it says that every nonzero real number has a multiplicative inverse. $\Diamond$

Thinking of Quantification as Loops

When the domain of the variables contains finitely many elements it is useful to think of quantification of more than one variable in terms of nested loops. This way of thinking is still helpful even when we have infinitely many elements in the domain. Below we discuss different ways of quantifying a binary predicate $P(x, y)$. We study the statements $\forall x \forall y P(x, y)$, and $\forall x \exists y P(x, y)$, and $\exists x \forall y P(x, y)$, and $\exists x \exists y P(x, y)$.

Moreover, for each of these statements we give an example of a scenario where it is true, and another one where it is false. In these examples, the variables domain consists of three people: a, b, c. And we interpret $P(x, y)$ as "x likes y". We present these scenarios by means of tables where the truth value in a cell of these tables is the truth value of the predicate $P(x, y)$ when assigning the corresponding row value for x and column value for y. Keep an eye on the order of quantifiers.

- $\forall x \forall y P(x, y)$ says "for any x, for any y, we have that $P(x, y)$ is true".

  To check whether $\forall x \forall y P(x, y)$ is true, we loop through all the values of x, and for each x we loop through the values of y. If for each value of x, we get $P(x, y)$ is true for all values of y, then $\forall x \forall y P(x, y)$ is true. However, if for a particular value of x we hit a value of y for which $P(x, y)$ is false, then we have determined that $\forall x \forall y P(x, y)$ is false.

  Example. When $P(x, y)$ is the statement "x likes y", then $\forall x \forall y P(x, y)$ means that "Everyone likes everyone". Apply thinking via loops to the following two scenarios of a domain containing three people a, b, c.

$\forall x \forall y P(x, y)$ is true

y x	a	b	c
a	T	T	T
b	T	T	T
c	T	T	T

$\forall x \forall y P(x, y)$ is false

y x	a	b	c
a	T	T	T
b	T	T	F
c	T	T	T

- $\forall x \exists y P(x, y)$ says "for any x, there exists some y such that $P(x, y)$ is true".

  To check whether $\forall x \exists y P(x, y)$ is true, we loop through all the values of x, and for each x we start looping through the values of y until we find some y for which $P(x, y)$ is true. If for each value of x, we hit a y for which $P(x, y)$ is true, then $\forall x \exists y P(x, y)$ is true. Otherwise, if for a particular x, we never hit a y for which $P(x, y)$ is true, then we have shown that $\forall x \exists y P(x, y)$ is false.

  Example. When $P(x, y)$ is the statement "x likes y", then $\forall x \exists y P(x, y)$ means that "Every person likes someone". Apply thinking via loops to the following two scenarios of a domain containing three people a, b, c.

$\forall x \exists y P(x, y)$ is true

y x	a	b	c
a	F	T	F
b	F	F	T
c	T	T	T

$\forall x \exists y P(x, y)$ is false

y x	a	b	c
a	T	T	F
b	T	T	T
c	F	F	F

- $\exists x \forall y P(x, y)$ says "there is an x such that for any y we have $P(x, y)$ is true".

  To check whether $\exists x \forall y P(x, y)$ is true, we start looping through the values of x until we find one x for which $P(x, y)$ is always true when we loop over all values of y. Once we find such x we know that $\exists x \forall y P(x, y)$ is true. However, if we never find such x, that is, for every x, we hit some y for which $P(x, y)$ is false, then we we know that $\exists x \forall y P(x, y)$ is false.

  Example. When $P(x, y)$ is the statement "x likes y", then $\exists x \forall y P(x, y)$ means that "There is a person who likes everyone". Apply thinking via loops to the following two scenarios of a domain containing three people a, b, c.

$\exists x \forall y P(x, y)$ is true

$x \backslash y$	a	b	c
a	T	F	F
b	T	T	F
c	T	T	T

$\exists x \forall y P(x, y)$ is false

$x \backslash y$	a	b	c
a	T	T	F
b	T	F	T
c	F	T	T

- $\exists x \exists y P(x, y)$ says "there are some x and y such that $P(x, y)$ is true".

  To check whether $\exists x \exists y P(x, y)$ is true, we start looping through the values of x, and for each x we start looping over the values of y until we hit some x for which we hit some y for which $P(x, y)$ is true. Once we find such x and y we know that $\exists x \exists y P(x, y)$ is true. However, if we never find one x and one y such that $P(x, y)$ is true, then $\exists x \exists y P(x, y)$ is false. That is, we loop through all values of x, and for each x we have $P(x, y)$ is always false when we loop through all values of y.

  Example. When $P(x, y)$ is the statement "x likes y", then $\exists x \exists y P(x, y)$ means that "There is a person who likes someone". Apply thinking via loops to the following two scenarios of a domain containing three people a, b, c.

$\exists x \exists y P(x, y)$ is true

$x \backslash y$	a	b	c
a	F	F	F
b	F	F	F
c	F	T	F

$\exists x \exists y P(x, y)$ is false

$x \backslash y$	a	b	c
a	F	F	F
b	F	F	F
c	F	F	F

As we have seen, thinking using loops is very useful to evaluate the truth value of a quantified statement. Use thinking via loops to investigate the following example.

Example. Let the domain of variables be $\{-2, -1, 0, 1, 2\}$, and the predicate $Q(x, y)$ be "$x + y = 0$". Think using loops and confirm that the truth values of the statements below are as given.

(i) $\forall x \forall y \, Q(x, y)$ is false.

(ii) $\forall x \exists y \, Q(x, y)$ is true.

(iii) $\exists x \forall y \, Q(x, y)$ is false.

(iv) $\exists x \exists y \, Q(x, y)$ is true. ◊

Whether the domain of the variables is finite or infinite, in the table below we sum up how to decide the truth value of a statement involving quantification of two variables. Let $P(x, y)$ be any binary predicate. We remark that $\forall x \forall y P(x, y) \equiv \forall y \forall x P(x, y)$ and $\exists x \exists y P(x, y) \equiv \exists y \exists x P(x, y)$.

Statement	True	False
$\forall x \forall y P(x, y)$	When for every pair x and y in the domain, we have $P(x, y)$ is true.	When there is some pair x and y such that $P(x, y)$ is false.
$\forall x \exists y P(x, y)$	When for each x we can find some y for which $P(x, y)$ is true.	When there is some x such that $P(x, y)$ is false for every y.
$\exists x \forall y P(x, y)$	When there is some x such that $P(x, y)$ is true for all y.	When for each x there is some y such that $P(x, y)$ is false.
$\exists x \exists y P(x, y)$	When there are some x and y such that $P(x, y)$ is true.	When for every pair x and y, we have $P(x, y)$ is false.

Next, we consider ternary predicates and quantification of three variables.

Example. Consider the ternary predicate $Q(x, y, z)$ which is the statement "$x+y = z$" where the domain is the set of real numbers $\mathbb{R}$. In words, a triple (x, y, z) of numbers has property Q if the sum of the first two numbers is equal to the third number.

(i) $Q(4, 3, 7)$ is true, while $Q(8, 5, 3)$ is false.

(ii) $\forall x \forall y \exists z Q(x, y, z)$ says that for every pair of numbers x and y, there exists a number z such that $Q(x, y, z)$ is true.
$\forall x \forall y \exists z Q(x, y, z)$ is true because for any pair of numbers x and y we actually can find a third number z equal to to their sum. Simply, whenever we are given any two numbers x and y, just choose z to be the number $x + y$. For example, when you are given 3 and 6, choose the third number to be 9 and see that $Q(3, 6, 9)$ is true. This strategy always works and so $\forall x \forall y \exists z Q(x, y, z)$ is true.

(iii) $\exists z \forall x \forall y Q(x, y, z)$ says that there exists a number z such that for any pair of numbers x and y we have that $Q(x, y, z)$ is true.
In other words, it says that there is a (special) number z with the property that whenever we sum any two numbers whatsoever the result will be z. Obviously, such number does not exist and so $\exists z \forall x \forall y Q(x, y, z)$ is false. Here is a rigorous argument why such z cannot exist: if there exists some number z such that the sum of any two numbers is equal to z, then for such z choose the two numbers to be z itself and 1, which leads to $z + 1 = z$ and so $1 = 0$, which is false in the world of real numbers. ♢

Example. Let the ternary predicate $P(x, y, z)$ be "$xy = z$" and the domain be $\mathbb{R}$, the set of all real numbers. So a triple (x, y, z) of numbers has property P if the product of the first two numbers is equal to the third. Find the truth value of the following statements.

- $\forall x \forall y \exists z P(x, y, z)$ is true.
- $\forall x \forall z \exists y P(x, y, z)$ is false. Take $x = 0$ and $z = 1$.
- $\forall x \forall y \forall z P(x, y, z)$ is false. Take $x = 2$, $y = 3$, and $z = 1$.
- $\forall z \exists x P(x, x, z)$ is false. Take $z = -1$.
- $\exists x \forall z P(x, z, z)$ is true. Take $x = 1$. ◊

Example. Translate the following English statements into predicate logic where the domain is the set $\mathbb{R}$ of all real numbers.

- The sum of any two positive real numbers is always positive.

$$\forall x \forall y \big((0 < x \wedge 0 < y) \rightarrow (0 < x + y)\big).$$

- Every nonnegative real number has a square root.

$$\forall x \exists y (x \geq 0 \rightarrow x = y^2).$$ ◊

Example. Let the predicate $C(x)$ be "x owns a car", and $F(x, y)$ be "x and y are friends". The domain of the variables is the set of all students of the university.

Translate the following logical expressions into English.

- $\forall x (C(x) \vee \exists y (C(y) \wedge F(x, y)))$.
  Every student either owns a car or has a friend who owns a car.
- $\forall x \exists y F(x, y)$
  Every student has at least one friend.
- $\exists x \forall y F(x, y)$
  There is a student who is a friend of every student.

Translate the sentence "Every student has exactly one friend" into predicate logic.

$$\forall x \exists y (F(x, y) \wedge (\forall z (F(x, z) \rightarrow y = z))).$$

As an exercise, translate "Every student has exactly two friends". ◊

1.6 Formal Proofs

One goal in this course is to study mathematical proofs. Proofs are valid arguments that establish the truth of mathematical statements. We use logic to write down proofs. Recall that logic has two parts: precise language and a list of deduction rules. In previous sections we learned how to write precise language using propositions, predicates, logical operators, and quantifiers. In this section we will get introduced to a collection of deduction rules, also called rules of inference. Both the language of logic together with these rules are the basic tools for constructing proofs.

Consider the following argument.

(1) "If you have a student card, then you can enter the university campus."
(2) "You have a student card."
Therefore,
(3) "You can enter the university campus."

Statements (1) and (2) are called the *premises* of the argument, and statement (3) is called the *conclusion.* This argument is a valid argument because the conclusion must be true whenever the premises are both true. Let p represents the proposition "You have a student card" and q represents "You can enter the university campus". Moreover, we will use the symbol $\therefore$ to denote "therefore". That being said, the argument above has the following logical form.

$$\begin{array}{ll} 1. & p \to q \\ 2. & p \\ \hline \therefore & q \end{array}$$

The above argument is valid because whenever $p \to q$ and p are both true, then q must also be true (check it). Recall that a conditional statement is a tautology means that whenever its hypothesis is true, then its conclusion must be true. Thus, another way to see that the argument above is valid is to check that the conditional statement $((p \to q) \wedge p) \to q$ is a tautology (check it).

Motivated by the example above, let us formally introduce the concepts involved.

Definition.

- An *argument* in propositional logic is a sequence of compound propositions,

$$(\alpha_1,\ \alpha_2,\ \alpha_3,\ \dots\ ,\ \alpha_n,\ \beta).$$

  Each α_i is called a *premise* or *hypothesis*, and the last proposition β is called the *conclusion.*

- An argument $(\alpha_1, \alpha_2, \dots, \alpha_n, \beta)$ is *valid* if whenever the premises $\alpha_1, \alpha_2, \dots, \alpha_n$ are all true, then the conclusion β must be true.

We can write an argument in a vertical form as follows.

$$\begin{array}{rl} & \alpha_1 \\ & \alpha_2 \\ & \vdots \\ & \alpha_n \\ \hline \therefore & \beta \end{array}$$

Recall that we have discussed before in Section 1.3 that a conditional statement is a tautology if and only if whenever its hypothesis is true, then its conclusion must also be true. Consequently, an argument $(\alpha_1, \alpha_2, \alpha_3, \ldots, \alpha_n,\ \beta)$ is valid exactly when the conjunction of all the premises implying the conclusion is a tautology. In other words, when the compound proposition

$$\big(\alpha_1 \wedge \alpha_2 \wedge \alpha_3 \wedge \cdots \wedge \alpha_n\big) \to \beta$$

is a tautology.

♣ Rules of Inference

A *rule of inference* is a short valid argument. We now present a list of rules of inference (i.e. valid arguments).

(1) Modus Ponens

Modus Ponens is the valid argument whose premises are $p \to q$ and p, and whose conclusion is q. That is, it is the argument $(p \to q,\ p,\ q)$. In vertical form we write it as follows.

$$\begin{array}{rl} & p \to q \\ & p \\ \hline \therefore & q \end{array}$$

To see that Modus Ponens is a valid argument check that $\big((p \to q) \wedge p\big) \to q$ is a tautology. The modus ponens rule states that if a conditional statement and its hypothesis are both true, then the conclusion must also be true.

(2) Modus Tollens

$$\begin{array}{rl} & p \to q \\ & \neg q \\ \hline \therefore & \neg p \end{array}$$

To see that Modus Tollens is a valid argument check that $\big(\neg q \wedge (p \to q)\big) \to \neg p$ is a tautology.

(3) Hypothetical Syllogism

$$\begin{array}{rl} & p \to q \\ & q \to r \\ \hline \therefore & p \to r \end{array}$$

To see that Hypothetical Syllogism is a valid argument we need to check that $\big((p \to q) \wedge (q \to r)\big) \to (p \to r)$ is a tautology.

(4) Disjunctive Syllogism

$$\begin{array}{rl} & p \vee q \\ & \neg p \\ \hline \therefore & q \end{array}$$

To see that Disjunctive Syllogism is a valid argument check that $\big((p \vee q) \wedge \neg p\big) \to q$ is a tautology.

(5) Addition Rule

$$\begin{array}{rl} & p \\ \hline \therefore & p \vee q \end{array}$$

To see that the Addition Rule is a valid argument check that $p \to (p \vee q)$ is a tautology.

(6) Simplification Rule

$$\begin{array}{rl} & p \wedge q \\ \hline \therefore & p \end{array}$$

To see that the Simplification Rule is a valid argument check that $(p \wedge q) \to p$ is a tautology.

(7) Conjunction Rule

$$\begin{array}{rl} & p \\ & q \\ \hline \therefore & p \wedge q \end{array}$$

To see that the Conjunction Rule is a valid argument check that $(p \wedge q) \to (p \wedge q)$ is a tautology.

(8) Resolution Rule

$$\begin{array}{rl} & p \vee q \\ & \neg p \vee r \\ \hline \therefore & q \vee r \end{array}$$

To see that the Resolution Rule is a valid argument we need to check that the proposition $\big((p \vee q) \wedge (\neg p \vee r)\big) \to (q \vee r)$ is a tautology. Here are two special cases of the resolution rule. Notice that the second case is Disjunctive Syllogism.

When $r = q$

$$\begin{array}{rl} & p \vee q \\ & \neg p \vee q \\ \hline \therefore & q \end{array}$$

When $r = \mathbf{F}$

$$\begin{array}{rl} & p \vee q \\ & \neg p \\ \hline \therefore & q \end{array}$$

Example. Which rules of inference are used in the arguments below?

- "It is sunny and windy outside. Therefore, it is sunny outside."
  This argument uses the Simplification Rule.
- "It is sunny outside. Therefore, it is sunny or windy outside."
  This argument uses the Addition Rule. ◊

Example. The following argument is *not* valid.

$$\begin{array}{rl} & p \to q \\ & q \\ \hline \therefore & p \end{array}$$

Because there is a scenario where both premises $p \to q$ and q are true, but the conclusion p is false, namely when $p = F$ and $q = T$. Alternatively, check that $((p \to q) \wedge q) \to p$ is not a tautology. Examine this argument when p states that "You have a job", and q states that "You have money". ◊

Our aim is to use the rules of inference introduced above to prove the validity of more complex arguments. We start by assuming the premises to be true and then continue by applying one rule of inference at each successive step until we reach the desired conclusion. We demonstrate this process in the next example.

Example. Show that

$$\big((\neg p \wedge q), (r \to p), (\neg r \to s), (s \to t),\ t\big)$$

is a valid argument.

1.	$\neg p \wedge q$	Premise
2.	$r \to p$	Premise
3.	$\neg r \to s$	Premise
4.	$s \to t$	Premise
5.	$\neg p$	Simplification using (1)
6.	$\neg r$	Modus Tollens using (2) and (5)
7.	s	Modus Ponens using (3) and (6)
8.	t	Modus Ponens using (4) and (7)

◊

The sequence above, consisting of 8 statements, is called a formal proof and we use it to prove that the conclusion follows from the premises, that is, to prove the validity of the argument in question. Let us state clearly the definition of a formal proof.

Definition. A *formal proof* of the validity of an argument $(\alpha_1,\ \alpha_2,\ \ldots,\ \alpha_n,\ \beta)$ is a sequence of statements of the form

$$\begin{array}{r|c}
1. & \phi_1 \\
2. & \phi_2 \\
\vdots & \vdots \\
i. & \phi_i \\
\vdots & \vdots \\
m. & \phi_m = \beta
\end{array}$$

where the last statement ϕ_m is the conclusion β of the argument and each statement ϕ_i in the sequence is either

(i) one of the premises $\alpha_1, \alpha_2, \ldots, \alpha_n$, or

(ii) inferred from previous statements in the sequence using an inference rule, or

(iii) logically equivalent to a previous statement, or

(iv) a tautology.

Exercise. Prove that if we construct a formal proof for the validity of an argument $(\alpha_1, \alpha_2, \ldots, \alpha_n,\ \beta)$, then the compound proposition $(\alpha_1 \wedge \alpha_2 \wedge \ldots \wedge \alpha_n) \to \beta$ must be a tautology.

♣ Rules of Inference for Quantified Statements

We introduce four more rules of inference which deal with universal and existential statements.

Universal Instantiation

$$\begin{array}{ll} & \forall x\, P(x) \\ \hline \therefore & P(c) \quad \text{where } c \text{ is a particular element} \end{array}$$

Universal Instantiation states that if every element in the domain has property P, and c is some particular element, then we can infer that c has property P.

Universal Generalisation

$$\begin{array}{ll} & P(a) \quad \text{for any arbitrary element } a \\ \hline \therefore & \forall x\, P(x) \end{array}$$

Universal Generalisation states that if we know that any arbitrary element in the domain has property P, then we can infer that $\forall x P(x)$ is a true statement. Arbitrary element a means that a must be any element of the domain, and *not* a specific one. More precisely, we cannot make any other assumptions about a other than it comes from the domain.

Existential Instantiation

$$\begin{array}{ll} & \exists x\, P(x) \\ \hline \therefore & P(c) \quad \text{for some element } c \end{array}$$

Existential Instantiation states that if we know that some element with property P exists in the domain, then we give it a name, say the name c, and infer that $P(c)$ is true. Usually we do not really know what the element is, we only know that it exists.

Existential Generalisation

$$\begin{array}{ll} & P(c) \quad \text{for some particular element } c \\ \hline \therefore & \exists x\, P(x) \end{array}$$

Existential Generalisation states that if we know that a particular element c in the domain has property P, then we infer that $\exists x P(x)$ is true.

Example. Write a formal proof to show that the premises “A student in the class has not read the book”, and ”Everyone in the class passed the exam” imply the conclusion “Someone who passed the exam has not read the book”.
We are asked to show that the following argument is valid.

1. A student in the class has not read the book.
2. Everyone in the class passed the exam.

$\therefore$ Someone who passed the exam has not read the book.

We firstly translate these sentences into predicate logic. So let the predicate $C(x)$ be the statement "x is in the class", and $R(x)$ be "x has read the book", and $P(x)$ be "x has passed the exam". Moreover, let the variable domain be all students in the university. Our task now boils down to show that the argument below is valid.

$$\exists x(C(x) \wedge \neg R(x)), \ \ \forall x(C(x) \to P(x)), \ \ \exists x(P(x) \wedge \neg R(x)).$$

Here is a formal proof for its validity.

1.	$\exists x(C(x) \wedge \neg R(x))$	Premise
2.	$\forall x(C(x) \to P(x))$	Premise
3.	$C(a) \wedge \neg R(a)$	Existential Instantiation (1)
4.	$C(a) \to P(a)$	Universal Instantiation
5.	$C(a)$	Simplification (3)
6.	$P(a)$	Modus Ponens (4) and (5)
7.	$\neg R(a)$	Simplification (3)
8.	$P(a) \wedge \neg R(a)$	Conjunction (6) and (7)
9.	$\exists x(P(x) \wedge \neg R(x))$	Existential Generalisation (8)

$\Diamond$

Example. Assume that "For all positive integers n, we have that if $n > 4$, then $n^2 < 2^n$" is true. Use this assumption to write a formal proof to show that $100^2 < 2^{100}$ must be true.
Let $G(n)$ represents "$n > 4$" and $Q(n)$ represents "$n^2 < 2^n$". Moreover, let the variable domain be the set of positive integers. Here is a formal proof. (Notice that 100 is an element of the domain.)

1.	$\forall n(G(n) \to Q(n))$	Premise
2.	$G(100) \to Q(100)$	Universal Instantiation (1)
3.	$G(100)$	Tautology
4.	$Q(100)$	Modus Ponens (2) and (3)

We concluded that $Q(100)$ is a true proposition, but $Q(100)$ means that $100^2 < 2^{100}$ as desired. $\Diamond$

Example. Write down a formal proof showing the validity of the following argument.

1. All men are mortal.
2. Socrates is a man.
3. Therefore, Socrates is mortal.

Let the predicate $M(x)$ be "x is a man" and $R(x)$ be "x is mortal". Moreover, let the universe of the variables be the set of all humans.

1.	$\forall x(M(x) \to R(x))$	Premise
2.	$M(\text{Socrates})$	Premise
3.	$M(\text{Socrates}) \to R(\text{Socrates})$	Universal Instantiation (1)
4.	$R(\text{Socrates})$	Modus Ponens (2) and (3)

$\Diamond$

Example. Prove the validity of the following argument. Assume $\exists x \forall y P(x, y)$ is true. Therefore, $\forall y \exists x P(x, y)$ is also true.

1.	$\exists x \forall y P(x, y)$	Premise
2.	$\forall y P(c, y)$ where c is some element	Existential Instantiation (1)
3.	$P(c, d)$ for any d	Universal Instantiation (2)
4.	$\exists x P(x, d)$	Existential Generalisation (3)
5.	$\forall y \exists x P(x, y)$	Universal Generalisation (4)

$\Diamond$

Example. The following argument is not valid.

$$\frac{\forall y \exists x P(x, y)}{\therefore\ \exists x \forall y P(x, y)}$$

To show it is invalid, we need to find a domain and interpretation for the predicate $P(x, y)$ where the premise is true but the conclusion is false. This is left as an exercise for the reader. Any domain will do, for instance, a domain consisting of 3 people where $P(x, y)$ means "x likes y". $\Diamond$

1.7 Mathematical Proofs

In this section we will delve deeper in the notion of a proof and learn several techniques for constructing mathematical proofs. Roughly speaking, a *proof* is a valid argument that establishes the truth of a mathematical statement. In the previous section we learned formal proofs to establish the validity of an argument. In formal proofs we supply all steps and justify in each step which rule of inference is used. Formal proofs can be extremely long and difficult to follow by humans. Formal proofs are important objects of study in the field of mathematical logic, and they are suitable for computers when they run automated reasoning systems in the area of artificial intelligence.

For smoothness in readability and better understanding, in this section we will instead write *informal proofs*. These are proofs written in our natural languages and suitable for human consumption. Within informal proofs we may apply more than one rule of inference at a single step, we may skip obvious steps, and we do not explicitly state which rule of inference is used.

We start by introducing important terminology in the field of mathematics.

- A *theorem* is a mathematical statement that is shown to be true by providing a proof. Theorems are also called *facts* or *results*.
- A *proposition* is a minor or less important theorem.
- A *lemma* is a less important theorem which is used in the proof of other important theorems.
- A *corollary* is a theorem that is an easy consequence of a theorem that has already been proved.
- A *conjecture* is a mathematical statement that is proposed to be true but has not been proved yet; it is an open question.
- An *axiom* is a mathematical statement that is assumed to be true without providing a proof.

Famous theorems in the field of mathematics include the *Pythagorean Theorem* and the *Fundamental Theorem of Calculus*. Another famous theorem which we will prove in this course states that “There are infinitely many prime numbers”.

Examples of axioms include the axioms of the world of real numbers $\mathbb{R}$ where we assume the following axioms to be true.

1. $\forall x\,(x + 0 = x)$ (Additive Identity)
2. $\forall x\,\exists y\,(x + y = 0)$ (Additive Inverse)
3. $\forall x\,\forall y\,(x + y = y + x)$ (Addition Commutativity)
4. $\forall x\,\forall y\,\forall z\,\big((x + y) + z = x + (y + z)\big)$ (Addition Associativity)

5. $\forall x\,(1 \cdot x = x)$ (Multiplicative Identity)
6. $\forall x\,\exists y\,(x \neq 0 \to x \cdot y = 1)$ (Multiplicative Inverse)
7. $\forall x\,\forall y\,(x \cdot y = y \cdot x)$ (Multiplication Commutativity)
8. $\forall x\,\forall y\,\forall z\,\big((x \cdot y) \cdot z = x \cdot (y \cdot z)\big)$ (Multiplication Associativity)
9. $\forall x\,\forall y\,\forall z\,\big(x \cdot (y + z) = x \cdot y + x \cdot z\big)$ (Distributivity)

Conjectures emerge when one observes a pattern that holds true for numerous cases or when a mathematician's intuition is telling that a statement has to be true. Nevertheless, checking a pattern holds true for a sample of the cases does not mean that the pattern will hold true for all cases; usually we have infinitely many cases! Also, it is possible that our intuition turns out to be false. To be certain that a conjecture holds true for all cases we need to construct a mathematical proof. Only when a conjecture is rigorously proved, it becomes a theorem. Many times a conjecture turns out to be false, so in this case it is not a theorem. One way to refute a conjecture is by finding or constructing a counter example for which the conjecture does not hold.

In mathematics, specifically in number theory, one of the most famous conjectures was *Fermat's Conjecture*. Before we introduce this conjecture let us have a short motivation. Observe that the equation $x + y = z$ has many solutions in the positive integers, for instance, $3 + 5 = 8$. Now, examine the case when the exponent of the variables in this equation is 2, that is, $x^2 + y^2 = z^2$. Can you find three positive integers satisfying this equation? A triple (x, y, z) of positive integers satisfying the equation $x^2 + y^2 = z^2$ is called a *Pythagorean triple.* For example, the triples $(3, 4, 5)$ and $(5, 12, 13)$ are Pythagorean. Moreover, observe that if (a, b, c) is a Pythagorean triple then so is (ka, kb, kc) for any positive integer k. Thus we have infinitely many Pythagorean triples, and this fact was known since antiquity. Next, can you find positive integers satisfying the equation $x^3 + y^3 = z^3$?

Fermat's Conjecture states that the equation $x^n + y^n = z^n$ has no positive integer solutions when the exponent $n \geq 3$, and it was conjectured in 1637 by the French lawyer and mathematician Pierre de Fermat who claimed to have a proof of this conjecture but never provided one. The quest for a proof of Fermat's Conjecture took 358 years during which generations of mathematicians worked hard on the problem, and as a result developed new fields and tools in mathematics such as algebraic number theory. Fermat's Conjecture has the largest number of unsuccessful attempts to prove it. In 1993, the English mathematician Andrew Wiles presented a proof of Fermat's Conjecture at the Isaac Newton Institute for Mathematical Sciences at the University of Cambridge. Unfortunately, when experts were refereeing Wile's manuscript found a critical error. Wiles spent the next year trying to repair his proof under stressful conditions. Together with his former student Richard Taylor, the error was fixed. In 1994, Wiles submitted the first successful proof of the conjecture, and so it became known as *Fermat's Last Theorem.* His work was published in 1995 in the journal of *Annals of Mathematics.*

Theorem 1.1 (Fermat's Last Theorem)**.** *The equation*

$$x^n + y^n = z^n$$

has no solution in the positive integers when $n \geq 3$.

Another famous conjecture in number theory dating back to 1742 was conjectured by the German mathematician Christian Goldbach in his letters to Leonhard Euler.

Conjecture (Goldbach's Conjecture)**.** *Every even integer greater than 2 is the sum of two primes.*

For instance, $4 = 2 + 2$ and $18 = 7 + 11$ and $22 = 11 + 11$ and $36 = 17 + 19$. The conjecture has been verified to hold up to very large numbers. However, it is still an unsolved problem up to this day. Mathematicians are still looking for a proof or a counter example.

Let us describe in accurate terms what is considered a proof. A mathematical proof is a valid argument which uses ingredients such as axioms, hypotheses, previously proven theorems together with rules of inference to establish the truth of a desired mathematical statement, which when proven becomes a theorem. More precisely, a proof is defined as follows.

Definition. A *proof* of a theorem β is a sequence of statements

$$\alpha_1, \alpha_2, \alpha_3, \ldots, \alpha_m$$

where the last statement α_m is the theorem β and each α_i is either an axiom, or a previously proved theorem, or a hypothesis (premise), or inferred from previous statements in the sequence using an inference rule.

A mathematical proof (sequence of statements as described above) is written using our natural languages in the form of complete sentences separated by punctuation marks forming understandable clear paragraphs.

Remark. We use the symbol $\Longrightarrow$ as a shorthand for "implies" and the symbol $\Longleftrightarrow$ is used in place of "if and only if". So we write $\alpha \Longrightarrow \beta$ to denote the statement "If α, then β". And we write $\alpha \Longleftrightarrow \beta$ to denote the statement " α if and only if β". We save the symbols $\rightarrow$ and $\leftrightarrow$ for statements written in propositional or predicate logic.

Proof Techniques

We describe different strategies of proofs and their overall structure. Understanding these strategies aids in learning how to read and how to construct a mathematical proof. Our goal now is to build a large arsenal of proof techniques that can be utilised to prove a broad variety of theorems.

♣ Universal Theorems

To prove universal theorems of the form

$$\forall x P(x)$$

we choose an arbitrary element c of the domain, and then prove that $P(c)$ is true. And we finish by applying Universal Generalisation to deduce that $\forall x P(x)$ is a true statement (a theorem). When writing a mathematical proof using this strategy we do not really need to state explicitly that we are applying the rule of Universal Generalisation. Similarly, to prove universal theorems of the form

$$\forall x \forall y Q(x, y)$$

the strategy is to prove that $Q(c, d)$ is true where c, d are general arbitrary elements of the domain. In the same fashion we think of universal theorems of the form

$$\forall x \big(P(x) \to Q(x)\big).$$

To prove such statements we need to show that $P(c) \to Q(c)$ is true for an arbitrary element c in the domain. In this case, we see the need for techniques to prove a conditional statement $p \to q$ is true. Shortly we present different methods fulfilling this need such as direct proofs and proof by contraposition.

Universal theorems are often stated without universal quantifiers. For instance, working in the domain $\mathbb{R}$ of real numbers we write the statement

$$x^2 - y^2 = (x - y)(x + y)$$

instead of writing

$$\forall x \forall y \big(x^2 - y^2 = (x - y)(x + y)\big).$$

Hooray! We are now ready to write our first (informal) mathematical proof. Let's prove our first theorem!

Theorem 1.2. *For any real number x, we have that $x \cdot 0 = 0$.*

Proof. Observe that we want to prove a universal statement of the form $\forall x (x \cdot 0 = 0)$ where the domain of the variable x is the set of all real numbers. Let's start!

Choose an arbitrary real number x. We need to show that $x \cdot 0 = 0$. Now, by the Additive Identity Axiom we have that $0 + 0 = 0$. Thus,

$$x \cdot 0 = x \cdot (0 + 0).$$

By Distributivity Axiom, we have that

$$x \cdot 0 = (x \cdot 0) + (x \cdot 0).$$

By the Additive Inverse Axiom, the number $x \cdot 0$ has an additive inverse, that is, there exists a real number b such that $(x \cdot 0) + b = 0$. Next, add this b to both sides of the last equation to get

$$(x \cdot 0) + b = \big((x \cdot 0) + (x \cdot 0)\big) + b.$$

By Addition Associativity Axiom, we have that

$$(x \cdot 0) + b = (x \cdot 0) + \big((x \cdot 0) + b\big).$$

By choice of b, we get that

$$0 = (x \cdot 0) + 0.$$

Finally, by the Additive Identity Axiom we infer that

$$0 = x \cdot 0$$

as desired, and the proof is complete. ■

♣ Direct Proofs

We have seen in Section 1.3 that a conditional statement is always true if whenever its hypothesis is true, then its conclusion is also true. Thus to prove a conditional statement $p \to q$ is always true we first assume that p is true, and then construct subsequent deductions using axioms, rules of inference, and known theorems, with the final step showing that q must be true. To sum up, in order to prove a conditional statement $p \to q$ is true using a *direct proof* we do the following.

- Assume p is true.
- Use axioms.
- Use rules of inference.
- Use previously proved theorems.

- Show q is true.
- Thus, $p \to q$ is true.

A direct proof shows that a conditional statement $p \to q$ is true by showing that if p is true, then q must also be true. This approach guarantees that the case where p is true and q is false never occurs, and thus $p \to q$ is never false.

Next we provide applications of the method of direct proofs in proving some mathematical statements. In order to formulate these statements we need to define some terminology.

Definition. An *integer* is a whole number. The set of all integers is denoted by $\mathbb{Z}$ standing for the German word "Zahlen" meaning "numbers".

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

- An integer n is *even* if there exists an integer k such that $n = 2k$.
- An integer n is *odd* if there exists an integer k such that $n = 2k + 1$.
- An integer n is a *perfect square* if there exists an integer k such that $n = k^2$.

Next, we will prove that the square of an odd integer is also odd.

Theorem 1.3. *If n is an odd integer, then n^2 is odd.*

Proof. First, note that this theorem is a universal statement of the form

$$\forall n \left(D(n) \to D(n^2)\right)$$

where $D(n)$ is the predicate "n is odd" and the variable domain is all integers.
Let n be an arbitrary integer. We want to prove that the conditional statement $D(n) \to D(n^2)$ is true. So assume n is odd. Our goal is to show that n^2 is odd as well. Since n is odd, there exists an integer k such that $n = 2k + 1$. We now have the following sequence of equalities.

$$\begin{aligned} n^2 &= (2k+1)^2 \\ &= 4k^2 + 4k + 1 \\ &= 2(2k^2 + 2k) + 1 \\ &= 2m + 1, \end{aligned}$$

where $m = 2k^2+2k$ is an integer. Obviously, we have used the axiom of Distributivity above. Since n^2 is of the form $2m + 1$ where m is an integer, it means that n^2 is odd. Therefore, we have proved that if an integer n is odd, then n^2 is also odd. ■

Since a conditional statement is logically equivalent to its contrapositive, an easy consequence of the above theorem is that its contrapositive is also true. While formulating the contrapositive we use the fact that if an integer is not odd, then it is even. The proof of this fact is left as an exercise for the reader (Hint: revise the definition of even and odd). Thus we have the following result.

Corollary 1.4. *If n^2 is an even integer, then n is even.*

Next, we present another application of a direct proof.

Theorem 1.5. *Let m, n be integers. If m and n are both perfect squares, then their product mn is also a perfect square.*

Proof. Choose any arbitrary integers m and n. Suppose that they are perfect squares. By definition of perfect square, there are integers a, b such that $m = a^2$ and $n = b^2$. Now, we have that

$$mn = a^2b^2 = aabb = (ab)(ab) = (ab)^2 = k^2,$$

where $k = ab$ is an integer. (For the third equality above we used the axioms of commutativity and associativity of multiplication.) So, $mn = k^2$ where k is an integer. Thus, mn is a perfect square. We have proved that the product of two perfect squares is also a perfect square. ■

♣ Proof by Contraposition

We have seen that direct proofs flow from the hypothesis of a conditional statement towards its conclusion. We might encounter cases where this path leads to a dead end, at which we need to rethink our strategy. Let us describe another approach to prove the truth of a conditional statement. We established earlier that a conditional statement is logically equivalent to its contrapositive, that is, $p \to q \equiv \neg q \to \neg p$ where p, q are compound propositions. Our new strategy to prove that a conditional statement $p \to q$ is true is by proving its contrapositive $\neg q \to \neg p$ is true via a direct proof strategy. Thus, we assume the $\neg q$ is true, and in subsequent steps we use axioms, previously proved theorems, and rules of inference, until we show that $\neg p$ must be true. This approach shows that $\neg q \to \neg p$ is true, but $\neg q \to \neg p \equiv p \to q$, and so $p \to q$ must be true as well. To sum up, in order to prove a conditional statement $p \to q$ is true using a *proof by contraposition* we do the following.

- Assume $\neg q$ is true.
- Use axioms.
- Use rules of inference.
- Use previously proved theorems.

- Show $\neg p$ is true.
- Thus, $\neg q \to \neg p$ is true.
- Equivalently, $p \to q$ is true.

Theorem 1.6. *Let n be an integer. If $3n + 2$ is odd, then n is odd.*

Proof. We will proceed by showing the contraposition. So assume n is not odd. It follows that n is even, and so there exists an integer k such that $n = 2k$. We find that

$$3n + 2 = 3(2k) + 2 = 6k + 2 = 2(3k + 1) = 2m,$$

where $m = 3k + 1$ is an integer. So $3n + 2$ is even, and hence not odd. So we proved that if n is even, then $3n + 2$ is even. Equivalently, we showed that if $3n + 2$ is odd, then n is odd. ■

Exercise. Try to prove the theorem above via a direct proof.

♣ Proof by Contradiction

Suppose our aim is to prove that a statement p is true. One approach to achieve this is that we assume p is false, and then use axioms, rules of inference, and previously proved theorems to deduce a contradiction (a statement which is always false, for example $r \wedge \neg r$ for some statement r). Once we derive a contradiction, it must be that p cannot be false, and so it must be true (remember a statement is either true or false).

To sum up, to prove that a statement p is true using *proof by contradiction* we do the following.

- Assume p is false. (Assume $\neg p$ is true.)
- Use axioms.
- Use rules of inference.
- Use previously proved theorems.

- Deduce a contradiction.
- Thus, p must be true.

Remark. Our aim is to show that some statement p is true. The strategy above proves that the conditional statement $\neg p \to \mathbf{F}$ is true using a direct proof. Here, $\mathbf{F}$ is a contradiction. Once we prove that $\neg p \to \mathbf{F}$ is true, we can conclude that $\neg p$ must be false (revise the truth table of a conditional statement). As $\neg p$ is false, it follows that the statement p is true as desired.

Example. Show that $\sqrt[3]{345} > 7$.
Suppose for the contrary that the statement $\sqrt[3]{345} > 7$ is false. So we have that $\sqrt[3]{345} \leq 7$. We now take the cube of both sides of the inequality to get that $345 \leq 7^3$, and so $345 \leq 343$ which is a contradiction. Therefore, we conclude that $\sqrt[3]{345} > 7$ cannot be false, and so it must be true as required.
Another way to show this is to use a calculator to compute that $\sqrt[3]{345} \approx 7.0135790835$. Nevertheless, we didn't need it! ♢

We will use proof by contradiction to show that irrational numbers exist.

Definition. A real number r is called *rational* if there exist integers m and n with $n \neq 0$ such that $r = \frac{m}{n}$. An *irrational* number is a real number that is not rational.

Recall that $\sqrt{2}$ is the real number which when multiplied with itself the result is equal to 2. In order to prove that $\sqrt{2}$ exists we need the *Completeness Axiom* for the real numbers which states that every nonempty subset of the real numbers that is bounded above has a least upper bound. The proof of existence is covered in a Real Analysis course. Note that since $\sqrt{2}$ exists, it follows that $\exists x\,(x \cdot x = 2)$ is a true statement in the real numbers. For our purpose, let us just take for granted that it exists and ask ourselves: Is $\sqrt{2}$ a rational number?

Theorem 1.7. *The number $\sqrt{2}$ is irrational.*

Proof. Our goal is to prove that $\sqrt{2}$ is irrational. For the sake of contradiction, assume that $\sqrt{2}$ is not irrational, and so assume that $\sqrt{2}$ is rational. We will show that this assumption leads to a contradiction. By definition of rational, there are integers m, n with $n \neq 0$ such that $\sqrt{2} = \frac{m}{n}$. We now use the fact that every rational number can be written in lowest terms to further assume that m and n have no common divisors other than 1. We now deduce the following implications.

$$\sqrt{2} = \frac{m}{n} \Longrightarrow 2 = \frac{m^2}{n^2} \Longrightarrow m^2 = 2n^2.$$

Thus $m^2 = 2n^2$ (Notice here we used Hypothetical Syllogism and Modus Ponens). By definition of an even integer it follows that m^2 is even. By a previous result, namely Corollary 1.4, we deduce that m is even as well. So there is an integer k such that $m = 2k$. Thus, we get that

$$m^2 = 2n^2 \Longrightarrow (2k)^2 = 2n^2 \Longrightarrow 4k^2 = 2n^2 \Longrightarrow 2k^2 = n^2.$$

This means that n^2 is even, and again by Corollary 1.4, we have that n is even as well. Therefore, 2 is a common divisor of both m and n, contradicting our original assumption that m and n have no common divisors except 1. Therefore, our assumption that $\sqrt{2}$ is rational is false, and so we proved that $\sqrt{2}$ is irrational. ■

Using Proof by Contradiction to Prove Conditional Statements

The method of proof by contradiction can also be used to prove conditional statements. Here is the strategy to prove that $p \to q$ is true: First we assume that p is true. Then we aim to show that q is true using proof by contradiction. Towards this, we assume that $\neg q$ is true and then using axioms, rules of inference, known theorems, together with the hypothesis p we deduce a contradiction. Thus $\neg q$ must be false, and so q is true as required. This strategy is justified by the following logical equivalence: $p \to q \equiv (p \wedge \neg q) \to \mathbf{F}$. Here is a summary of the aforementioned strategy to prove that $p \to q$ is true.

- ▶ Assume p is true.
- ▶ Towards a contradiction, assume q is false.
- ▶ Use axioms.
- ▶ Use rules of inference.
- ▶ Use previously proved theorems.
- ▶ Use the hypothesis p.

- ▶ Deduce a contradiction.
- ▶ Thus, q must be true.
- ▶ Therefore, $p \to q$ is true.

We use proof by contradiction to prove the theorem below.

Theorem 1.8. *Let n be an integer. If $5n$ is odd, then n is odd.*

Proof. Suppose that $5n$ is odd. We aim to prove that n is odd as well. For the sake of contradiction, assume that n is not odd, and so n is even. Thus, there exists an integer k such that $n = 2k$. Hence,

$$5n = 5(2k) = 2(5k) = 2m,$$

where $m = 5k$ is an integer. This means that $5n$ is even, contradicting that $5n$ is odd. Thus, n must be odd as required. ■

♣ Proof of Equivalences

We now discuss a method to prove biconditional statements. A strategy to prove that a theorem having the form of a biconditional statement $p \leftrightarrow q$ is true is to prove that the two conditional statements $p \to q$ and $q \to p$ are both true. Indeed, this approach is justified by the logical equivalence: $p \leftrightarrow q \equiv (p \to q) \wedge (q \to p)$.

When proving the conditional statements $p \to q$ and $q \to p$ we are free to use any method in our arsenal such as direct proof, proof by contraposition, or proof by contradiction. To sum up, in order to prove a biconditional statement $p \leftrightarrow q$ we proceed as follows.

- ▶ Prove $p \to q$ is true.
- ▶ Prove $q \to p$ is true.

- ▶ Thus, $p \leftrightarrow q$ must be true.

We will prove next that an integer is odd is equivalent for its square being odd.

Theorem 1.9. *Let n be an integer. Then n is odd if and only if n^2 is odd.*

Proof. Choose any arbitrary integer n. We will show that n is odd if and only if n^2 is odd. To prove this, we need to prove both directions: First, to show that if n is odd, then n^2 is odd. Second, to show that if n^2 is odd, then n is odd. The forward direction, if n is odd, then n^2 is odd, was proved earlier in Theorem 1.3.
For the reverse direction, we need to show that if n^2 is odd, then n is odd. So assume that n^2 is odd. We need to show that n is odd. For the sake of contradiction, assume that n is not odd, and so even. Thus, there is an integer k such that $n = 2k$. Hence,

$$n^2 = (2k)^2 = 4k^2 = 2(2k^2) = 2m$$

where $m = 2k^2$ is an integer. By definition of an even integer, it follows that n^2 is even, contradicting that n^2 is odd. So the assumption that n is not odd is false, meaning that n is odd as desired.
Therefore, we have shown that n is odd if and only if n^2 is odd. ■

Using the logical equivalence $p \leftrightarrow q \equiv \neg p \leftrightarrow \neg q$ we are able to obtain an easy consequence of the above theorem.

Corollary 1.10. *Let n be an integer. Then n is even if and only if n^2 is even.*

Recall that $p \leftrightarrow q$ is true exactly when both p and q are true or both are false. Thus when we prove a biconditional statement $p \leftrightarrow q$ is true, it means that we proved that statements p and q are equivalent (have the same truth value). Often, we encounter theorems which states that several statements are equivalent. For instance, a theorem stating that statements p, q, r are equivalent means that all of them have the same truth value: either the three of them are true or the three of them are false. This can be proved by showing that $(p \leftrightarrow q) \wedge (q \leftrightarrow r) \wedge (p \leftrightarrow r)$ is true. Notice that by Hypothetical Syllogism it is enough to show that $(p \leftrightarrow q) \wedge (q \leftrightarrow r)$ is true. We can even do better since by Hypothetical Syllogism again we have that

$$(p \leftrightarrow q) \wedge (q \leftrightarrow r) \equiv (p \rightarrow q) \wedge (q \rightarrow r) \wedge (r \rightarrow p).$$

Thus to prove a theorem stating that statements p, q, r are equivalent we need to prove three conditional statements as follows.

- ▶ Prove $p \rightarrow q$ is true.
- ▶ Prove $q \rightarrow r$ is true.
- ▶ Prove $r \rightarrow p$ is true.

- ▶ Thus p, q, r are equivalent.

Similarly, to prove that statements p, q, r, t are equivalent it is enough to prove four conditional statements, namely, $p \rightarrow q$ and $q \rightarrow r$ and $r \rightarrow t$ and $t \rightarrow p$.

A common practice in mathematics is to abbreviate the sentence “The following are equivalent” by the acronym “TFAE”.

Theorem 1.11. *Let n be an integer. The following statements are equivalent.*

(i) n *is even.*

(ii) $n+3$ *is odd.*

(iii) n^2 *is even.*

Proof. Fix any arbitrary integer n. To prove that the three statements about n are equivalent we will prove three conditional statements: We show that Statement (i) implies Statement (ii), and Statement (ii) implies Statement (iii), and Statement (iii) implies Statement (i).

$(i) \implies (ii)$. We will show that if n is even, then $n+3$ is odd. Suppose that n is even. So there is an integer k such that $n = 2k$. It follows that

$$n+3 = 2k+3 = 2k+2+1 = 2(k+1)+1.$$

Thus we showed that $n+3$ is odd.

$(ii) \implies (iii)$. We will prove that if $n+3$ is odd, then n^2 is even. So suppose that $n+3$ is odd. It follows that there is an integer t such that $n+3 = 2t+1$, and so $n = 2t-2$. We now have that

$$n^2 = (2t-2)^2 = 4t^2 - 8t + 4 = 2(2t^2 - 4t + 2).$$

Since $2t^2 - 4t + 2$ is an integer, it follows that n^2 is even as desired.

$(iii) \implies (i)$. We need to show that if n^2 is even, then n is even. This implication has been already established in Corollary 1.4. This completes the proof. ■

♣ Counterexamples

We discussed above techniques for proving the truth of universal statements. However, a universal statement in hand might be false. To disprove a universal statement of the form $\forall x P(x)$ we need to find at least one element c in the domain not having the property P, that is, an element c with $P(c)$ is false. Such an element is called a *counterexample.* That is, to disprove a universal statement $\forall x P(x)$ we prove that its negation $\neg \forall x\, P(x)$ is true, equivalently, we show that $\exists x \neg P(x)$ is true. So we need to find some element c in the domain such that $\neg P(c)$ holds.

Let us illustrate this by an example. Recall that a positive integer $p > 1$ is *prime* if its only positive divisors are 1 and itself.

Example. Prove or disprove that for any positive integer p, if p is prime, then $2^p - 1$ is also prime.
Before starting proving this statement let us test it in few cases. When $p = 2$ the conclusion is true since $2^2 - 1$ is prime. When $p = 3$, the conclusion is also true as

$2^3 - 1 = 7$ is prime. When $p = 5$, we have that $2^5 - 1 = 31$ which is prime. When $p = 7$, we have that $2^7 - 1 = 127$ which is prime as well. The next prime turns out to be a counterexample. Thus, the conditional statement, if p is prime, then $2^p - 1$ is prime, is false when $p = 11$ because its hypothesis is true since 11 is prime but its conclusion is false since $2^{11} - 1 = 2047$ is not prime since $2047 = 23 \times 89$. ♢

Exercise. Prove or disprove the following universal statement.

"For every integer $n \geq 0$, we have that $n^2 + n + 41$ is prime."

Start checking whether $n^2 + n + 41$ is prime for $n = 0, 1, 2, 3, 4, \ldots$ searching for a counterexample, or find a proof that the statement is true. ♢

Chapter 2

Naive Set Theory

Set theory is a branch of pure mathematics that serves as a foundation of mathematics. The language of set theory can be used to define nearly all mathematical objects.

2.1 Sets

Sets are of great importance in mathematics.

Definition. A *set* is a collection of objects.

Such objects are called the elements or members of the set. We write $a \in A$ to denote that a is an element of the set A or the set A contains the element a. We write $a \notin A$ if a is not an element of A.

Example. Here are some examples of sets.

1. The set V of all vowels in the English alphabet is $V = \{a, e, i, o, u\}$.
2. The set of prime numbers less than 20 is $\{2, 3, 5, 7, 11, 13, 17, 19\}$.
3. The set $\{\text{Beirut}, \text{Cairo}, \text{Damascus}, \text{Jerusalem}, \text{Rabat}, \text{Tunisia}\}$ has 6 members.
4. The set of positive integers less than 100 is written as $\{1, 2, 3, \ldots, 98, 99\}$. $\Diamond$

In the above examples we described sets by listing their elements between curly brackets. We can also describe a set by stating the property that all of its elements must have. We define a set by declaring which property grants membership to the set. In this aspect, a set S is written in the form

$$S = \{x \mid x \text{ has property } P\}.$$

We read it as *"S is the set of all objects x such that x has property P".*

Notice that the vertical bar "$\mid$" is read as "such that". In this context, we use the bar "$\mid$" and the colon "$:$" interchangeably. We may also use a predicate $P(x)$ to represent the statement "x has property P". So a set S may be expressed as

$$S = \{x \mid P(x)\} \quad \text{or} \quad S = \{x : P(x)\}.$$

Example. The set E of all even positive integers less than 100 can be written as

$$E = \{x \mid x \text{ is an even positive integer and } x < 100\} = \{2, 4, 6, 8, \ldots, 98\}.$$

We can also specify the universe of the object x, so the previous set is expressed as

$$E = \{x \text{ is an integer} \mid x > 0 \text{ and } x \text{ is even and } x < 100\}. \qquad \Diamond$$

Remark. The members of sets could be themselves sets. For example, the set

$$A = \Big\{u, \alpha, \{a, b, c\}, \{x, y\}, v, \{1, 2, 3, 4\}\Big\}$$

has exactly 6 members. We have that $u \in A$ and $\{x, y\} \in A$. However, $y \notin A$ because none of the members of A is y.

Remark. In computer science, the concept of a *datatype* is built upon the concept of a set. For example, the type *boolean* is the name of the set $\{0, 1\}$.

♣ Important Sets

- The set of natural numbers,

$$\mathbb{N} = \{0,\ 1,\ 2,\ 3,\ 4,\ 5,\ 6,\ 7,\ 8,\ 9,\ 10,\ \ldots\}.$$

- The set of all integers,

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

- The set of positive integers,

$$\mathbb{Z}^{+} = \{1,\ 2,\ 3,\ 4,\ 5,\ 6,\ 7,\ 8,\ \ldots\}.$$

- The set of all rational numbers,

$$\mathbb{Q} = \Big\{\frac{m}{n} \mid m, n \in \mathbb{Z} \text{ and } n \neq 0\Big\}.$$

- The set of all real numbers $\mathbb{R}$.
- The set of all complex numbers,

$$\mathbb{C} = \big\{x + iy \mid x, y \in \mathbb{R} \text{ and } i = \sqrt{-1}\,\big\}.$$

Given two sets, when should we consider them equal?

Definition. Two sets are *equal* if and only if they contain the same elements.

This means that we know a set precisely when we know the elements it contains. We can express the definition above in predicate logic for sets A, B as follows

$$A = B \longleftrightarrow \forall x\,(x \in A \leftrightarrow x \in B).$$

Example.

- $\{x \in \mathbb{Z} \mid x \text{ is odd and } 2 \leq x \leq 8\} = \{n \in \mathbb{Z} \mid n \text{ is prime and } 2 < n \leq 10\}$. The reason that these two sets are equal is that both of them contain precisely the elements $3, 5, 7$.

- $\{1, 2, 3\} = \{3, 2, 1\} = \{2, 1, 3\} = \{1, 1, 1, 3, 2, 3, 2, 3, 3, 3\}$. All of these sets are equal because they all contain precisely the elements $1, 2, 3$. Thus, in sets we do not care about how the elements of a set are ordered, neither about repetitions of elements. In conclusion, we know what the set is when we know what elements it contains. $\diamondsuit$

Definition.

- The *empty set* or *null set* is the set that contains no elements. The empty set is denoted by the symbol $\emptyset$ or $\{\}$.
- A set containing exactly one element is called a *singleton.*

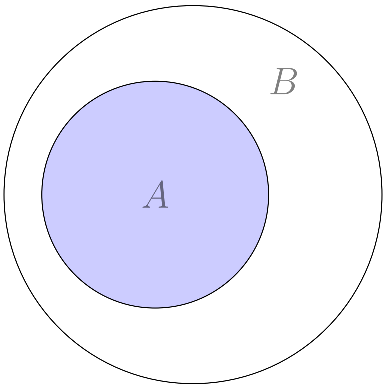
Observe that $\emptyset \neq \{\emptyset\}$ because they do not contain the same elements. Well, the set $\emptyset$ has no elements, while the set $\{\emptyset\}$ contains one element, namely, $\emptyset$. Think of $\emptyset$ as an empty bag and the set $\{\emptyset\}$ as a bag containing an empty bag.

Definition. A set A is a *subset* of a set B if and only if every element of A is also an element of B. We write $A \subseteq B$ to indicate that A is a subset of B.

If $A \subseteq B$ we also say that B is a *superset* of A and we write $B \supseteq A$. Using predicate logic we express $A \subseteq B$ as follows

$$\forall x(x \in A \rightarrow x \in B).$$

Below is the Venn diagram expressing that the set A is a subset of B.



To show that $A \nsubseteq B$, we need to find one element $a \in A$ such that $a \notin B$.

Example. Check that every element in a set below also belongs to the next set.

$$\{1\} \subseteq \{1,2,3\} \subseteq \mathbb{N} \subseteq \mathbb{Z} \subseteq \mathbb{Q} \subseteq \mathbb{R} \subseteq \mathbb{C}. \qquad \lozenge$$

Example. Let P be the set of all prime numbers and D be the set of all positive odd integers. Then $P \nsubseteq D$ because $2 \in P$ as 2 is prime but $2 \notin D$ as 2 is not odd, thus, not every element of P is an element of D. ◊

Example. Let

$$V = \{3, 1, x, 8, u, \{3, x\}, w, \{z, 4\}, z\}.$$

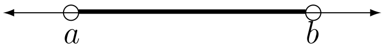
Then the following are true.

- V contains exactly 9 elements.
- $1 \in V$, and $3 \in V$, and $u \in V$, and $4 \notin V$.
- $\{3, x\} \in V$
- $\{3, x\} \subseteq V$.
- $\{z, 4\} \in V$.
- $\{z, 4\} \nsubseteq V$.
- $\{\{z, 4\}\} \subseteq V$.
- $\{8, w, z, u\} \subseteq V$.
- $\{w, \{x, 3\}, 1, u, 8\} \subseteq V$.
- $\emptyset \subseteq V$. ◊

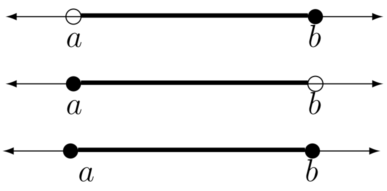
Lemma 2.1. *Let A, B be sets. Then $A = B$ if and only if $A \subseteq B$ and $B \subseteq A$.*

We introduce common notation for special subsets of real numbers, called intervals of real numbers. Let $a, b \in \mathbb{R}$, we define the following sets of real numbers.

- $(a, b) = \{x \in \mathbb{R} \mid a < x < b\}$



- $(a, b] = \{x \in \mathbb{R} \mid a < x \leq b\}$
- $[a, b) = \{x \in \mathbb{R} \mid a \leq x < b\}$
- $[a, b] = \{x \in \mathbb{R} \mid a \leq x \leq b\}$
- $(a, \infty) = \{x \in \mathbb{R} \mid a < x\}$
- $[a, \infty) = \{x \in \mathbb{R} \mid a \leq x\}$
- $(-\infty, b) = \{x \in \mathbb{R} \mid x < b\}$
- $(-\infty, b] = \{x \in \mathbb{R} \mid x \leq b\}$



The next lemma says that the empty set is a subset of any set. Moreover, any set is a subset of itself.

Lemma 2.2. *Let S be any set. Then the following are true.*

(i) $\emptyset \subseteq S$.

(ii) $S \subseteq S$.

Proof. We will prove the two statements above.

(i) To show that $\emptyset \subseteq S$, we need to show that $\forall x(x \in \emptyset \to x \in S)$. Let x be any object. As $\emptyset$ contains no elements, the hypothesis $x \in \emptyset$ is false, and thus the conditional statement $x \in \emptyset \to x \in S$ is true. Thus, $\forall x(x \in \emptyset \to x \in S)$ is true and so $\emptyset \subseteq S$.

(ii) Clearly, $S \subseteq S$ is true, because every element of S is also an element of S. For a rigorous proof, choose any arbitrary object x. Suppose that $x \in S$ is true. Then $x \in S \to x \in S$ is also true. So we showed that $\forall x(x \in S \to x \in S)$ is true. This means that $S \subseteq S$.

■

Definition. We say that A is *a proper subset* of B if $A \subseteq B$ and $A \neq B$. We write $A \subset B$ when A is a proper subset of B.

Using predicate logic we express $A \subset B$ by saying that $A \subseteq B \wedge \exists x(x \in B \wedge x \notin A)$. More precisely,

$$\forall x(x \in A \to x \in B) \wedge \exists z(z \in B \wedge z \notin A).$$

For instance, we have that $\mathbb{N} \subset \mathbb{Z}$ because every natural number is also an integer, and so $\mathbb{N} \subseteq \mathbb{Z}$, however, $\mathbb{N} \neq \mathbb{Z}$ because $-1 \in \mathbb{Z}$ but $-1 \notin \mathbb{N}$.

Notation. We may use set notation to restrict the domain of the variables in predicate logic. For example, the statement

$$\forall x \in S \, (P(x))$$

is the universal quantification of the predicate $P(x)$ over all elements in the set S, and it states that $P(x)$ is true for all elements in S. In other words, $\forall x \in S\,(P(x))$ is shorthand for $\forall x\,(x \in S \to P(x))$. Similarly, the existential quantification of $P(x)$ over all elements in S is written as

$$\exists x \in S\,(P(x))$$

and it is shorthand for $\exists x\,(x \in S \land P(x))$.

♣ Cardinality

Definition. Let A be a set and $n \in \mathbb{N}$. We say that the *cardinality* of A is equal to n if A has n many distinct elements. And we write $|A| = n$ and say A is a *finite* set in this case.

Example.

- Let $S = \{x \in \mathbb{Z}^+ \mid x \text{ is odd and } x < 10\}$. Then $|S| = 5$.
- Let A be the set of letters in the Arabic alphabet. Then $|A| = 28$.
- $|\emptyset| = 0$. ◊

A set S is not finite when there does not exist a natural number n such that S contains n elements. In this case, we say that the set is *infinite*. For example, all the sets $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, and $\mathbb{R}$ are infinite.

2.2 Set Operations

We will introduce several operations performed on sets to obtain new sets.

♣ Power Sets

We start by introducing an important set, called the power set. Let us do an exercise by listing all subsets of $A = \{1, 2\}$. The four subsets of A are $\emptyset$, $\{1\}$, $\{2\}$, $\{1, 2\}$. We collect all these subsets in one set, and call it the power set of A.

Definition. Let S be a set. The *power set* of S is the set of all subsets of S. The power set of S is denoted by $\mathcal{P}(S)$.

In other words,

$$\mathcal{P}(S) = \big\{x \mid x \subseteq S\big\}.$$

Example. Some examples of power sets.

- $\mathcal{P}(\emptyset) = \{\emptyset\}$.
- $\mathcal{P}(\{1\}) = \Big\{\emptyset, \{1\}\Big\}$.
- $\mathcal{P}(\{1, 2\}) = \Big\{\emptyset, \{1\}, \{2\}, \{1, 2\}\Big\}$.
- $\mathcal{P}(\{1, 2, 3\}) = \Big\{\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}\Big\}$.
- $\mathcal{P}(\{\emptyset\}) = \Big\{\emptyset, \{\emptyset\}\Big\}$.
- $\mathcal{P}(\{\emptyset, z\}) = \Big\{\emptyset, \{\emptyset\}, \{z\}, \{\emptyset, z\}\Big\}$. $\Diamond$

Examine the examples above. What is the cardinality of the power set in terms of the cardinality of the set?

Theorem 2.3. *Let S be a finite set. Then*

$$|\mathcal{P}(S)| = 2^{|S|}.$$

Proof. Let $S = \{s_1, s_2, \ldots, s_n\}$ be a finite set with $|S| = n$ for some natural number n. We can associate to each subset $A \subseteq S$ a binary string $x_1 x_2 \cdots x_n$ of length n defined by setting the i^{th} digit x_i as follows.

$$x_i = \begin{cases} 0 & \text{if } s_i \notin A \\ 1 & \text{if } s_i \in A \end{cases}$$

For example, the binary string $000 \cdots 00$ of n many 0s is associated to the empty subset, and the string $100 \cdots 00$ of 1 followed by $n - 1$ many 0s is associated with the singleton $\{s_1\}$, and the string $111\cdots 11$ of n many 1s is associated with the whole set S. On the other hand, and in a similar fashion every binary string of length n corresponds to some subset of S. Therefore, the number of subsets of S is equal to the number of binary strings of length exactly n. But there are 2^n many binary strings of length n. Thus, there are 2^n different subsets of S. Therefore, we proved that $|\mathcal{P}(S)| = 2^n$. ■

♣ Cartesian Products

Members of sets are not ordered. We will introduce a new structure to represent ordered collections of objects. The *ordered n-tuple*

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

is the ordered list that has a_1 as the first element, a_2 as its second element, and so on, up to a_n its n^{th} and last element.

In other words, a tuple is a finite sequence of elements, where the order on which the objects appear in the tuple is significant. For example, (α, β) is a 2-tuple, while (β, α) is a different 2-tuple. Furthermore, $(8, 3, 9)$ is a 3-tuple, and $(2, 4, 6, 8)$ is a 4-tuple. We call 2-tuples *pairs* or *ordered pairs*, and 3-tuples are called *triples.*

We declare two n-tuples to be *equal* if and only if their corresponding elements are equal. More precisely,

$$(a_1, a_2, \ldots, a_n) = (b_1, b_2, \ldots, b_n) \iff (a_1 = b_1) \wedge (a_2 = b_2) \wedge \ldots \wedge (a_n = b_n).$$

For example, $(2, 3) \neq (3, 2)$ because the first element in the first tuple is 2 but the first element in the second tuple is 3. However, remember that $\{2, 3\} = \{3, 2\}$. Similarly, $(1, 2, 3) \neq (1, 3, 3)$ because the two triples differ in their second element.

Remark. As we mentioned earlier, most objects in mathematics can be defined using sets. We can use sets to define an ordered pair as follows. We think of the ordered pair (x, y) as the set $\big\{\{x\}, \{x, y\}\big\}$. This definition is due to the Polish mathematician Kazimierz Kuratowski. To justify this definition, check that if two pairs are equal, then their corresponding elements are also equal. That is, show that

$$\big\{\{a\}, \{a, b\}\big\} = \big\{\{x\}, \{x, y\}\big\} \text{ if and only if } a = x \wedge b = y.$$

How would you use ordered pairs to define a triple (x, y, z) using sets? One way would be $(x, y, z) = ((x, y), z)$. And in general, we define recursively an n-tuple as

$$(x_1, x_2, \ldots, x_{n-1}, x_n) = ((x_1, x_2, \ldots, x_{n-1}), x_n).$$

Next we use n-tuples to define another operation on sets.

Definition. The *cartesian product* $A \times B$ of sets A and B is the set of all ordered pairs whose first element belongs to A and second element belongs to B.

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

The term "Cartesian" is after the French philosopher and mathematician René Descartes.

Example. Let $A = \{1, 2, 3\}$ and $B = \{x, y\}$. Then

- $A \times B = \big\{(1, x), (1, y), (2, x), (2, y), (3, x), (3, y)\big\}$.
- $B \times A = \big\{(x, 1), (x, 2), (x, 3), (y, 1), (y, 2), (y, 3)\big\}$.
- $A \times \emptyset = \emptyset \times A = \emptyset$. $\diamondsuit$

Similarly, we define the cartesian product $A \times B \times C$ of three sets A, B, C to be the set of 3-tuples whose first element belongs to A, second element belongs to B, and third element belongs to C. That is,

$$A \times B \times C = \big\{(a, b, c) \mid a \in A \wedge b \in B \wedge c \in C\big\}.$$

Example. Let $A = \{0, 1\}$, $B = \{a, b\}$, and $C = \{\triangle, \square\}$. Then

$$A \times B \times C = \big\{(0, a, \triangle), (0, a, \square), (0, b, \triangle), (0, b, \square), (1, a, \triangle), (1, a, \square), (1, b, \triangle), (1, b, \square)\big\}.$$

$$A \times C \times B = \big\{(0, \triangle, a), (0, \triangle, b), (0, \square, a), (0, \square, b), (1, \triangle, a), (1, \triangle, b), (1, \square, a), (1, \square, b)\big\}.$$

$\diamondsuit$

We may generalise the cartesian product to any number of sets. Let $A_1, A_2, \ldots, A_n$ be sets. The cartesian product $A_1 \times A_2 \times \cdots \times A_n$ is the set of all n-tuples $(a_1, a_2, \ldots, a_n)$ such that the i^{th} element a_i belongs to the i^{th} set A_i for every $1 \leq i \leq n$.

$$A_1 \times A_2 \times \cdots \times A_n = \big\{(a_1, a_2, \ldots, a_n) \mid a_i \in A_i \text{ for } i = 1, 2, \ldots, n\big\}.$$

We define the n^{th} power of a set A, where n is a positive integer, as follows.

$$A^n = \underbrace{A \times A \times \cdots \times A}_{n \text{ times}} = \big\{(a_1, a_2, \ldots, a_n) \mid a_i \in A \text{ for } i = 1, 2, \ldots, n\big\}.$$

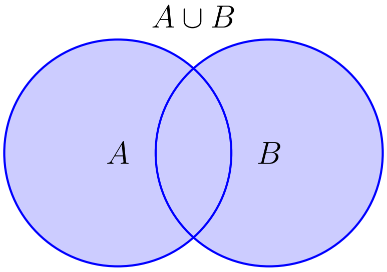
Example. Let $A = \{1, 2\}$. Compute A^2 and A^3.

$$\begin{aligned} A^2 &= A \times A = \{(1, 1), (1, 2), (2, 1), (2, 2)\}. \\ A^3 &= A \times A \times A \\ &= \{(1, 1, 1), (1, 1, 2), (1, 2, 1), (1, 2, 2), (2, 1, 1), (2, 1, 2), (2, 2, 1), (2, 2, 2)\}. \end{aligned}$$ $\diamondsuit$

♣ Union of Sets

The *union* $A \cup B$ of sets A and B is the set that contains precisely those elements that are either in A *or* in B, or in both.

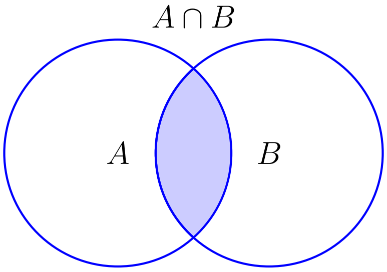
$$A \cup B = \{x \mid x \in A \vee x \in B\}.$$



♣ Intersection of Sets

The *intersection* $A \cap B$ of sets A and B is the set containing precisely those elements that are in both A *and* in B.

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



Example.

- $\{1, 3, 5\} \cup \{1, 2, 3\} = \{1, 2, 3, 5\}$.
- $\{1, 3, 5\} \cap \{1, 2, 3\} = \{1, 3\}$.
- $\{1, 3, 5\} \cap \{2, 4\} = \emptyset$. ◊

Definition. Two sets A and B are called *disjoint* if $A \cap B = \emptyset$.

How many elements in $A \cup B$ are there?

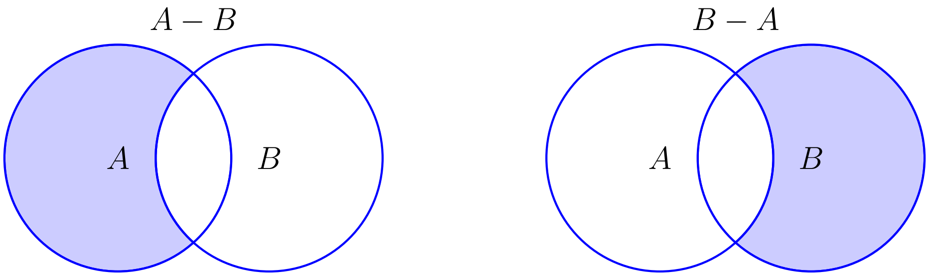
Lemma 2.4. *Let A, B be finite sets. Then*

$$|A \cup B| = |A| + |B| - |A \cap B|$$

♣ Difference of Sets

The *difference* of sets A and B denoted by $A - B$ or $A \setminus B$ is the set containing precisely those elements which are in A but not in B.

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



Example.

- $\{1, 3, 5\} - \{1, 2, 3\} = \{5\}$.
- $\{1, 2, 3\} - \{1, 3, 5\} = \{2\}$.
- $\{1, a, 2, b, 3, c\} \setminus \{1, b, 2, w, t, z\} = \{a, 3, c\}$. ◊

♣ Complement of Sets

To define complements of sets we need to work in a large universal set U which is the set that contains all objects under consideration.

Definition. The *complement* of A, denoted by $\overline{A}$, is the set of elements that are in the universal set U but not in A. That is, $\overline{A} = U - A$.

$$\overline{A} = \{x \in U \mid x \notin A\}$$

Example. Consider the universal set $U = \mathbb{Z}^+$ and the set $A = \{n \in \mathbb{Z}^+ \mid n > 7\}$. Then $\overline{A} = \{1, 2, 3, 4, 5, 6, 7\}$. ◊

♣ Set Identities

Let A, B, C be sets and U be the universal set which contains all objects under consideration. In particular, each of A, B, C is a subset of U.

Identity Laws	$A \cap U = A$ $A \cup \emptyset = A$
Domination Laws	$A \cup U = U$ $A \cap \emptyset = \emptyset$
Idempotent Laws	$A \cup A = A$ $A \cap A = A$
Complementation Laws	$\overline{(\overline{A})} = A$
Commutative Laws	$A \cup B = B \cup A$ $A \cap B = B \cap A$
Associative Laws	$A \cup (B \cup C) = (A \cup B) \cup C$ $A \cap (B \cap C) = (A \cap B) \cap C$
Distributive Laws	$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$ $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
Complement Laws	$A \cup \overline{A} = U$ $A \cap \overline{A} = \emptyset$
De Morgan's Laws	$\overline{A \cap B} = \overline{A} \cup \overline{B}$ $\overline{A \cup B} = \overline{A} \cap \overline{B}$

All of the identities above are true statements and the reader is advised to write a proof for each of them. Let us prove one of the De Morgan's laws.

Lemma 2.5. *Let A and B be sets. Then the complement of their intersection is equal to the union of their complements. That is,*

$$\overline{A \cap B} = \overline{A} \cup \overline{B}.$$

Proof. Using the definition of complement, intersection and union we have,

$$\begin{aligned}\overline{A \cap B} &= \{x \in U \mid x \notin A \cap B\}\\ &= \{x \in U \mid \neg(x \in A \cap B)\}\\ &= \{x \in U \mid \neg(x \in A \wedge x \in B)\}\\ &= \{x \in U \mid x \notin A \vee x \notin B\}\\ &= \{x \in U \mid x \in \overline{A} \vee x \in \overline{B}\}\\ &= \overline{A} \cup \overline{B}.\end{aligned}$$

■

Next, we prove one of the Distributive Laws.

Lemma 2.6. *Let A and B be sets. Then $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.*

Proof. Using the definition of intersection and union, we have

$$\begin{aligned}A \cap (B \cup C) &= \{x \mid x \in A \wedge x \in B \cup C\}\\ &= \{x \mid x \in A \wedge (x \in B \vee x \in C)\}\\ &= \{x \mid (x \in A \wedge x \in B) \vee (x \in A \wedge x \in C)\}\\ &= \{x \mid (x \in A \cap B) \vee (x \in A \cap C)\}\\ &= (A \cap B) \cup (A \cap C).\end{aligned}$$

■

Example. Show that $\overline{A \cup (B \cap C)} = (\overline{C} \cup \overline{B}) \cap \overline{A}$.
We will use the set identities.

$$\begin{aligned}\overline{A \cup (B \cap C)} &= \overline{A} \cap \overline{(B \cap C)} && \text{(By De Morgan's Law)}\\ &= \overline{A} \cap (\overline{B} \cup \overline{C}) && \text{(By De Morgan's Law)}\\ &= (\overline{B} \cup \overline{C}) \cap \overline{A} && \text{(By Commutative Law)}\\ &= (\overline{C} \cup \overline{B}) \cap \overline{A}. && \text{(By Commutative Law)}\end{aligned}$$

◊

Another way to show that two sets are equal is using Lemma 2.1 which states that two sets are equal if and only if they are subsets of each other.

Example. Let A and B be sets. Prove that $A \cap (A \cup B) = A$.

Proof. First, we show that $A \cap (A \cup B) \subseteq A$. So let $x \in A \cap (A \cup B)$. By definition of intersection, it follows that $x \in A$ and $x \in (A \cup B)$. So $x \in A$. Thus every element in $A \cap (A \cup B)$ is also in A, and so $A \cap (A \cup B) \subseteq A$.

Second, we show that $A \subseteq A \cap (A \cup B)$. Suppose that $x \in A$. By definition of union, it follows that $x \in A \cup B$. Since, $x \in A$ and $x \in A \cup B$, we conclude that $x \in A \cap (A \cup B)$. Thus, $A \subseteq A \cap (A \cup B)$. Therefore, since A and $A \cap (A \cup B)$ are subsets of each other, we have that $A \cap (A \cup B) = A$ as desired. ■

Since unions and intersections satisfy the associative law, we use $A \cup B \cup C$ and $A \cap B \cap C$ to denote the union and intersection, respectively, of the sets A, B, C with no ambiguity.

Example. Let $A = \{0, 2, 4, 6, 8\}$, $B = \{0, 1, 2, 3, 4\}$, $C = \{0, 3, 6, 9\}$.

- $A \cup B \cup C = (A \cup B) \cup C = A \cup (B \cup C) = \{0, 1, 2, 3, 4, 6, 8, 9\}$.
- $A \cap B \cap C = (A \cap B) \cap C = A \cap (B \cap C) = \{0\}$. ♢

We may consider the unions and intersections of any collection of sets. Let us examine these sets more precisely.

Definition. The *union* of a collection of sets is the set which contains precisely those elements which belong to at least one set in the collection.

We use the following notation to express the union of n many sets $A_1, A_2, \ldots, A_n$.

$$\bigcup_{i=1}^{n} A_i = A_1 \cup A_2 \cup \cdots \cup A_n = \{x \mid x \in A_1 \vee x \in A_2 \vee \cdots \vee x \in A_n\}.$$

Definition. The *intersection* of a collection of sets is the set which contains precisely those elements which belong to every set in the collection.

We use the following notation to express the intersection of n many sets.

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

Example. For each $i \in \mathbb{Z}^+$ let $A_i = \{n \in \mathbb{Z} \mid i \leq n \leq 3i\}$.

- $A_1 = \{1, 2, 3\}$.
- $A_2 = \{2, 3, 4, 5, 6\}$.
- $\bigcup_{i=1}^{3} A_i = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$.

- $\bigcap\limits_{i=1}^{3} A_i = \{3\}$. ◊

Our definitions of unions and intersections above apply also to infinite collections of sets. Given an infinite family $\{A_i \mid i \in \mathbb{Z}\}$ of sets, we use the following notation to express their union.

$$\bigcup_{i\in\mathbb{Z}^+} A_i = \bigcup_{i=1}^{\infty} A_i = A_1 \cup A_2 \cup A_3 \cup \cdots = \big\{x \mid \exists i \in \mathbb{Z}^+ (x \in A_i)\big\}.$$

Similarly, the intersection of this family of sets is denoted by

$$\bigcap_{i\in\mathbb{Z}^+} A_i = \bigcap_{i=1}^{\infty} A_i = A_1 \cap A_2 \cap A_3 \cap \cdots = \big\{x \mid \forall i \in \mathbb{Z}^+ (x \in A_i)\big\}.$$

Example. For each $i \in \mathbb{Z}^+$, let $A_i = \{0, 1, 2, \ldots, i\}$ and $B_i = \{2i, 3i\}$

(i) $\bigcup\limits_{i=1}^{\infty} A_i = \mathbb{N}$.

(ii) $\bigcap\limits_{i=1}^{\infty} A_i = \{0, 1\}$.

(iii) $\bigcup\limits_{i=1}^{\infty} B_i = \{2, 3, 4, 6, 8, 9, 10, 12, 14, 15, 16, 18, \ldots\}$.

(iv) $\bigcap\limits_{i=1}^{\infty} B_i = \emptyset$. ◊

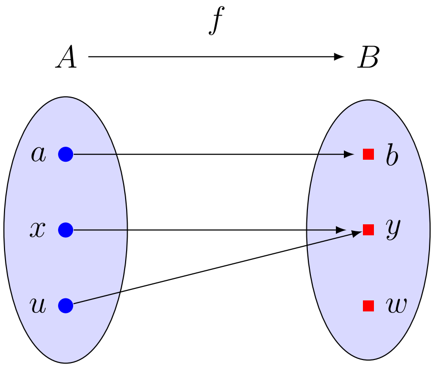
2.3 Functions

A function is a very important mathematical concept that was developed in the study of infinitesimal calculus at the end of the 17^{th} century. Functions are central objects in many fields of mathematics and are ubiquitous in mathematics in general. They are also widely used in science and economics to model physical and social phenomena. A function is also called a *map* or *mapping*.

Definition. A *function* $f : A \to B$ from a set A to a set B is an assignment which for each element $a \in A$, it assigns exactly one element $b \in B$.

We write $f(a) = b$ when the function f assigns the unique element $b \in B$ to the element $a \in A$. In this case, we say that b is the *image* of a, and we also say that a is the *preimage* of b. We also write $a \mapsto b$ when $f(a) = b$.

Given the function $f : A \to B$ represented by the diagram below, we have that $f(a) = b$, and $f(x) = y$, and $f(u) = y$. In other words, the image of a is b, and the image of x is y, and the image of u is also y. Notice that every element in A must have an image in B, however, $w \in B$ has no preimage in A, that is fine. Moreover, no element in A has two images, this is not allowed in functions.



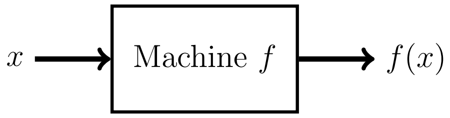
Let $f : A \to B$ be a function from set A to B.

- The *domain* of f is the set A.
- The *codomain* of f is the set B.
- The *range* of f is the set of the images of all elements in the domain. That is,

  $$\text{range}(f) = \{f(a) \mid a \in A\}.$$

  Thus we always have that $\text{range}(f) \subseteq \text{codom}(f)$.

We learned that a function from a set to another is an assignment that associates to every element of the first set (domain) exactly one element from the second set (codomain). This assignment could be described by a mathematical formula, rule, table, natural language, computer program, or any other process. That being said, it is useful to think of a function as a machine. This machine takes in an input from the domain and produces exactly one corresponding output in the codomain.



Example. Let $g : \mathbb{Z} \to \mathbb{Z}$ which assigns to each integer its square. That is, $g(n) = n^2$ for every integer n.

- $g(0) = 0,\ g(1) = 1,\ g(2) = 4,\ g(-3) = 9,\ g(-7) = 49.$
- $\mathrm{dom}(g) = \mathrm{codom}(g) = \mathbb{Z}.$
- $\mathrm{range}(g) = \{0, 1, 4, 9, 16, 25, 36, 49, \ldots\}.$ $\Diamond$

Example. Let $\Gamma = \{$Omar, Faten, Mona, Khaled, Yasmin, Hind$\}$, and Λ be the set of English alphabet. Consider the function $f : \Gamma \to \Lambda$ which assigns for each name in Γ its first letter.

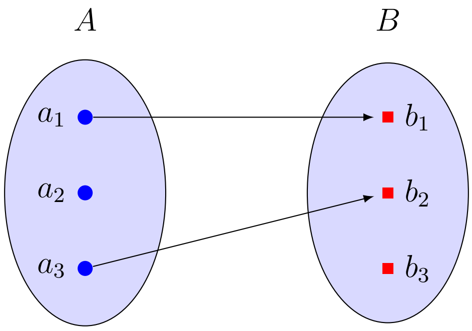
- $f(\text{Omar}) = O,\ f(\text{Faten}) = F,\ f(\text{Mona}) = M.$
- $f(\text{Khaled}) = K,\ f(\text{Yasmin}) = Y,\ f(\text{Hind}) = H.$
- $\mathrm{dom}(f) = \Gamma$ and $\mathrm{codom}(f) = \Lambda = \{A, B, C, \ldots, X, Y, Z\}.$
- $\mathrm{range}(f) = \{O, F, M, K, Y, H\}.$ $\Diamond$

Example. Here are real life functions. What is a suitable domain and codomain?

- Think of the function which assigns to every fruit in a supermarket its price.
- Think of the function which assigns to every person their biological mother.
- Think of the function which assigns to every student their ID number. $\Diamond$

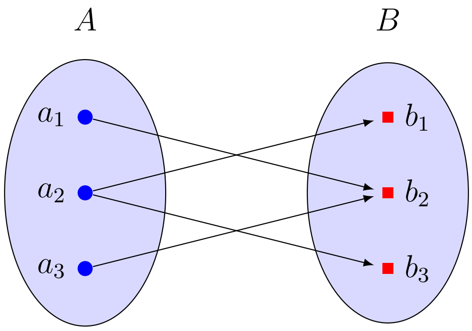
Example. We may think of a unary predicate $P(x)$ as a function with domain equals to the domain of the variable and codomain equals to the set $\{T, F\}$. This function assigns to every value in the domain the truth value of the resulting proposition when this value is substituted in the predicate. For example, if $P(x)$ says "x is prime" where x is a positive integer, then thinking of $P(x)$ as the function $P : \mathbb{Z}^+ \to \{T, F\}$ we have that $P(5) = T$, $P(7) = T$, $P(9) = F$ and $P(29) = T$. $\Diamond$

Example. The following assignment is *not* a function from A to B because $a_2 \in A$ is not assigned any image in B. This is a violation of the definition of a function since a function assigns an image to *every* element in the domain.



$\Diamond$

Example. The following assignment is *not* a function from A to B because $a_2 \in A$ is assigned two images in B, namely b_1 and b_3. This violates the definition of a function since a function assigns *exactly one* image to every element in the domain.



$\Diamond$

Remember when we said that almost every object in mathematics can be defined using sets? Sets give us another way to define functions. We may think of a function as the set of all pairs of domain elements and their corresponding images. We call this set the graph of the function.

Definition. The *graph* of a function $f : A \to B$ is the set

$$\text{Graph}(f) = \big\{(a, b) \mid a \in A \wedge b \in B \wedge f(a) = b\big\}.$$

Clearly, $\text{Graph}(f) \subseteq A \times B$.

Example. Consider the function $f : \{-2, -1, 0, 1, 2, 3\} \to \mathbb{Z}$ where $f(n) = n^2$. Then its graph is

$$\begin{aligned} \text{Graph}(f) &= \{(n, n^2) \mid n \in \text{dom}(f)\} \\ &= \{(0, 0), (1, 1), (2, 4), (3, 9), (-1, 1), (-2, 4)\}. \end{aligned} \qquad \Diamond$$

Given a function, we can assign to any subset of the domain a corresponding subset of the codomain. Let S be some subset of the domain, we define the image of S under the function to be the set of images of all the elements in S. More precisely, we have the following definition.

Definition. Let $f : A \to B$ be a function and $S \subseteq A$. The image of S under f is the set of the images of all elements in S, that is,

$$f(S) = \{f(s) \mid s \in S\}.$$

Note that $f(S) \subseteq B$ and $f(A) = \text{range}(f)$.

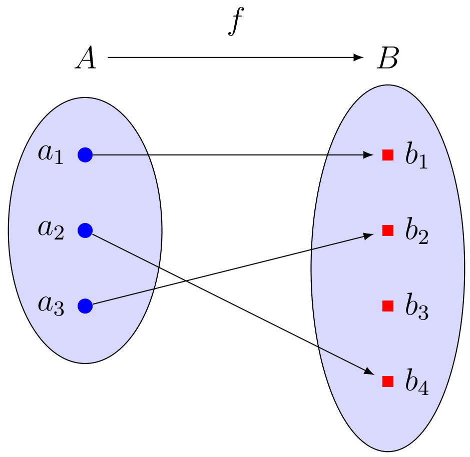
Example. Let $f : \mathbb{Z} \to \mathbb{Z}$ where $f(n) = n^2$, and let $S = \{-2, -1, 0, 1, 2, 6, 9\}$. Then

$$f(S) = \{f(-2), f(-1), f(0), f(1), f(2), f(6), f(9)\} = \{4, 1, 0, 36, 81\}. \qquad \Diamond$$

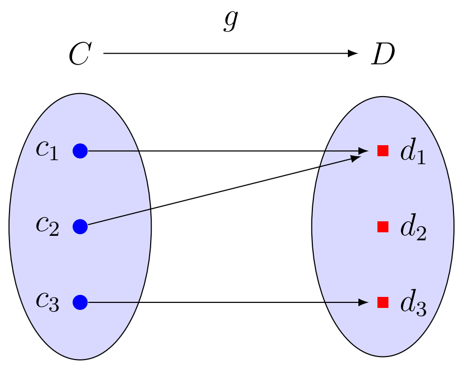
♣ Injective Functions

Definition. A function $f : A \to B$ is called *injective* if for all $a_1, a_2 \in A$, we have that if $a_1 \neq a_2$ then $f(a_1) \neq f(a_2)$.

An injective function is also called *one-to-one*. A function f being injective means that f never assigns the same image to two different elements in the domain. By the contrapositive, a function is injective if and only if whenever $f(a_1) = f(a_2)$ then $a_1 = a_2$ for every a_1, a_2 in the domain of f. Finally, we can also say that a function is injective if and only if every element in the range has exactly one preimage in the domain. The function $f : A \to B$ represented below is injective.



On the other hand, the function $g : C \to D$ below is not injective because $c_1 \neq c_2$ but $f(c_1) = d_1 = f(c_2)$.



Given a function f, using predicate logic with variables x_1, x_2 whose domain is the same as the domain of f, we have that the function f is injective if and only if

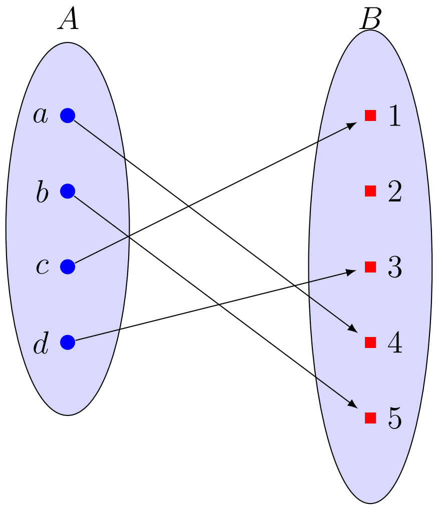
$$\forall x_1 \forall x_2 \big(x_1 \neq x_2 \to f(x_1) \neq f(x_2)\big).$$

Equivalently,

$$\forall x_1 \forall x_2 \big(f(x_1) = f(x_2) \to x_1 = x_2\big).$$

Warning. A function being one-to-one (injective) does not mean that every element in the domain has a unique image in the codomain; this is true for all functions (read the definition of a function). One-to-one means no two distinct elements of the domain are assigned the same image.

Example. The function below is injective (or one-to-one), since no two different elements in the domain are assigned the same image in the codomain.



◊

Example. Is the function $f : \mathbb{Z} \to \mathbb{Z}$ given by $f(n) = n^2$ injective?

No, it is not injective because $2 \neq -2$, however, $f(2) = 4 = f(-2)$. ◊

Example. Show that the function $f : \mathbb{Z}^+ \to \mathbb{Z}^+$ given by $f(n) = n^2$ is injective.

Let $m, n \in \mathbb{Z}^+ = \{1, 2, 3, \ldots\}$. We will show that $m \neq n$ implies $f(m) \neq f(n)$. So suppose that $m \neq n$. Without loss of generality, assume that $m < n$. Since $m > 0$, multiplying the inequality $m < n$ by m we get $m^2 < mn$. Next, multiply $m < n$ by n, again as $n > 0$, we get $mn < n^2$. Thus, $m^2 < mn < n^2$, and so $m^2 < n^2$. Therefore, $m^2 \neq n^2$, that is, $f(m) \neq f(n)$. This shows that any distinct elements of the domain of f have distinct images in the codomain, thus, f is injective. ◊

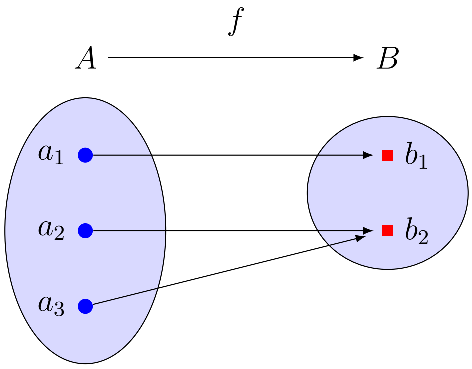
♣ Surjective Functions

Definition. A function $f : A \to B$ is *surjective* if for every $b \in B$ there exists $a \in A$ such that $f(a) = b$.

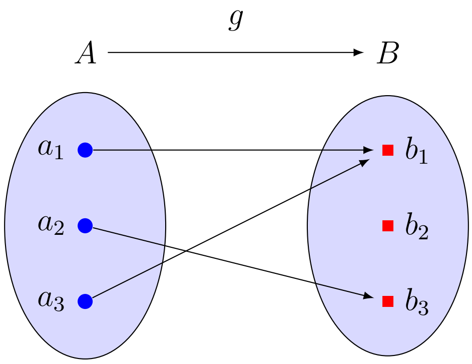
So a function is surjective if every element in the codomain has at least one preimage in the domain. Surjective functions are also called *onto*. Using predicate logic we can express that a function $f : A \to B$ is surjective by writing

$$\forall y \in B \, \exists x \in A \, (f(x) = y).$$

Note that a function $f : A \to B$ is surjective if and only if its range is equal to its codomain, that is, $f(A) = B$. The function f below is surjective because all elements in the codomain, namely b_1, b_2, have preimages in the domain A.



The function g below is not surjective because the element b_2 in the codomain has no preimage.



Example. Is $f : \mathbb{Z} \to \mathbb{Z}$ given by $f(n) = n^2$ surjective?

No, it is not surjective, because 2 in the codomain has no preimage. In other words, there is no $n \in \mathbb{Z}$ such that $f(n) = 2$. $\diamondsuit$

Example. Is $f : \mathbb{Z} \to \mathbb{Z}$ given by $f(n) = n + 1$ surjective?

Yes, it is surjective, because for every integer $n \in \mathbb{Z}$ there is an integer $m \in \mathbb{Z}$ such that $f(m) = n$. To see this, choose any integer n in the codomain, and take $m = n - 1$ which is in the domain. Now see that m is the preimage of n since $f(m) = m + 1 = (n - 1) + 1 = n$. $\diamondsuit$

♣ Bijective Functions

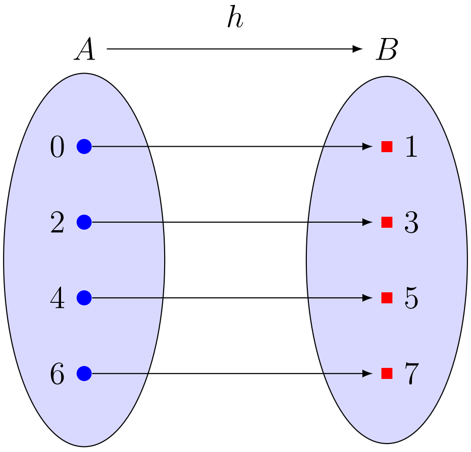
Definition. A function is *bijective* if it is both injective and surjective.

We express that $f : A \to B$ is bijective in predicate logic as follows.

$$\forall x \in A \, \forall y \in A \, (x \neq y \to f(x) \neq f(y)) \land \forall z \in B \, \exists w \in A \, (f(w) = z).$$

When a function is bijective, we also say it is a *bijection* or *one-to-one correspondence*. How do bijections look like? Let us analyse them. A bijection f is an injective surjective function. First, as f is a function each element in the domain is assigned exactly one image in the codomain. Second, as f is surjective, each element in the codomain has at least one preimage in the domain. Moreover, since f is injective it follows that each element in the codomain must have exactly one preimage. To sum up, a function is bijective if and only if every element in the codomain has exactly one preimage in the domain. Therefore, when we have a bijective function we know that we have an assignment in which every element in the domain has exactly one image (function) and every element in the codomain has exactly one preimage (bijective).

Example. Let $A = \{0, 2, 4, 6\}$ and $B = \{1, 3, 5, 7\}$. The function $h : A \to B$ given by $h(n) = n + 1$ is a bijection.

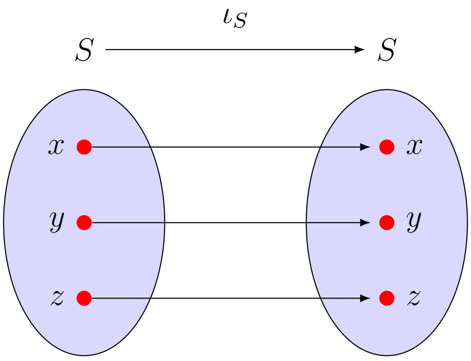


$\diamondsuit$

Definition. Let A be a set. The *identity function* on A is the function $\iota_A : A \to A$ given by $\iota_A(x) = x$ for all $x \in A$.

It should be clear that the identity function on any set is a bijective function.

Example. Let $S = \{x, y, z\}$. The identity function on S is represented as follows.

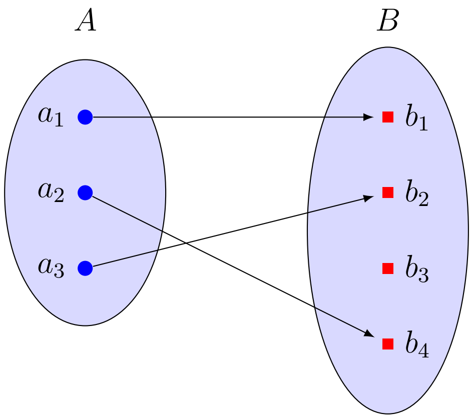


Find all other bijections from S to S. ◊

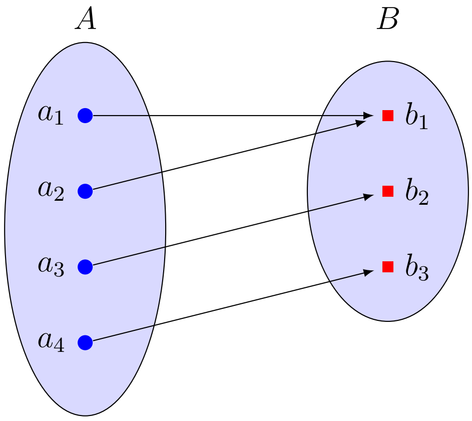
Using the properties of injectivity and surjectivity we can classify functions into four different categories as described below.

Four Types of Functions

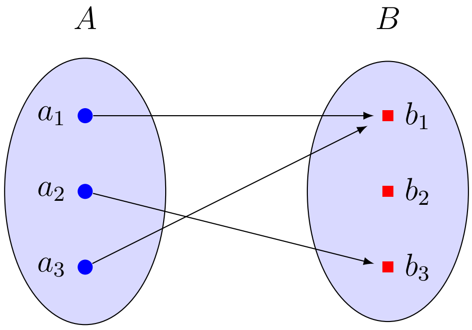
1. Injective and not surjective functions.



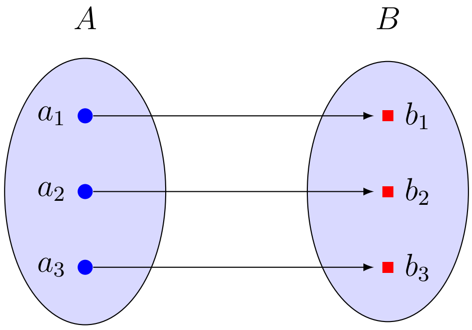
2. Surjective and not injective functions.



3. Not injective and not surjective functions.



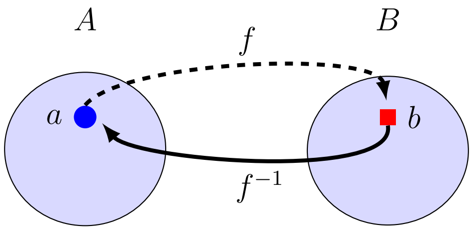
4. Bijective functions.



♣ Inverse of a Function

In the world of real numbers we have the ordinary operation of multiplication. We call the number 1 the multiplicative identity and every nonzero real number has a multiplicative inverse. For instance, the inverse of 3 is $\frac{1}{3}$ since $3 \times \frac{1}{3} = 1$ and we call $\frac{1}{3}$ the inverse of 3. Moreover, the number zero has no inverse. Now, let's turn our attention to the world of functions, what operation should we define on functions? Which functions play the role of the identity? And which functions have inverses, and which don't?

Given a bijection $f : A \to B$, choose any element b in the codomain B. As f is bijective, we know that b has exactly one preimage $a \in A$ as depicted by the dashed arrow below. This fact allows us to define a new function f^{-1} from B to A where the image of b under f^{-1} is the *unique* preimage of b under f.



Definition. Let $f : A \to B$ be a bijection. The *inverse* of f is the function $f^{-1} : B \to A$ that assigns to every $b \in B$ the unique element $a \in A$ such that $f(a) = b$.

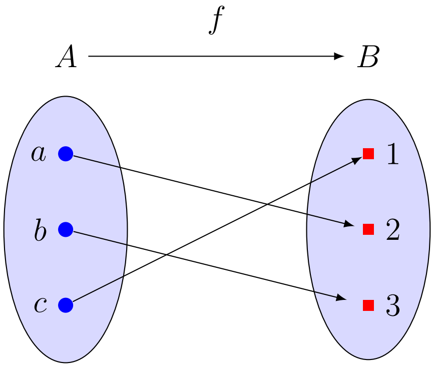
Remark. The definition tells us that $f(a) = b$ if and only if $f^{-1}(b) = a$ for any $a \in A$ and $b \in B$.

Note that $\text{dom}(f^{-1}) = \text{codom}(f)$ and $\text{codom}(f^{-1}) = \text{dom}(f)$. When f^{-1} exists, we call f *invertible*. So a function is invertible exactly means that it is bijective. Let us explain how the inverse function works in steps.

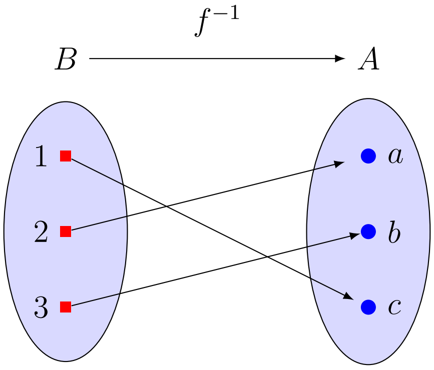
1. We start with a bijection $f : A \to B$.
2. Our aim is to define a function f^{-1} from B to A using f.
3. This means we need to assign to every $b \in B$ some unique image in A.
4. So choose any element $b \in B$. (Remember that $\text{dom}(f^{-1}) = B$.)
5. Since f is a bijection, our b has a unique preimage $a \in A$. In other words, there exists exactly one $a \in A$ such that $f(a) = b$, see the dashed arrow above.
6. We declare this unique a to be the image of b under the new function f^{-1}. Therefore, $f^{-1}(b) = a$ where $f(a) = b$ (i.e. a is the preimage of b under f).

Question. What goes wrong in defining the inverse function when the function we start with is not bijective in the first place?

Example. Let $A = \{a, b, c\}$ and $B = \{1, 2, 3\}$. Let $f : A \to B$ where $f(a) = 2$, and $f(b) = 3$, and $f(c) = 1$.



Then f is a bijection, and so invertible. The inverse function is $f^{-1} : B \to A$ where $f^{-1}(1) = c$ because $f(c) = 1$, and $f^{-1}(2) = a$ because $f(a) = 2$, and $f^{-1}(3) = b$ because $f(b) = 3$.



◊

Example. Consider the successor function $f : \mathbb{Z} \to \mathbb{Z}$ given by $f(n) = n + 1$. We know that f is bijective. Find its inverse $f^{-1} : \mathbb{Z} \to \mathbb{Z}$.

Let $m \in \text{dom}(f^{-1}) = \mathbb{Z}$. Since f is bijective and $m \in \text{codom}(f)$, there exists a unique $n \in \text{dom}(f)$ such that $f(n) = m$. Thus, $n + 1 = m$. By definition of the inverse function, this n is the image of m under f^{-1}, that is, $f^{-1}(m) = n = m - 1$. Therefore, the inverse function $f^{-1} : \mathbb{Z} \to \mathbb{Z}$ is given by $f^{-1}(m) = m - 1$ for every $m \in \text{dom}(f^{-1})$. ♢

Example. Is $f : \mathbb{R} \to \mathbb{R}$ where $f(x) = x^2$ invertible?

No, because f is not injective as $f(2) = f(-2)$ but $2 \neq -2$. (Also f is not surjective.) To be able to define the inverse, the function must be bijective. ♢

We can restrict the domain and codomain of the previous function to obtain a bijective function. Recall that $\mathbb{R}^{\geq 0} = \{x \in \mathbb{R} \mid x \geq 0\}$, the set of non-negative real numbers.

Example. Show that $f : \mathbb{R}^{\geq 0} \to \mathbb{R}^{\geq 0}$ where $f(x) = x^2$ is bijective. And then find its inverse $f^{-1} : \mathbb{R}^{\geq 0} \to \mathbb{R}^{\geq 0}$.

First, we show that f is injective. Suppose that $x, y \in \mathbb{R}^{\geq 0}$ such that $f(x) = f(y)$. So $x^2 = y^2$, which means that $x^2 - y^2 = 0$. And so $(x - y)(x + y) = 0$. This implies that either $x - y = 0$ or $x + y = 0$. If the former holds then $x = y$ as desired. Otherwise, if $x + y = 0$ holds, then since both $x, y \geq 0$, we must have that $x = y = 0$, and so $x = y$ as well. So f is injective.
Second we show that f is onto. Choose any $y \in \mathbb{R}^{\geq 0}$ (the codomain of f). We need to search for a preimage of y. Suppose the preimage of y is some x. Then $f(x) = y$, implying that $x^2 = y$, and so $x = \sqrt{y}$. Note that since $y \geq 0$, we have that $\sqrt{y}$ is a defined real number and $\sqrt{y} \geq 0$, meaning that x belongs to the domain $\mathbb{R}^{\geq 0}$ of f. Now let us check that x is indeed the preimage of y under f.

$$f(x) = x^2 = (\sqrt{y})^2 = y.$$

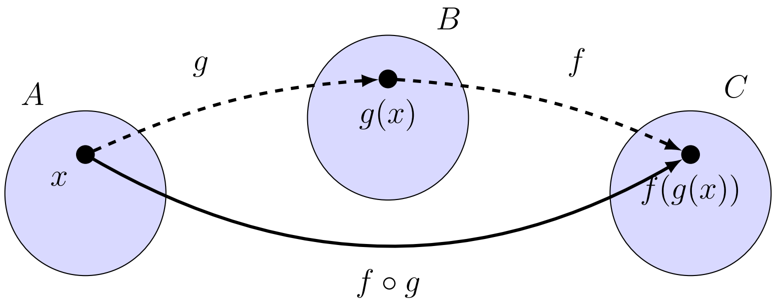
This shows that f is onto.
Now it is easy to find the inverse function since $f^{-1}(y)$ is the unique preimage of y under f, which we found to be $\sqrt{y}$. Therefore, the inverse function is $f^{-1} : \mathbb{R}^{\geq 0} \to \mathbb{R}^{\geq 0}$ where $f^{-1}(y) = \sqrt{y}$. ♢

♣ Composition of Functions

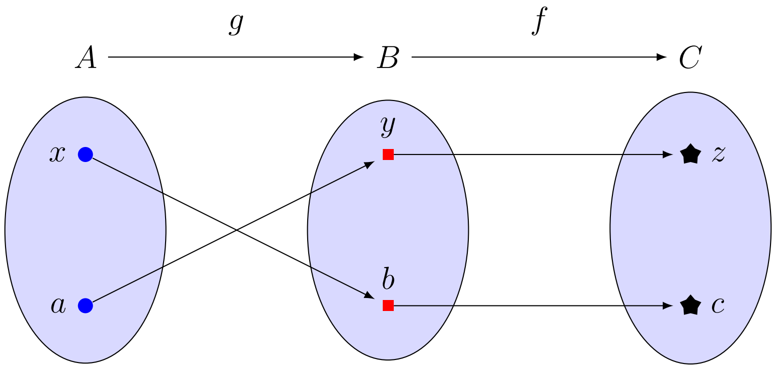
Composition is an operation which takes two functions as an input and produces a new function as an output.

Definition. Let $g : A \to B$ and $f : B \to C$ be functions. The *composition of f after g* is the function $f \circ g : A \to C$ where for each $x \in A$ we have that

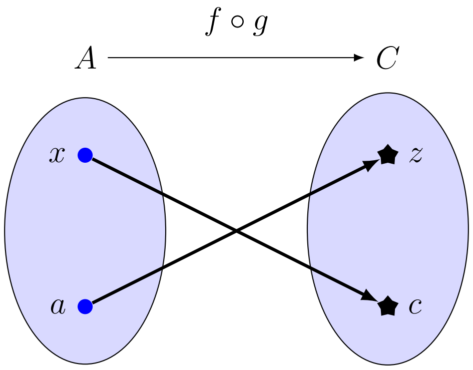
$$(f \circ g)(x) = f(g(x)).$$



Example. Find the composition function $f \circ g : A \to C$ where the functions g and f are as described below.

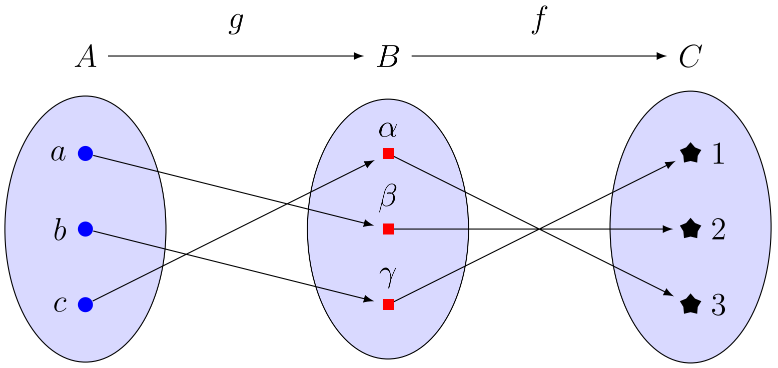


We have that $(f \circ g)(x) = f(g(x)) = f(b) = c$ and, similarly, $(f \circ g)(a) = z$. Thus, the composition function $f \circ g : A \to C$ is given by the diagram below.



In general, for two function $g : A \to B$ and $f : B \to C$ we must have that $\text{range}(g) \subseteq \text{dom}(f)$ for the composition $f \circ g$ to be defined. Moreover, the domain of $f \circ g$ is the domain of g, and the range of $f \circ g$ is the image of the range of g under f, that is, the range of $(f \circ g)$ is the set $f(g(A))$.

Example. Find the graph of the composition function $f \circ g$ where the functions $g : A \to B$ and $f : B \to C$ are given below.



The images under the composition function $f \circ g : A \to C$ are:

- $(f \circ g)(a) = f(g(a)) = f(\beta) = 2$.
- $(f \circ g)(b) = f(g(b)) = f(\gamma) = 1$.
- $(f \circ g)(c) = f(g(c)) = f(\alpha) = 3$.

Thus,

$$\text{Graph}(f \circ g) = \{(a, 2), (b, 1), (c, 3)\}. \qquad \Diamond$$

Example. Let $f : \mathbb{N} \to \mathbb{N}$ where $f(n) = 2n+3$ and $g : \mathbb{N} \to \mathbb{N}$ where $g(n) = 3n+2$. Then we have the composition functions $f \circ g : \mathbb{N} \to \mathbb{N}$ and $g \circ f : \mathbb{N} \to \mathbb{N}$ where

- $(f \circ g)(n) = f(g(n)) = f(3n + 2) = 2(3n + 2) + 3 = 6n + 7$, and
- $(g \circ f)(n) = g(f(n)) = g(2n + 3) = 3(2n + 3) + 2 = 6n + 11$. $\Diamond$

Example. Consider the functions $f : \mathbb{R} \to \mathbb{R}^{\geq 0}$ where $f(x) = x^2$ and $g : \mathbb{R}^{\geq 0} \to \mathbb{R}$ where $g(x) = \sqrt{x}$. Then the composition function $f \circ g : \mathbb{R}^{\geq 0} \to \mathbb{R}^{\geq 0}$ is given by

$$(f \circ g)(x) = f(g(x)) = f(\sqrt{x}) = (\sqrt{x})^2 = x.$$

And the composition function $g \circ f : \mathbb{R} \to \mathbb{R}$ is given by

$$(g \circ f)(x) = g(f(x)) = g(x^2) = \sqrt{x^2} = |x|. \qquad \Diamond$$

At this point we have the composition operation on functions, inverse functions, and the identity functions. We will now show that when we compose a function with its inverse we get the identity function. This is similar to when we multiply a real number with its multiplicative inverse and get 1.

Suppose that $f : A \to B$ is a bijection, and let $f^{-1} : B \to A$ be its its inverse. From the way the inverse function was defined it follows that for every $a \in A$ and $b \in B$ we have that

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

This fact implies that $f^{-1} \circ f : A \to A$ is the identity function ι_A on A. To see this, let $a \in A$ and suppose that $f(a) = b$, then

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

Similarly, $f \circ f^{-1} : B \to B$ is the identity function ι_B on B. Let $b \in B$ and suppose that $f^{-1}(b) = a$, then

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

Gladly every thing worked well, that is, $f^{-1} \circ f = \iota_A$ and $f \circ f^{-1} = \iota_B$.

♣ Special Functions

In the remaining part of this functions we examine well-known functions in mathematics such as the factorial function, the ceiling and floor functions, and real functions.

Factorial Function

The *factorial function* $f : \mathbb{N} \to \mathbb{Z}^+$ is defined as follows.

- $f(0) = 1$,
- $f(n+1) = (n+1) \cdot f(n)$, for $n \geq 0$.

The image $f(n)$ is denoted by $n!$. Thus it follows that $f(0) = 0! = 1$ and for $n \geq 1$ we have that

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

Example.

- $f(1) = 1! = 1$.
- $f(2) = 2! = 2 \cdot 1 = 2$.
- $f(6) = 6! = 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 720$.
- $f(10) = 10! = 10 \cdot 9 \cdot 8 \cdot 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 3\,628\,800$. ♢

Floor and Ceiling Functions

The *floor function* assigns to every real number x the maximum integer $\lfloor x \rfloor$ less than or equal to x. In other words, the floor function $\lfloor\ \rfloor : \mathbb{R} \to \mathbb{Z}$ is given by

$$\lfloor x \rfloor = \max\{n \in \mathbb{Z} \mid n \leq x\}.$$

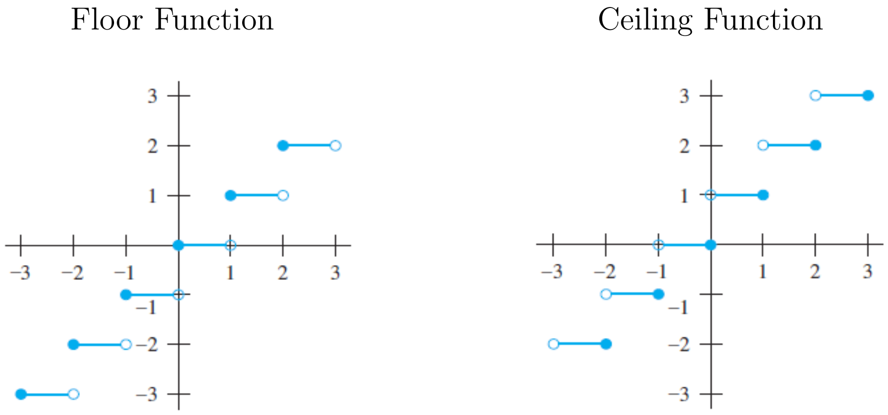
The *ceiling function* assigns to every real number x the smallest integer $\lceil x \rceil$ greater than or equal to x. In other words, the ceiling function $\lceil\ \rceil : \mathbb{R} \to \mathbb{Z}$ is given by

$$\lceil x \rceil = \min\{n \in \mathbb{Z} \mid x \leq n\}.$$

Example. We have that,

$$\lfloor 0.5 \rfloor = 0,\ \ \lfloor -0.5 \rfloor = -1,\ \ \lfloor 3.1 \rfloor = 3,\ \ \lfloor 7 \rfloor = 7,\ \ \lfloor -4.8 \rfloor = -5,$$

$$\lceil 0.5 \rceil = 1,\ \ \lceil -0.5 \rceil = 0,\ \ \lceil 3.1 \rceil = 4,\ \ \lceil 7 \rceil = 7,\ \ \lceil -4.8 \rceil = -4. \qquad \Diamond$$



Theorem 2.7 (Properties of the Floor and Ceiling functions.)**.** *Let x be a real number and n be an integer. Then the following are true.*

(i) $\lfloor x \rfloor = n$ *if and only if* $n \leq x < n + 1$

(ii) $\lceil x \rceil = n$ *if and only if* $n - 1 < x \leq n$

(iii) $\lfloor x \rfloor = n$ *if and only if* $x - 1 < n \leq x$

(iv) $\lceil x \rceil = n$ *if and only if* $x \leq n < x + 1$

(v) $x - 1 < \lfloor x \rfloor \leq x \leq \lceil x \rceil < x + 1$

(vi) $\lfloor -x \rfloor = -\lceil x \rceil$

(vii) $\lceil -x \rceil = -\lfloor x \rfloor$

(viii) $\lfloor x+n \rfloor = \lfloor x \rfloor + n$

(ix) $\lceil x+n \rceil = \lceil x \rceil + n$

Proof. We only prove Property (i) and Property(viii) and leave the proof of the remaining properties as an exercise for the reader. Towards this, choose any real number x and any integer n.

Property (i): $\lfloor x \rfloor = n$ if and only if $n \leq x < n+1$.
For the forward direction, suppose $\lfloor x \rfloor = n = \max\{k \in \mathbb{Z} \mid k \leq x\}$. So by definition of the floor function, $n \leq x$. Now, we need to show that $x < n+1$. Suppose for the sake of contradiction that $x \geq n+1$. This means that $\lfloor x \rfloor \geq n+1 > n$ which contradicts our assumption that $\lfloor x \rfloor = n$. So it must be that $x < n+1$.
For the reverse direction, assume that $n \leq x < n+1$. Then n is the greatest integer less than or equal to x. That is, $\lfloor x \rfloor = n$.

Property (viii): $\lfloor x+n \rfloor = \lfloor x \rfloor + n$.
Let $\lfloor x \rfloor = m$. Then using both directions of Property (i) we have that

$$\begin{aligned}\lfloor x \rfloor = m &\Longrightarrow m \leq x < m+1\\ &\Longrightarrow m+n \leq x+n < (m+n)+1\\ &\Longrightarrow \lfloor x+n \rfloor = m+n\\ &\Longrightarrow \lfloor x+n \rfloor = \lfloor x \rfloor + n.\end{aligned}$$

■

Remark. To prove facts about the floor and ceiling functions, it is useful to write a real number x in the form $x = n + \epsilon$ where $n \in \mathbb{Z}$ and $0 \leq \epsilon < 1$. For example, if $x = 3.67$, then $x = 3 + 0.67$ where $n = 3$ and $\epsilon = 0.67$.

Real Functions

Real functions are functions whose domain and codomain are subsets of the set of real numbers. You encountered these functions intensively in your calculus courses, such as polynomial functions, trigonometric functions, and exponential functions to name a few.

Given two functions $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$, we define two new functions as follows.

The *sum function* $(f+g):\mathbb{R}\to\mathbb{R}$ where

$$(f+g)(x)=f(x)+g(x).$$

And the *product function* $(fg):\mathbb{R}\to\mathbb{R}$ where

$$(fg)(x)=f(x)\cdot g(x).$$

Example. Let $f:\mathbb{R}\to\mathbb{R}$ where $f(x)=x^2$, and $g:\mathbb{R}\to\mathbb{R}$ where $g(x)=x-x^2$.

- $(f+g)(3)=f(3)+g(3)=3^2+(3-3^2)=9-6=3.$

- $(fg)(2)=f(2)\cdot g(2)=(2^2)(2-2^2)=(4)(-2)=-8.$

In general, for a real number x we have that

$$(f+g)(x)=f(x)+g(x)=x^2+(x-x^2)=x,$$

and

$$(fg)(x)=f(x)\cdot g(x)=x^2(x-x^2)=x^3-x^4. \qquad \Diamond$$

We may tell from the graph of a real function when sketched in the cartesian plane whether the function is injective, surjective, or both. The next theorem elaborates on the details.

Theorem 2.8. *Let $f:\mathbb{R}\to\mathbb{R}$ be a function, then the following holds.*

1. *The function f is injective if and only if every horizontal line intersects the graph of f at most once.*
2. *The function f is surjective if and only if every horizontal line intersects the graph of f at least once.*
3. *The function f is bijective if and only if every horizontal line intersects the graph of f exactly once.*

We present below some properties of real functions.

Definition. Let $f:\mathbb{R}\to\mathbb{R}$ be a function. And let the domain of the variables x and y be the domain of f which is $\mathbb{R}$.

- The function f is *increasing* if $\forall x\,\forall y\,\big(x<y\to f(x)\le f(y)\big)$.
- The function f is *strictly increasing* if $\forall x\,\forall y\,\big(x<y\to f(x)<f(y)\big)$.
- The function f is *decreasing* if $\forall x\,\forall y\,\big(x<y\to f(x)\ge f(y)\big)$.

- The function f is *strictly decreasing* if $\forall x \, \forall y \, \big(x < y \to f(x) > f(y)\big)$.

The definitions above generalise to functions $f : A \to B$ where $A, B \subseteq \mathbb{R}$. Observe that if f is strictly increasing, then f is increasing. However, the converse is false.

Exercise. Prove that if a function $f : \mathbb{R} \to \mathbb{R}$ is strictly increasing, then f is injective.

2.4 Sequences and Summations

Sequences are discrete structures used to represent finite or infinite ordered lists of objects. For instance, below is a sequence listing the positive powers of 2.

$$2,\ 4,\ 8,\ 16,\ 32,\ 64,\ 128,\ 256,\ 512,\ 1024,\ \ldots$$

Strictly speaking a sequence is just a function as explained by the next definition.

Definition.

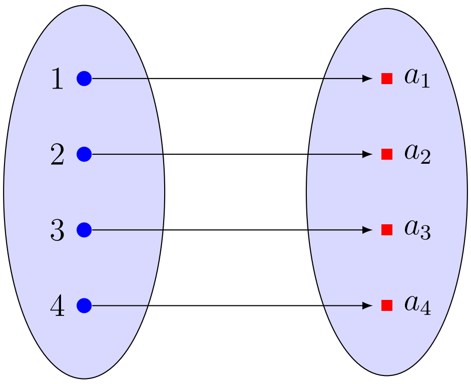
- A *finite sequence* of length n is a function f with domain $\{1, 2, \ldots, n\}$. We write $f(k) = a_k$ where k belongs to the domain.
- An *infinite sequence* is a function f with domain $\mathbb{Z}^+$. We write $f(k) = a_k$ where $k \in \mathbb{Z}^+$. We may also take the domain to be $\mathbb{N}$.

The notation a_k denotes the image of the integer k under the sequence f. So $f(1) = a_1$, $f(2) = a_2$, $f(3) = a_3$, and so on. When we just say a "sequence" we mean an infinite sequence. The idea here is that the domains of the functions mentioned the definition above consist of natural numbers, and natural numbers themselves are ordered. We use the order of the natural numbers in the domain to project an order on elements in the range of these functions. So $f(1)$ is the first element of the range, $f(2)$ is the second element, and in general $f(k)$ is the k^{th} element. As $f(k) = a_k$, it follows that a_k denotes the k^{th} element in the range. Consequently, a finite sequence of length n as defined above may be represented by listing the elements in the range in order as $f(1), f(2), f(3), \ldots, f(n)$. Thus, the range of the sequence f can be expressed as an n-tuple (ordered list) whose k^{th} element is $f(k)$ where $k \in \text{dom}(f) = \{1, 2, \ldots, n\}$ as follows,

$$(a_1, a_2, a_3, \ldots, a_n).$$

Note that this tuple lists *all* the elements in range(f). We write $(a_k)_{k=1}^{n}$ to denote the sequence $(a_1, a_2, \ldots, a_n)$.

Example. A finite sequence of length 4 is a function $f(k) = a_k$ with domain $\{1, 2, 3, 4\}$ as represented in the diagram below. Alternatively, such function can be thought of as a 4-tuple (a_1, a_2, a_3, a_4) listing the elements of its range in order.



Similarly, an infinite sequence g where $g(i) = b_i$ for $i \in \mathbb{Z}^+$ can be represented by listing all the elements in the range through an infinite ordered list whose i^{th} element is $g(i)$ where $i \in \operatorname{dom}(g) = \mathbb{Z}^+$ as follows,

$$(b_1, b_2, b_3, b_4, \ldots).$$

This list is denoted by $(b_k)_{k=1}^{\infty}$. In general, whether it is finite or infinite, we denote a sequence f where $f(k) = a_k$ by (a_k), and we call a_k the k^{th} *term* of the sequence.

Example. Consider the following infinite sequence (a_n) given by

$$(2,\ 5,\ 10,\ 17,\ 26,\ \ldots).$$

So $a_1 = 2, a_2 = 5, a_3 = 10$. In general, $a_n = n^2 + 1$ where $n \in \mathbb{Z}^+$. ♢

Next, we will introduce two special types of sequences.

Definition. Fix some real numbers a and d. An *arithmetic progression* is a sequence of the form

$$(a,\ a+d,\ a+2d,\ a+3d,\ \ldots).$$

The first term a is called the *initial term*, and d is called the *common difference.*

Notice that an arithmetic sequence is a function $f : \mathbb{N} \to \mathbb{R}$ given by $f(n) = a + nd$ where a, d are some fixed real numbers. We denote this sequence by (a_n) where the n^{th} term is $a_n = a + nd$ where $n \in \mathbb{N}$.

Example. The sequence (s_n) where $s_n = -1 + 4n$ is an arithmetic progression,

$$(-1,\ 3,\ 7,\ 11,\ 15,\ 19,\ \ldots)$$

whose initial term is -1, and common difference is 4. ♢

Example. Consider the sequence (t_n) below,

$$(7,\ 4,\ 1, -2, -5,\ \ldots).$$

This sequence is an arithmetic progression whose initial term is 7, and common difference is -3. The n^{th} term is given by $t_n = 7 - 3n$. ♢

Definition. Fix some real numbers a and r. A *geometric progression* is a sequence of the form

$$(a,\ ar,\ ar^2,\ ar^3,\ \ldots).$$

We call a the *initial term* and r the *common ratio.*

Notice that a geometric progression is a function $f : \mathbb{N} \to \mathbb{R}$ given by $f(n) = ar^n$, where a and r are some fixed real numbers. We denote this sequence by (a_n) where the n^{th} term is $a_n = ar^n$ where $n \in \mathbb{N}$.

Example. The following are geometric progressions.

- $(a_n) = (\ 1, -1,\ 1, -1,\ 1, -1, \ldots)$. Here $a = 1$, $r = -1$, and $a_n = (-1)^n$.
- $(b_n) = (2, 10, 50, 250, 1250, \ldots)$. Here $a = 2$, $r = 5$, and $b_n = 2 \cdot (5)^n$.
- $(c_n) = (6, 2, \frac{2}{3}, \frac{2}{9}, \frac{2}{27}, \ldots)$. Here $a = 6$, $r = \frac{1}{3}$, and $c_n = 6 \cdot (\frac{1}{3})^n$. ♢

Recursive Definition of a Sequence

One can describe a sequence by a *recurrence relation*, that is, the n^{th} term a_n is defined in terms of previous terms. When a recurrence relation is used to define a sequence, we say that the sequence is defined *recursively.*

Example. Let (a_n) be the sequence given by the initial condition $a_0 = 2$ and the recurrence relation $a_n = a_{n-1} + 3$ for $n \geq 1$.
Then $a_1 = a_0 + 3 = 2 + 3 = 5$ and $a_2 = a_1 + 3 = 5 + 3 = 8$. The first terms of the sequence (a_n) are:

$$(2,\ 5,\ 8,\ 11,\ 14,\ 17,\ \ldots).$$

An explicit formula for the sequence (a_n) is $a_n = 2 + 3n$ for $n \in \mathbb{N}$. Note that this sequence is an arithmetic progression. ♢

Example. Let (b_n) be given by $b_0 = 3$, $b_1 = 5$, and the recurrence relation $b_n = b_{n-1} - b_{n-2}$ for $n \geq 2$.
Then $b_2 = b_1 - b_0 = 5 - 3 = 2$ and $b_3 = b_2 - b_1 = 2 - 5 = -3$ and $b_4 = b_3 - b_2 = -3 - 2 = -5$. Here are the first terms of the sequence:

$$(b_n) = (3,\ 5,\ 2, -3, -5, -2,\ \ldots)$$ ♢

Example. The Fibonacci sequence (f_n) is given by the initial conditions $f_0 = 0, f_1 = 1$ and the recurrence relation $f_n = f_{n-1} + f_{n-2}$ for $n \geq 2$.
So the n^{th} term is obtained by summing the previous two terms. We find that $f_2 = f_1 + f_0 = 1 + 0 = 1$, and $f_3 = f_2 + f_1 = 1 + 1 = 2$, and $f_4 = f_3 + f_2 = 2 + 1 = 3$. The first terms of the Fibonacci sequence are given below.

$$(0,\ 1,\ 1,\ 2,\ 3,\ 5,\ 8,\ 13,\ 21,\ 34,\ 55,\ 89,\ 144,\ \ldots). \qquad \Diamond$$

Example. Let (a_n) be the sequence given by $a_1 = 1$ and $a_n = na_{n-1}$ for $n \geq 2$.
In other words, the n^{th} term of the sequence is obtained by multiplying n by the previous term. Here are the first terms of the sequence:

$$(a_n) = (1, 2, 6, 24, 120, 720, \ldots).$$

Note that the formula $a_n = n!$ for $n \geq 1$ determines the terms of this sequence. To prove this we need to use a proof technique called mathematical induction. $\Diamond$

Example. Find an explicit formula for the following sequences.

- $(a_n) = (1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \ldots)$. The formula is $a_n = \frac{1}{2^n}$ for $n \geq 0$.
- $(b_n) = (1, 3, 5, 7, 9, \ldots)$. The formula is $b_n = 1 + 2n$ for $n \geq 0$.
- $(c_n) = (1, -1, 1, -1, 1, -1, \ldots)$. The formula is $c_n = (-1)^n$ for $n \geq 0$. $\Diamond$

> Check *The On-Line Encyclopedia of Integer Sequences.* This is an online database of integer sequences developed by Neil Sloane in the 1960s. Neil Sloane is a British-American mathematician who was born in 1939. He studied mathematics and electrical engineering at the University of Melbourne in Australia. He received his PhD from Cornell University where his thesis was on what is now called neural networks. He took a job at AT&T Bell Labs in 1969, working in many areas: network design, coding theory, and sphere packing. His books include Sphere Packings, Lattices and Groups (with John Conway); The Theory of Error-Correcting Codes (with Jessie MacWilliams); and The Rock-Climbing Guide to New Jersey Crags (with Paul Nick).

♣ Summations

Given a finite sequence all of its terms are real numbers, we may obtain a new number by adding all the terms of the sequence. Let m and n be integers where $m \leq n$, and consider a finite sequence of real numbers, say,

$$a_m,\ a_{m+1},\ a_{m+2}, \ldots,\ a_n.$$

The sum of the sequence terms is

$$a_m + a_{m+1} + a_{m+2} + \ldots + a_n,$$

and it is denoted by

$$\sum_{k=m}^{n} a_k \quad \text{or} \quad \sum_{m \leq k \leq n} a_k.$$

where

- k is called the index of the summation,
- m is the lower limit of the summation,
- n is the upper limit of the summation,
- k runs through all integers starting with m and ending with n,
- Σ is the Greek letter sigma, and
- we read $\sum_{k=m}^{n} a_k$ as "the sum of a_k from $k = m$ to $k = n$".

Lemma 2.9. *Let (a_n) and (b_n) be sequences of real numbers and $c \in \mathbb{R}$. Then the following are true.*

(i) $\sum_{k=1}^{n} (a_k + b_k) = \sum_{k=1}^{n} a_k + \sum_{k=1}^{n} b_k.$

(ii) $\sum_{k=1}^{n} ca_k = c \sum_{k=1}^{n} a_k.$

(iii) $\sum_{k=1}^{n} c = n \cdot c.$

Example. Consider the sequence $(a_k) = (1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \ldots)$ where $a_k = \frac{1}{k}$. The sum of the first 6 terms is,

$$\sum_{k=1}^{6} a_k = \sum_{k=1}^{6} \frac{1}{k} = \frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \frac{1}{6} = \frac{49}{20}. \qquad \lozenge$$

Example. Find $\sum_{k=1}^{5} k^2$.

$$\sum_{k=1}^{5} k^2 = 1^2 + 2^2 + 3^2 + 4^2 + 5^2 = 55. \qquad \lozenge$$

We now find the summation of the first n terms in an arithmetic progression with initial term 1 and common difference 1.

Theorem 2.10. *Let n be a positive integer, then*

$$\sum_{k=1}^{n} k = 1 + 2 + 3 + \cdots + n = \frac{n(n+1)}{2}.$$

Proof. Fix some positive integer n, then we have that,

$$\begin{aligned}
2\sum_{k=1}^{n} k &= \sum_{k=1}^{n} k + \sum_{k=1}^{n} k \\
&= \big(1+2+\cdots+(n-1)+n\big) + \big(1+2+\cdots+(n-1)+n\big) \\
&= \big(1+2+\cdots+(n-1)+n\big) + \big(n+(n-1)+\cdots+2+1\big) \\
&= \big(1+n\big) + \big(2+(n-1)\big) + \big(3+(n-2)\big) + \cdots + \big((n-1)+2\big) + \big(n+1\big) \\
&= \underbrace{(n+1)+(n+1)+\cdots+(n+1)+(n+1)}_{n \text{ times}} \\
&= n(n+1).
\end{aligned}$$

Therefore, we have that $\sum_{k=1}^{n} k = \frac{n(n+1)}{2}$ as desired. ■

We next find a formula for the summation of terms in a geometric progression.

Theorem 2.11. *Let n be a positive integer and $a, r \in \mathbb{R}$ with $r \neq 0$, then*

$$\sum_{k=0}^{n} ar^k = \begin{cases} \dfrac{a(r^{n+1}-1)}{r-1} & \textit{if } r \neq 1; \\ (n+1)a & \textit{if } r = 1. \end{cases}$$

Proof. When $r = 1$, we have that

$$\sum_{k=0}^{n} ar^k = \sum_{k=0}^{n} a = \underbrace{a + a + \cdots + a}_{n+1 \text{ times}} = (n+1)a.$$

So suppose $r \neq 1$, and let us start by multiplying the summation by $r - 1$.

$$\begin{aligned}
(r-1)\sum_{k=0}^{n} ar^k &= r\sum_{k=0}^{n} ar^k - \sum_{k=0}^{n} ar^k \\
&= r(a + ar + ar^2 + \cdots + ar^n) - (a + ar + ar^2 + \cdots + ar^n) \\
&= ar + ar^2 + \cdots + ar^n + ar^{n+1} - a - ar - ar^2 - \cdots - ar^n \\
&= ar^{n+1} - a \\
&= a(r^{n+1} - 1).
\end{aligned}$$

Since $r - 1 \neq 0$, it follows from these equalities that $\sum_{k=0}^{n} ar^k = \frac{a(r^{n+1}-1)}{r-1}$ as required. ■

Lemma 2.12 (Summation Formulas)**.** *Let n be a positive integer. Then the following are true.*

(i) $\sum\limits_{k=1}^{n} k = \dfrac{n(n+1)}{2}$.

(ii) $\sum\limits_{k=1}^{n} k^2 = \dfrac{n(n+1)(2n+1)}{6}$.

(iii) $\sum\limits_{k=1}^{n} k^3 = \dfrac{n^2(n+1)^2}{4}$.

(iv) $\sum\limits_{k=0}^{n} ar^k = \dfrac{a(r^{n+1}-1)}{r-1}$ *where $a, r \in \mathbb{R}$ with $r \neq 1$.*

Example. Find $\sum\limits_{i=1}^{4}\sum\limits_{j=1}^{3} ij$.
We start by evaluating the inner summation.

$$\sum_{i=1}^{4}\sum_{j=1}^{3} ij = \sum_{i=1}^{4}(i+2i+3i) = \sum_{i=1}^{4} 6i = 6\sum_{i=1}^{4} i = 6\Big(\frac{4(5)}{2}\Big) = 60. \qquad \Diamond$$

Example. Find $\sum\limits_{k=50}^{100} k^2$.
Since $\sum\limits_{k=1}^{100} k^2 = \sum\limits_{k=1}^{49} k^2 + \sum\limits_{k=50}^{100} k^2$ we have that

$$\begin{aligned}\sum_{k=50}^{100} k^2 &= \sum_{k=1}^{100} k^2 - \sum_{k=1}^{49} k^2 \\ &= \frac{100(101)(201)}{6} - \frac{(49)(50)(99)}{6} \\ &= 338\,350 - 40\,425 = 297\,925.\end{aligned} \qquad \Diamond$$

Example. Express the following sum in sigma notation and evaluate it.

$$-100 + (-97) + (-94) + (-91) + \cdots + 101$$

Observe that these terms come from an arithmetic progression (a_k) with initial term -100 and common difference 3. The general term is $a_k = -100 + 3k$ for $k \geq 0$. It remains to find the upper limit n which occurs when $a_n = 101$. Thus,

$$a_n = 101 \iff -100 + 3n = 101 \iff 3n = 201 \iff n = 67.$$

Therefore, the value of the sum is,

$$\begin{aligned}
-100 + (-97) + (-94) + \cdots + 101 &= \sum_{n=0}^{67}(-100 + 3n) \\
&= -\sum_{n=0}^{67} 100 + 3\sum_{n=0}^{67} n \\
&= -(68 \times 100) + 3\Big(\frac{67 \times 68}{2}\Big) \\
&= -6800 + 6834 = 34. \qquad \Diamond
\end{aligned}$$

2.5 Cardinality of Infinite Sets

Recall that the cardinality of a *finite* set is its size or the number of elements in the set. In this section we extend the notion of cardinality to infinite sets.

Observe that when A and B are finite sets, then the following statements are equivalent (always have the same truth value).

(i) The number of elements in A is equal to the number of elements in B.

(ii) There exists a bijection from A to B.

However, when A and B are infinite, the first point above is not very useful because it is unclear what is meant by the "number of elements in an infinite set". On the other hand, the second point works perfectly in comparing sizes of infinite sets. Consequently, we adopt the second point and use bijections as a device to measure sizes of sets. The idea is clarified in the next definition which applies to both finite and infinite sets.

Definition. We say that sets A and B have *equal cardinality* if and only if there exists a bijection from A to B. We write $|A| = |B|$ for this.

Clearly, the statement $|A| \neq |B|$ means that there exists no bijection from A to B.

Those sets which have equal cardinality to that of the set of the natural numbers are given a special name.

Definition. A set S is *countably infinite* if there exists a bijection from $\mathbb{N}$ to S.

In short, a set S is countably infinite if $|\mathbb{N}| = |S|$. Another way to express that S is countably infinite is by writing $|S| = \aleph_0$. You may think of $\aleph_0$ as the first number which comes after all natural numbers!

Clearly, the set of natural numbers is countably infinite since the identity function $\iota : \mathbb{N} \to \mathbb{N}$ is a bijection.

How can we think of countably infinite sets? Let S be a countably infinite set. Consequently, there is at least one bijection $f : \mathbb{N} \to S$. So the set S can be put in one-to-one correspondence with the natural numbers as demonstrated below.

$$\begin{tikzcd}
0 \arrow[d] & 1 \arrow[d] & 2 \arrow[d] & 3 \arrow[d] & 4 \arrow[d] & \cdots \\
f(0) & f(1) & f(2) & f(3) & f(4) & \cdots
\end{tikzcd}$$

Since f is surjective, the range is equal to the codomain S, and it follows that S is a "mirror image" of $\mathbb{N}$. To elaborate more, consider the range of f where

$$\text{range}(f) = \{f(n) \mid n \in \mathbb{N}\} = \{f(0), f(1), f(2), f(3), f(4), \ldots\}.$$

If we let $f(n) = s_n$ for every n in the domain $\mathbb{N}$, then we express the range of f by the sequence whose n^{th} term is s_n given below

$$(s_0, s_1, s_2, s_3, s_4, \ldots).$$

Every element in this sequence belongs to S because S is the codomain of f. What else can we say about this sequence? First, since f is injective, all the terms in the sequence are pairwise distinct, so there are no repetitions. It follows that S must be an infinite set. Second, since f is surjective, every element in the codomain S must appear somewhere in this sequence. Therefore, the sequence of images of the function f lists all the elements of S, that is,

$$S = \{s_0, s_1, s_2, s_3, \ldots\}.$$

The moral of the story is that S being countably infinite means that S is infinite and it is possible to list *all* its elements in a sequence indexed by the natural numbers. Although, the process of listing the elements of S will take forever, any particular element of S will appear in the list after a finite amount of time. Just imagine a machine which spits out all the elements of S one after the other! The summary of our discussion can be stated as follows.

Lemma 2.13. *Let S be a set. Then the following are equivalent.*

(i) S is countably infinite.

(ii) S can be enumerated in an infinite sequence whose terms are pairwise distinct.

David Hilbert (1862–1943) was a German mathematician born in the city of Königsberg which is famous in mathematics for its seven bridges. He worked in several areas including the calculus of variations, geometry, algebra, number theory, logic, and mathematical physics. Many concepts are named after him, for example, Hilbert spaces. Hilbert is remembered for his important and influential list of 23 unsolved problems which he presented at the 1900 International Congress of Mathematicians to keep mathematicians busy in the twentieth century. One of these problems, the *Riemann Hypothesis*, is still unsolved till now.

In the 1920's Hilbert devised a thought experiment about a grand hotel — a hotel with countably infinite number of rooms, each occupied by a guest. One day, a new guest arrives into the hotel and asks for a room. The smart hotel manager assigned a room for the new guest without asking any of the current guests to leave the hotel! How?

Let us see more examples of countably infinite sets. Remember that one way to show that an infinite set S is countably infinite is by constructing a bijection from $\mathbb{N}$ to S. Another way is to describe a way to list all the elements in S. Note that these two ways are two faces of the same coin.

Example. Show that the set D of positive odd integers is a countably infinite.
We need to find some bijection $f : \mathbb{N} \to D$. We propose the function $f(n) = 2n+1$. We will show that f is indeed bijective.

$$\begin{array}{ccccccc}
0 & 1 & 2 & 3 & 4 & 5 & \cdots \\
\downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \\
1 & 3 & 5 & 7 & 9 & 11 & \cdots
\end{array}$$

First, we will show that f is injective. Choose any $m, n \in \mathbb{N}$ and suppose that $f(n) = f(m)$. So we have that $2n+1 = 2m+1$, and so $2n = 2m$. Thus, it must be that $n = m$, showing that f is injective.

Second, we show that f is surjective. Choose any m in the codomain D. Since m is a positive odd integer, it follows that $n = (m-1)/2$ belongs to the domain $\mathbb{N}$. We claim that n is the preimage of m. To see this,

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

Therefore, f is surjective. This shows that f is bijective, and so D is countably infinite.

Alternatively, to show that D is countably infinite we can can list all the elements of D as in the sequence

$$1, 3, 5, 7, 9, 11 \ldots. \qquad \Diamond$$

Example. Show that the set $\mathbb{Z}^+$ of positive integers is countably infinite.

One possible bijection is $f : \mathbb{N} \to \mathbb{Z}^+$ given by $f(n) = n + 1$. Alternatively, we can list all the elements of $\mathbb{Z}^+$ in a sequence as

$$1, 2, 3, 4, 5, 6, \ldots. \qquad \Diamond$$

Lemma 2.14. *The set of all integers is countably infinite. That is,* $|\mathbb{N}| = |\mathbb{Z}|$.

Proof. A quick way to show that $\mathbb{Z}$ is countably infinite is to list all the integers, one after the other, in a sequence indexed by the natural numbers. One such sequence is given below.

$$0,\ 1, -1,\ 2, -2,\ 3, -3,\ 4, -4,\ 5, -5, \ldots.$$

Alternatively, using this sequence we may present a bijection $f : \mathbb{N} \to \mathbb{Z}$ as follows

$$\begin{array}{ccccccccc} 0 & 1 & 2 & 3 & 4 & 5 & 6 & \cdots \\ \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \\ 0 & 1 & -1 & 2 & -2 & 3 & -3 & \cdots \end{array}$$

This function may also be given using the following formula,

$$f(n) = \begin{cases} \frac{-n}{2} & \text{if } n \text{ is even;} \\ \frac{n+1}{2} & \text{if } n \text{ is odd.} \end{cases}$$

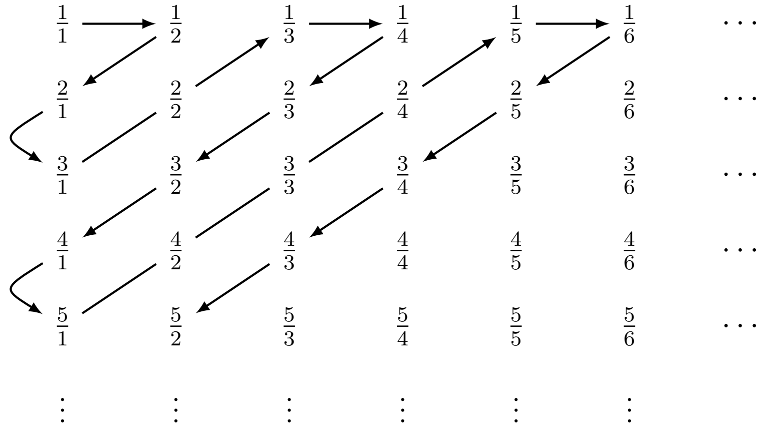
One needs to check that f is bijective. ■

We next present a quite surprising fact. We will show that $|\mathbb{N}| = |\mathbb{Q}|$.

Theorem 2.15. *The set of rational numbers $\mathbb{Q}$ is countably infinite.*

Proof. We will first list all the positive rational numbers in a sequence indexed by the natural numbers, and then using this sequence we will be able to list all rational numbers. We start by arranging all positive rational numbers in an infinite array where the first row contains all positive rational numbers with numerator equal to 1, and the second row contains all positive rational numbers with numerator equal to

2, and so on. Clearly, this array contains all positive rational numbers, for instance, the rational number $\frac{m}{n}$ lies in the m^{th} row and n^{th} column. After that we start our sequence with $\frac{1}{1}$ and then move in a snake-like motion covering all numbers in the array while skipping numbers already listed, forming our desired sequence as shown below.



The sequence produced is

$$\frac{1}{1}, \frac{1}{2}, \frac{2}{1}, \frac{3}{1}, \frac{1}{3}, \frac{1}{4}, \frac{2}{3}, \frac{3}{2}, \frac{4}{1}, \frac{5}{1}, \frac{1}{5}, \frac{1}{6}, \frac{2}{5}, \frac{3}{4}, \frac{4}{3}, \frac{5}{2}, \frac{6}{1}, \cdots$$

We can describe this sequence as follows. First we list positive rational numbers where the sum of their numerator and denominator is 2, followed by those where the sum of their numerator and denominator is 3, followed by those whose numerator and denominator add up to 4, and so on. Of course, we avoid repetitions.

Now it is easy to list all rational numbers in a sequence by alternating between positive and negative rational numbers.

$$0, \frac{1}{1}, \frac{-1}{1}, \frac{1}{2}, \frac{-1}{2}, \frac{2}{1}, \frac{-2}{1}, \frac{3}{1}, \frac{-3}{1}, \frac{1}{3}, \frac{-1}{3}, \frac{1}{4}, \frac{-1}{4}, \frac{2}{3}, \frac{-2}{3}, \frac{3}{2}, \frac{-3}{2}, \frac{4}{1}, \frac{-4}{1}, \cdots$$

This proves that the set of rational numbers is countably infinite. ■

As we proved that both the set of integers and the set of rationals are both countably infinite we can state the following corollary.

Corollary 2.16.

$$|\mathbb{N}| = |\mathbb{Z}| = |\mathbb{Q}|.$$

We are able to list all the elements of a finite set or countably infinite set, although in the latter case the process will take forever. These sets which we can list all their elements in a finite or an infinite sequence are considered of small size. We now divide sets into two categories: "small" sets and "large" sets.

Definition.

- A set is called *countable* if it is either finite or countably infinite.
- A set is called *uncountable* if it is not countable.

A countable set is a set which we can enumerate all its elements in a finite sequence or an infinite sequence indexed by the natural numbers. For example, the empty set, the set $\{0, 1, 2, 3, 4, 5, 6\}$, the set of all cities, the set of positive even integers, the set of natural numbers, the set of integers, and the set of rational numbers are all countable sets. Moreover, whenever we have two countable sets and we form their union we obtain another countable set as we show below.

Theorem 2.17. *Suppose that A and B are countable sets. Then their union $A \cup B$ is also countable.*

Proof. For simplicity, we may assume that A and B are disjoint sets. If both A and B are finite, then their union is also finite and so countable. If A is finite, say $A = \{a_1, a_2, \ldots, a_n\}$ for some positive integer n, and B is countably infinite, say $B = \{b_0, b_1, b_2, \ldots\}$, then we can list all the elements of $A \cup B$ one after the other by first listing the finitely many elements of A followed by the list of the elements of B as shown in the sequence below,

$$a_1, a_2, \ldots, a_n, b_0, b_1, b_2, b_3, \ldots$$

And thus $A \cup B$ is countably infinite in this case, and so countable.

Finally, we are left with the case when both A and B are countably infinite. Let A be enumerated by the sequence $a_0, a_1, a_2, \ldots$ and B be enumerated by the sequence $b_0, b_1, b_2, \ldots$. Then we can enumerate their union $A \cup B$ as

$$a_0, b_0, a_1, b_1, a_2, b_2, a_3, b_3, \ldots$$

This proves that the union of two countably infinite sets is countably infinite. ■

Read the definition of uncountability again. A set is uncountable if it is not countable, meaning that it is neither finite nor countably infinite. Thus, an uncountable set is infinite and not countably infinite. This means that an uncountable set is infinite and it is impossible to have any bijection from $\mathbb{N}$ to the set. In other words, an uncountable set is an infinite set which we cannot list its elements in a sequence indexed by the natural numbers. But do uncountable sets really exist? The German mathematician Georg Cantor in 1879 introduced his “Diagonalisation Argument” to prove that the set $\mathbb{R}$ of real numbers is uncountable!

Georg Cantor (1845–1918) was a German mathematician. He was born in Saint Petersburg, Russia, where his father was a successful merchant. In 1863, he studied at the University of Berlin under the eminent mathematicians Weierstrass, Kummer, and Kronecker. He received his doctorate degree in 1867, after submitting a dissertation on number theory. Cantor held a position at the University of Halle in 1869, where he spent his entire career. Cantor married Vally Guttmann in 1874 and had six children.

Cantor is considered the founder of set theory. His contributions include the discovery that the set of real numbers is uncountable. He has important contributions to real analysis as well. Cantor was interested in philosophy and discussed the relation between his theory of sets with metaphysics. He was poorly paid as a professor, and he tried to obtain a better position at the University of Berlin, however, his appointment there was hindered by Kronecker, who refused Cantor’s views on set theory. Cantor suffered from depression and severe mental illness throughout his last years. He died in 1918 from a heart attack.

Can you list *all* real numbers between 0 and 1 as we list the natural numbers? We shall see that this is impossible! Recall that $(0,1)=\{x\in\mathbb{R}\mid 0<x<1\}$. Next, we will show that the interval $(0,1)$ is uncountable.

Theorem 2.18. *The set* $(0,1)$ *of real numbers between 0 and 1 is uncountable.*

Proof. For the sake of contradiction assume that the set $(0,1)$ is countable. So it is countably infinite and thus we can list all real numbers in $(0,1)$ in an infinite sequence as follows

$$r_1, r_2, r_3, r_4, \ldots$$

So we are assuming that any real number between 0 and 1 appears somewhere in the list above. Now we express each number in this sequence using its decimal representation. Let us say that $r_i = 0.d_{i1}d_{i2}d_{i3}\ldots$ where d_{ij} is the j^{th} decimal digit of r_i and $d_{ij}\in\{0,1,2,3,4,5,6,7,8,9\}$.

$$\begin{aligned}
r_1 &= 0\,.\,\underline{d_{11}}\,d_{12}\,d_{13}\,d_{14}\,d_{15}\,d_{16}\;\ldots\\
r_2 &= 0\,.\,d_{21}\,\underline{d_{22}}\,d_{23}\,d_{24}\,d_{25}\,d_{26}\;\ldots\\
r_3 &= 0\,.\,d_{31}\,d_{32}\,\underline{d_{33}}\,d_{34}\,d_{35}\,d_{36}\;\ldots\\
r_4 &= 0\,.\,d_{41}\,d_{42}\,d_{43}\,\underline{d_{44}}\,d_{45}\,d_{46}\;\ldots\\
r_5 &= 0\,.\,d_{51}\,d_{52}\,d_{53}\,d_{54}\underline{d_{55}}\,d_{56}\;\ldots\\
&\;\vdots
\end{aligned}$$

We now form a new real number s with decimal expansion $s = 0.s_1s_2s_3\ldots$ where its decimal digits are determined by the following rule which depends on the diagonal digits in the configuration above.

$$s_i = \begin{cases} 1 & \text{if } d_{ii} \neq 1 \\ 2 & \text{if } d_{ii} = 1 \end{cases}$$

This new number s is not in our sequence $r_1, r_2, r_3, \ldots$. To see this, observe that for every i, we have that $s \neq r_i$ because in the way we defined s we ensured that $s_i \neq d_{ii}$ and so the decimal expansion of s differs from that of r_i in the i^{th} place to the right of the decimal point. Moreover, the decimal digits of s are 1's and 2's (for example $s = 0.11211222212\ldots$) and so s belongs to the set $(0, 1)$. Thus s must be in the sequence $r_1, r_2, r_3, \ldots$ since by our assumption in the beginning all such numbers appear there, nevertheless, s cannot be in the sequence. We have a contradiction! Therefore, it is impossible for $(0, 1)$ to be countable. This proves that the set $(0, 1)$ is uncountable. ■

Example. The previous argument in an example. Suppose a list of real numbers between 0 and 1 starts as follows.

$$\begin{aligned}
r_1 &= 0.037142556987\ldots\\
r_2 &= 0.416527901478\ldots\\
r_3 &= 0.962561478521\ldots\\
r_4 &= 0.120869853333\ldots\\
r_5 &= 0.457315708365\ldots\\
&\;\vdots
\end{aligned}$$

Let $r_i = 0.d_{i1}d_{i2}d_{i3}\ldots$. We have $d_{11} = 0$, $d_{12} = 3$, $d_{13} = 7$, $d_{14} = 1$. Moreover, $d_{22} = 1$, $d_{33} = 2$, and $d_{44} = 8$ and $d_{55} = 1$. The new number is $s = 0.s_1s_2s_3s_4\ldots$ where $s_1 = 1$ because $d_{11} \neq 1$, and $s_2 = 2$ because $d_{22} = 1$, and $s_3 = 1$ because $d_{33} \neq 1$, and $s_4 = 1$ because $d_{44} \neq 1$, and $s_5 = 2$ because $d_{33} = 1$. Thus

$$s = 0.12112\ldots.$$

Note that $s \neq r_i$ for each i. $\Diamond$

Lemma 2.19. *Let A and B be sets with $A \subseteq B$. If B is countable, then A is countable as well.*

Proof. Suppose B is countable and $A \subseteq B$. If B is finite, then so is A, and so A is countable. If B is countably infinite, then B can be enumerated by a sequence indexed by the natural numbers. From this sequence remove all the elements which are not in A obtaining a (finite or infinite) sequence listing all the elements of A. Thus, A is countable. ■

Corollary 2.20. *The set of real numbers $\mathbb{R}$ is uncountable.*

Proof. The contrapositive of Lemma 2.19 states that any set which contains an uncountable subset is also uncountable. By Theorem 2.18 we know that the interval $(0, 1)$ is uncountable. Since $(0, 1) \subseteq \mathbb{R}$, it follows that $\mathbb{R}$ is uncountable. ■

Next we introduce another relations comparing cardinalities of sets.

Definition. Let A and B be sets.

- We write $|A| \leq |B|$ when there is an injective function from A to B.
- We write $|A| < |B|$ when there is an injective function from A to B, but there is no bijection from A to B.

Remark. The second point of the definition above can be expressed by saying that $|A| < |B|$ if and only if $|A| \leq |B|$ and $|A| \neq |B|$.

Recall that $|A| = |B|$ means that there is a bijection from A to B. Clearly, if $|A| = |B|$, then $|A| \leq |B|$ because a bijective function is also injective. Notice that we can express the uncountability of a set as stated in the lemma below. The proof is left as an exercise.

Lemma 2.21. *A set S is uncountable if and only if $|\mathbb{N}| < |S|$.*

Corollary 2.22. *We have that $|\mathbb{N}| < |\mathbb{R}|$.*

Example. The following statements are true.

- $|\{1, 2, 3\}| \leq |\{a, b, c, d, e\}|$
- $|\{1, 2, 3\}| < |\{a, b, c, d, e\}|$
- $|\{1, 2, 3\}| \leq |\mathbb{N}|$

- $|\mathbb{Z}| \leq |\mathbb{N}|$
- $|\mathbb{Q}| \leq |\mathbb{N}|$
- $|\mathbb{N}| = |\mathbb{Z}| = |\mathbb{Q}|$
- $|\mathbb{N}| \leq |\mathbb{R}|$
- $|\mathbb{N}| < |\mathbb{R}|$ ◊

In 1891, Cantor proved one of his famous theorems which states that the cardinality of any set is strictly less than the cardinality of its power set. While this result is clear for finite sets, the fact it holds for infinite ones is challenging to prove.

Theorem 2.23 (Cantor's Theorem)**.** *Let S be any set. Then*

$$|S| < |\mathcal{P}(S)|.$$

We now state another important theorem from year 1897.

Theorem 2.24 (Schröder–Bernstein Theorem)**.** *Let A and B be sets. If $|A| \leq |B|$ and $|B| \leq |A|$, then $|A| = |B|$.*

The theorem says that if there is an injective function from A to B and an injective function from B to A, then there exists a bijective function from A to B. We note that the proof is tricky and not easy. Let us see an application of this theorem.

Example. Show that $|(0,1)| = |(0,1]|$.
It is difficult to give a bijection $h : (0,1) \to (0,1]$. Alternatively, we will use Schröder–Bernstein Theorem. Consider the functions $f : (0,1) \to (0,1]$ given by $f(x) = x$, and $g : (0,1] \to (0,1)$ given by $g(x) = \frac{x}{2}$. Since both f and g are injective, there must be a bijection from $(0,1)$ to $(0,1]$. Therefore, $|(0,1)| = |(0,1]|$. ◊

Continuum Hypothesis. In 1877 Georg Cantor stated his famous continuum hypothesis which asserts that there is no subset $A \subseteq \mathbb{R}$ such that $|\mathbb{N}| < |A| < |\mathbb{R}|$. In other words, there is no set whose cardinality is strictly between the cardinality of the natural numbers and that of the real numbers. Alternatively, there is no uncountable set whose cardinality is strictly less than that of the set of real numbers. Cantor worked very hard to prove the continuum hypothesis, but unfortunately he was unsuccessful. In 1900, the continuum hypothesis was declared by David Hilbert to be the first of the most important unsolved 23 problems in mathematics. Later, the problem was settled down by the work of Kurt Gödel in 1940 and Paul Cohen in 1963, for which Cohen was awarded a Fields Medal.

2.6 Russell's Paradox

In naive set theory, we defined any collection of objects to be a set. In 1901, the British mathematician and philosopher Bertrand Russell discovered a problem with this definition.

Let us discuss Russell's Paradox. Suppose that any collection of objects is a set (as we do in naive set theory). Consider the collection $\mathcal{R}$ of all sets that are not members of themselves.

$$\mathcal{R} = \{S \text{ is a set} \mid S \notin S\}.$$

For instance, $\emptyset \in \mathcal{R}$ and $\mathbb{N} \in \mathcal{R}$. As $\mathcal{R}$ is a set, we may ask ourselves: Does $\mathcal{R}$ belong to itself? Suppose that $\mathcal{R} \in \mathcal{R}$. Then by definition of $\mathcal{R}$, the set $\mathcal{R}$ must not contain itself. So $\mathcal{R} \notin \mathcal{R}$. We deduce that $\mathcal{R} \in \mathcal{R} \to \mathcal{R} \notin \mathcal{R}$ is true. Conversely, suppose that $\mathcal{R} \notin \mathcal{R}$. Again, by definition of $\mathcal{R}$, the set $\mathcal{R}$ must be a member of itself, so $\mathcal{R} \in \mathcal{R}$. Thus, we showed that $\mathcal{R} \notin \mathcal{R} \to \mathcal{R} \in \mathcal{R}$ is true. Therefore, we established that

$$\mathcal{R} \in \mathcal{R} \leftrightarrow \mathcal{R} \notin \mathcal{R},$$

which is a contradiction. (Check that a compound proposition of the form $p \leftrightarrow \neg p$ is always false; it is a contradiction.) It follows that we cannot regard the collection $\mathcal{R}$ as a set. The moral of the story is that not all collections of objects can be regarded as sets.

To avoid such paradoxes of naive set theory we build set theory starting with some list of axioms describing sets, their properties, and their behaviour. This approach is called *Axiomatic Set Theory*, and it uses Zermelo-Franenkel (ZFC) axioms. Here are some of these axioms.

- **Axiom of Empty Set**
  An empty set exists.
- **Axiom of Extensionality**
  Two sets are equal if and only if they contain the same elements.
- **Axiom of Pairing**
  For any two sets, there is a set whose members are exactly these two sets.
- **Axiom of Union**
  For any set x there is a set y which contains precisely the elements of the elements of x.
- **Axiom of Power Set**
  The power set of any set exists.
- **Axiom of Infinity**
  There exists an infinite set.

- **Axiom of Choice**
  Given any collection of nonempty sets, there is a set which contains one element from each set in the collection.

Bertrand Russell (1872 – 1970) was a British philosopher, mathematician, logician, historian, writer, and political activist. In 1910 Trinity College of the University of Cambridge appointed him to a lectureship in logic and the philosophy of mathematics. Russell held strong pacifist views (opposition to war and violence), and his protests against World War I led to dismissal from Trinity College. He was imprisoned for 6 months in 1918 because of an article he wrote that was considered seditious. Russell fought for women's suffrage (right to vote) in Great Britain. In 1961, at the age of 89, he was imprisoned for the second time for his protests advocating nuclear disarmament. Russell's greatest work was in his development of principles that could be used as a foundation for all of mathematics. His most famous work is Principia Mathematica, written with Alfred North Whitehead, which attempts to deduce all of mathematics using a set of primitive axioms. He wrote many books on philosophy, physics, and his political ideas. Russell won the Nobel Prize for Literature in 1950.

♣ Barber Paradox

The *Barber Paradox* is derived from Russell's Paradox. There is a town with the rule that every man must be clean-shaven. Moreover, in this town there is one barber who shaves all those and only those men who do not shave themselves.

Question. Who shaves the barber?

If the barber does not shave himself, then he should shave himself! And if he does shave himself, then he should not!

Chapter 3

Mathematical Induction

3.1 Mathematical Induction

In this section we will work with a predicate $P(n)$ where the domain of the variable n is the set of natural numbers $\mathbb{N}$. Recall that such predicate $P(n)$ is a statement stating a property about a natural number n. And the universal statement $\forall n P(n)$ states that for all $n \in \mathbb{N}$, the statement $P(n)$ is true. We have seen before that one approach to prove that $\forall n P(n)$ is true is to show that $P(n)$ is true for any arbitrary natural number n, and then apply Universal Generalisation. Our aim in this chapter is to introduce another proof technique called *mathematical induction* used to prove statements of the form $\forall n P(n)$ where n is a natural number. More generally, the domain of n could be a subset of the integers of the form $\{m \in \mathbb{Z} \mid m \geq k\}$ for some fixed integer k. The principle of mathematical induction states that if the number 0 has property P and whenever a natural number has property P, its successor has property P too, then we can conclude that every single natural number has property P. More precisely, we have the following.

The Principle of Mathematical Induction. To show that $\forall n P(n)$ is true we complete two steps.

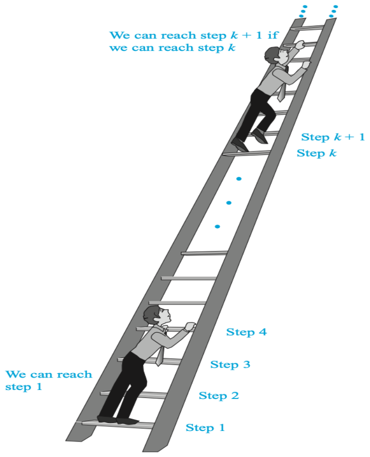
1. Base Case: Show that $P(0)$ is true.

2. Induction Step: Show that $\forall n \in \mathbb{N} \big(P(n) \to P(n+1)\big)$ is true.

In the induction step, we choose any arbitrary natural number n, and we assume that $P(n)$ is true. Then we need to show that $P(n+1)$ must be true. The assumption $P(n)$ is called the *Induction Hypothesis (IH)*. In the induction step we usually point out where we use the induction hypothesis.

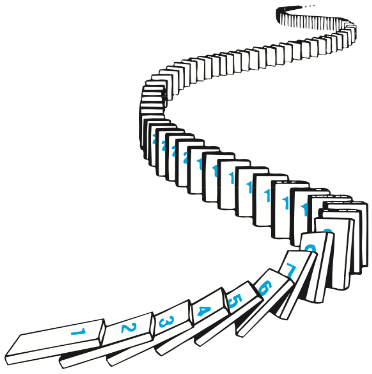
Example. Think of an infinite ladder. Mathematical induction tells us that you can reach every step of the ladder if you can do the following.

1. You can reach the first step.
2. Whenever you can reach any step of the ladder, you can reach the next step.



Example. Think of an infinite queue of dominoes where each domino is standing up. Mathematical induction tells us that you can knock over every domino if you can do the following.

1. You can knock over the first domino.
2. Whenever a domino is knocked over, it also knocks over the next domino.



We will provide a proof of the principle of mathematical induction at the end of this section. Observe that it can be stated as follows.

Let $P(n)$ be a predicate where the variable domain is $\mathbb{N}$. Then the following holds.

$$\big(P(0) \wedge \forall n\big(P(n) \to P(n+1)\big)\big) \implies \forall n P(n).$$

In the remaining part of this section we apply the principle of mathematical induction to prove a variety of theorems, the first of which gives a formula for the sum of powers of 2.

Theorem 3.1. *Let n be a natural number, then*

$$\sum_{k=0}^{n} 2^k = 2^{n+1} - 1.$$

Proof. We will prove the theorem by mathematical induction. Let $P(n)$ be the statement $\sum_{k=0}^{n} 2^k = 2^{n+1} - 1$. So $P(n)$ says that "$1 + 2 + 2^2 + \ldots + 2^n = 2^{(n+1)} - 1$".
Base Case: We show the statement $P(n)$ is true for $n = 0$. Clearly $P(0)$ is true since the left hand side is $2^0 = 1$ which is equal to the right hand side which is $2^{0+1} - 1 = 2 - 1 = 1$.
Induction Step: Choose an arbitrary $n \in \mathbb{N}$, and assume that $P(n)$ is true. We now need to show that $P(n + 1)$ is true, that is, to show that

$$\sum_{k=0}^{n+1} 2^k = 2^{n+2} - 1.$$

We proceed as follows.

$$\begin{aligned}\sum_{k=0}^{n+1} 2^k &= 1 + 2 + 2^2 + \ldots + 2^n + 2^{n+1} \\ &= \sum_{k=0}^{n} 2^k + 2^{n+1} \overset{IH}{=} 2^{n+1} - 1 + 2^{n+1} \\ &= 2 \cdot 2^{n+1} - 1 = 2^{n+2} - 1.\end{aligned}$$

Thus $P(n + 1)$ is true and the induction step is complete. Therefore, we have shown by mathematical induction that for every natural number n we have that $1 + 2 + 2^2 + \ldots + 2^n = 2^{(n+1)} - 1$. ■

The next theorem has been already proved in Section 2.4, it involves the sum of an initial segment of positive integers.

Theorem 3.2. *Let n be a positive integer, then the following holds.*

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

Proof. We will proceed by mathematical induction. Let $P(n)$ be the statement

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

Note that $P(n)$ says that "$1 + 2 + 3 + \ldots + n = n(n+1)/2$", in words, $P(n)$ says that the sum of the first n positive integers is equal to $n(n+1)/2$.
Base Case: We show that the statement $P(n)$ holds for $n = 1$. This is clear since $1 = \frac{1(1+1)}{2}$. Thus $P(1)$ is true.
Induction Step: We show that $\forall n(P(n) \to P(n+1))$ is true. So let n be an arbitrary positive integer. Assume that $P(n)$ is true, we must show that $P(n+1)$ is also true. Thus, we need to show that $\sum\limits_{k=1}^{n+1} k = (n+1)(n+2)/2$.

$$\begin{aligned}\sum_{k=1}^{n+1} k &= 1 + 2 + \ldots + n + (n+1) = \sum_{k=1}^{n} k + (n+1)\\ &\overset{IH}{=} \frac{n(n+1)}{2} + (n+1) = \frac{n(n+1) + 2(n+1)}{2}\\ &= \frac{(n+1)(n+2)}{2}.\end{aligned}$$

Therefore $P(n+1)$ is true completing the induction step. Thus by mathematical induction we proved $\forall n P(n)$, that is, for all $n \in \mathbb{N}$, we have that $\sum\limits_{k=1}^{n} k = \frac{n(n+1)}{2}$. ■

Theorem 3.3. *For any positive integer n, we have that*

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

Proof. Let $P(n)$ be the statement $\sum\limits_{k=1}^{n}(2k-1) = n^2$. Note that $P(n)$ says that "$1+3+5+7+\ldots+(2n-1) = n^2$", in other words, it says that the sum of the first n positive odd integers is equal to n^2. We will prove $\forall n P(n)$ is true by induction.
Base Case: We show the statement $P(n)$ is true for $n = 1$. We have that $P(1)$ is true since $1 = 1^2$.
Induction Step: We need to prove that $P(n) \to P(n+1)$ for every integer $n \geq 1$. So let n be an arbitrary positive integer. Suppose that $P(n)$ is true. We need to show that $P(n+1)$ is true as well. That is, we will show that

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

Starting with the left hand side, we have that

$$\begin{aligned}\sum_{k=1}^{n+1}(2k-1) &= 1 + 3 + 5 + \ldots + (2n-1) + (2(n+1)-1)\\ &= \sum_{k=1}^{n}(2k-1) + (2n+1) \overset{IH}{=} n^2 + 2n + 1\\ &= (n+1)^2.\end{aligned}$$

Therefore, $P(n+1)$ is true and the induction step is complete. Thus by mathematical induction we proved that $P(n)$ is true for all positive integers. ■

The next theorem has been encountered in Section 2.4, it gives a formula for the sum of the terms in a geometric progression.

Theorem 3.4. *Let $a, r \in \mathbb{R}$ with $r \neq 1$. Then for all $n \in \mathbb{N}$, we have that*

$$\sum_{k=0}^{n} ar^k = \frac{a(r^{n+1}-1)}{r-1}.$$

Proof. We will prove the theorem by induction.
Base Case: We show that the statement is true for $n = 0$. This holds since left hand side is $\sum\limits_{k=0}^{0} ar^k = ar^0 = a$ which is equal to the right hand side as $\frac{a(r-1)}{(r-1)} = a$.
Induction Step: Choose any arbitrary $n \in \mathbb{N}$ and assume that the statement is true for this n, that is, assume that $\sum\limits_{k=0}^{n} ar^k = \frac{a(r^{n+1}-1)}{r-1}$. We need to show that the statement holds for $n+1$, that is,

$$\sum_{k=0}^{n+1} ar^k = \frac{a(r^{n+2}-1)}{r-1}.$$

We have that,

$$\begin{aligned}\sum_{k=0}^{n+1} ar^k &= \sum_{k=0}^{n} ar^k + ar^{n+1} \overset{IH}{=} \frac{a(r^{n+1}-1)}{r-1} + ar^{n+1}\\ &= \frac{ar^{n+1} - a + ar^{n+1}(r-1)}{r-1} = \frac{ar^{n+1} - a + ar^{n+2} - ar^{n+1}}{r-1}\\ &= \frac{ar^{n+2} - a}{r-1} = \frac{a(r^{n+2}-1)}{r-1}.\end{aligned}$$

This completes the induction step. Therefore, by mathematical induction the statement $\sum\limits_{k=0}^{n} ar^k = \frac{a(r^{n+1}-1)}{r-1}$ is true for every natural number n. ■

Theorem 3.5. *For every integer n with $n \geq 4$, we have that $2^n < n!$.*

Proof. We will proceed by induction. Let $P(n)$ be the statement "$2^n < n!$".
Base Case: We show that $P(n)$ is true for $n = 4$. Since $2^4 = 16 < 24 = 4!$, we have that $P(4)$ is true.
Induction Step: Choose an arbitrary integer n with $n \geq 4$, and assume that $P(n)$ is true. We will show that $P(n+1)$ is true, that is, we show that $2^{n+1} < (n+1)!$ must

be true. We start with the left hand side and then use the induction hypothesis and the fact that $2 < n+1$ since $4 \leq n$.

$$2^{n+1} = 2 \cdot 2^n \overset{IH}{<} 2 \cdot n! < (n+1) \cdot n! = (n+1)!.$$

By mathematical induction we conclude that $2^n < n!$ for all $n \geq 4$. ■

Theorem 3.6. For all $n \in \mathbb{N}$, the following inequality holds.

$$\frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{2^n} \geq 1 + \frac{n}{2}.$$

Proof. We will prove the inequality by induction on n. For the base case, the inequality holds true when $n = 0$ since its left hand side is 1 and its right hand side is $1 + \frac{0}{2} = 1$, and $1 \geq 1$.
For the induction step, fix an arbitrary integer n and assume the induction hypothesis $\frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{2^n} \geq 1 + \frac{n}{2}$ is true. Our goal is to show that

$$\frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{2^n} + \frac{1}{2^n+1} + \ldots + \frac{1}{2^{n+1}} \geq 1 + \frac{n+1}{2}.$$

We now proceed as follows.

$$\begin{aligned}\frac{1}{1} + \frac{1}{2} + \ldots + \frac{1}{2^n} + \frac{1}{2^n+1} + \ldots + \frac{1}{2^{n+1}} &\overset{IH}{\geq} 1 + \frac{n}{2} + \overbrace{\frac{1}{2^n+1} + \frac{1}{2^n+2} + \ldots + \frac{1}{2^n+2^n}}^{2^n \text{ terms, the last being the least}} \\ &\geq 1 + \frac{n}{2} + 2^n\Big(\frac{1}{2^{n+1}}\Big) = 1 + \frac{n}{2} + \frac{1}{2} \\ &= 1 + \frac{n+1}{2}.\end{aligned}$$

Thus $\frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{2^{n+1}} \geq 1 + \frac{n+1}{2}$ establishing the induction step. Therefore, by mathematical induction the proof is complete. ■

Exercise. Use mathematical induction to prove that for every positive integer n, the following hold.

(i) $\sum\limits_{k=1}^{n} \frac{1}{k(k+1)} = \frac{n}{n+1}$.

(ii) $\sum\limits_{k=1}^{n} k(k!) = 1(1!) + 2(2!) + 3(3!) + \ldots + n(n!) = (n+1)! - 1$.

We now present the proof of the principle of mathematical induction which is based on the *well-ordering axiom* for the natural numbers.

Well-ordering Axiom. Every nonempty subset of the set of natural numbers has a least element.

Theorem 3.7. *Let $P(n)$ be a predicate where the variable domain is $\mathbb{N}$. Then the following holds true.*

$$\big(P(0) \wedge \forall n\big(P(n) \to P(n+1)\big)\big) \Longleftrightarrow \forall n P(n).$$

Proof. For the forward direction, suppose that both $P(0)$ and $\forall n(P(n) \to P(n+1))$ are true. We need to show that $\forall n P(n)$ is true. For the contrary, assume that $\forall n P(n)$ is false. It follows that $P(n)$ is not true for all $n \in \mathbb{N}$, and so there is some $c \in \mathbb{N}$ such that $P(c)$ is false. Now let A be the set of all natural numbers which do not have property P, that is,

$$A = \{n \in \mathbb{N} \mid \neg P(n)\}.$$

Thus $c \in A$, and so A is a nonempty subset of $\mathbb{N}$. By the well-ordering axiom, we deduce that A has a least element, call it m. Since m is the least such that it does not have property P, it follows that its predecessor $m-1$ has property P, that is, $P(m-1)$ is true. Moreover, $m \neq 0$, because m does not have property P, while 0 does since we assumed $P(0)$ is true. So $m \geq 1$, and so $m-1 \geq 0$ meaning that $m-1$ is a natural number such that $P(m-1)$ is true. But, by assumption, we have that $P(m-1) \to P(m)$. Since $P(m-1) \to P(m)$ and $P(m-1)$ are true, we have that $P(m)$ is true, contradicting that $P(m)$ is false. Therefore $\forall n P(n)$ must be true, that is, every natural number has property P.

The reverse direction is trivial, since if $\forall n P(n)$ is true, it follows that $P(0)$ is true, and $\forall n(P(n) \to P(n+1))$ is true as well. ■

3.2 Strong Induction

We introduce another form of mathematical induction called *strong induction* where its induction hypothesis assumes more information than we do in usual mathematical induction making strong induction a more flexible proof technique. Let us describe the strategy to prove that $\forall n P(n)$ is true. As before, $P(n)$ is a predicate involving the variable n which ranges over the domain of natural numbers.

Strong Induction. To prove a statement $P(n)$ is true for all $n \in \mathbb{N}$, we complete two steps.

1. Base Case: We show that $P(0)$ is true.
2. Induction Step: We show that

$$\forall n \in \mathbb{N}\Big(\big(P(0) \wedge P(1) \wedge P(2) \wedge \cdots \wedge P(n)\big) \to P(n+1)\Big).$$

   That is, we show that for any $n \in \mathbb{N}$, if $P(k)$ is true for all $0 \leq k \leq n$, then $P(n+1)$ is true.

Notice that any proof using mathematical induction may be considered a proof by strong induction simply because the induction hypothesis of mathematical induction is included in the induction hypothesis of a proof by strong induction.

Theorem 3.8. *Consider the sequence (a_n) where $n \in \mathbb{N}$ defined recursively by the initial conditions $a_0 = 1$ and $a_1 = 3$ and the recurrence relation $a_n = 2a_{n-1} + a_{n-2}$ for $n \geq 2$. Then a_n is an odd integer for every natural number n.*

Proof. We will prove the theorem by strong induction. Let $P(n)$ be the statement

$$\text{“}a_n \text{ is odd”}.$$

Base Case: By definition of the sequence both $P(0)$ and $P(1)$ are true since a_0 and a_1 are both odd integers.
Induction Step: Choose any arbitrary natural number $n \geq 1$, and suppose that the statements

$$P(0), P(1), P(2), \ldots, P(n-1), P(n)$$

are all true. We will show that $P(n+1)$ is also true. So we need to show that a_{n+1} is odd. By the recurrence relation, we have that $a_{n+1} = 2a_n + a_{n-1}$. Since $1 \leq n$, it follows that $0 \leq n-1 \leq n$ and so by induction hypothesis we have that $P(n)$ and $P(n-1)$ are both true. Thus, a_n and a_{n-1} are odd. So there are integers k and l such that $a_n = 2k+1$ and $a_{n-1} = 2l+1$. Thus,

$$a_{n+1} = 2a_n + a_{n-1} = 2(2k+1) + (2l+1) = 4k + 2 + 2l + 1 = 2(2k+1+l) + 1.$$

Thus, a_{n+1} is odd meaning that $P(n+1)$ is true completing the induction step. By strong induction, we showed that a_n is odd for every integer $n \geq 0$. ■

We will use strong induction to prove that any integer greater than or equal to 12 can be written as the sum of nonnegative multiples of 4 and 5.

Theorem 3.9. *For any $n \geq 12$ there are $a, b \in \mathbb{N}$ such that $n = 4a + 5b$.*

Proof. Let $P(n)$ be the statement

$$\text{“}\exists a \in \mathbb{N}\, \exists b \in \mathbb{N}\, (n = 4a + 5b)\text{”}.$$

Base Case: We show that $P(12), P(13), P(14), P(15)$ are all true (the reason will be clear in the induction step).

$$\begin{aligned} 12 &= 4(3) + 5(0), \\ 13 &= 4(2) + 5(1), \\ 14 &= 4(1) + 5(2), \\ 15 &= 4(0) + 5(3). \end{aligned}$$

Now choose any integer $n \geq 15$ and suppose that the statements

$$P(12), P(13), P(14), \ldots, P(n)$$

are all true. We need to show that $P(n+1)$ is true. That is, we need to find some natural numbers a and b such that $n + 1 = 4a + 5b$. Observe that

$$n + 1 = n + (4 - 3) = (n - 3) + 4.$$

Since $15 \leq n$ it follows that $12 \leq n - 3 \leq n$. Thus, by the induction hypothesis $P(n-3)$ is true, and so there are $c, d \in \mathbb{N}$ such that $n - 3 = 4c + 5d$. Thus,

$$n + 1 = (n - 3) + 4 = (4c + 5d) + 4 = 4(c + 1) + 5d.$$

Clearly $c + 1$ and d are natural numbers and thus $P(n+1)$ is true, completing the induction step. Thus, by strong induction we have that $P(n)$ is true for all $n \geq 12$. ■

Motivated by the above example we state below a general form of strong induction.

General Strong Induction. To show that $P(n)$ is true for $n \geq b$ for some integer b, we establish the following.

1. Base Step: We show that $P(b), P(b+1), P(b+2), \ldots, P(b+k)$ are all true for some $k \geq 0$.
2. Induction Step: We show that for every $n \geq b + k$ we have that

$$\big(P(b) \wedge P(b+1) \wedge P(b+2) \wedge \cdots \wedge P(n)\big) \rightarrow P(n+1).$$

Chapter 4

Number Theory

Number theory is a branch of pure mathematics concerned with the study of integers and their properties, in particular prime numbers. The German mathematician Carl Fredrich Gauss once said "*Mathematics is the queen of the sciences, and number theory is the queen of mathematics*".

4.1 Divisibility and Congruence

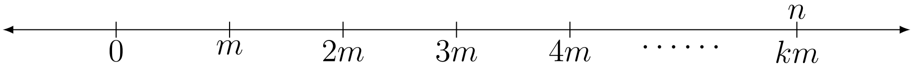
We work with the set of all integers $\mathbb{Z}$ where

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

Furthermore, the set of integers is equipped with the operations of addition and multiplication. And the integers are ordered in the usual sense. We now introduce the relation of divisibility on the integers.

Definition. Given integers m and n with $m \neq 0$, we say that m *divides* n if there exists an integer k such that $n = km$. When m divides n we write $m \mid n$.

Let us elaborate more on the meaning of divisibility. For simplicity, suppose that m, n are positive integers and $m \mid n$. Then $n = km$ for some integer k. Now, imagine yourself standing up at 0 on the number line facing the number n. You then start walking in the positive direction by taking steps of length m. After k many such steps you will hit the integer n. *Thus m divides n means that n can be reached starting from 0 by taking some number of steps each of length m.*



We can express divisibility using predicate logic where $m \mid n$ means that

$$\exists k \in \mathbb{Z}\,(n = km).$$

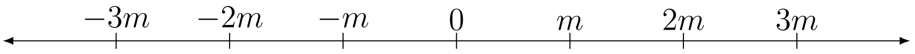
When m divides n we say that m is a *divisor* of n, or n is a *multiple* of m, or that n is *divisible by* m. When m does not divide n, we write $m \nmid n$. Notice that m does not divide n means that it is impossible to find an integer k such that $n = km$. Alternatively, $m \nmid n$ means that one can never hit n by taking steps of length m starting from 0.

Example. It is true that $4 \mid 28$ because we can find an integer, namely 7, such that $28 = 7 \cdot 4$. Note that we can reach 28 starting from 0 by taking 7 steps of length 4. However, $5 \mid 12$ is false because there is no integer k such that $12 = 5k$, so we write $5 \nmid 12$. Here are more examples.

$$\begin{array}{lllllllll}
3 \mid 12, & 1 \mid 5, & 6 \nmid 3, & -2 \mid 6, & 4 \mid 12, & 2 \mid 6, & 7 \mid 14, & -4 \mid 8, & 7 \nmid 19, \\
5 \mid 25, & 3 \mid 6, & 14 \nmid 7, & 4 \mid 8, & 5 \mid 20, & 4 \nmid 6, & 8 \mid 0, & 4 \mid -8, & 19 \mid 323.
\end{array}$$

$\Diamond$

Example. What are the integers divisible by a positive integer m? In other words, which integers m divides?
These are the integers of the form km where $k \in \mathbb{Z}$. That is, they are the integers which can be reached starting from 0 and then moving forward or backward in steps of length m.



For instance, the integers divisible by 5 are

$$0,\ 5, -5,\ 10, -10,\ 15, -15,\ 20, -20,\ 25, \ldots.$$

These integers are also called the multiples of 5. $\Diamond$

Next we prove some basic properties of the divisibility relation on the integers.

Theorem 4.1. *Let a, b, c be integers. Then the following hold.*

(i) If $a \mid b$ and $a \mid c$, then $a \mid (b + c)$.

(ii) If $a \mid b$, then $a \mid bc$.

(iii) If $a \mid b$ and $b \mid c$, then $a \mid c$.

Proof. These statements are conditional statements. We will use a direct proof.

(i) Suppose $a \mid b$ and $a \mid c$. Then there are integers k and l such that $b = ka$ and $c = la$. We have that, $b + c = ka + la = (k + l)a$. Since $k + l$ is an integer, it follows that $b + c$ is equal to an integer multiplied by a, and so by definition of divisibility we have that $a \mid (b + c)$ as desired.

(ii) Suppose $a \mid b$. Then $b = ka$ for some $k \in \mathbb{Z}$. Hence, $bc = (ka)c = (kc)a$. Since kc is an integer, we have that $a \mid bc$ as required.

(iii) Suppose $a \mid b$ and $b \mid c$. Then there exist integers k and l such that $b = ka$ and $c = lb$. We now have that $c = lb = l(ka) = (lk)a$. Since lk is an integer we conclude that $a \mid c$.

The proof is complete. ■

Suppose that m and n are integers. An integer of the form $km + ln$ where $k, l \in \mathbb{Z}$ is called a *linear combination* of m and n. For example 26 is a linear combination of 4 and 7 because $26 = 3 \cdot 4 + 2 \cdot 7$. A consequence of Theorem 4.1 is that if an integer divides m and n, then it divides any linear combination of m and n. See the next corollary.

Corollary 4.2. Let $a, m, n \in \mathbb{Z}$. Suppose that $a \mid m$ and $a \mid n$. Then $a \mid (km + ln)$ for any integers k and l.

Proof. Let a, m, n be integers and suppose that $a \mid m$ and $a \mid n$. Choose arbitrary integers k and l. By Theorem 4.1(ii) we have that $a \mid km$ and $a \mid ln$. By Theorem 4.1(i) we have that $a \mid (km + ln)$ as required. ■

Theorem 4.3. *For any positive integer n, we have that $3 \mid (n^3 - n)$.*

Proof. We will prove the theorem by induction on n. For the base case, suppose that $n = 1$. Then $1^3 - 1 = 1 - 1 = 0 = 0 \cdot 3$, meaning that 3 divides $1^3 - 1$.
For the induction step, choose any integer $n \geq 1$ and assume that $3 \mid (n^3 - n)$. We need to show that 3 divides $(n + 1)^3 - (n + 1)$. Since $3 \mid n^3 - n$, there is an integer k such that $n^3 - n = 3k$. Now, we have that

$$\begin{aligned}(n+1)^3 - (n+1) &= (n+1)((n+1)^2 - 1) \\ &= (n+1)(n^2 + 2n + 1 - 1) = (n+1)(n^2 + 2n) \\ &= n^3 + 3n^2 + 2n = (n^3 - n) + 3n^2 + 3n \\ &\overset{IH}{=} 3k + 3n^2 + 3n = 3(k + n^2 + n).\end{aligned}$$

Therefore, $3 \mid (n + 1)^3 - (n + 1)$ establishing the induction step. By mathematical induction we proved that $3 \mid n^3 - n$ for any integer $n \geq 1$. ■

♣ Division Algorithm

Given an integer n and a positive integer m, the Division Algorithm guarantees the existence of unique integers q and r satisfying two properties:

- $n = qm + r$,
- $0 \leq r < m$.

We call n the *dividend*, and m the *divisor*, and q the *quotient*, and r the *remainder*. We say that r is the remainder when n is divided by m. The remainder r is denoted by "$n \bmod m$".

The discussion above can be described as follows. Imagine yourself standing on the number line at 0. Then you can reach an integer n by taking q steps, each of length m, until you get very close to n, and finally hitting n by taking one last shorter step of length r.

Example. When we divide 30 by 4 the Division Algorithm produces a quotient of 7 and a remainder of 2 because

- $30 = 7(4) + 2$ and
- $0 \leq 2 < 4$.

Here we have that $30 \bmod 4 = 2$. Now imagine you stand up at 0 on the number line aiming to reach 30 by steps of length 4. You start walking from 0 taking 7 steps of length 4 where you arrive at 28. At this point, if you take one more step of length 4 you will jump over 30 missing your goal. In this situation you are forced to take a shorter last step of length 2 to arrive to 30. In conclusion, to arrive to 30 starting from 0, first we take 7 steps of length 4 and then followed with a single remaining shorter step of length 2. ◊

Remark. We have that $m \mid n$ if and only if the remainder when n is divided by m is 0 if and only if $n \bmod m = 0$.

Example. Find the remainder in each of the following divisions.

- Divide 101 by 11.
  $101 = 9(11) + 2$. Thus, $101 \bmod 11 = 2$.
- Divide -11 by 3.
  $-11 = -4(3) + 1$. Thus, $-11 \bmod 3 = 1$.
- Divide 19 by 5.
  $19 = 3(5) + 4$. Thus, $19 \bmod 5 = 4$.

- Divide -19 by 5.
  $-19 = -4(5) + 1$. Thus, $-19 \bmod 5 = 1$.
- Divide -27 by 5.
  $-27 = -6(5) + 3$. Thus, $-27 \bmod 5 = 3$.
- Divide 7 by 11.
  $7 = 0(11) + 7$. Thus, $7 \bmod 11 = 7$.
- Divide -7 by 11.
  $-7 = -1(11) + 4$. Thus, $-7 \bmod 11 = 4$. ◊

We next prove the Division Algorithm. The proof uses the well-ordering axiom which states that any nonempty subset of the set of natural numbers has a least element.

Theorem 4.4 (Division Algorithm). *Let n be any integer and $m \in \mathbb{Z}^+$. Then there are unique integers q and r such that $n = qm + r$ and $0 \leq r < m$.*

Proof. Fix some $n \in \mathbb{Z}$ and $m \in \mathbb{Z}^+$. We will show that q and r as required exist and, moreover, we will show that they are unique. Towards showing their existence consider the set

$$S = \{t \in \mathbb{Z} \mid t \geq 0 \text{ and } \exists k \in \mathbb{Z}\,(t = n - km)\}.$$

So an element in S is a nonnegative integer of the form $n - km$ for some integer k. Since every element in S is a nonnegative integer, it follows that $S \subseteq \mathbb{N}$. We now show that S is nonempty. To see this, we consider two cases. If $n \geq 0$, then $n \in S$ because in this case n is in the required form, i.e. $n = n - 0 \cdot m$. Otherwise, suppose that $n < 0$ and consider the integer $n - nm$. We have that $n - nm = n(1 - m)$, and since $m \geq 1$ it follows that $1 - m \leq 0$. Therefore, we have that n is negative and $1 - m$ is negative or zero, so their product $n(1 - m) \geq 0$. Hence, $n - nm$ belongs to S. In both cases, S is a nonempty set. Since S is a nonempty subset of the natural numbers it follows by the well-ordering axiom that S contains a least element, call this least element r.

As $r \in S$ it follows that $r \geq 0$ and $r = n - qm$ for some integer q. Thus, $n = qm + r$ as required in the Division Algorithm. It remains to show that $r < m$. For the sake of contradiction assume that $r \geq m$. This implies that $0 \leq r - m$. Moreover, as $m > 0$ it follows that $r - m < r$. Thus, $0 \leq r - m < r$. So $r - m$ is a nonnegative integer strictly less than r. Moreover, $r - m = n - qm - m = n - (q + 1)m$, and so $r - m$ satisfies the membership conditions of S which implies that $r - m$ belongs to S. But $r - m < r$ contradicting that r is the least element in S. Therefore, it must be that $r < m$. Therefore, we showed that there are integers q and r such that $n = qm + r$ and $0 \leq r < m$.

It remains to show that q and r are unique. Suppose that there are integers q' and

r' such that $n = q'm + r'$ and $0 \le r' < m$. We will show that $r' = r$ and $q' = q$. Without loss of generality, assume that $r' \le r$. Since $0 \le r, r' < m$, it follows that $0 \le r - r' < m$. From $n = qm + r = q'm + r'$ we deduce that $r - r' = m(q' - q)$, and so m divides $r - r'$. But since $0 \le r - r' < m$, we must have that $r - r' = 0$ because 0 is the only nonnegative integer less than m which is divisible by m. Therefore, $r = r'$. We now have that $0 = r - r' = m(q' - q)$. But $m > 0$, which implies that $q' - q = 0$, and so $q = q'$. Therefore, q and r are unique. This finishes the proof of the Division Algorithm. ■

♣ Congruences

We use divisibility to introduce another relation among the integers called congruence.

Definition. Let $a, b \in \mathbb{Z}$ and $m \in \mathbb{Z}^+$. We say that a is *congruent* to b modulo m if m divides $a - b$. We write $a \equiv b \pmod{m}$ when a is congruent to b modulo m.

We call the statement $a \equiv b \pmod{m}$ a *congruence*. When a is not congruent to b modulo m we write $a \not\equiv b \pmod{m}$ which means that $m \nmid (a - b)$.

Example. It is true that $29 \equiv 11 \pmod{6}$ because 6 divides $29 - 11 = 18$ since $18 = 3 \cdot 6$. Study the following congruences and find a relation between a, b being congruent modulo m and their remainders after division by m.

- $17 \equiv 11 \pmod{6}$. Note that $17 \bmod 6 = 5 = 11 \bmod 6$.
- $17 \equiv 5 \pmod{6}$. Note that $17 \bmod 6 = 5 = 5 \bmod 6$.
- $28 \equiv 19 \pmod{3}$. Note that $28 \bmod 3 = 1 = 19 \bmod 3$.
- $33 \not\equiv 18 \pmod{2}$. Note that $33 \bmod 2 = 1$ but $18 \bmod 2 = 0$.
- $24 \not\equiv 14 \pmod{6}$. Note that $24 \bmod 6 = 0$ but $14 \bmod 6 = 2$.
- $10 \equiv 45 \pmod{7}$. Note that $10 \bmod 7 = 3 = 45 \bmod 7$.
- $33 \equiv 0 \pmod{11}$. Note that $33 \bmod 11 = 0 = 0 \bmod 11$. ♢

The examples above suggest that two integers being congruent modulo m implies that they have the same remainder when divided by m. The next theorem shows that these two properties are in fact equivalent.

Theorem 4.5. *Let a, b be integers and $m \in \mathbb{Z}^+$. Then,*

$$a \equiv b \pmod{m} \quad \textit{if and only if} \quad a \bmod m = b \bmod m.$$

Proof. For the forward direction suppose that $a \equiv b \pmod{m}$. This means that $m \mid a - b$ and so there is $k \in \mathbb{Z}$ such that $a - b = km$. We now apply the Division Algorithm to divide both a and b by m. We obtain integers q_1, r_1, q_2, r_2 such that $a = q_1 m + r_1$ where $0 \leq r_1 < m$ and $b = q_2 m + r_2$ where $0 \leq r_2 < m$.
We need to show that $a \bmod m = b \bmod m$, in other words, to show that $r_1 = r_2$. Without loss of generality, assume that $r_2 \leq r_1$. Hence, $0 \leq r_1 - r_2 < m$. Then we have that

$$km = a - b = (q_1 m + r_1) - (q_2 m + r_2) = (q_1 - q_2)m + (r_1 - r_2).$$

Thus,

$$(r_1 - r_2) = m(k - q_1 + q_2).$$

Hence, $m \mid (r_1 - r_2)$ and since $0 \leq r_1 - r_2 < m$, it must be that $r_1 - r_2 = 0$. Therefore, we showed that $r_1 = r_2$.

For the reverse direction assume that a mod $m = b$ mod m. So a and b have the same remainder when divided by m. Applying the Division Algorithm to divide a and b by m we get integers q_1, q_2, and r such that $a = q_1 m + r$ and $b = q_2 m + r$ where $0 \leq r < m$. It follows that,

$$a - b = (q_1 m + r) - (q_2 m + r) = (q_1 - q_2)m.$$

Thus, m divides $a - b$, and so $a \equiv b \pmod{m}$ as required. ■

The next lemma shows that we can add and multiply the sides of congruences.

Lemma 4.6. *Let a, b, c, d be integers and $m \in \mathbb{Z}^+$. Suppose that $a \equiv b \pmod{m}$ and $c \equiv d \pmod{m}$. Then the following are true.*

(i) $a + c \equiv b + d \pmod{m}$

(ii) $ac \equiv bd \pmod{m}$.

Proof. Since $a \equiv b \pmod{m}$ and $c \equiv d \pmod{m}$, there are $k, l \in \mathbb{Z}$ such that $a - b = km$ and $c - d = lm$. For the first congruence we have that,

$$(a + c) - (b + d) = (a - b) + (c - d) = km + lm = (k + l)m.$$

So $a + c \equiv b + d \pmod{m}$.
For the second congruence, we have that

$$\begin{aligned} ac - bd &= (b + km)(d + lm) - bd \\ &= bd + blm + kmd + klm^2 - bd \\ &= (bl + kd + klm)m. \end{aligned}$$

This shows that $m \mid (ac - bd)$, and so $ac \equiv bd \pmod{m}$. ■

Example. Consider the congruences $7 \equiv 2 \pmod 5$ and $11 \equiv 1 \pmod 5$. And clearly, $4 \equiv 4 \pmod 5$. Then using the lemma above we have the following. Check that they are true statements.

- $7 + 11 \equiv 2 + 1 \pmod 5$, and so $18 \equiv 3 \pmod 5$.
- $7 \cdot 11 \equiv 2 \cdot 1 \pmod 5$, and so $77 \equiv 2 \pmod 5$.
- $7 + 4 \equiv 2 + 4 \pmod 5$, and so $11 \equiv 6 \pmod 5$.
- $7 \cdot 4 \equiv 2 \cdot 4 \pmod 5$, and so $28 \equiv 8 \pmod 5$. ◊

Example. Note that $14 \equiv 8 \pmod 6$ but $7 \not\equiv 4 \pmod 6$. Thus we cannot always divide both sides of a congruence. ◊

Lemma 4.7. *Let a and b be integers and m be a positive integer. Then*

$$(a+b) \bmod m = \big((a \bmod m) + (b \bmod m)\big) \bmod m$$

and

$$(ab) \bmod m = \big((a \bmod m)(b \bmod m)\big) \bmod m.$$

♣ Modular Arithmetic

Given a positive integer m, consider the set

$$\mathbb{Z}_m = \{0, 1, 2, \ldots, m-1\}.$$

Note that $\mathbb{Z}_m$ is the set of all possible remainders when dividing by m. We aim to develop a number system on $\mathbb{Z}_m$ by defining arithmetic operations on $\mathbb{Z}_m$. Let $a, b \in \mathbb{Z}_m$. First, we define *addition modulo m* as follows,

$$a +_m b = a + b \bmod m.$$

Second, we define *multiplication modulo m* as follows,

$$a \cdot_m b = ab \bmod m.$$

So $a +_m b$ is the remainder when $a + b$ is divided by m. And $a \cdot_m b$ is the remainder when ab is divided by m.

Example. Consider $\mathbb{Z}_{11} = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$. Below are examples of addition and multiplication modulo 11 on elements from $\mathbb{Z}_{11}$.

- $7 +_{11} 9 = (7 + 9) \bmod 11 = 16 \bmod 11 = 5$.

- $4 +_{11} 6 = (4 + 6) \bmod 11 = 10 \bmod 11 = 10$.
- $9 +_{11} 10 = (9 + 10) \bmod 11 = 19 \bmod 11 = 8$.
- $8 \cdot_{11} 3 = (8 \cdot 3) \bmod 11 = 24 \bmod 11 = 2$.
- $4 \cdot_{11} 10 = (4 \cdot 10) \bmod 11 = 40 \bmod 11 = 7$.
- $7 \cdot_{11} 9 = (7 \cdot 9) \bmod 11 = 63 \bmod 11 = 8$. ♢

Remark. Working in $\mathbb{Z}_{12}$ with addition modulo 12 is what we do in *clock arithmetic* when we add the hours on a clock. For instance, if the time now is 10 o'clock, then 5 hours later it will be $10 +_{12} 5 = 15 \bmod 12 = 3$ o'clock.

The next theorem states several properties of the number system $(\mathbb{Z}_m, +_m, \cdot_m)$.

Theorem 4.8. *Let m be a positive integer, and $a, b, c \in \mathbb{Z}_m$. Then the following properties hold.*

- *Both integers $a +_m b$ and $a \cdot_m b$ belong to $\mathbb{Z}_m$.* *(Closure)*
- *$a +_m b = b +_m a$ and $a \cdot_m b = b \cdot_m a$.* *(Commutativity)*
- *$(a +_m b) +_m c = a +_m (b +_m c)$ and $(a \cdot_m b) \cdot_m c = a \cdot_m (b \cdot_m c)$.* *(Associativity)*
- *$a +_m 0 = a$.* *(Additive Identity)*
- *$a \cdot_m 1 = a$.* *(Multiplicative Identity)*
- *If $a \neq 0$, then $a +_m (m - a) = 0$.* *(Additive Inverse)*
- *$a \cdot_m (b +_m c) = (a \cdot_m b) +_m (a \cdot_m c)$.* *(Distributivity)*

Remark. Taking into account the properties above, the mathematical structure $(\mathbb{Z}_m, +_m)$ is an example of an “Abelian group”, and $(\mathbb{Z}_m, +_m, \cdot_m)$ is an example of a “commutative ring”. Groups and rings are objects studied in the field of modern algebra.

4.2 Prime Numbers

Observe that every positive integer greater than 1 is divisible by at least two positive integers, namely 1 and itself. Those integers which have no other positive divisors are special. They are called prime numbers and they were studied since ancient times. Prime numbers have important applications in cryptography. The concept of prime numbers is based on the concept of divisibility.

Definition. An integer $p > 1$ is *prime* if its only positive divisors are 1 and p. An integer $n > 1$ that is not a prime is called *composite.*

Here are the first prime numbers.

2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, 53, 59, 61, 67, 71, 73, 79, 83, 89, 97, 101, 103, 107, 109, 113, 127, 131, 137, 139, 149, 151, 157, 163, 167, 173, 179, 181, 191, 193, 197, 199, 211, 223, 227, 229, 233, 239, 241, 251, 257, 263, 269, 271, 277, 281, 283, 293, 307, 311, 313, 317, 331, 337, 347, 349, 353, 359, 367, 373, 379, 383, 389, 397, 401, 409, 419, 421, 431, 433, 439, 443, 449, 457, 461, 463, 467, 479, 487, 491, 499, 503, 509, 521, 523, 541, 547, 557, 563, 569, 571, 577, 587, 593, 599, 601, 607, 613, 617, 619, 631, 641, 643, 647, 653, 659, 661, 673, 677, 683, 691, 701, 709, 719, 727, 733, 739, 743, 751, 757, 761, 769, 773, 787, 797, 809, 811, 821, 823, 827, 829, 839, 853, 857, 859, 863, 877, 881, 883, 887, 907, 911, 919, 929, 937, 941, 947, 953, 967, 971, 977, 983, 991, 997, 1009, 1013, 1019, 1021, 1031, 1033, 1039, 1049, 1051, 1061, 1063, 1069, 1087, 1091, 1093, 1097, 1103, 1109, 1117, 1123, 1129, 1151, 1153, 1163, 1171, 1181, 1187, 1193, 1201, 1213, 1217, 1223, 1229, 1231, 1237, 1249, 1259, 1277, 1279, 1283, 1289, 1291, 1297, 1301, 1303, 1307, 1319, 1321, 1327, 1361, 1367, 1373, 1381, 1399, 1409, 1423, 1427, 1429, 1433, 1439, 1447, 1451, 1453, 1459, 1471, 1481, 1483, 1487, 1489, 1493, 1499, 1511, 1523, 1531, 1543, 1549, 1553, 1559, 1567, 1571, 1579, 1583, 1597, 1601, 1607, 1609, 1613, 1619, 1621, 1627, 1637, 1657, 1663, 1667, 1669, 1693, 1697, 1699, 1709, 1721, 1723, 1733, 1741, 1747, 1753, 1759, 1777, 1783, 1787, 1789, 1801, 1811, 1823, 1831, 1847, 1861, 1867, 1871, 1873, 1877, 1879, 1889, 1901, 1907, 1913, 1931, 1933, 1949, 1951, 1973, 1979, 1987, ...

Exercise. At which point this sequence of prime numbers will stop?

Lemma 4.9. *An integer $n > 1$ is composite if and only if there is an integer k with $1 < k < n$ such that $k \mid n$.*

Proof. For the forward direction suppose that $n > 1$ is composite. By definition of composite, n is not prime meaning that it has a positive divisor k which is not 1 nor n. Thus $k \mid n$ and $1 < k < n$.

For the reverse direction, if there is an integer k such that $1 < k < n$ and $k \mid n$, then n is not a prime, that is, n is composite. ■

The prime numbers are the building blocks of integers, they are the atoms from which integers are built up. This is what the Fundamental Theorem of Arithmetic tells us. We will prove this theorem in the end of this chapter.

Theorem 4.10 (Fundamental Theorem of Arithmetic)**.** *Every integer greater than 1 can be written uniquely as a prime or a product of primes written in a nondecreasing order.*

Example. Here are some examples showing the decompostion of integers into product of primes. Can you find different representations of the integers below other than the ones given?

- $210 = 2 \cdot 3 \cdot 5 \cdot 7$.
- $18 = 2 \cdot 3 \cdot 3$.
- $30 = 2 \cdot 3 \cdot 5$.
- $32 = 2 \cdot 2 \cdot 2 \cdot 2 \cdot 2$.
- $2079 = 3 \cdot 3 \cdot 3 \cdot 7 \cdot 11$.
- $1024 = 2^{10}$. ◊

Lemma 4.11. *If n is composite, then there is a prime p that divides n where $p \leq \sqrt{n}$.*

Proof. Since n is composite, by Lemma 4.9 there are integers k and l such that $n = kl$ and $1 < k, l < n$. We claim that either k or l is at most $\sqrt{n}$. For the sake of contradiction, assume that both $k > \sqrt{n}$ and $l > \sqrt{n}$. Then

$$n = kl > \sqrt{n} \cdot \sqrt{n} = n.$$

It follows that $n > n$, a clear contradiction. Consequently, either $k \leq \sqrt{n}$ or $l \leq \sqrt{n}$. Without loss of generality, assume that $k \leq \sqrt{n}$. By the Fundamental Theorem of Arithmetic k has a prime divisor p. Since $p \mid k$ and $k \mid n$, by Theorem 4.1(iii) it follows that $p \mid n$, and since $p \leq k$ and $k \leq \sqrt{n}$, we have that $p \leq \sqrt{n}$ as desired. ■

The contrapositive of the lemma above gives us a clever test to check whether an integer is prime or composite.

Corollary 4.12. Given a positive integer $n > 1$, if every prime $p \leq \sqrt{n}$ does not divide n, then n is a prime number.

Example. Is 107 prime?
To know whether 107 is prime or not, check if $p \mid 107$ for each prime $p \leq \sqrt{107}$. If the answer is always negative then 107 is prime. Since $\sqrt{107} \approx 10.344$, the primes less than or equal to $\sqrt{107}$ are $2, 3, 5$, and 7. Since $2 \nmid 107$, and $3 \nmid 107$, and $5 \nmid 107$, and $7 \nmid 107$, we conclude that 107 is indeed prime. ◊

Exercise. Write a program using your favourite programming language to test whether an integer input is prime or not.

> Euclid of Alexandria (3^{rd} century BC) was a Greek mathematician known as the founder of geometry. He lived in Alexandria where he wrote the mathematical treatise "*Elements*", one of the most influential works in the history of mathematics which served as a main textbook in teaching mathematics since its publication until the nineteenth century. Euclid used the axiomatic method in the Elements, where he deduced theorems on geometry from a set of axioms. This work of his is known as Euclidean Geometry. He also worked on conic sections, spherical geometry, and number theory.

Given any prime number, no matter how large it is, there is another prime number greater than it. The sequence of prime numbers never stop! The proof given below of this fact is found in the text "*Elements*" written by Euclid of Alexandria. This proof is considered among the most beautiful and elegant proofs in mathematics.

Theorem 4.13. *There are infinitely many primes.*

Proof. For the sake of contradiction, assume that there are only finitely many primes, say, $p_1, p_2, p_3, \ldots, p_n$. Consider the integer N which is one more than the product of all these primes. That is,

$$N = p_1 p_2 p_3 \cdots p_n + 1.$$

By the Fundamental Theorem of Arithmetic, N is a product of a collection of these primes, and thus one of these primes divides N. So $p_i \mid N$ for some $1 \leq i \leq n$. Thus, there is an integer k such that $N = kp_i$. We have that,

$$\begin{aligned} N = p_1 p_2 p_3 \ldots p_n + 1 &\Longrightarrow kp_i = p_1 p_2 p_3 \ldots p_n + 1 \\ &\Longrightarrow kp_i - p_1 p_2 \ldots p_i \ldots p_n = 1 \\ &\Longrightarrow p_i(k - p_1 p_2 \ldots p_{i-1} p_{i+1} \ldots p_n) = 1. \end{aligned}$$

The last statement implies that the prime p_i divides 1, which is a contradiction since no prime divides 1. So our original assumption is false, meaning that there are infinitely many primes. ■

In the seventeenth century, the French monk Marin Mersenne studied primes of special form which were named after him.

Definition. A *Mersenne prime* is a prime number of the form $2^p - 1$ where p is a prime.

Exercise. Prove that if r is composite, then $2^r - 1$ is also composite.
Hint: Use the identity $2^{mn} - 1 = (2^m - 1)(2^{m(n-1)} + 2^{m(n-2)} + \cdots + 2^m + 1)$.

The integer 3 is a Mersenne prime because 3 is prime and $3 = 2^2 - 1$. Also 7 is a Mersenne prime since $7 = 2^3 - 1$, and 31 is a Mersenne prime since $31 = 2^5 - 1$. On the other hand, although $2047 = 2^{11} - 1$ but 2047 is not a Mersenne prime because 2047 is not prime since $2047 = 23 \cdot 89$. Thus, if p is prime, $2^p - 1$ need not to be prime. One of the efficient tests used to determine the primality of $2^p - 1$ is known as Lucas-Lehmer primality test. Testing whether $2^p - 1$ is prime or not is one of the methods used to test supercomputers.

We finish this section with a famous open question in number theory. A *twin prime* is a pair (p, q) of primes such that $p + 2 = q$. Here are some examples of twin primes.

$$(3,5),\ (5,7),\ (11,13),\ (17,19),\ (41,43),\ (59,61),\ (101,103),\ (107,109), \ldots$$

We still do not know whether this sequence continues for ever. We have the following conjecture.

Conjecture (Twin Prime Conjecture)**.** *There are infinitely many twin primes.*

In 2013, the Chinese mathematician Yitang Zhang published a paper in the journal *Annals of Mathematics* proving that there are infinitely many pairs of primes which differ by some integer less than 70 million. One year later, the bound has been reduced to 246. Further work by the British mathematician James Maynard and the Australian-American Terence Tao accomplished substantial progress towards proving the Twin Prime Conjecture, but at present it remains unsolved.

4.3 Greatest Common Divisor

Definition. The *greatest common divisor* of nonzero integers m and n is the largest integer which divides both m and n. Such integer is denoted by $\gcd(m, n)$.

$$\gcd(m, n) = \max\{d \in \mathbb{Z} : d \mid m \wedge d \mid n\}.$$

Example. What is $\gcd(24, 36)$?
The divisors of 24 are $1, 2, 3, 4, 6, 8, 12, 24$.
The divisors of 36 are $1, 2, 3, 4, 6, 9, 12, 18, 36$.
Their common divisors are $1, 2, 3, 4, 6, 12$. Hence $\gcd(24, 36) = 12$. $\Diamond$

Example. What is $\gcd(21, 34)$?
The divisors of 21 are $1, 3, 7$.
The divisors of 34 are $1, 2, 17$.
The only common divisor is 1. Hence $\gcd(21, 34) = 1$. $\Diamond$

Definition. We call integers m and n *coprime* or *relatively prime* if $\gcd(m, n) = 1$.

Definition. The *least common multiple* of nonzero integers m and n is the smallest positive integer which is divisible by both m and n.

$$\operatorname{lcm}(m, n) = \min\{l \in \mathbb{Z}^+ : m \mid l \wedge n \mid l\}.$$

We can use the prime factorisation of m and n to find $\gcd(m, n)$ and $\operatorname{lcm}(m, n)$ as the theorem below explains.

Theorem 4.14. *Suppose the prime factorisation of $m, n \in \mathbb{Z}^+$ are given as*

$$m = p_1^{m_1} p_2^{m_2} p_3^{m_3} \dots p_k^{m_k} \qquad \textit{and} \qquad n = p_1^{n_1} p_2^{n_2} p_3^{n_3} \dots p_k^{n_k}$$

where $p_1, p_2, \dots, p_k$ are distinct primes, $m_i \geq 0$, and $n_i \geq 0$. Then

$$\gcd(m, n) = p_1^{\min(m_1,n_1)} \times p_2^{\min(m_2,n_2)} \times \cdots \times p_k^{\min(m_k,n_k)}$$

and

$$\operatorname{lcm}(m, n) = p_1^{\max(m_1,n_1)} \times p_2^{\max(m_2,n_2)} \times \cdots \times p_k^{\max(m_k,n_k)}.$$

The reader is encouraged to prove the theorem above. Let us apply this theorem in an example.

Example. Use prime factorisation to find $\gcd(120, 500)$ and $\operatorname{lcm}(120, 500)$.

We first find the prime factorisations of 120 and 500.

$$120 = 2^3 \cdot 3^1 \cdot 5^1 \qquad \text{and} \qquad 500 = 2^2 \cdot 3^0 \cdot 5^3.$$

Therefore,

$$\gcd(120,500) = 2^2 \cdot 3^0 \cdot 5^1 = 20 \quad \text{and} \quad \operatorname{lcm}(120,500) = 2^3 \cdot 3^1 \cdot 5^3 = 3000.$$

Moreover, observe that

$$120 \times 500 = 60\,000 = 20 \times 3000 = \gcd(120,500) \cdot \operatorname{lcm}(120,500). \qquad \diamondsuit$$

The last observation is not a coincidence; it always holds.

Lemma 4.15. *Let $m, n \in \mathbb{Z}^+$. Then*

$$m \cdot n = \gcd(m,n) \cdot \operatorname{lcm}(m,n).$$

♣ Euclidean Algorithm

Computing the prime factorisation of an integer is time consuming, which makes computing the greatest common divisor (GCD) using prime factorisation inefficient. The Euclidean algorithm is a more efficient way to compute the GCD. Recall that when we divide n by m using the Division Algorithm we get q and r such that $n = qm + r$, and we call m the divisor and r the remainder.

Euclidean Algorithm. To find the greatest common divisor of two positive integers we successively apply the Division Algorithm.

- Start by dividing the larger integer by the smaller one.
- Is the remainder 0? If not, obtain the next iteration through dividing the divisor by the remainder of the current iteration. Repeat.
- Stop when the remainder is 0.
- Declare the GCD to be the divisor of the last iteration.

Example. Find $\gcd(665, 287)$ using the Euclidean Algorithm.

Iteration	Computation
1	$665 = 2(287) + 91$
2	$287 = 3(91) + 14$
3	$91 = 6(14) + 7$
4	$14 = 2(\mathbf{7}) + 0$

We conclude that $\gcd(665, 287) = 7$. To check that this is true, we look at their prime factorisation: $665 = 5 \cdot 7 \cdot 19$ and $287 = 7 \cdot 41$. $\Diamond$

Our goal now is to prove that the Euclidean algorithm is correct, that is, it always terminates and spits out the greatest common divisor. Towards this goal, we need a little lemma.

Lemma 4.16. *Suppose that $n = qm + r$ where m, n, q, r are nonzero integers. Then $\gcd(n, m) = \gcd(m, r)$.*

Proof. Let $d = \gcd(n, m)$ and $k = \gcd(m, r)$. Since $d \mid n$ and $d \mid m$, we get by Corollary 4.2 that d divides any linear combination of m and n, in particular, $d \mid n - qm$, and so $d \mid r$. Since d is a common divisor of m and r, it follows that $d \leq k$ because k is the greatest common divisor of m and r.
Similarly, since $k \mid m$ and $k \mid r$, we have that $k \mid qm + r$ since $qm + r$ is a linear combination of m and r, so $k \mid n$. Thus, k is a common divisor of m and n, implying that $k \leq d$. Since $k \leq d$ and $d \leq k$ it follows that $k = d$ as required. $\blacksquare$

When we apply this lemma to the computations we did above to find $\gcd(665, 287)$ we get the following facts.

Iteration	Computation	Observation
1	$665 = 2(287) + 91$	$\gcd(665, 287) = \gcd(287, 91)$
2	$287 = 3(91) + 14$	$\gcd(287, 91) = \gcd(91, 14)$
3	$91 = 6(14) + 7$	$\gcd(91, 14) = \gcd(14, 7)$
4	$14 = 2(\mathbf{7}) + 0$	$\gcd(14, 7) = 7$, as $7 \mid 14$

Putting the equalities together we get,

$$\gcd(665, 287) = \gcd(287, 91) = \gcd(91, 14) = \gcd(14, 7) = 7.$$

Observe that the Euclidean Algorithm works by reducing the problem of finding the GCD of two integers to finding the GCD of some smaller integers. In the previous example, the problem of finding $\gcd(665, 287)$ was reduced to finding $\gcd(14, 7)$.

We now describe the mechanism of the Euclidean Algorithm in its general form and explain why it works. Let m, n be positive integers with $m \leq n$. We successively apply the Division Algorithm to obtain a sequence of iterations.

Iteration	Computation	Remainder
1	$n = q_1 m + r_1$	$0 \leq r_1 < m$
2	$m = q_2 r_1 + r_2$	$0 \leq r_2 < r_1$
3	$r_1 = q_3 r_2 + r_3$	$0 \leq r_3 < r_2$
4	$r_2 = q_4 r_3 + r_4$	$0 \leq r_4 < r_3$
$\vdots$	$\vdots$	$\vdots$
k	$r_{k-2} = q_k r_{k-1} + r_k$	$0 \leq r_k < r_{k-1}$
$k+1$	$r_{k-1} = q_{k+1}\mathbf{r_k} + 0$	$r_{k+1} = 0$

The algorithm output is $\gcd(n, m) = r_k$.

First, the algorithm must eventually terminate because the remainders are nonnegative and strictly decreasing. Thus, after at most m iterations of successive divisions a remainder of 0 is obtained because

$$m > r_1 > r_2 > r_3 > \cdots > r_k > r_{k+1} = 0.$$

Second, by Lemma 4.16 we have that,

$$\begin{aligned} \gcd(n, m) &= \gcd(m, r_1) = \gcd(r_1, r_2) \\ &= \gcd(r_2, r_3) = \ldots = \gcd(r_{k-2}, r_{k-1}) \\ &= \gcd(r_{k-1}, r_k) = r_k. \end{aligned}$$

The last equality holds because by the last iteration of the algorithm, where the remainder is 0, we have that $r_{k-1} = q_{k+1} r_k$ meaning that r_k is a divisor of r_{k-1} and so $\gcd(r_{k-1}, r_k) = r_k$.

♣ Bezout's Theorem

Bezout's Theorem tells us that the greatest common divisor of two integers can always be written as a linear combination of these integers.

Theorem 4.17 (Bezout's Theorem)**.** *Suppose that m and n are nonzero integers. Then there are $s, t \in \mathbb{Z}$ such that $gcd(m, n) = sm + tn$.*

Proof. Let $d = \gcd(m, n)$. Define l to be the least positive integer such that l is a linear combination of m and n. That is,

$$l = \min\{k \in \mathbb{Z}^+ : \exists a \in \mathbb{Z}\, \exists b \in \mathbb{Z}\, (k = am + bn)\}.$$

Notice that l exists by the well-ordering axiom. By definition of l, we know that $l > 0$ and $l = sm + tn$ for some integers s, t. We will show that $d = l$, meaning that $\gcd(m, n)$ is the least positive linear combination of m and n, which proves the theorem.

First, we prove that $d \leq l$. To see this, as $d \mid m$ and $d \mid n$, by Corollary 4.2 we know that d divides any linear combination of m and n, in particular $d \mid l$. Thus $d \leq l$.

Second, we show that l is a common divisor of m and n. To show that $l \mid m$, we apply the Division Algorithm to divide m by l and obtain integers q, r such that $m = ql + r$ where $0 \leq r < l$. Now,

$$\begin{aligned} r = m - ql &= m - q(sm + tn) \\ &= m - qsm - qtn \\ &= (1 - qs)m + (-qt)n. \end{aligned}$$

It follows that $r = am + bn$ for some $a, b \in \mathbb{Z}$, i.e. r is a linear combination of m and n. Moreover, we know that r is either 0 or positive. If r were positive, then r is a positive linear combination of m and n and $r < l$, but this contradicts the choice of l being the least such integer. Therefore, $r = 0$ implying that $m = ql$ and so $l \mid m$. Similarly, we show that $l \mid n$. Therefore, l is a common divisor of m and n, which implies that $l \leq d$ because d is the greatest common divisor. Since $d \leq l$ and $l \leq d$, we conclude that $d = l = sm + tn$ as desired. ■

We now know that $\gcd(m, n)$ can be expressed as a linear combination of m and n. Next we describe how to accomplish this. The method involves working backward through the iterations of the Euclidean Algorithm.

Example. Use the Euclidean Algorithm to find $\gcd(252, 198)$ and express it as a linear combination of 252 and 198.

Iteration	Computation
1	$252 = 1(198) + 54$
2	$198 = 3(54) + 36$
3	$54 = 1(36) + \mathbf{18}$
4	$36 = 2(18) + 0$

Therefore, $\gcd(252, 198) = 18$. Now, to express 18 as a linear combination of 252 and 198 we start working from the second last iteration (Iteration 3) proceeding our way backward.

$$\begin{aligned}
18 &= 54 - 1(36) && \text{Using Iteration 3}\\
&= 54 - (198 - 3(54)) && \text{Using Iteration 2}\\
&= 54 - 198 + 3(54)\\
&= -198 + 4(54)\\
&= -198 + 4(252 - 198) && \text{Using Iteration 1}\\
&= 4(252) - 5(198).
\end{aligned}$$

Thus, $\gcd(252, 198) = 18 = 4(252) - 5(198)$. ◊

4.4 Fundamental Theorem of Arithmetic

In this section we prove the Fundamental Theorem of Arithmetic. Let us start by showing an important property of prime numbers. We will show that when a prime divides a product of two integers, then it must divide at least one of these integers as well. Note that this property is not true for integers in general, for instance 6 divides the product $3 \cdot 8$ but 6 neither divides 3 nor 8.

Lemma 4.18 (Euclid's Lemma)**.** *Suppose that a and b are integers and p is a prime. If $p \mid ab$, then either $p \mid a$ or $p \mid b$.*

Proof. Let $a, b \in \mathbb{Z}$ and p be a prime number. Suppose that $p \mid ab$. It follows that $ab = kp$ for some integer k. If $p \mid a$, then we are done. Otherwise, we have that p does not divide a, and so we must show that $p \mid b$. As p is prime and $p \nmid a$, the only common divisor between p and a is 1. Thus p and a are coprime, i.e. gcd(p, a)= 1. By Bezout's theorem, we can express 1 as a linear combination of p and a, that is, there exist $s, t \in \mathbb{Z}$ such that $1 = \gcd(p, a) = sp + ta$. We now have that,

$$\begin{aligned} 1 = sp + ta &\iff b = spb + tab \\ &\iff b = spb + t(kp) \\ &\iff b = (sb + tk)p. \end{aligned}$$

Therefore, p divides b and the proof is complete. ■

Using mathematical induction we can generalise the lemma above to show that if a prime p divides a product of integers, then p must divide one of these integers.

Lemma 4.19. Let $a_1, a_2, \ldots, a_n \in \mathbb{Z}$, and p be prime. If $p \mid a_1 a_2 \cdots a_n$, then $p \mid a_i$ for some $1 \leq i \leq n$.

We now have all the ingredients needed to prove the main theorem of this chapter.

Theorem 4.20 (Fundamental Theorem of Arithmetic)**.** *Any integer $n \geq 2$ can be written as a prime or a product of primes. Moreover, such prime factorisation is unique when written in a nondecreasing order.*

Proof. We will prove by strong induction that any integer $n \geq 2$ can be written as a prime or a product of primes. For the base case, the integer 2 is a prime itself. For the induction step, choose any arbitrary integer $n \geq 2$ and assume that any integer k where $2 \leq k \leq n$ can be written as a product of primes. We will show that $n + 1$ can also be written as a prime or a product of primes. Now, if $n + 1$ is a

prime number itself, then we are done. Otherwise, suppose that $n+1$ is a composite number. Thus, there are integers a, b with $2 \leq a, b \leq n$ such that $n+1 = a \cdot b$.
By the induction hypothesis, each of a and b can be written as a prime or product of primes. So we have that $a = p_1 p_2 \cdots p_s$ and $b = q_1 q_2 \cdots q_t$ where each p_i and q_i is a prime number. Thus

$$n + 1 = a \cdot b = p_1 p_2 \cdots p_s q_1 q_2 \cdots q_t.$$

This establishes that $n+1$ is a product of primes proving the induction step. Therefore, by strong induction, we showed that every integer $n \geq 2$ can be written as a prime or a product of primes.
For the uniqueness part, assume for the sake of contradiction that there exists a positive integer $m \geq 2$ which can be written as the product of primes in two different ways, say $m = p_1 p_2 \cdots p_k c_1 c_2 \cdots c_j$ and also $m = q_1 q_2 \cdots q_l c_1 c_2 \cdots c_j$ where each of p_i, q_i, and c_i is a prime number and $\{p_1, \ldots, p_k\} \cap \{q_1, \ldots, q_l\} = \emptyset$, meaning that $c_1, \ldots, c_j$ are the only common primes between the two representations. Moreover, since the two representations are different it must be that $k, l \geq 1$. We now have,

$$p_1 p_2 \cdots p_k c_1 c_2 \cdots c_j = q_1 q_2 \cdots q_l c_1 c_2 \cdots c_j.$$

By removing all the common primes from both sides we get that,

$$p_1 p_2 \cdots p_k = q_1 q_2 \cdots q_l.$$

It follows that the prime p_1 divides the product $q_1 q_2 \cdots q_l$. So for some $1 \leq i \leq l$, we have that $p_1 \mid q_i$. Since both p_1 and q_i are primes, it must be that $p_1 = q_i$, which is a contradiction as we assumed that p_1 is different from every $q_1, \ldots, q_l$. Thus, the prime factorisation of every integer is unique completing the proof of the Fundamental Theorem of Arithmetic. ■

Chapter 5

Combinatorics

Combinatorics is the art of counting. It is the area of mathematics concerned with counting and arrangements of objects.

5.1 Basics of Counting

We present our first counting technique.

Product Rule. Suppose we are asked to perform a mission which can be broken down into a sequence of two tasks. If there are m ways to do the first task, and there are n ways to do the second task, then there are $m \times n$ ways to do the whole mission.

Example. Suppose we have 12 available offices. How many ways are there to assign different offices to Radwa and Mourid?
We will use the Product Rule. So we need to divide the mission of assigning the offices into a sequence of tasks.

- Task 1. Assign an office to Radwa. There are 12 available offices.
- Task 2. Assign an office to Mourid. There are 11 ways to do this since one of the offices was already assigned to Radwa.

By the Product Rule, there are $11 \times 12 = 132$ ways to assign offices to Radwa and Mourid. $\Diamond$

Example. How many words are there whose first part is an uppercase English letter followed by a positive integer not exceeding 10?

To use the Product Rule we need to break down how to create such a word into a sequence of tasks.

- Task 1. Choose an English letter. There are 26 such letters.
- Task 2. Choose a number from $\{1, 2, 3, \ldots, 10\}$. There are 10 choices.

So, by the Product Rule, there are $26 \times 10 = 260$ such words. For instance, $Z4$, $D3$, $Q8$, and $R10$ to name a few. $\Diamond$

Example. How many binary strings of length 5 are there?
We are looking for strings such as 10101 and 01001. To create such a string we need to decide whether each of the five digits is 0 or 1. So there are 5 tasks to complete, where in each task there are 2 different ways to choose from, namely either choose 0 or 1. So, by the Product Rule, the number of binary strings of length 5 is $2 \times 2 \times 2 \times 2 \times 2 = 2^5 = 32$ binary strings of length 5. $\Diamond$

Example. How many different car plates are there in Egypt?
An Egyptian car plate is made of a sequence of 3 Arabic letters followed by 3 decimal digits. Thus, to create a car plate we need to accomplish six tasks: in each of the first three tasks we need to choose a letter from 28 Arabic letters, and in each of the last three tasks we need to choose a number from $0, 1, 2, 3, 4, 5, 6, 7, 8, 9$. By the Product Rule, there are $28 \times 28 \times 28 \times 10 \times 10 \times 10 = 21\,952\,000$ car plates. $\Diamond$

Example. Let A and B be finite sets where $|A| = m$ and $|B| = n$. How many functions $f : A \to B$ are there?
To construct a function $f : A \to B$ we need to assign an image to every element in the domain A. For each element in A, there are n choices in the codomain B from which we choose its image; we are allowed to choose any element from B to be the image. Since there are m elements in A we need to perform a sequence of m tasks, and each one of them can be completed in n ways. Therefore, by the Product Rule, there are $\underbrace{n \times n \times \cdots \times n}_{m \text{ times}} = n^m$ many functions. $\Diamond$

Example. Let A and B be finite sets where $|A| = m$ and $|B| = n$. How many injective functions $f : A \to B$ are there?
In case $m > n$, there are no injective functions. Otherwise assume that $m \leq n$. Let us say that the domain is $A = \{a_1, a_2, \ldots, a_m\}$. To create an injective function $f : A \to B$ we need to assign an image from the codomain B to every element in A where no two distinct elements in A are assigned the same image. Thus we need to perform m many tasks as follows.

- Task 1. Assign an image to the first element a_1 of the domain. There are n choices available in the codomain B.
- Task 2. Assign an image to a_2. There are $(n-1)$ choices available because we cannot choose the image assigned to a_1.
- Task 3. Assign an image to a_3. There are $(n-2)$ choices available because we cannot choose the two images assigned to a_1 and a_2.

$\vdots$ $\vdots$

- Task m. Assign an image to a_m. There are $n-(m-1)$ choices available because we cannot choose any of the $m-1$ images already assigned to the previous elements $a_1, a_2, \ldots, a_{m-1}$.

So, by the Product Rule, there are

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

injective functions from A to B. $\Diamond$

Example. Let $A = \{a, b, c\}$ and $B = \{\alpha, \beta, \gamma, \delta, \epsilon\}$.
There are $5^3 = 125$ functions from A to B. One of these functions is

$$a \mapsto \beta, \ b \mapsto \epsilon, \ c \mapsto \epsilon.$$

There are $5 \times 4 \times 3 = 60$ injective functions from A to B. One of these functions is

$$a \mapsto \delta, \ b \mapsto \gamma, \ c \mapsto \alpha.$$

Write down more examples of these functions. $\Diamond$

The Product Rule can be phrased in terms of sets as follows.

Theorem 5.1. *The cardinality of a Cartesian product of finite sets is equal to the product of the cardinalities of these sets. More precisely, given some finite sets $A_1, A_2, \ldots, A_n$, then*

$$|A_1 \times A_2 \times \cdots \times A_n| = |A_1| \times |A_2| \times \cdots \times |A_n|.$$

To see how the theorem is related to the Product Rule, observe that in order to construct an n-tuple $(a_1, a_2, \ldots, a_n)$ in the set $A_1 \times A_2 \times \cdots \times A_n$ we need to complete n tasks: first we choose an element from A_1, then an element from A_2, and so on, until we choose the n^{th} element of the tuple from A_n. The theorem may be proved by induction where the base case is $n = 2$.

We next introduce another counting technique.

Sum Rule. If a single task can be done either in one of m ways or in one of *different* n ways, then there are $m+n$ ways to complete the task.

Example. How many different choices are there if we are asked to choose a word which is weekday or a month?
Either we choose one day from the 7 weekdays or one month from the 12 months. Since the days are different from the months, then by the sum rule we have $7+12=19$ such words. $\Diamond$

We say that sets $A_1, A_2, \ldots, A_n$ are *pairwise disjoint* if $A_i \cap A_j = \emptyset$ whenever $i \neq j$. The sum rule can be phrased in terms of sets as follows.

Theorem 5.2. *Suppose that $A_1, A_2, \ldots, A_n$ are finite pairwise disjoint sets. Then the cardinality of the union of these sets is equal to the sum of their individual cardinalities. That is,*

$$|A_1 \cup A_2 \cup \cdots \cup A_n| = |A_1| + |A_2| + \cdots + |A_n|.$$

Below we discuss a combinatorial problem where we use both the Product Rule and the Sum Rule.

Example. In a computer system, a password must contain six to eight characters, where each character is an uppercase English letter or a decimal digit. Moreover, each password must contain at least one decimal digit. How many possible passwords are there?
We may choose a password of length 6, 7, or 8. Let N_6, N_7, N_8 denote the number of passwords of length $6, 7$, and 8, respectively. Obviously passwords of length 6 are different from those of length 7 and 8. Similarly for passwords of other lengths. Thus, by the Sum Rule, the total number of passwords is $N_6 + N_7 + N_8$.
Let us compute N_6 which is the number of strings with six characters (letters or decimal digits) which contain at least one decimal digit. For example, $PAL4S8$ is a valid password, however, $EKQZTV$ is not. Observe that N_6 is equal to the total number of strings with six characters (letters or digits) minus the number of strings of length 6 containing no decimal digits at all. Now, to choose a character we have 36 choices (26 English letters plus 10 decimal digits). By the Product Rule, the total number of strings of six characters is 36^6. In the same fashion, the total number of strings of six English letters is 26^6. Thus $N_6 = 36^6 - 26^6$.
Similarly, $N_7 = 36^7 - 26^7$, and $N_8 = 36^8 - 26^8$. Therefore the number of possible passwords is $36^6 + 36^7 + 36^8 - 26^6 - 26^7 - 26^8$. $\Diamond$

Next, we introduce a generalisation of the Sum Rule which takes into account the case when the first set has common elements with the second one.

Inclusion-Exclusion Principle. Let A and B be finite sets each containing items to choose from. If a task can be done by choosing an item from A or B, then the number of different ways to complete the task is

$$|A \cup B| = |A| + |B| - |A \cap B|.$$

Observe that when the items in A are different from B, we have that $|A \cap B| = |\emptyset| = 0$, and so we obtain the Sum Rule.

Example. How many binary strings of length 8 which either start with a 1 bit or end with the two bits 00?
Let A be the set of binary strings of length 8 which start with 1, and B be the set of strings which end with 00. Obviously, the number of the required strings is equal to $|A \cup B|$. To create a string in A we have one choice for the first bit (it must be 1) and two choices (0 or 1) for each of the remaining 7 bits. Thus, by the Product Rule, we have that $|A| = 2^7$. To create an element in B we have two choices for each of the first 6 bits and only one choice (just 0) for each of the 7^{th} and 8^{th} bits, so $|B| = 2^6$. An element in $A \cap B$ must start with 1 and end with 00, and so $|A \cap B| = 2^5$. We now apply the Inclusion-Exclusion Principle to get that $|A \cup B| = |A| + |B| - |A \cap B| = 2^7 + 2^6 - 2^5 = 160$ strings. $\Diamond$

♣ Pigeonhole Principle

We discuss the following wonderful principle.

Theorem 5.3 (Pigeonhole Principle). *Suppose there are n pigeons flying into m pigeonholes in any fashion. If $n > m$, then at least one pigeonhole must have at least two pigeons in it.*

Proof. For the sake of contradiction, suppose that no pigeonhole contains two or more pigeons. This means that each pigeonhole contains one or no pigeons. Since there are m pigeonholes, we have at most m pigeons, and so $n \leq m$. This contradicts that $n > m$. Thus, our initial assumption is false, and there must be one pigeonhole with two pigeons or more. ■

Example. Show that among any group of 367 people, there must be at least two people with the same birthday.
Think of the 367 people as the pigeons, and the different birthdays as the pigeonholes. There are 365 days in a year and so the number of people is strictly more than the number of possible birthdays, it follows by the Pigeonhole Principle that there is at least one birthday which contain at least two people. $\Diamond$

Example. Given any eleven students, at least two of them must have the same last digit in their university ID numbers.
To see this, notice that the last digit could be $0, 1, 2, 3, 4, 5, 6, 7, 8$ or 9. Think of these digits as the pigeonholes, and the students as the pigeons. Since the number of students is strictly greater than the number of options of the last digit of an ID number, by the Pigeonhole Principle at least two students must have the same last digit in their ID numbers. $\Diamond$

Below is a general form of the Pigeonhole Principle. Its proof is left for the reader.

Theorem 5.4 (Generalised Pigeonhole Principle)**.** *Suppose there are n pigeons flying into m pigeonholes in any fashion. Then there is at least one pigeonhole containing at least $\lceil n/m \rceil$ pigeons.*

5.2 Permutations and Combinations

Our first aim is to count the number of sequences of distinct objects coming from a given a set. The second aim is to count the number of subsets of a fixed size of some given set. Let us start with the first aim.

♣ Permutations

Definition. Let A be a set, and let k be a natural number.

- A *k-permutation* of A is a sequence of length k of *distinct* elements from A.
- The number of all k-permutations of a set containing n elements is denoted by ${}^nP_k$ or $P(n, k)$.

Example. Consider the set $A = \{a, b\}$. Here $n = |A| = 2$.

- 1-permutations of A: (a) and (b).
- 2-permutations of A: (a, b) and (b, a).
- Therefore, ${}^2P_1 = 2$ and ${}^2P_2 = 2$.

Now consider the set $B = \{a, b, c\}$. Here $n = |B| = 3$.

- 1-permutations of B: (a), (b), and (c).
- 2-permutations of B: (a, b), (a, c), (b, a), (b, c), (c, a), and (c, b).
- 3-permutations of B: (a, b, c), (a, c, b), (b, a, c), (b, c, a), (c, a, b), and (c, b, a).
- Therefore, ${}^3P_1 = 3$, and ${}^3P_2 = 6$, and ${}^3P_3 = 6$.

Observe that each of these sequences consists of distinct elements, that is, no element is repeated in the same sequence. This condition is required by the definition of a permutation. ◊

Note that ${}^nP_n$ is the number of sequences of length n from a set containing n elements. In other words, ${}^nP_n$ is the number of ways of arranging n objects in a sequence one after the other. From this point of view and using the Product Rule it follows that ${}^nP_n = n!$. Now, let us find a general formula for ${}^nP_k$.

Let n and k be natural numbers with $n \geq k$. When either $n = 0$ or $k = 0$, we have that ${}^nP_k = 1$ since in this case we only have one permutation, namely, the empty sequence of length 0. Next, when $n \geq k \geq 1$, in order to create a k-permutation we need to perform k tasks, namely choosing the first element from the set (we have n choices), choosing the second element which should be different from the first (we have $n-1$ choices), and so on, until we choose the k^{th} element which must be different from the previous $k-1$ elements in the sequence. We now apply the Product Rule to find that

$$^{n}P_k = n \times (n-1) \times (n-2) \times \cdots \times (n-(k-1)).$$

Therefore, for natural numbers n and k with $n \geq k$ we have that

$$^{n}P_k = \frac{n!}{(n-k)!}\,.$$

Let us use permutations to solve some combinatorial problems.

Example. In a contest of 10 people, how many ways are there to get a first prize winner, a second prize winner, and a third prize winner?
These are the sequences of length 3 of distinct people from the contest. In other words, it is the number of 3-permutations from the set of the contestants. Thus, the number of ways to get different winners is

$$^{10}P_3 = \frac{10!}{(10-3)!} = \frac{10!}{7!} = 10 \times 9 \times 8 = 720. \qquad \Diamond$$

Example. How many possible orders one can use to visit four cities?
Let us call the cities A, B, C, D. A trip visiting the four cities can be represented by a 4-permutation of the set of these cities. For instance, the permutation (D, A, B, C) represents a trip which starts at city D, then moves to A, then to B, and finally finishes at city C. Thus, the total number of ways to visit four cities is

$$^{4}P_4 = \frac{4!}{(4-4)!} = \frac{4!}{0!} = 4! = 4 \times 3 \times 2 \times 1 = 24. \qquad \Diamond$$

Example. How many strings of 4 distinct English letters are there?
Such a string is a 4-permutation from the set of English alphabet which consists of 26 letters. Thus, the number of such strings is

$$P(26,4) = \frac{26!}{(26-4)!} = \frac{26!}{22!} = 26 \times 25 \times 24 \times 23 = 358\,800. \qquad \Diamond$$

Example. How many permutations of the letters A, B, C, D contain the string BC?
The trick here is to think of BC as one object, thus the problem boils down to finding the number of 3-permutations of a set containing 3 objects, namely A and (BC) and D. Therefore, the number of such strings is

$$P(3,3) = 3! = 6.$$

Here are they: $ABCD$, $ADBC$, $BCAD$, $BCDA$, $DABC$, and $DBCA$. $\Diamond$

♣ Combinations

We have seen permutations which are ordered collections of objects, we now turn our attention to count the number of unordered collections (i.e. subsets) of objects. Towards this aim, we discuss the following problem.

In a class of four students: A, B, C, D, the teacher is asked to choose a group of three students. How many different groups can the teacher form from this class? The possible groups are $\{A, B, C\}$, $\{A, B, D\}$, $\{A, C, D\}$, and $\{B, C, D\}$. So there are four different groups of three students that can be chosen by the teacher. Observe that these are the subsets with three elements from the set $\{A, B, C, D\}$. Of course, the teacher is concerned with the group as a whole disregarding any ordering of the chosen students. These groups of three students are called 3-combinations.

Definition. Let A be a set, and let k be a natural number.

- A *k-combination* of a set A is a subset of A containing exactly k elements.
- The number of all k-combinations of a set with n elements is denoted by ${}^nC_k$ or $C(n, k)$.

This means that ${}^nC_k$ is the number of subsets with exactly k elements from a set containing n elements. Simply, ${}^nC_k$ is the number of ways of choosing k objects from n objects. We read ${}^nC_k$ as "n choose k".

Example. Compute ${}^4C_2$.
Consider any set of 4 elements, say $A = \{1, 2, 3, 4\}$. The are six 2-combinations of A, namely: $\{1, 2\}$, $\{1, 3\}$, $\{1, 4\}$, $\{2, 3\}$, $\{2, 4\}$, and $\{3, 4\}$. Thus ${}^4C_2 = 6$. ♢

Let us find a formula to compute ${}^nC_k$. In other words, we want to find how many subsets of cardinality k of a set of cardinality n are there.

Theorem 5.5. Let n and k be natural numbers with $n \geq k$. Then

$${}^nC_k = \frac{n!}{(n-k)! \times k!}.$$

Proof. Let A be a set with n elements and fix some k-combination X of A. Since $X \subseteq A$, it follows that every k-permutation of X is also a k-permutation of A. Since $|X| = k$, there are ${}^kP_k = k!$ many k-permutations of X. This means that every k-combination of A gives rise to $k!$ many k-permutations of A. Thus, the total number of k-permutations of A is equal to ${}^nC_k \times k!$. That is, ${}^nP_k = {}^nC_k \times k!$. Therefore,

$${}^nC_k = \frac{{}^nP_k}{k!} = \frac{n!}{(n-k)! \times k!}.$$

And the proof is complete. ∎

Example. How many football teams we can form from 15 people?
A football team needs 11 players. So the number of football teams we can form from 15 people is equal to the number of subsets with 11 people from the set of these 15 people. Said differently, it is the number of ways of choosing 11 people from 15 people. Thus, there are

$$^{15}C_{11} = \frac{15!}{(15-11)! \times 11!} = \frac{15!}{4! \times 11!} = \frac{15 \times 14 \times 13 \times 12}{4 \times 3 \times 2} = 1365 \text{ teams.} \qquad \Diamond$$

Example. How many binary strings of length 5 contain exactly three 1s?
Our strategy will adopt the Product Rule. To construct such a string the mission is to fill 5 places with 0s and 1s such that exactly 3 1s are used. We will break down this into a sequence of tasks.

- Task 1. Choose 3 places from the 5 places and fill them with 1s. To complete this task we have $^5C_3$ ways.
- Task 2. Fill the remaining two places with 0s. There is only one way to complete this step.

Therefore, there are $^5C_3 \times 1 = \frac{5!}{2! \times 3!} = 10$ binary strings of length 5 containing exactly three 1s. $\Diamond$

Example. There are 9 faculty members in the Mathematics Department and 11 faculty members in the Computer Science Department. How many ways are there to select a committee made of 3 mathematicians and 4 computer scientists?
We will divide the mission of forming the committee into a sequence of tasks.

- Task 1. Choose 3 mathematicians. There are $^9C_3$ ways to do this.
- Task 2. Choose 4 computer scientists. There are $^{11}C_4$ ways to do this.

Thus, to select the desired committee there are

$$^9C_3 \times {}^{11}C_4 = \frac{9!}{6! \times 3!} \times \frac{11!}{7! \times 4!} = 27\,720 \text{ ways.} \qquad \Diamond$$

Example. Let A be a finite set with n elements. The cardinality of any subset of A is either $0, 1, 2, \dots, n$. Thus $|\mathcal{P}(A)|$ which is the number of all subsets of A is equal to the number of subsets with no elements plus the number of subsets with 1 element plus the number of subsets with 2 elements, keeping the summation up to adding the number of subsets with n elements. Since $|\mathcal{P}(A)| = 2^n$, we therefore obtain the following beautiful equality.

$$^nC_0 + {}^nC_1 + {}^nC_2 + \cdots + {}^nC_n = 2^n$$

Chapter 6

Matrices

In this chapter we give a brief introduction to matrices. A more detailed treatment of matrices is covered in a linear algebra course where matrices are used in solving systems of linear equations and they also provide examples of vector spaces (A vector space is a mathematical object consisting of a set of elements together with two operations called “addition” and “scalar multiplication” satisfying a certain list of axioms). Apart from mathematics, matrices have wide applications in physics, numerical analysis, economics, Google search, and computer graphics. The numerical computing software MATLAB (abbreviation for “Matrix Laboratory”) uses matrices extensively.

6.1 Real Matrices

Definition. An $m \times n$ *matrix* is a rectangular array of real numbers with m rows and n columns, where m, n are positive integers. In this case, we say that $m \times n$ is the *dimension* or the *size* of the matrix.

The plural of “matrix” is “matrices”. When the number of rows of a matrix is equal to the number of columns, we say it is a *square matrix.* An $n \times n$ matrix is called a square matrix of order n.

Example.

- Matrix A is a 2×2 matrix. It is a square matrix of order 2.

$$A = \begin{bmatrix} 1 & 2 \\ 5 & 7 \end{bmatrix}.$$

- Matrix B is a 3×2 matrix.

$$B = \begin{bmatrix} 1 & 4 \\ 6 & 2 \\ 7 & 8 \end{bmatrix}.$$

- Matrix C is a 1×3 matrix.

$$C = \begin{bmatrix} 4 & 6 & 8 \end{bmatrix}. \qquad \lozenge$$

We represent a matrix A of dimension $m \times n$ in the following way.

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \dots & a_{mn} \end{bmatrix}.$$

For a shortcut notation of the above representation we write,

$$A = [a_{ij}].$$

The entry a_{ij} is located in the i^{th} row and j^{th} column of the matrix A where $1 \leq i \leq m$ and $1 \leq j \leq n$. We say that a_{ij} is the (i,j)-entry. For example, a_{23} is the entry in the 2^{nd} row and 3^{rd} column. When $A = [a_{ij}]$ is a square matrix, the entries $a_{11}, a_{22}, a_{33}, \dots, a_{nn}$ are the entries which lie on the main diagonal of A.

♣ Matrix Arithmetic

Imagine the world of all matrices. Our goal in the world of matrices is to define matrix arithmetic, such as addition and multiplication, in a similar way to what we have in the world of real numbers. But first we need to define when two matrices are considered equal.

Definition. Two matrices are *equal* if and only if they have the same dimension and the corresponding entries in every position are equal.

Let us elaborate more. Suppose that $A = [a_{ij}]$ and $B = [b_{ij}]$ are matrices of the same dimension $m \times n$. Then we say that $A = B$ when $a_{ij} = b_{ij}$ for each i, j where $1 \leq i \leq m$ and $1 \leq j \leq n$.

Example.

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix} \neq \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 2 & 3 \\ 2 & 3 \end{bmatrix} \neq \begin{bmatrix} 2 & 3 \\ 3 & 3 \end{bmatrix}. \qquad \lozenge$$

♣ Matrix Addition

Let $A = [a_{ij}]$ and $B = [b_{ij}]$ be $m \times n$ matrices. The *sum* of A and B is the $m \times n$ matrix $A + B$ whose (i, j)-entry is obtained by adding the corresponding entries a_{ij} and b_{ij} in A and B, respectively. That is,

$$A + B = [c_{ij}] \text{ where } c_{ij} = a_{ij} + b_{ij}.$$

Two matrices can be added only when they have the same number of rows and columns. Otherwise, if they have different dimensions then their addition is undefined.

Example. Addition of two 2 × 2 matrices.

$$\begin{bmatrix} 1 & 2 \\ 3 & 7 \end{bmatrix} + \begin{bmatrix} 0 & -2 \\ -3 & -6 \end{bmatrix} = \begin{bmatrix} 1+0 & 2-2 \\ 3-3 & 7-6 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \qquad \lozenge$$

Example. Addition of two 3 × 3 matrices.

$$\begin{bmatrix} 1 & 0 & 2 \\ 4 & 5 & 6 \\ -1 & -2 & 3 \end{bmatrix} + \begin{bmatrix} 4 & 5 & 7 \\ -1 & -3 & 2 \\ 0 & 1 & -1 \end{bmatrix} = \begin{bmatrix} 1+4 & 0+5 & 2+7 \\ 4-1 & 5-3 & 6+2 \\ -1+0 & -2+1 & 3-1 \end{bmatrix} = \begin{bmatrix} 5 & 5 & 9 \\ 3 & 2 & 8 \\ -1 & -1 & 2 \end{bmatrix}. \ \lozenge$$

An $m \times n$ matrix whose all of its entries are zeros is denoted by $\mathbf{0}_{m \times n}$ or just $\mathbf{0}$ if the dimension is known. For example,

$$\mathbf{0}_{2\times 3} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Theorem 6.1. *Matrix addition is commutative and associative. Moreover, the zero matrix is the additive identity. More precisely, given any matrices A, B, C of the same dimension, the following hold.*

(i) $A + B = B + A$.

(ii) $(A+B)+C = A+(B+C)$.

(iii) $A + \mathbf{0} = A$.

Proof. We prove the first one and leave the other two for the reader. Suppose that $A = [a_{ij}]$ and $B = [b_{ij}]$ are matrices of the same dimension. Then using the definition of matrix addition and the fact that addition of real numbers is commutative we have the following.

$$A + B = [a_{ij} + b_{ij}] = [b_{ij} + a_{ij}] = B + A.$$

Thus $A + B = B + A$. So we have shown that matrix addition is commutative. ■

♣ Matrix Multiplication

Let $A = [a_{ij}]$ be an $m \times k$ matrix and $B = [b_{ij}]$ be an $k \times n$ matrix. Their *product* AB is the $m \times n$ matrix whose (i, j)-entry is equal to the sum of the products of the corresponding entries from the i^{th} row of A and j^{th} column of B.

More precisely, let c_{ij} denote the (i, j)-entry in the product, so $AB = [c_{ij}]$, then

$$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{ik}b_{kj} = \sum_{r=1}^{k} a_{ir}b_{rj}.$$

$$\begin{array}{cccc}
A & B & & AB \\
\begin{bmatrix}
a_{11} & a_{12} & \cdots & a_{1k} \\
a_{21} & a_{22} & \cdots & a_{2k} \\
\vdots & \vdots & \ddots & \vdots \\
\mathbf{a_{i1}} & \mathbf{a_{i2}} & \cdots & \mathbf{a_{ik}} \\
\vdots & \vdots & \ddots & \vdots \\
a_{m1} & a_{m2} & \cdots & a_{mk}
\end{bmatrix}
&
\begin{bmatrix}
b_{11} & b_{12} & \cdots & \mathbf{b_{1j}} & \cdots & b_{1n} \\
b_{21} & b_{22} & \cdots & \mathbf{b_{2j}} & \cdots & b_{2n} \\
b_{31} & b_{32} & \cdots & \mathbf{b_{3j}} & \cdots & b_{3n} \\
\vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
b_{k1} & b_{k2} & \cdots & \mathbf{b_{kj}} & \cdots & b_{kn}
\end{bmatrix}
& = &
\begin{bmatrix}
c_{11} & \cdots & c_{1j} & \cdots & c_{1n} \\
c_{21} & \cdots & c_{2j} & \cdots & c_{2n} \\
\vdots & \ddots & \vdots & \ddots & \vdots \\
c_{i1} & \cdots & \mathbf{c_{ij}} & \cdots & c_{in} \\
\vdots & \ddots & \vdots & \ddots & \vdots \\
c_{m1} & \cdots & c_{mj} & \cdots & c_{mn}
\end{bmatrix}
\end{array}$$

The slogan for matrix multiplication is “row × column” since the (i, j)-entry of the product AB is calculated using the i^{th} row of A and j^{th} column of B. The product AB is defined only when the number of columns of A is equal to the number of rows of B. Otherwise, we say that the product is undefined.

Example. Compute the product of the following matrices.

$$A = \begin{bmatrix} 1 & 0 & 4 \\ 2 & 1 & 1 \\ 3 & 1 & 0 \\ 0 & 2 & 2 \end{bmatrix} \text{ and } B = \begin{bmatrix} 2 & 4 \\ 1 & 1 \\ 3 & 0 \end{bmatrix}.$$

Matrix A has dimension 4×3 and B has dimension 3×2. Since the number of columns of A is equal to the number of rows of B, the product AB is defined and has dimension 4×2. Let $AB = [c_{ij}]$. To compute c_{11} we sum the products of the corresponding elements from the 1^{st} row of A and the 1^{st} column of B. So

$$c_{11} = (1 \cdot 2) + (0 \cdot 1) + (4 \cdot 3) = 2 + 0 + 12 = 14.$$

To compute c_{12} we sum the products of the corresponding elements from the 1^{st} row of A and the 2^{nd} column of B. So

$$c_{12} = (1 \cdot 4) + (0 \cdot 1) + (4 \cdot 0) = 4 + 0 + 0 = 4.$$

Similarly we compute c_{32} using the 3^{rd} row of A and the 2^{nd} column of B.

$$c_{32} = (3 \cdot 4) + (1 \cdot 1) + (0 \cdot 0) = 12 + 1 + 0 = 13.$$

We continue in this manner until computing all the entries of AB.

$$AB = \begin{bmatrix} 1 & 0 & 4 \\ 2 & 1 & 1 \\ 3 & 1 & 0 \\ 0 & 2 & 2 \end{bmatrix}_{4\times 3} \begin{bmatrix} 2 & 4 \\ 1 & 1 \\ 3 & 0 \end{bmatrix}_{3\times 2} = \begin{bmatrix} 14 & 4 \\ 8 & 9 \\ 7 & 13 \\ 8 & 2 \end{bmatrix}_{4\times 2}.$$

Note that BA is undefined as the number of columns of B (2 columns in B) is not equal to the number of rows of A (4 rows in A). $\Diamond$

When A and B are square matrices of the same order then both AB and BA are defined, however, they are not necessarily equal as the next example demonstrates.

Example. Consider the matrices $A = \begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix}$ and $B = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}$. We have that

$$AB = \begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 2 \\ 5 & 3 \end{bmatrix}, \text{ while } BA = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix} = \begin{bmatrix} 4 & 3 \\ 3 & 2 \end{bmatrix}.$$

Observe that $AB \neq BA$. In general, matrix multiplication is not commutative. $\Diamond$

Theorem 6.2. *Matrix multiplication is associative. Moreover, the distributive property holds for matrices. More precisely, given matrices A, B, C of suitable dimensions, the following hold.*

(i) $(AB)C = A(BC)$.

(ii) $A(B + C) = AB + AC$.

(iii) $(B + C)A = BA + CA$.

Exercise. Write down a clear proof of the theorem above. To prove these equalities, one needs to show that every entry in the matrix on the left hand side is equal to its corresponding entry in the matrix on the right hand side.

In the world of real numbers, the number 1 plays the role of multiplicative identity in the sense that if we multiply 1 by any nonzero real number we get the number itself. In matrix multiplication, which matrix plays the role of multiplicative identity?

Definition. The *identity matrix* I_n of order n is the $n \times n$ matrix $I_n = [\delta_{ij}]$ where

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases}$$

It follows that the identity matrix of order n is an $n \times n$ matrix whose main diagonal entries are 1s and all the remaining entries are 0s.

$$I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \qquad I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad I_n = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}.$$

The identity matrix plays the role of the multiplicative identity in the world of matrices as the next lemma states. The proof is left for the reader.

Lemma 6.3. *Let A be an $m \times n$ matrix. Then*

$$AI_n = A \quad \textit{and} \quad I_m A = A.$$

Definition (Power of a Matrix)**.** Let A be an $n \times n$ matrix. We define

- $A^0 = I_n$,
- $A^{k+1} = A^k A$ for $k \geq 0$.

It follows that $A^1 = A^0A = I_nA = A$, and $A^2 = A^1A = AA$, and $A^3 = A^2A = AAA$, and $A^4 = A^3A = AAAA$, and so on. In general, for positive k we have that

$$A^k = \underbrace{AAA \cdots A}_{k \text{ times}}.$$

Theorem 6.4. *Let A be a square matrix of order n. Then the following hold where $m, k \in \mathbb{N}$.*

(i) $A^mA^k = A^{m+k}$

(ii) $(A^m)^k = A^{mk}$

Proof. We prove the first equality and leave the second for the reader. Fix an arbitrary natural number m. We aim to prove that $A^mA^k = A^{m+k}$ by induction on k. For the base case, using the definition $A^0 = I_n$ and the fact that I_n is a multiplicative identity, we have that,

$$A^mA^0 = A^mI_n = A^m = A^{m+0}.$$

For the induction step, suppose that $A^mA^k = A^{m+k}$ is true, our goal is to prove that $A^mA^{k+1} = A^{m+(k+1)}$. Using the definition of the power of a matrix, and the associativity of matrix multiplication, and the induction hypothesis we have that,

$$A^mA^{k+1} = A^m(A^kA) = (A^mA^k)A \overset{IH}{=} A^{m+k}A = A^{m+k+1}.$$

Thus, the induction step holds and the proof is complete. ■

♣ Matrix Transpose

Definition. Let $A = [a_{ij}]$ be an $m \times n$ matrix. The *transpose* of A, denoted by A^T, is the $n \times m$ matrix $A^T = [\hat{a}_{ij}]$ where $\hat{a}_{ij} = a_{ji}$ for each $1 \leq i \leq n$ and $1 \leq j \leq m$. In other words, A^T is the matrix whose columns are the rows of A.

Example. Consider the 2×3 matrix $A = [a_{ij}] = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$. Then its transpose is the 3×2 matrix,

$$A^T = [\hat{a}_{ij}] = \begin{bmatrix} \hat{a}_{11} & \hat{a}_{12} \\ \hat{a}_{21} & \hat{a}_{22} \\ \hat{a}_{31} & \hat{a}_{32} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \\ a_{13} & a_{23} \end{bmatrix} = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}. \qquad \Diamond$$

Definition. A square matrix A is *symmetric* if $A = A^T$.

Observe that if $A = [a_{ij}]$ is a symmetric matrix of order n, then $a_{ij} = a_{ji}$ for every $1 \leq i, j \leq n$. This means that the main diagonal acts as a mirror where every entry on the left hand side of the main diagonal is equal to its mirror image on the right hand side of the main diagonal.

Example. The following matrices are symmetric. Notice the symmetry across the main diagonal.

$$A = \begin{bmatrix} 1 & 4 & 5 \\ 4 & 2 & 7 \\ 5 & 7 & 3 \end{bmatrix}, \qquad B = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \qquad C = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 1 & 4 \\ 3 & 4 & 0 \end{bmatrix}. \qquad \Diamond$$

Theorem 6.5. *Let A, B be matrices of suitable dimensions.*

(i) $(A^T)^T = A$.

(ii) $(A + B)^T = A^T + B^T$.

(iii) $(AB)^T = B^T A^T$.

6.2 Boolean Matrices

Definition. A *Boolean matrix* is a matrix where each of its entries is either 0 or 1.

There are interesting operations on Boolean matrices called the *meet operation*, the *join operation*, and the *Boolean product*. In order to define these matrix operations we need to introduce the meet and the join functions. The *meet function* is the function $\wedge : \{0,1\} \times \{0,1\} \to \{0,1\}$ given by the following table. Notice that the domain of the meet function is $\{(0,0),(0,1),(1,0),(1,1)\}$ and the codomain is $\{0,1\}$. We denote the image $\wedge(x,y)$ by the notation $x \wedge y$.

x	y	$x \wedge y$
1	1	1
1	0	0
0	1	0
0	0	0

The *join function* is the function $\vee : \{0,1\} \times \{0,1\} \to \{0,1\}$ given by the following table. As before, we denote the image $\vee(x,y)$ by $x \vee y$.

x	y	$x \vee y$
1	1	1
1	0	1
0	1	1
0	0	0

We use the meet and join functions to define matrix operations on Boolean matrices.

Let $A = [a_{ij}]$ and $B = [b_{ij}]$ be $m \times n$ Boolean matrices (notice that each of a_{ij} and b_{ij} is either 0 or 1). We define the *meet* of A and B to be the $m \times n$ Boolean matrix $A \wedge B$ whose (i,j)-entry is the meet of the corresponding entries in A and B.

$$A \wedge B = [a_{ij} \wedge b_{ij}].$$

Similarly, the *join* of A and B is the $m \times n$ Boolean matrix $A \vee B$ whose (i,j)-entry is the join of the corresponding entries in A and B.

$$A \vee B = [a_{ij} \vee b_{ij}].$$

Example. Consider the 2×3 matrices $A = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$ and $B = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 1 & 0 \end{bmatrix}$. Their meet is

$$A \wedge B = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \wedge \begin{bmatrix} 0 & 1 & 0 \\ 1 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 \wedge 0 & 0 \wedge 1 & 1 \wedge 0 \\ 0 \wedge 1 & 1 \wedge 1 & 0 \wedge 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}.$$

And their join is

$$A \vee B = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \vee \begin{bmatrix} 0 & 1 & 0 \\ 1 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 \vee 0 & 0 \vee 1 & 1 \vee 0 \\ 0 \vee 1 & 1 \vee 1 & 0 \vee 0 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \end{bmatrix}. \qquad \diamondsuit$$

Definition. Let $A = [a_{ij}]$ be an $m \times k$ Boolean matrix, and $B = [b_{ij}]$ be an $k \times n$ Boolean matrix. The *Boolean product* $A \odot B$ is the $m \times n$ Boolean matrix whose (i, j)-entry is equal to

$$(a_{i1} \wedge b_{1j}) \vee (a_{i2} \wedge b_{2j}) \vee \cdots \vee (a_{ik} \wedge b_{kj}).$$

Example. Consider the matrices $A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}$. Since the number of columns of A is equal to the number of rows of B, the Boolean product is defined.

$$\begin{aligned} A \odot B &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} \odot \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \\ &= \begin{bmatrix} (1 \wedge 1) \vee (0 \wedge 0) & (1 \wedge 1) \vee (0 \wedge 1) & (1 \wedge 0) \vee (0 \wedge 1) \\ (0 \wedge 1) \vee (1 \wedge 0) & (0 \wedge 1) \vee (1 \wedge 1) & (0 \wedge 0) \vee (1 \wedge 1) \\ (1 \wedge 1) \vee (0 \wedge 0) & (1 \wedge 1) \vee (0 \wedge 1) & (1 \wedge 0) \vee (0 \wedge 1) \end{bmatrix} \\ &= \begin{bmatrix} 1 \vee 0 & 1 \vee 0 & 0 \vee 0 \\ 0 \vee 0 & 0 \vee 1 & 0 \vee 1 \\ 1 \vee 0 & 1 \vee 0 & 0 \vee 0 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \end{bmatrix}. \qquad \diamondsuit \end{aligned}$$

Exercise. Prove that the Boolean product of matrices is associative. That is, show that for any Boolean matrices A, B, C with suitable dimensions we have that

$$(A \odot B) \odot C = A \odot (B \odot C).$$

Definition (Boolean Power)**.** Let A be a square Boolean matrix. We define the Boolean power of A as follows.

- $A^{[1]} = A$,
- $A^{[k+1]} = A^{[k]} \odot A$ for $k \geq 1$.

As the Boolean product of matrices is associative, it follows that the k^{th} Boolean power of A is

$$A^{[k]} = \underbrace{A \odot A \odot A \odot \cdots \odot A}_{k \text{ times}}.$$

Warning. There is a difference between A^k and $A^{[k]}$. The former is computed using standard matrix multiplication, while the latter using Boolean product.

Example. Compute all Boolean powers of $A = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \end{bmatrix}$.

$$A^{[2]} = A \odot A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix}.$$

$$A^{[3]} = A^{[2]} \odot A = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}.$$

$$A^{[4]} = A^{[3]} \odot A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix}.$$

$$A^{[5]} = A^{[4]} \odot A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}.$$

Consequently, for each $k \geq 5$ we have that,

$$A^{[k]} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}.$$

◊

Chapter 7

Relations

We will study relationships between elements of sets. We have encountered many relations in the previous chapters. For instance, we have the order relation between real numbers, the divisibility and congruence relations between integers, the logical equivalence relation between compound propositions, and the subset relation between sets. In this chapter, we define in a precise way the concept of a "relation" and study several interesting properties of relations such as reflexivity, symmetry, and transitivity.

7.1 Relations

Recall that the Cartesian product $A \times B$ of sets A and B is the set of all ordered pairs (a, b) where $a \in A$ and $b \in B$.

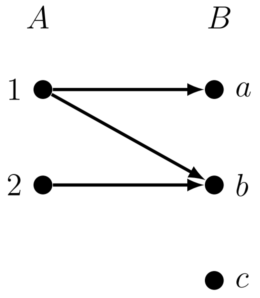
Definition. A *relation* from a set A to a set B is a subset of $A \times B$.

Suppose that R is a relation from A to B. By definition, this means that R is a set such that $R \subseteq A \times B$. And so R is a set of ordered pairs (a, b) where $a \in A$ and $b \in B$. When a pair (a, b) belongs to the relation R, we say that "*a is related to b by R*", and we write $(a, b) \in R$ or aRb or $R(a, b)$ or $a \to b$. On the other hand, if $(a, b) \notin R$, then a is not related to b and we write $a \not R b$.

Definition. A relation R on A is a relation from A to A. So R is a relation on A if and only if $R \subseteq A \times A$.

Example. Let $A = \{1, 2\}$ and $B = \{a, b, c\}$. One example of a relation from A to B is $R = \{(1, a), (1, b), (2, b)\}$. Notice that R is a subset of $A \times B$.

So 1 is related to a, and 1 is related to b, and 2 is related to b by the relation R, and no other elements are related. We can also write $1Ra$, $1Rb$, and $2Rb$. Alternatively, we write $R(1,a)$, $R(1,b)$, $R(2,b)$. Note that $1\not{R}c$, $2\not{R}a$, and $2\not{R}c$.
When using arrows to represent the relation R we obtain the following diagram. Such an arrow diagram is also called a *directed graph*, or *digraph.*

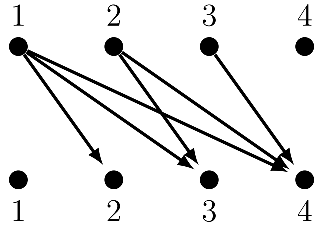


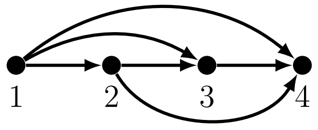
◊

Example. Let $A = \{1,2,3,4\}$. The strictly-less-than relation $R_<$ on A is given by

$$R_< = \{(1,2),(1,3),(1,4),(2,3),(2,4),(3,4)\} \subseteq A \times A.$$

For simplicity, let us use $<$ instead of $R_<$. Then the relation says that $1 < 2$, $1 < 3$, $1 < 4$, $2 < 3$, $2 < 4$, and $3 < 4$. As this relation is a relation from A to A, when drawing its arrow diagram we have two options, one where we use two copies of A (as shown below on the left), and the other using a single copy of A (as shown on the right).





◊

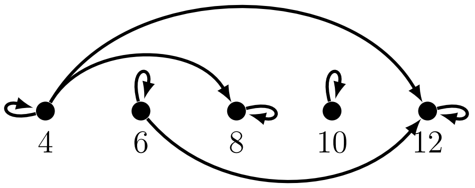
Example. Let A and B be sets. Both the empty set and $A \times B$ are obvious examples of relations from A to B since both are subsets of $A \times B$. When the relation is $\emptyset$, no element in A is related to any element in B. However, when the relation is $A \times B$, every element in A is related to every element in B. ◊

Example. Let M be the set of all men, and W be the set of all women. We can express the marriage relation between people using the relation $R \subseteq M \times W$ where

$$R = \{(m,w) \mid m \in M, w \in W, \text{ and } m \text{ is married to } w\}.$$ ◊

Example. Write down the divisibility relation D on the set $A = \{4, 6, 8, 10, 12\}$ and express it using an arrow diagram.
We have that aDb if and only if $a \mid b$ where $a, b \in A$. The relation is

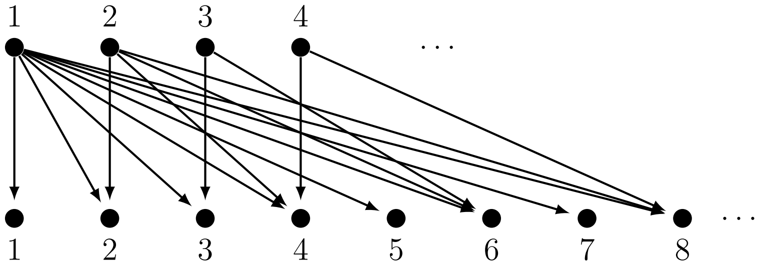
$$D = \{(4,4), (4,8), (4,12), (6,6), (6,12), (8,8), (10,10), (12,12)\} \subseteq A \times A.$$



Example. We also consider the divisibility relation on the set of positive integers.

$$D = \{(m,n) \mid m, n \in \mathbb{Z}^+ \text{ and } m \mid n\}.$$

Clearly D is a relation on $\mathbb{Z}^+$ since $D \subseteq \mathbb{Z}^+ \times \mathbb{Z}^+$. Moreover, D has infinitley many elements. For instance we have that $1D3$, $2D18$, $4D16$, $5D35$, $7D49$, $9D27$, $6D36$, $5\not{D}22$, and $10\not{D}95$. Below is the beginning of the arrow diagram of this relation.



◊

Example. Let $f : A \to B$ be a function. Then,

$$\text{Graph}(f) = \{(a,b) \mid a \in A,\ b \in B, \text{ and } f(a) = b\}$$

is a relation from A to B because $\text{Graph}(f) \subseteq A \times B$. So every function can be seen as a relation through its graph, but not every relation is a function. ◊

Example. Consider the relation C on $\mathbb{R}$ where

$$C = \{(x,y) \in \mathbb{R}^2 \mid x^2 + y^2 = 1\}.$$

So a real number x is related to y by the relation C if $x^2 + y^2 = 1$. In other words, xCy if and only if the point (x,y) lies on the unit circle in the Cartesian plane. ◊

Example. Let A be a set with $|A| = n$. How many relations are there on A? Since a relation on A is a subset of $A \times A$, the number of relations on A is equal to the number of subsets of $A \times A$. We know that $|A \times A| = n^2$, and so the number of relations on A is equal to

$$|\mathcal{P}(A \times A)| = 2^{|A \times A|} = 2^{(n^2)}.$$

For instance, on the set $\{a, b\}$ there are $2^{(2^2)} = 2^4 = 16$ relations. One of them is the relation $\{(a, b), (b, b)\}$. List them all as an exercise. $\Diamond$

♣ Properties of Relations

Definition. Let R be a relation on a set A. Moreover, let x, y, z be variables with domain A.

- The relation R is *reflexive* if $\forall x\,(xRx)$.
- The relation R is *irreflexive* if $\forall x\,\neg(xRx)$.
- The relation R is *symmetric* if $\forall x\,\forall y\,(xRy \to yRx)$.
- The relation R is *asymmetric* if $\forall x\,\forall y\,(xRy \to \neg(yRx))$.
- The relation R is *antisymmetric* if $\forall x\,\forall y\,(x \neq y \wedge xRy \to \neg(yRx))$.
- The relation R is *transitive* if $\forall x\,\forall y\,\forall z\,(xRy \wedge yRz \to xRz)$.

We explain these properties for a relation R on a set A.

- R being reflexive means that every element of A is related to itself. In the arrow diagram of a reflexive relation there is an arrow from every element to itself (a *loop*).
- R being irreflexive means that no element of A is related to itself. In the arrow diagram of an irreflexive relation there is no arrow from any element to itself; there are no loops whatsoever.
- R being symmetric means that if an element a is related to b, then b is also related to a. In the arrow diagram of a symmetric relation whenever there is an arrow, then there is an arrow in the opposite direction.
- R being asymmetric means that if an element a is related to b, then b is not related to a. In the arrow diagram of an asymmetric relation whenever there is an arrow, then there is no arrow in the opposite direction. Note that this implies that no element can be related to itself, and so an asymmetric relation is also irreflexive.

- R being antisymmetric means that if an element a is related to a *different* element b, then b is not related to a. Here we do not care whether an element is related to itself or not. In the arrow diagram of an antisymmetric relation whenever there is an arrow between *distinct* elements, then there is no arrow in the opposite direction. Here, it is fine if an element is related to itself.

- R being transitive means that if an element a is related to b, and b is related to c, then a is related to c. In the arrow diagram of a transitive relation whenever there are two consecutive arrows ($a \to b \to c$), then there must be an arrow from the first element to the third element ($a \longrightarrow c$).

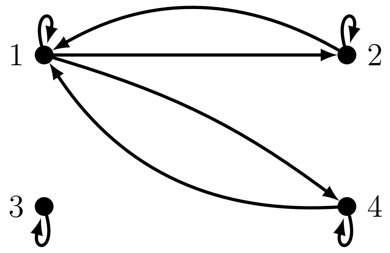
Exercise. Prove that a relation R is asymmetric if and only if R is irreflexive and antisymmetric. Moreover, show that R being antisymmetric is equivalent to

$$\forall x \, \forall y \, (xRy \wedge yRx \to x = y).$$

Example. Consider the following relations on $A = \{1, 2, 3, 4\}$.

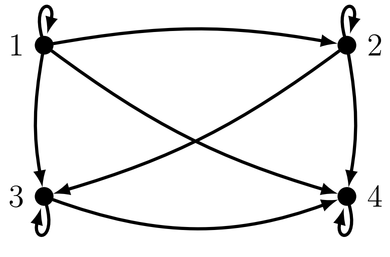
(a) The relation R on A given below is reflexive and symmetric. It is neither irreflexive, nor asymmetric, nor antisymmetric, nor transitive.

$$R = \{(1,1), (1,2), (1,4), (2,1), (2,2), (3,3), (4,1), (4,4)\}.$$



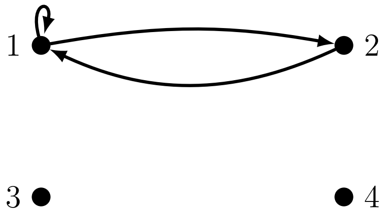
(b) The relation S on A given below is reflexive, antisymmetric, and transitive. It is neither irreflexive, nor symmetric, nor asymmetric.

$$S = \{(1,1), (1,2), (1,3), (1,4), (2,2), (2,3), (2,4), (3,3), (3,4), (4,4)\}.$$



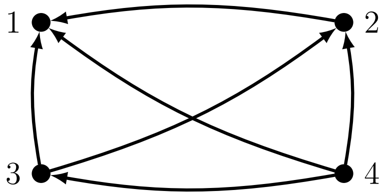
(c) The relation T on A given below is only symmetric. It is neither reflexive, nor irreflexive, nor asymmetric, nor antisymmetric, nor transitive.

$$T = \{(1,1), (1,2), (2,1)\}.$$

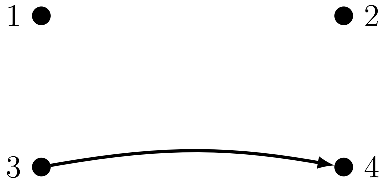


(d) The relation U on A given below is irreflexive, asymmetric, antisymmetric and transitive. It is not reflexive nor symmetric.

$$U = \{(2,1),(3,1),(3,2),(4,1),(4,2),(4,3)\}.$$



(e) The relation $V = \{(3,4)\}$ on A is irreflexive, asymmetric, antisymmetric, and transitive. It is not reflexive, nor symmetric.



◊

Example. Check that the following relations on $\mathbb{Z}$ satisfy the mentioned properties.

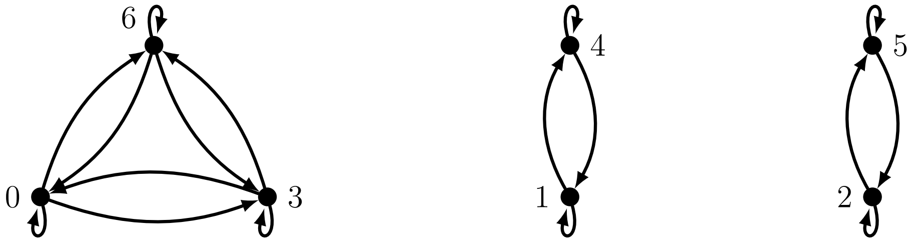
- $L = \{(a,b) \in \mathbb{Z}^2 \mid a \leq b\}$. Reflexive, antisymmetric, transitive.
- $G = \{(a,b) \in \mathbb{Z}^2 \mid a > b\}$. Irreflexive, asymmetric, antisymmetric, transitive.
- $A = \{(a,b) \in \mathbb{Z}^2 \mid a = b \text{ or } a = -b\}$. Reflexive, symmetric, transitive.
- $E = \{(a,b) \in \mathbb{Z}^2 \mid a = b\}$. Reflexive, symmetric, antisymmetric, transitive.
- $S = \{(a,b) \in \mathbb{Z}^2 \mid a = b+1\}$. Irreflexive, asymmetric, antisymmetric.
- $T = \{(a,b) \in \mathbb{Z}^2 \mid a+b \leq 3\}$. Symmetric. ◊

Example. Check that the divisibility relation on the set of positive integers is reflexive, antisymmetric, and transitive. For transitivity see Theorem 4.1(iii). ◊

Example. Let $R = \{(x,y) \mid x,y \in \mathbb{Z}, \text{ and } y-x \text{ is even}\}$.
Then R is reflexive, symmetric, and transitive. ◊

7.2 Equivalence Relations

In this section we study an important type of a relation, called an equivalence relation. Let us start by examining a motivating example, the congruence relation. Let $A = \{0, 1, 2, 3, 4, 5, 6\}$, and let R be a relation on A defined by setting, mRn if and only if $m \equiv n \pmod 3$. Let us draw the arrow diagram representing R.



What properties does R satisfy? The arrow diagram is telling us that the relation R is reflexive, symmetric, and transitive. Reflexive because there is a loop (an arrow from an element to itself) at every element in A. Symmetric because every arrow has a corresponding reverse arrow. Transitive because for any path of two consecutive arrows, there is a path from the departure point to the destination point. This example motivates the following definition.

Definition. A relation on a set A is called an *equivalence relation* if it satisfies the following three properties:

- For any $x \in A$, we have that xRx. (Reflexivity)
- For any $x, y \in A$, if xRy, then yRx. (Symmetry)
- For any $x, y, z \in A$, if xRy and yRz, then xRz. (Transitivity)

Thus, an equivalence relation is a reflexive, symmetric, and transitive relation. In mathematics, objects which are considered different in one context may be viewed the same in another context. For instance, both numbers 3 and 6 are considered the same in the eyes of the modulo 3 relation because they both leave a remainder of 0 when divided by 3. We make the idea of "viewed the same" precise using the concept of an equivalence relation which is a generalisation of the notion of equality. Observe that the equality relation is, indeed, an example of an equivalence relation.

When an element x is related to y by an equivalence relation R we say that "*x is equivalent to y*". And since an equivalence relation is a generalisation of equality instead of writing xRy we usually write $x \sim y$, or $x \approx y$, or $x \equiv y$ when x is equivalent to y.

We will see that an equivalence relation on a set partitions the set into disjoint subsets. For example, the diagram of the example above shows how the elements of A clustered into three different groups. Each group contains those elements who are related to each other.

$$A = \{0, 3, 6\} \cup \{1, 4\} \cup \{2, 5\}.$$

We now prove that the congruence relation on the set of integers is indeed an equivalence relation.

Example. Fix some $m \in \mathbb{Z}^+$. Define the relation R on $\mathbb{Z}$ by setting, aRb if and only if $a \equiv b \pmod{m}$ for any $a, b \in \mathbb{Z}$. Show that R is an equivalence relation.
We need to show that R is reflexive, symmetric, and transitive.

- R is reflexive. Pick $a \in \mathbb{Z}$. Then since $a - a = 0 = 0 \cdot m$, we have that m divides $a - a$, and thus $a \equiv a \pmod{m}$. Therefore, aRa. This shows that every element in $\mathbb{Z}$ is related to itself, and so R is reflexive.
- R is symmetric. Choose $a, b \in \mathbb{Z}$, we need to show that if aRb, then bRa. So suppose that aRb. Thus $a \equiv b \pmod{m}$, which means that $m \mid a - b$, and so $a - b = km$ for some integer k. This implies that $b - a = (-k)m$, and so $m \mid b - a$, meaning that $b \equiv a \pmod{m}$, and so bRa as desired.
- R is transitive. Let $a, b, c \in \mathbb{Z}$. We prove that aRb and bRc implies aRc. So suppose aRb and bRc. This means that $a \equiv b \pmod{m}$ and $b \equiv c \pmod{m}$. So there are integers k, l such that $a - b = km$ and $b - c = lm$. We now have that $a - c = (a - b) + (b - c) = km + lm = (k + l)m$. Thus, $m \mid a - c$ and so $a \equiv c \pmod{m}$. Thus aRc as required. $\Diamond$

Example. Let S be the set of all students of some university where every student is studying exactly one major. Define a relation R on S by setting, xRy if and only if "x is studying the same major as y". Then R is an equivalence relation (check that R is reflexive, symmetric, and transitive). Note that R divides the students into disjoint groups of major, like the group of mathematics students, the group of computer science students, the group of physics students, etc. $\Diamond$

Definition. Let R be an equivalence relation on a set A, and let $x \in A$. The *equivalence class* of x is the set of all elements in A equivalent to x. The equivalent class of x is denoted by $[x]_R$ or, simply, $[x]$.

$$[x] = \{y \in A \mid y \sim x\}.$$

Clearly, $[x] \subseteq A$ for any element $x \in A$.

Example. Let R be the congruence relation modulo 4 on $\mathbb{N}$. Find all the equivalence classes of the relation R. We have exactly 4 different equivalence classes.

- $[0] = \{n \in \mathbb{N} \mid n \sim 0\} = \{0, 4, 8, 12, \ldots\} = \{n \in \mathbb{N} \mid n \bmod 4 = 0\}$.
- $[1] = \{n \in \mathbb{N} \mid n \sim 1\} = \{1, 5, 9, 13, \ldots\} = \{n \in \mathbb{N} \mid n \bmod 4 = 1\}$.
- $[2] = \{n \in \mathbb{N} \mid n \sim 2\} = \{2, 6, 10, 14, \ldots\} = \{n \in \mathbb{N} \mid n \bmod 4 = 2\}$.
- $[3] = \{n \in \mathbb{N} \mid n \sim 3\} = \{3, 7, 11, 15, \ldots\} = \{n \in \mathbb{N} \mid n \bmod 4 = 3\}$.

There are no other equivalence classes since, by the Division Algorithm, any $n \in \mathbb{N}$ can be written as $n = 4q + r$ where $r \in \{0, 1, 2, 3\}$. This yields $n \equiv r \pmod 4$, and so $n \sim r$, implying that $[n] = [r]$. So for any natural number n, we have either $[n] = [0]$, or $[n] = [1]$, or $[n] = [2]$, or $[n] = [3]$.

- $[0] = [4] = [8] = [12] = [4q]$ for any $q \in \mathbb{N}$.
- $[1] = [5] = [9] = [13] = [4q + 1]$ for any $q \in \mathbb{N}$.
- $[2] = [6] = [10] = [14] = [4q + 2]$ for any $q \in \mathbb{N}$.
- $[3] = [7] = [11] = [15] = [4q + 3]$ for any $q \in \mathbb{N}$. ◊

In the example above, notice that different equivalence classes are disjoint (they do not have any elements in common). We only had two situations: either $[m] = [n]$ or $[m] \cap [n] = \emptyset$ for any $m, n \in \mathbb{N}$. This property holds for any equivalence relation in general. We establish this fact and more in the next theorem where we prove that every element is contained in its own equivalence class, and that two elements are equivalent precisely when they have equal equivalence classes.

Theorem 7.1. *Suppose that R is an equivalence relation on a set A. Then for any $x, y \in A$ we have the following.*

(i) $x \in [x]$.

(ii) $x \sim y \iff [x] = [y]$.

(iii) $[x] \neq [y] \iff [x] \cap [y] = \emptyset$.

(iv) $x \not\sim y \iff [x] \cap [y] = \emptyset$.

Proof. Choose any arbitrary $x, y \in A$.

(i) Since R is reflexive, we have that $x \sim x$, and so $x \in [x]$.

(ii) For the forward direction, suppose that $x \sim y$. We will show first that $[x] \subseteq [y]$. So choose any element $z \in [x]$. It follows that $z \sim x$. Since R is transitive and $z \sim x$ and $x \sim y$, it follows that $z \sim y$, and so $z \in [y]$. Thus, $[x] \subseteq [y]$. We now show that $[y] \subseteq [x]$. So let $w \in [y]$. It follows that $w \sim y$. Since $x \sim y$ and R is symmetric we have that $y \sim x$. Since R is transitive and $w \sim y$ and $y \sim x$, we conclude that $w \sim x$, and thus $w \in [x]$. Thus $[y] \subseteq [x]$. Finally, as

$[x] \subseteq [y]$ and $[y] \subseteq [x]$, we have that $[x] = [y]$.
For the reverse direction, suppose that $[x] = [y]$. By Part(i), we know that $x \in [x]$, so $x \in [y]$ as well. Thus $x \sim y$ as desired.

(iii) For the forward direction, we will show the contrapositive. Suppose that $[x] \cap [y] \neq \emptyset$. Then there exists some $z \in [x] \cap [y]$. So $z \in [x]$ and $z \in [y]$, which means that $z \sim x$ and $z \sim y$. By symmetry, we have that $x \sim z$. By transitivity, since $x \sim z$ and $z \sim y$, we obtain that $x \sim y$. By Part(ii) we conclude that $[x] = [y]$ as desired.
For the reverse direction, suppose that $[x] \cap [y] = \emptyset$. By Part(i), we have that $x \in [x]$. Since the intersection is empty, $x \notin [y]$. Thus, $[x]$ and $[y]$ do not contain the same elements, and so $[x] \neq [y]$.

(iv) Using Parts (ii) and (iii) we obtain the following equivalences.

$$x \not\sim y \iff [x] \neq [y] \iff [x] \cap [y] = \emptyset.$$

This completes the proof. ■

The following is an immediate corollary of the theorem above.

Corollary 7.2. *Let R be an equivalence relation on a set A, and choose elements $x, y \in A$. Then either $[x] = [y]$ or $[x] \cap [y] = \emptyset$.*

The discussion above says that a set with an equivalence relation can be divided into disjoint subsets where each subset is an equivalence class, i.e, each subset contains those elements which are equivalent to each other.

Corollary 7.3. *Let R be an equivalence relation on A. Then we can express A as*

$$A = [x_1] \cup [x_2] \cup \cdots \cup [x_n]$$

where $x_1, \ldots, x_n \in A$ and the equivalence classes $[x_1], [x_2], \ldots, [x_n]$ are pairwise disjoint. In other words, A is the union of the distinct equivalence classes of R.

Example. Consider an equivalence relation we examined earlier, the one given by congruence modulo 4 on the natural numbers. We can express $\mathbb{N}$ as follows.

$$\mathbb{N} = [0] \cup [1] \cup [2] \cup [3].$$

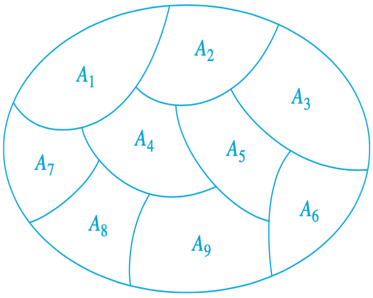
Check that any natural number belongs to one of the sets [0], [1], [2], or [3]. And check that these sets are pairwise disjoint. ◊

The action of an equivalence relation of splitting the underlying set into disjoint subsets motivates the following concept.

Definition. Suppose that S is a nonempty set. A *partition* of S is a collection $P = \{A_1, A_2, \cdots, A_n\}$ of subsets of S such that,

- For each i we have that $A_i \subseteq S$ and $A_i \neq \emptyset$, and
- $S = A_1 \cup A_2 \cup \cdots \cup A_n$, and
- $A_i \cap A_j = \emptyset$ if $i \neq j$.

The sets in P are called the *parts* of the partition.



A partition of a set.

Example. Let $S = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$, and let $A_1 = \{1, 4\}$, $A_2 = \{0, 5, 6, 9\}$, $A_3 = \{7\}$, $A_4 = \{2, 3, 8\}$.
Then $P = \{A_1, A_2, A_3, A_4\}$ is a partition of S because $S = A_1 \cup A_2 \cup A_3 \cup A_4$ and the parts A_1, A_2, A_3, A_4 are pairwise disjoint. $\Diamond$

The next example shows that any equivalence relation on a set induces a partition of the set whose parts are the equivalence classes of the equivalence relation.

Example. Let R be any equivalence relation on A. Then the collection P_R of all distinct equivalence classes is a partition of A.

$$P_R = \{[x] \mid x \in A\}.$$

The union of all equivalence classes of R is A because every element $x \in A$ belongs to its own equivalence class $[x]$. Moreover, different equivalence classes are disjoint. (See Theorem 7.1.) $\Diamond$

We next show that given a partition of a set we can define an equivalence relation on the set whose equivalence classes are the parts of the partition. Let $P = \{A_1, A_2, \ldots, A_n\}$ be a partition of a set S. Define a relation R_P on S by declaring that for any $x, y \in S$ we have that

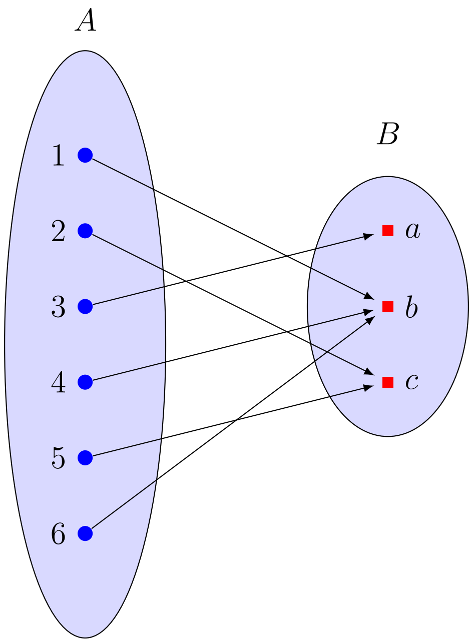
$$(x, y) \in R_P \text{ if and only if } \exists i(x \in A_i \wedge y \in A_i).$$

So two elements in S are related by R_P if they belong to the same part of the partition P. The reader is encouraged to verify that R_P is an equivalence relation on S, that is, R_P is reflexive, symmetric, and transitive. Moreover, the equivalence classes of R_P are precisely the parts of P.

Moral of the Story. Equivalence relations and partitions are two faces of the same coin. They express the same concept. On one hand, equivalence classes can be seen as parts of a partition. On the other hand, parts of a partition are seen as equivalence classes of the induced equivalence relation.

Here is an interesting way to obtain a partition on a set using functions. Any function induces a partition on its domain where each part consists of all those elements having the same image. Let us elaborate this fact through an example.

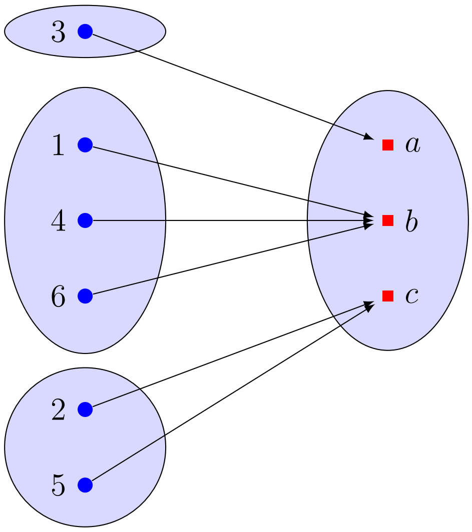
Example. Let $A = \{1, 2, 3, 4, 5, 6\}$ and $B = \{a, b, c\}$, and let $f : A \to B$ be the function given by the diagram below.



We can define an equivalence relation R on the domain A by setting xRy if and only if $f(x) = f(y)$ for any $x, y \in \text{dom}(f)$. So two elements of the domain are equivalent if they have the same image under f. The reader has to verify that R is an equivalence relation on A. The equivalence classes of R are $[1] = \{1, 4, 6\}$, and $[2] = \{2, 5\}$, and $[3] = \{3\}$. Thus $P = \{[1], [2], [3]\}$ is a partition of the domain A. Observe that,

$$A = [1] \cup [2] \cup [3] = \{1, 4, 6\} \cup \{2, 5\} \cup \{3\}.$$

The figure below shows how the domain is partitioned by R where each part consists of all those elements which have the same image under the function f. $\Diamond$


3
1
4
6
2
5
a
b
c

7.3 Relations and Boolean Matrices

In this section we utilise Boolean matrices to represent and manipulate relations.

Consider finite sets $A = \{a_1, a_2, \ldots, a_m\}$ and $B = \{b_1, b_2, \ldots, b_n\}$. A relation R from A to B can be represented by a Boolean matrix as follows. The *Boolean matrix representing* R is the $m \times n$ matrix $M_R = [m_{ij}]$ where

$$m_{ij} = \begin{cases} 1 & \text{if } (a_i, b_j) \in R, \\ 0 & \text{if } (a_i, b_j) \notin R. \end{cases}$$

This means that the (i, j)-entry m_{ij} of the matrix M_R is 1 if the i^{th} element in A is related to the j^{th} element in B, otherwise m_{ij} is 0. Notice that the matrix M_R depends on the ordering we choose for the elements in A and B. So different orderings of A and B result in different matrices representing the same relation.

Example. Represent the relation R from $\{1, 2\}$ to $\{a, b, c\}$ by a Boolean matrix where

$$R = \{(1, a), (1, b), (2, b), (2, c)\}.$$

The Boolean matrix representing R is $M_R = [m_{ij}]$ where

$$M_R = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}.$$

The first row of M_R describes the relation between 1 and the elements a, b, c. While the second row describes the relation between 2 and the elements a, b, c. For instance, we have that $m_{12} = 1$ because $1Rb$, and $m_{13} = 0$ because $1 \not R c$. ◊

Example. Let $A = \{\alpha, \beta, \gamma\}$ and $B = \{a, b, c, d, e\}$. Find the relation $R \subseteq A \times B$ which is represented by the Boolean matrix

$$M_R = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \end{bmatrix}.$$

The first row of M_R shows that $\alpha \in A$ is only related to $b \in B$. The second row shows that $\beta \in A$ is only related to $a, c, d \in B$. The third row shows that $\gamma \in A$ is related only to $a, c, e \in B$. So the relation R from A to B is

$$R = \{(\alpha, b), (\beta, a), (\beta, c), (\beta, d), (\gamma, a), (\gamma, c), (\gamma, e)\}.$$ ◊

Example. Write down the Boolean matrix representing the divisibility relation D on the set $A = \{2, 4, 8\}$.
The relation is

$$D = \{(2, 2), (2, 4), (2, 8), (4, 4), (4, 8), (8, 8)\} \subseteq A \times A.$$

The Boolean matrix representing D is,

$$M_D = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}. \qquad \Diamond$$

How do properties of relations such as reflexivity, symmetry, and antisymmetry reflect on the Boolean matrices representing them. Clearly, the Boolean matrix of a relation on a set A of n elements is a square matrix of dimension $n \times n$. Let R be a relation on a finite set A, and let M_R be the matrix representing R.

- R is reflexive if and only if all the elements on the main diagonal of M_R are equal to 1. For instance, the matrix below represents a reflexive relation on a set of cardinality 4 where the symbol $*$ is either 0 or 1.

$$M_R = \begin{bmatrix} 1 & * & * & * \\ * & 1 & * & * \\ * & * & 1 & * \\ * & * & * & 1 \end{bmatrix}.$$

- R is symmetric if and only if $m_{ij} = m_{ji}$ for all $1 \leq i \leq n$ and $1 \leq j \leq n$ if and only if M_R is a symmetric matrix (i.e. M_R is equal to its transpose M_R^T). Thus, R is symmetric if and only if every entry in the matrix M_R is equal to its mirror image across the main diagonal. The matrix below represents a symmetric relation.

$$M_R = \begin{bmatrix} * & 1 & 0 & 1 \\ 1 & * & 0 & 1 \\ 0 & 0 & * & 1 \\ 1 & 1 & 1 & * \end{bmatrix}.$$

- R is antisymmetric if and only if whenever $m_{ij} = 1$ and $i \neq j$ then its mirror image $m_{ji} = 0$. Thus, R is antisymmetric if and only if every 1 not on the main diagonal has 0 as its mirror image. The matrix below represents an antisymmetric relation.

$$M_R = \begin{bmatrix} * & 1 & 0 & 0 \\ 0 & * & 0 & 1 \\ 1 & 1 & * & 0 \\ 0 & 0 & 0 & * \end{bmatrix}.$$

Example. Consider the relation R whose Boolean matrix is

$$M_R = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix}.$$

Then the relation R is reflexive and symmetric. It is not antisymmetric because there is a 1 off the main diagonal whose mirror image is not 0. ♢

♣ New Relations from Old Relations

Suppose that R and S are relations from a set A to a set B. Let M_R and M_S be the Boolean matrices representing R and S, respectively.

The union $U = R \cup S$ is also a relation from A to B (because $U \subseteq A \times B$). It follows that a is related to b by U if either a is related to b by R or by S. In other words, we have that aUb if and only if aRb or aSb. The Boolean matrix representing U is obtained by taking the join of M_R and M_S.

$$M_{R \cup S} = M_R \vee M_S.$$

The intersection $I = R \cap S$ is a relation from A to B. It follows that a is related to b by I if a is related to b by R and by S. In other words, we have that aIb if and only if aRb and aSb. The Boolean matrix representing I is obtained by taking the meet of M_R and M_S.

$$M_{R \cap S} = M_R \wedge M_S.$$

The complement of R is the relation $\overline{R} = (A \times B) - R$. So $\overline{R}$ consists of all pairs which are in $A \times B$ but not in R. Thus, we have that $a\overline{R}b$ if and only if $a\not{R}b$. In other words, a is related to b by $\overline{R}$ precisely when they are not related by R. The Boolean matrix representing $\overline{R}$ is,

$$M_{\overline{R}} = M_{A \times B} - M_R.$$

Notice that $M_{A \times B}$ is the Boolean matrix representing the relation $A \times B$, thus all of its entries are 1s because $A \times B$ consists of all ordered pairs (a, b) where $a \in A, b \in B$. The reader is advised to verify the three matrix equations above.

Example. Let R and S be relations on $\{a, b, c\}$ where $R = \{(a, a), (a, c), (b, a), (c, b)\}$ and $S = \{(a, a), (a, c), (b, b), (b, c), (c, a)\}$. Compute the union and intersection of R and S, and the complement of R, as well as their corresponding Boolean matrices.

- $R \cup S = \{(a,a),(a,c),(b,a),(c,b),(b,b),(b,c),(c,a)\}$.

- $R \cap S = \{(a,a),(a,c)\}$.

- $\overline{R} = \{(a,b),(b,b),(b,c),(c,a),(c,c)\}$.

- $M_R = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$ and $M_S = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \end{bmatrix}$.

- $M_{R \cup S} = M_R \vee M_S = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 0 \end{bmatrix}$.

- $M_{R \cap S} = M_R \wedge M_S = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$.

- $M_{\overline{R}} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} - \begin{bmatrix} 1 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix}$. ◊

Example. Below we have the Boolean matrices representing the equality relation $R_=$, the strictly-less-than relation $R_<$, and the less-than-or-equal-to relation $R_\leq$ on the set $A = \{1,2,3,4\}$. Notice that $R_\leq = R_= \cup R_<$.

- $R_= = \{(1,1),(2,2),(3,3),(4,4)\}$.

- $R_< = \{(1,2),(1,3),(1,4),(2,3),(2,4),(3,4)\}$.

- $M_= = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$.

- $M_< = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$.

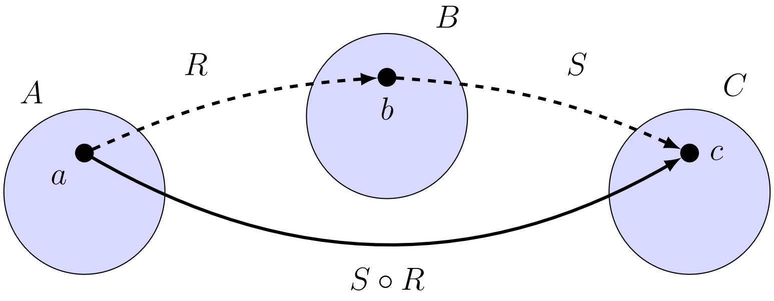
- $M_{\leq} = M_{<} \vee M_{=} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$. ◊

We next introduce another method of obtaining a new relation from older ones.

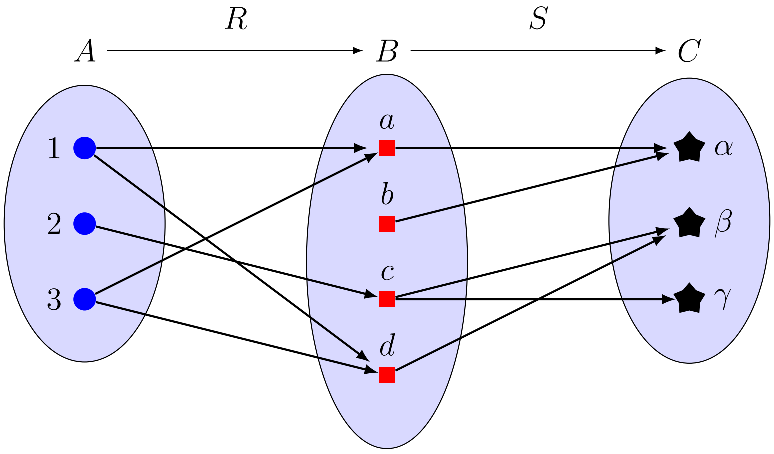
Definition. Consider the relations $R \subseteq A \times B$ and $S \subseteq B \times C$. The *composition of S after R* is the relation from A to C given by,

$$S \circ R = \{(a, c) \in A \times C \mid \exists b \in B(aRb \wedge bSc)\}.$$

So an element $a \in A$ is related to an element $c \in C$ via the composition relation $S \circ R$ if we can find an intermediate element $b \in B$ such that aRb and bSc.



Example. Given the relations R and S below, find the composition S after R.



For instance, we have that $2 \in A$ is related to $\beta \in C$ by the composition relation $S \circ R$ because $2Rc$ and $cS\beta$ where $c \in B$. Thus, $(2, \beta) \in S \circ R$. Working in this fashion we get that,

$$S \circ R = \{(1, \alpha), (1, \beta), (2, \beta), (2, \gamma), (3, \alpha), (3, \beta)\}. \qquad \Diamond$$

A special case of composing relations is when we compose a relation with itself as described by the next definition.

Definition. Let R be a relation on a set A. Then we define the n^{th} *power* R^n *of the relation* R as follows.

- $R^1 = R$, and
- $R^{n+1} = R^n \circ R$ for $n \in \mathbb{Z}^+$.

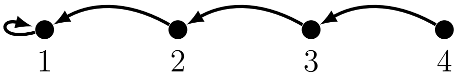
By the definition above, the second power of R is $R^2 = R \circ R$. Now by the definition of composition it follows that for any $a, b \in A$ we have that aR^2b if and only if there exists some $c \in A$ such that aRc and cRb. In other words,

$$aR^2b \iff \exists c\,(aRc \wedge cRb).$$

The third power of R is $R^3 = R^2 \circ R$. So for any $a, b \in A$ we have that aR^3b if and only if there exists some $c \in A$ such that aRc and cR^2b if and only if there exist some $c, d \in A$ such that aRc and cRd and dRb. In other words,

$$aR^3b \iff \exists c\,\exists d\,(aRc \wedge cRd \wedge dRb).$$

Example. Consider the relation $R = \{(1,1),(2,1),(3,2),(4,3)\}$ on the set $\{1,2,3,4\}$.



We have that $3R^21$ because $3R2$ and $2R1$. And $4R^31$ because $4R3$ and $3R^21$.

- $R^2 = R \circ R = \{(1,1),(2,1),(3,1),(4,2)\}$.
- $R^3 = R^2 \circ R = \{(1,1),(2,1),(3,1),(4,1)\}$. ♢

Example. Consider the relation S on the set of people where xSy holds if "x is the son of y" for any people x and y.
Then xS^2z if and only if $x(S \circ S)z$ if and only if $\exists y(xSy \wedge ySz)$ if and only if x is the grandson of z. ♢

The proof of the theorem below is left as an exercise for the reader.

Theorem 7.4. *Let R be a relation on a set A. Then R is transitive if and only if $R^n \subseteq R$ for every $n \in \mathbb{Z}^+$.*

We now turn our attention to the task of calculating the Boolean matrix representing the composition of relations. Recall that $A \odot B$ is the Boolean product of matrices A and B. See Section 6.2.

Lemma 7.5. *Let $R \subseteq A \times B$ and $S \subseteq B \times C$ be relations. Let M_R and M_S be their corresponding Boolean matrices, respectively. Then, the Boolean matrix of the composition relation $S \circ R$ is*

$$M_{S \circ R} = M_R \odot M_S.$$

Proof. To show that $M_{S \circ R} = M_R \odot M_S$, we need to show that the matrices $M_{S \circ R}$ and $M_R \odot M_S$ are equal, so we need to establish that they have the same dimension and their corresponding entries are equal.
Let $A = \{a_1, \ldots, a_m\}$, and $B = \{b_1, \ldots, b_k\}$, and $C = \{c_1, \ldots, c_n\}$, where $|A| = m$, $|B| = k$, and $|C| = n$. Since $S \circ R \subseteq A \times C$ we get that $M_{S \circ R}$ is of dimension $m \times n$. Since M_R has dimension $m \times k$ and M_S has dimension $k \times n$, by definition of the Boolean product, $M_R \odot M_S$ has dimension $m \times n$. Thus both $M_{S \circ R}$ and $M_R \odot M_S$ are of the same dimension.
We now show that an entry in $M_R \odot M_S$ is 1 if and only if its corresponding entry in $M_{S \circ R}$ is also 1. We proceed as follows, the (i, j)-entry in $M_R \odot M_S$ is 1 if and only if for some $1 \leq t \leq k$ the (i, t)-entry in M_R is 1 and (t, j)-entry in M_S is also 1 if and only if there is some $b_t \in B$ such that $a_i R b_t$ and $b_t S c_j$ if and only if a_i is related to c_j by $S \circ R$ if and only if (i, j)-entry of $M_{S \circ R}$ is 1. Since in Boolean matrices an entry is either 0 or 1, the previous argument implies that an entry in $M_R \odot M_S$ is 0 if and only if its corresponding entry in $M_{S \circ R}$ is also 0. Therefore, $M_{S \circ R} = M_R \odot M_S$. ■

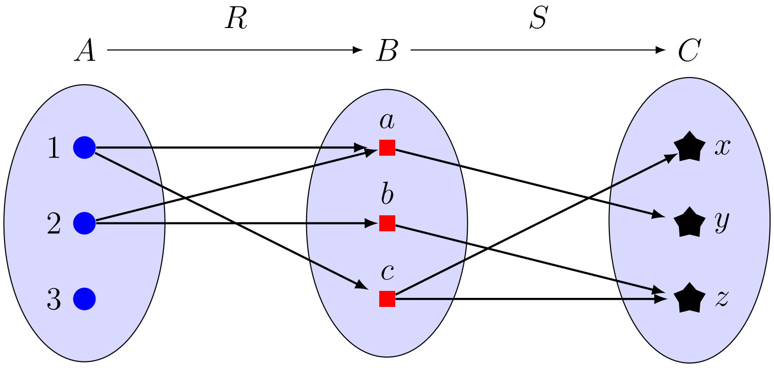
Recall from Section 6.2 that $M^{[n]}$ is the n^{th} Boolean power of a matrix M. Use the above lemma and mathematical induction to establish the following result which states that the matrix representing the n^{th} power of a relation is equal to the n^{th} Boolean power of the matrix representing the relation.

Corollary 7.6. *Let R be a relation on a set A. Then,*

$$M_{R^n} = M_R^{[n]}.$$

Example. Let $A = \{1, 2, 3\}$, $B = \{a, b, c\}$, and $C = \{x, y, z\}$. And let R be a relation from A to B, and S a relation from B to C that are represented by the Boolean matrices below. Sketch the arrow diagrams of R and S and compute the Boolean matrix representing the composition $S \circ R$, and compare them.

$$M_R = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \text{ and } M_S = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix}.$$



The composition $S \circ R$ is a relation from A to C represented by

$$M_{S \circ R} = M_R \odot M_S = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

For instance the $(2,2)$-entry in $M_{S \circ R}$ is 1 which means that $2 \in A$ is related to $y \in C$ by $S \circ R$, and that is indeed true because $2Ra$ and aSy where $a \in B$. ◊

7.4 Partial and Total Orders

In this section we study relations which impose an order on the elements of their sets. For instance think of ordering your family members by asking them to stand up in a line from the youngest to the oldest. In this situation the relation used on the set of family members is that one member is related to another if the first is younger than than the second. Another example of ordering is the arrangement of words in a dictionary. A third example is the ordering of real numbers on the real number line. Which relations induce an order on the elements of their sets? The next definition encapsulates the concept of "order".

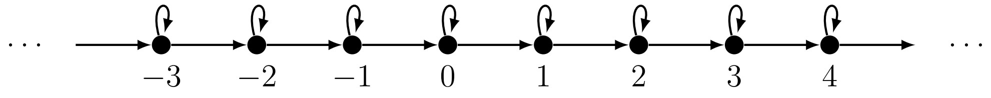
Definition. A *partial order* is a relation $\preceq$ on a set S which is reflexive, antisymmetric, and transitive. We call the pair $(S, \preceq)$ a *partially ordered set* or *poset.*

Let us unravel this definition. A relation $\preceq$ on a set S is a partial order if it satisfies the following properties where the variables domain is S.

- $\forall x\,(x \preceq x)$. (Reflexivity)
- $\forall x \forall y\,(x \preceq y \wedge y \preceq x \to x = y)$. (Antisymmetry)
- $\forall x \forall y \forall z\,(x \preceq y \wedge y \preceq z \to x \preceq z)$. (Transitivity)

We remark that antisymmetry is also equivalent to $\forall x \forall y\,(x \neq y \wedge x \preceq y \to y \not\preceq x)$.

Example. The pair $(\mathbb{Z}, \leq)$ is a poset.
Check that the less-than-or-equal relation $(\leq)$ on the set of integers is a partial order, that is, check that it is reflexive, antisymmetric, and transitive. The arrow diagram helps where $m \to n$ means $m \leq n$.



Note that by transitivity there must be an arrow from an integer on the line above to any integer to its right, for instance, there should be the arrows $0 \to 2$, $0 \to 3$, and $0 \to 4$, but we omit these arrows for the clarity of the diagram. ◊

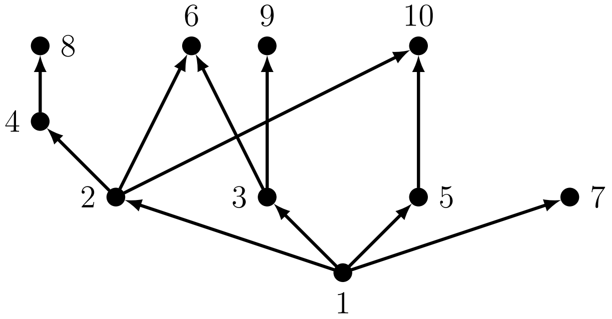
Remark. In the previous example we considered the partial order given by the less-than-or-equal-to relation denoted by $\leq$. In the next two examples we consider the divisibility relation denoted by $|$, and the subset relation denoted by $\subseteq$. In general, we use the symbol $\preceq$ to denote any partial order. So if $(S, \preceq)$ is a partial order, then $a \preceq b$ simply means that a is related to b by some relation $\preceq$ which is a partial order (reflexive, antisymmetric, and transitive). We usually read $a \preceq b$ as "a is less than or equal to b".

Observe that in the poset $(\mathbb{Z}, \leq)$ any two integers a, b are related by the relation $\leq$, that is, for any integers a and b either $a \leq b$ or $b \leq a$. However, in a general poset $(S, \preceq)$ it is *not* necessary that any two elements must be related. This leads to the next definition.

Definition. Suppose that $(S, \preceq)$ is a poset. We call elements $a, b \in S$ *comparable* if either $a \preceq b$ or $b \preceq a$. Otherwise, if $a \not\preceq b$ and $b \not\preceq a$, then we call a and b *incomparable.*

Thus, in the poset $(\mathbb{Z}, \leq)$ any pair of integers are comparable. For instance, 9 and 4 are comparable because $4 \leq 9$.

Example. The pair $(\mathbb{Z}^+, |)$ is a poset.
Check that the divisibility relation $(|)$ on the set of positive integers is a partial order. The reader needs to verify that $|$ is a reflexive, antisymmetric, and transitive relation (See Theorem 4.1).
For simplicity we sketch the arrow diagram of the divisibility relation on the set $\{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$ where $m \to n$ means $m \mid n$.



Note that in the diagram above there must be an arrow from an integer to itself (a loop) because every integer divides itself, however, we omitted all loops for clarity. Moreover, the arrows for transitivity are also missing, for example there must be an arrow $2 \to 8$ because $2 \mid 8$ (also because of transitivity as $2 \mid 4$ and $4 \mid 8$). The other transitivity arrows are $1 \to 4$, $1 \to 8$, $1 \to 6$, $1 \to 9$, and $1 \to 10$.

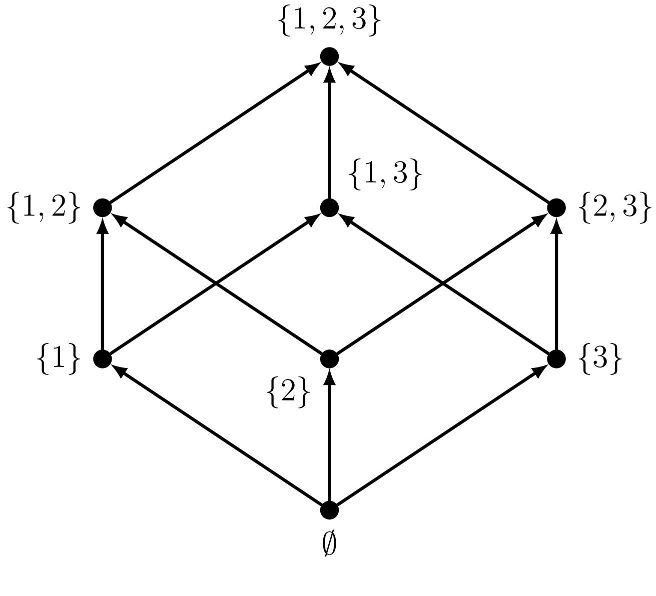
Observe that 2 and 10 are comparable because $2 \mid 10$, similarly, 6 and 3 are comparable because $3 \mid 6$. Also 1 is comparable with any positive integer because 1 divides all positive integers including itself. On the other hand, 3 and 7 are incomparable because $3 \nmid 7$ and $7 \nmid 3$. ♢

Remark. The arrow diagrams we use here to represent partially ordered sets are called *Hasse Diagrams*, named after the German mathematician Helmut Hasse (1898 – 1979) who was working in algebraic number theory.

Example. The pair $(\mathcal{P}(S), \subseteq)$ is a poset for any set S.
In other words, the subset relation $(\subseteq)$ is a partial order on the power set $\mathcal{P}(S)$ of a set S. Let us check that.

- First, since every set is a subset of itself, it follows that the relation $\subseteq$ is reflexive. That is, for every $A \in \mathcal{P}(S)$ we have $A \subseteq A$.
- Second, the relation $\subseteq$ is antisymmetric because for any $A, B \in \mathcal{P}(S)$ if $A \subseteq B$ and $B \subseteq A$, then $A = B$.
- Third, the relation $\subseteq$ is transitive because for any $A, B, C \in \mathcal{P}(S)$ it is easy to prove that if $A \subseteq B$ and $B \subseteq C$, then $A \subseteq C$.

For instance suppose that $S = \{1, 2, 3\}$. Then we have the subset relation on the power set $\mathcal{P}(S) = \big\{\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{2, 3\}, \{1, 3\}, \{1, 2, 3\}\big\}$. The Hasse diagram of the poset $(\mathcal{P}(S), \subseteq)$ is shown below. Note that the arrows due to reflexivity and transitivity are omitted.



Observe that $\emptyset$ is comparable with all sets in $\mathcal{P}(S)$ because it is a subset of any set. Moreover, $\{3\}$ is comparable with $\{1, 3\}$ because $\{3\} \subseteq \{1, 3\}$. On the other hand, $\{1, 2\}$ and $\{2, 3\}$ are incomparable because $\{1, 2\} \not\subseteq \{2, 3\}$ and $\{2, 3\} \not\subseteq \{1, 2\}$. $\Diamond$

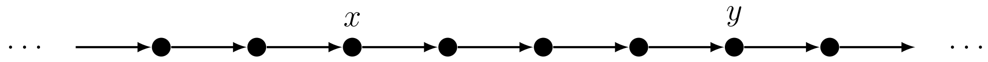
We have seen that it is possible to find elements in a poset which are incomparable. It is also possible that any pair of elements are comparable like the case of $(\mathbb{Z}, \leq)$. Such partial orders deserve a special name.

Definition. A partially order set $(S, \preceq)$ where any two elements are comparable is called a *totally ordered set* or a *linearly ordered set.*

In other words, a relation $\preceq$ on a set S is called a *total order* (or *a linear order*) if it satisfies the following properties where the variables domain is S.

- $\forall x\,(x \preceq x)$. (Reflexivity)
- $\forall x \forall y\,(x \preceq y \wedge y \preceq x \to x = y)$. (Antisymmetry)
- $\forall x \forall y \forall z\,(x \preceq y \wedge y \preceq z \to x \preceq z)$. (Transitivity)
- $\forall x \forall y\,(x \preceq y \vee y \preceq x)$. (Totality)

As any two elements are comparable in a totally ordered set $(S, \preceq)$, the elements of the set can be thought of as arranged in a line, one element after the other, where x being before y on the line means that $x \preceq y$.



Remark. A totally ordered set is also called a *chain.*

Obviously, $(\mathbb{Z}, \leq)$ is a total order. However, $(\mathbb{Z}^+, |)$ is not a total order because for instance 5 and 11 are incomparable.

The next example is an example of a total order used to order strings of letters or numbers, it is called a “*lexicographic order*” or ”*dictionary order*”. Lexicographic order is the order we use to arrange the words in a dictionary one after the other. For instance, the word “go” comes before “me” because, in the English alphabet, the first letter (g) in “go” comes before the first letter (m) in “me”. Also “me” appears before “my” in the dictionary because they have the same first letter (m) and the second letter (e) in “me” comes before the second letter (y) in “my”. In other words, go $\preceq$ me and me $\preceq$ my.

Example. We define a lexicographic relation on the Cartesian product $\mathbb{N} \times \mathbb{N}$ as follows. Given (a, b) and (x, y) in $\mathbb{N} \times \mathbb{N}$, we define

$$(a, b) \preceq (x, y) \iff a < x \vee (a = x \wedge b \leq y).$$

Prove that $(\mathbb{N} \times \mathbb{N}, \preceq)$ is a totally ordered set. $\Diamond$

Definition. A *well-ordered set* is a totally ordered set $(S, \preceq)$ such that every nonempty subset of S has a least element.

We note that $(\mathbb{N}, \leq)$ is a well-ordered set. However, $(\mathbb{Z}, \leq)$ is a totally ordered set which is not well-ordered because, for example, the subset $\{m \in \mathbb{Z} \mid m \leq 3\}$, which is a subset of $\mathbb{Z}$, has no least element.

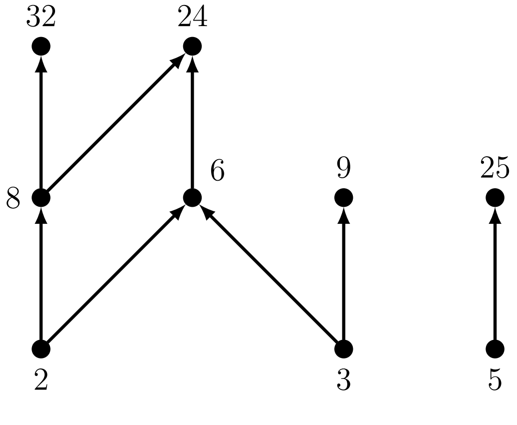
Remark. Every well-ordered set is a totally ordered set, and every totally-ordered set is a partially ordered set.

In a partially ordered set, certain elements have special properties.

Definition. Suppose that $(S, \preceq)$ is a poset.

- We call $a \in S$ *minimal* if there is no $x \in S$ such that $x \preceq a$ and $x \neq a$.
- We call $b \in S$ *maximal* if there is no $x \in S$ such that $b \preceq x$ and $x \neq b$.

Example. Consider the poset consisting of the set $S = \{2, 3, 5, 6, 8, 9, 24, 25, 32\}$ together with the divisibility relation. Draw the Hasse diagram for the poset $(S, |)$ and determine the minimal and maximal elements.



The minimal elements are 2, 3, and 5 since there are no elements below them. And the maximal elements are 9, 24, 25, and 32 as there are no elements above them. ♢

Chapter 8

Graph Theory

Think of cities as points and the roads between them as lines. Hooray! You have a graph!

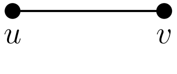
Graph theory is a central area of study in discrete mathematics and it has important applications within mathematics itself. In the real world, graphs have a wide variety of applications in computer science, and they are used as models to represent physical, social, and biological systems and networks.

8.1 Graphs

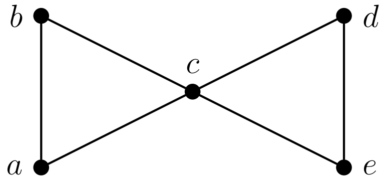
We first present an abstract definition of a graph using sets.

Definition. A *graph* G is a pair of sets V and E, where the elements of V are called *vertices* and the elements of E are called *edges*, where each edge is a set of exactly two distinct vertices.

Pictorially, we represent vertices by points, and an edge $e = \{u, v\}$ where u, v are two distinct vertices is represented by the diagram below.



Example. The *bowtie graph* (or the *butterfly graph)* has vertex set $V = \{a, b, c, d, e\}$ and edge set $E = \big\{\{a, b\}, \{a, c\}, \{b, c\}, \{c, d\}, \{c, e\}, \{d, e\}\big\}$.



Bowtie Graph

The *order* of a graph G is the cardinality of its vertex set, and the *size* of G is the cardinality of its edge set. For example, the bowtie graph has order 5, and size 6.

Remark. We state some comments on the graph definition above.

1. Our definition of a graph produces an undirected graph since edges are sets and not ordered pairs, moreover, our graphs have no loops (a loop is an edge from a vertex to itself).
2. A graph $G = (V, E)$ is essentially a relation E on the set V which is irreflexive (no loops) and symmetric (undirected edges). To see this, an edge $\{u, v\}$ corresponds to the two pairs (u, v) and (v, u) which belong to the relation E.

Let us introduce some important notation and terminology in graph theory. Suppose that $G = (V, E)$ is a graph and $u, v \in V$ are any two distinct vertices of G.

- We denote the set $\{u, v\}$ by uv.
- When the set uv is an edge (that is, $uv \in E$), we say that u and v are *adjacent* or *neighbours.*
- When e is an edge and $e = uv$, we say that u and v are the *end vertices* of e.
- If $uv \notin E$, then we say that u and v are *nonadjacent.*
- If v is an end vertex of an edge e, then we say that e is *incident* with v.
- The *neighborhood* $N(v)$ of a vertex v is the set of all vertices adjacent to v.

$$N(v) = \{w \in V \mid vw \in E\}.$$

- The *degree* of v is the number of edges incident with v,

$$\deg(v) = |\{e \in E \mid v \in e\}| = |N(v)|.$$

- The maximum degree of a graph G is denoted by $\Delta(G)$.

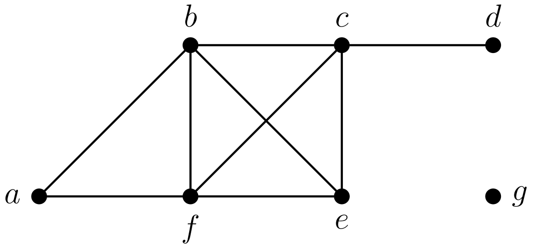
$$\Delta(G) = \max\{\deg(v) \mid v \in V_G\}.$$

- The minimum degree of a graph G is denoted by $\delta(G)$.

$$\delta(G) = \min\{\deg(v) \mid v \in V_G\}.$$

- The *degree sequence* of a graph of order n is the sequence of length n listing all the vertex degrees in decreasing order.

Example. Consider the following graph with 7 vertices and 9 edges.



$\deg(a) = 2$	$\deg(b) = 4$	$\deg(c) = 4$	$\deg(d) = 1$
$\deg(e) = 3$	$\deg(f) = 4$	$\deg(g) = 0$	$N(a) = \{b, f\}$
$N(b) = \{a, c, e, f\}$	$N(c) = \{b, d, e, f\}$	$N(d) = \{c\}$	$N(e) = \{b, c, f\}$
$N(f) = \{a, b, c, e\}$	$N(g) = \emptyset$		$\Diamond$

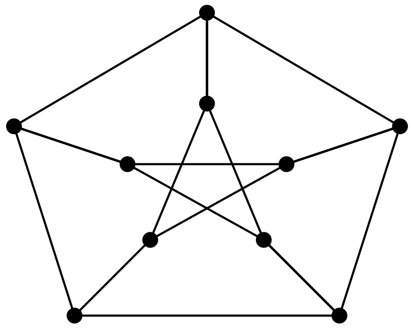
In the graph above, the sum of all the vertex degrees is 18 which is twice the number of edges. This is explained by the next result which is considered the first result in graph theory.

Theorem 8.1 (The Handshaking Theorem)**.** *Let $G = (V, E)$ be a graph. Then, we have that*

$$\sum_{v \in V} \deg(v) = 2|E|.$$

Proof. When we sum the degrees of all the vertices, every single edge is counted exactly twice; it is counted once for each of its end vertices. Therefore, the sum of the degrees of all the vertices is twice the number of edges. ■

Example. The graph below is very famous in graph theory, it is called *The Petersen Graph.* It has been used as a counterexample to many conjectures in graph theory.

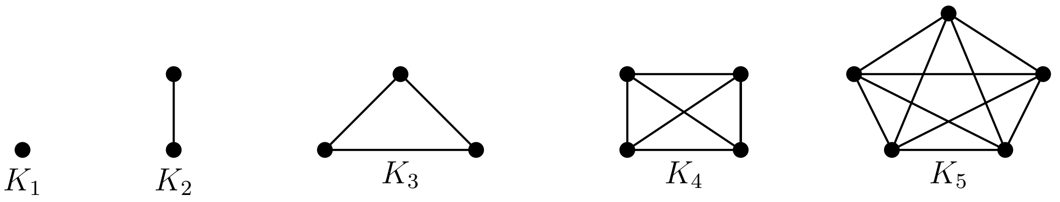


The Petersen Graph

The Petersen graph has 10 vertices and 15 edges, where every vertex has degree 3. It is named after the Danish mathematician Julius Petersen (1839 – 1910). ◊

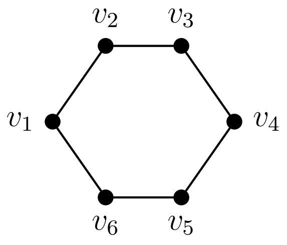
We now introduce several special graphs.

Definition. The *complete graph* K_n on n vertices is the graph with n vertices and such that any pair of distinct vertices are adjacent.



Note that for any vertex v in K_n we have that $\deg(v) = n-1$. Consequently, by the Handshaking Theorem, the graph K_n has exactly $n(n-1)/2$ edges. For example, K_5 has 10 edges.

Definition. The *cycle* C_n of length n is the graph whose vertex set is given by $V = \{v_1, v_2, \cdots, v_n\}$ and edge set is $E = \{\{v_1, v_2\}, \{v_2, v_3\}, \ldots, \{v_{n-1}, v_n\}, \{v_n, v_1\}\}$.

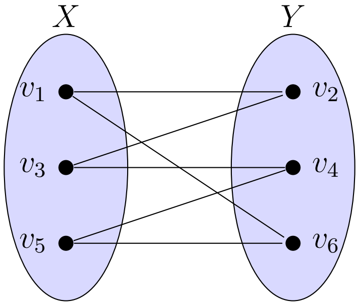


The cycle C_6

Definition. A *bipartite graph* $G = (V, E)$ is a graph such that its vertex set V can be partitioned into two disjoint nonempty subsets X and Y such that for every edge $\{u, v\}$ in E either $u \in X$ and $v \in Y$, or $v \in X$ and $u \in Y$. The sets X and Y are called partite sets.

Notice that if a graph G is bipartite with partite sets X and Y, then there is no edge in G whose end vertices both lie in X or both lie in Y.

Example. The cycle C_6 is bipartite because its vertex set $V = \{v_1, v_2, v_3, v_4, v_5, v_6\}$ may be partitioned into $X = \{v_1, v_3, v_5\}$ and $Y = \{v_2, v_4, v_6\}$ where every edge has one end vertex in X and the other in Y as illustrated below.



◊

Example. The complete graph K_3 is not bipartite.
To prove this we need to show that it is impossible to partition the vertex set $\{v_1, v_2, v_3\}$ as in the definition of a bipartite graph. For the sake of contradiction, suppose we can and let X and Y be partite sets of K_3. One of X or Y must have exactly 1 vertex and the other has 2 vertices. But then the two vertices in the same partite set are adjacent since every pair of vertices are adjacent in K_3, meaning that there is an edge in the same partite set, thus we have a contradiction. Therefore, K_3 cannot be bipartite. ◊

We give below a characterisation of bipartite graphs and leave the reader to think about its proof.

Theorem 8.2. *A graph G is bipartite if and only if G contains no odd cycles.*