Mathematics Course 111: Algebra I
Part I: Algebraic Structures, Sets and Permutations

D. R. Wilkins

Academic Year 1996-7

1 Number Systems and Matrix Algebra

Integers

The 'whole numbers' $0, \pm 1, \pm 2, \pm 3, \pm 4, \ldots$ are referred to as *integers*. The set of all integers is denoted by $\mathbb{Z}$. The set $\{1, 2, 3, 4, \ldots\}$ of all positive integers is denoted by $\mathbb{N}$. Positive integers are often referred to as *natural numbers*. (Some authors include zero in the set of natural numbers).

Here are some basic properties of the integers under addition:—

- the sum $x + y$ of two integers x and y is itself an integer;
- $x + y = y + x$ for all integers x and y (the *Commutative Law* for addition);
- $x + (y + z) = (x + y) + z$ for all integers x, y and z (the *Associative Law* for addition);
- there is an integer 0 with the property that $x + 0 = x$ for all integers x;
- given any integer x, there exists an integer u satisfying $x + u = 0$.

Many other basic properties of integers under addition can be derived formally from those listed above.

The first four of the basic properties of integers under addition listed above have analogues for multiplication:—

- the product $x \times y$ of two integers x and y is itself an integer;
- $x \times y = y \times x$ for all integers x and y (the *Commutative Law* for multiplication);
- $x \times (y \times z) = (x \times y) \times z$ for all integers x, y and z (the *Associative Law* for multiplication);
- there is an integer 1 with the property that $x \times 1 = x$ for all integers x;

Another basic property of the integers is the *Distributive Law* relating multiplication and addition. This states that $(x + y) \times z = (x \times z) + (y \times z)$ for all integers x, y and z.

The Principle of Mathematical Induction

In order to prove that certain properties hold for all positive integers, the *Principle of Mathematical Induction* is often employed. This principle is a consequence of the following important property of subsets of the set $\mathbb{N}$ of positive integers:

The Principle of Mathematical Induction (for subsets of $\mathbb{N}$). Let T be a set of positive integers which satisfies the following conditions:

- $1 \in T$;
- if $k \in T$ for any positive integer k then $k + 1 \in T$.

Then $T = \mathbb{N}$.

Indeed, repeated use of these conditions ensures that $1 \in T$, hence $2 \in T$, hence $3 \in T$, hence $4 \in T$, and so on.

Suppose that one wishes to prove that a certain property holds for all positive integers. Then the set T can be taken as the set of all positive integers with the specified property, and the above induction principle can be applied. This leads to the following formulation of the *Principle of Mathematical Induction*.

The Principle of Mathematical Induction. Suppose that some property of positive integers is specified and that the following conditions are satisfied:

- 1 has the specified property;
- if a positive integer k has the specified property, then so does $k + 1$.

Then all positive integers have the specified property.

Example. The Principle of Mathematical Induction can be used to show that $\sum_{i=1}^{n} i = \frac{1}{2}n(n+1)$ for all positive integers n. Clearly this equality holds when $n = 1$. We must show that if equality holds when $n = k$, then equality holds when $n = k + 1$. Suppose then that $\sum_{i=1}^{k} i = \frac{1}{2}k(k+1)$. Then

$$\sum_{i=1}^{k+1} i = \tfrac{1}{2}k(k+1) + (k+1) = \tfrac{1}{2}k^2 + \tfrac{3}{2}k + 1 = \tfrac{1}{2}(k+1)(k+2).$$

Thus if equality holds when $n = k$ then equality holds when $n = k + 1$. It now follows from the Principle of Mathematical Induction that $\displaystyle\sum_{i=1}^{n} i = \frac{1}{2}n(n+1)$ for all positive integers n.

The following principle is a useful variant of the Principle of Mathematical Induction.

The Principle of Complete Induction. Suppose that some property of positive integers is specified and that the following conditions are satisfied:

- 1 has the specified property;
- if k is an integer greater than one, and if all positive integers j satisfying $j < k$ have the specified property, then so does k.

Then all positive integers have the specified property.

Thus suppose that some property of positive integers satisfies these conditions. Then 1 has the property. Hence 2 has the property. Then 1 and 2 have the property, hence 3 has the property. Then 1, 2 and 3 have the property, hence 4 has the property, and so on.

The Principle of Complete Induction can be justified more formally as follows. Let T be the subset of the set $\mathbb{N}$ of positive integers which consists of 1 together with those positive integers k greater than one for which $1, 2, \ldots, k-1$ have the specified property. If the hypotheses of the Principle of Complete Induction are satisfied and if $k \in T$ for some positive integer k then k has the specified property, and hence $k+1 \in T$. It now follows from the Principle of Mathematical Induction that $T \in \mathbb{N}$, and thus all positive integers have the specified property.

Rational, Real and Complex Numbers

A *rational number* is a number of the form m/n, where m and n are both integers ('whole numbers'). The notation $\mathbb{Q}$ is used to denote the set of all rational numbers.

The system of *real numbers* includes not only rational numbers but irrational numbers such as $\sqrt{2}$ and π. (The ancient Greeks proved that there do not exist non-zero integers p and q satisfying $p^2 = 2q^2$: this demonstrates that $\sqrt{2}$ is not a rational number. One can think of the real numbers as comprising all numbers representing displacements along a straight line from some given point. Formal definitions of the real number system were given by Dedekind and Cantor in 1872. The notation $\mathbb{R}$ is used to denote the set of all real numbers.

A *complex number* is a number representable in the form $x + iy$, where x and y are real numbers and i satisfies $i^2 = -1$. The notation $\mathbb{C}$ is used to denote the set of all complex numbers.

The system of rational numbers, the system of real numbers and the system of complex numbers are examples of *fields*. A field is a set on which are defined operations of addition and multiplication with properties corresponding to those of addition and multiplication on the rational, real and complex numbers.

Matrix algebra

Let us now consider algebraic operations on the set of all 2×2 matrices with real coefficients.

Matrix addition is defined for 2×2 matrices as follows:—

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} + \begin{pmatrix} e & f \\ g & h \end{pmatrix} = \begin{pmatrix} a+e & b+f \\ c+g & d+h \end{pmatrix}.$$

Matrix addition is commutative (i.e., $A + B = B + A$ for all 2×2 matrices A and B) and associative (i.e., $A + (B + C) = (A + B) + C$ for all 2×2 matrices A, B and C). There is a *zero matrix* Z with the property that $A + Z = A$ (and $Z + A = A$) for all 2×2 matrices A. This matrix is given by $Z = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$. Moreover, given any 2×2 matrix A, there is a 2×2 matrix B satisfying $A + B = Z$ (and $B + A = Z$). If $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ then $B = \begin{pmatrix} -a & -b \\ -c & -d \end{pmatrix}$.

Matrix multiplication is defined for 2×2 matrices as follows:—

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} e & f \\ g & h \end{pmatrix} = \begin{pmatrix} ae+bg & af+bh \\ ce+dg & cf+dh \end{pmatrix}.$$

The rule for matrix multiplication can be described as follows: the element in the ith row and jth column of the product is obtained by multiplying each element of the ith row of the first matrix with the corresponding element of the jth column of the second matrix, and then adding up all these

products. Thus if A_{ij}, B_{ij} and M_{ij} denote the element in the ith row and jth column of matrices A, B and M respectively, where $M = AB$, then $M_{ij} = A_{i1}B_{1j} + A_{i2}B_{2j}$.

A straightforward calculation shows that multiplication of 2×2 matrices is associative (i.e., $(AB)C = A(BC)$ for all 2×2 matrices A, B and C. However multiplication of 2×2 matrices is *not* commutative. For example,

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \text{ whereas } \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

Let I be the *identity matrix* $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. Then $AI = A = IA$ for all 2×2 matrices A.

If $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, and if $ad - bc \neq 0$, then the matrix A has a well-defined *inverse* A^{-1} with the property that $AA^{-1} = I = A^{-1}A$. The inverse matrix A^{-1} is given by the formula

$$A^{-1} = \begin{pmatrix} d/(ad-bc) & -b/(ad-bc) \\ -c/(ad-bc) & a/(ad-bc) \end{pmatrix}.$$

The quantity $ad - bc$ is known as the *determinant* of a 2×2 matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$, and is denoted by $\begin{vmatrix} a & b \\ c & d \end{vmatrix}$. We also denote the determinant of a matrix A by $\det A$. A straightforward calculation shows that $\det A \det B = \det(AB)$ for all 2×2 matrices A and B. Note that if A and B are 2×2 matrices, and if $\det A \neq 0$, then there always exist a 2×2 matrices X and Y satisfying $AX = B$ and $YA = B$. These matrices are given by $X = A^{-1}B$ and $Y = BA^{-1}$, where A^{-1} is the inverse of the matrix A.

The operations of matrix addition and matrix multiplication satisfy the *distributive law*: $C(A + B) = CA + CB$ and $(A + B)C = AC + BC$ for all 2×2 matrices A, B and C.

Matrices and Transformations

Matrices can be used to represent an important class of transformations of the plane known as *linear transformations*. These transformations include rotations about the origin $(0, 0)$ of Cartesian coordinates, and reflections in lines passing through the origin.

A transformation sending points (x, y) of the plane to points (x', y') is said to be *linear* if there exist coefficients a, b, c and d such that $x' = ax + by$ and $y' = cx + dy$. Note that these equations defining the transformation can be written in the form

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}.$$

We can therefore represent linear transformations of the plane that fix the origin by corresponding 2×2 matrices. For example an anticlockwise rotation about the origin through an angle θ sends a point (x, y) to the point

$$(x\cos\theta - y\sin\theta, x\sin\theta + y\cos\theta).$$

This rotation is therefore represented by the matrix $\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$. A straightforward calculation shows that if linear transformations S and T are represented by matrices A and B in this way, then the composition $S \circ T$ of the two transformations is represented by the product AB of the matrices. (Recall that $S \circ T$ is the transformation obtained by first applying T and then applying S.)

2 Sets

Sets

A *set* is a collection of objects; these objects are known as *elements* of the set. If an element x belongs to a set X then we denote this fact by writing $x \in X$. Sets with small numbers of elements can be specified by listing the elements of the set enclosed within 'curly brackets'. For example $\{a, b, c, d\}$ is the set consisting of the elements a, b, c and d. Two sets are equal if and only if they have the same elements.

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

Standard notations $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$ are adopted for the following sets:

- the set $\mathbb{N}$ of positive integers;
- the set $\mathbb{Z}$ of integers;
- the set $\mathbb{Q}$ of rational numbers;
- the set $\mathbb{R}$ of real numbers;
- the set $\mathbb{C}$ of complex numbers.

A set A is said to be a *subset* of a set B if every element of A is also an element of B. If A is a subset of B but is not equal to B, then we say that A is a *proper subset* of B. If A is a subset of a set B then we denote this fact by writing $A \subset B$. Note that $A = B$ if and only if $A \subset B$ and $B \subset A$.

Given a set X and a *condition* that may or may not be satisfied by elements of X, the subset of X consisting of all elements of X that satisfy the stated *condition* is represented using the notation

$$\{x \in X : \textit{condition}\}.$$

Thus for example $\{n \in \mathbb{Z} : n > 0\}$ is the subset of the set $\mathbb{Z}$ of integers which consists of all strictly positive integers. (In certain contexts it is possible to simplify the above notation to $\{x : \textit{condition}\}$ if it is clear from the context what the set is to which the elements x in question belong.)

Let a and b be real numbers satisfying $a \leq b$. Then intervals in the set of real numbers are denoted using the following standard notation:

- $[a, b]$ denotes the set $\{x \in \mathbb{R} : a \leq x \leq b\}$;
- (a, b) denotes the set $\{x \in \mathbb{R} : a < x < b\}$;
- $[a, b)$ denotes the set $\{x \in \mathbb{R} : a \leq x < b\}$;
- $(a, b]$ denotes the set $\{x \in \mathbb{R} : a < x \leq b\}$;
- $[a, +\infty)$ denotes the set $\{x \in \mathbb{R} : x \geq a\}$;
- $(a, +\infty)$ denotes the set $\{x \in \mathbb{R} : x > a\}$;
- $(-\infty, a]$ denotes the set $\{x \in \mathbb{R} : x \leq a\}$;
- $(-\infty, a)$ denotes the set $\{x \in \mathbb{R} : x < a\}$.

The *union*, *intersection* and *difference* of two sets are defined as follows:—

- the *union* $X \cup Y$ of two sets X and Y is the set consisting of all elements that belong to X or to Y (or to both);
- the *intersection* $X \cap Y$ of two sets X and Y is the set consisting of all elements that belong to both X and Y;
- the *difference* $X \setminus Y$ of two sets X and Y is the set consisting of all elements that belong to X but not to Y.

The sets X and Y are said to be *disjoint* if no element belongs to both X and Y (i.e., $X \cap Y = \emptyset$.)

Note that $X \cup Y$ is the union of the three sets $X \cap Y$, $X \setminus Y$ and $Y \setminus X$. Moreover these three sets are pairwise disjoint (i.e., each pair is disjoint).

We can also consider unions and intersections of more than two sets. The *union* of a given collection of sets is the set consisting of all elements that belong to at least one of the given sets. The *intersection* of a given collection of sets is the set consisting of all elements that belong to every one of the given sets.

Let $X_1, X_2, X_3, \ldots, X_n$ be sets. We denote the union and intersection of these sets by $X_1 \cup X_2 \cup X_3 \cup \cdots \cup X_n$ and $X_1 \cap X_2 \cap X_3 \cap \cdots \cap X_n$ respectively.

The union and intersection of an infinite sequence $X_1, X_2, X_3, \ldots$ of sets are denoted by $\bigcup_{i=1}^{\infty} X_i$ and by $\bigcap_{i=1}^{\infty} X_i$ respectively. More generally, given any collection $\mathcal{C}$ of sets, the union and intersection of the sets in the collection are denoted by $\bigcup_{X \in \mathcal{C}} X$ and $\bigcap_{X \in \mathcal{C}} X$ respectively.

Let X be a set, and let A be a subset of X. The *complement* of A (in X) is the set $X \setminus A$ of all elements of X that do not belong to A.

For each subset A of a given set X, let A^c denote the complement of A in X. Then $(A \cup B)^c = A^c \cap B^c$ and $(A \cap B)^c = A^c \cup B^c$ for all subsets A and B of X. These identities generalize to situations where the number of subsets of X involved is greater than two: the complement of the intersection of any collection of subsets of X is the union of the complements of those subsets; the complement of the union of any collection of subsets of X is the intersection of the complements of those subsets.

Cartesian Products of Sets

Let X and Y be sets. An element x of X and an element y of Y together specify an *ordered pair* (x, y). Ordered pairs (x, y) are characterized by the following property:

$(x, y) = (u, v)$ if and only if $x = u$ and $y = v$.

The set of all ordered pairs (x, y) with $x \in X$ and $y \in Y$ is referred to as the *Cartesian product* of the sets X and Y, and is denoted by $X \times Y$.

Example. The Cartesian product $\mathbb{R} \times \mathbb{R}$ consists of all ordered pairs (x, y) where x and y are real numbers. This set is denoted by $\mathbb{R}^2$.

Example. Let $X = \{1, 2, 3\}$ and $Y = \{2, 4\}$. Then

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

The Cartesian product $X_1 \times X_2 \times X_3 \times \cdots \times X_n$ of n sets $X_1, X_2, X_3, \ldots, X_n$ consists of all ordered n-tuples $(x_1, x_2, \ldots, x_n)$ with $x_i \in X_i$ for $i = 1, 2, 3, \ldots, n$.

Example. Points of 3-dimensional space are represented with respect to a Cartesian co-ordinate system as ordered triples (x, y, z), where x, y and z are real numbers. The set of all such ordered triples is the Cartesian product $\mathbb{R} \times \mathbb{R} \times \mathbb{R}$ (denoted by $\mathbb{R}^3$).

Note that if X_i is a finite set with m_i elements for $i = 1, 2, \ldots, n$, then the Cartesian product $X_1 \times X_2 \times X_3 \times \cdots \times X_n$ has $m_1 m_2 m_3 \cdots m_n$ elements.

Relations

Let X be a set. A *binary relation* on X determines, for elements u and v of X, whether or not u is related to v. For example, there is a binary relation on the set of real numbers, where two real numbers x and y are related if and only if x is less than y.

It is traditional to denote binary relations by inserting the symbol for the relation between any two elements that are related. Thus if $\sim$ is a relation on a set X then $u \sim v$ is true for elements u and v of X if and only if u and v are related. Familiar examples of this notation are provided by the relations $=$ ('equals'), $<$ ('less than') and $\leq$ ('less than or equal to') on sets of numbers.

Any binary relation $\sim$ on a set X determines a corresponding subset $\{(u, v) \in X \times X : u \sim v\}$ of the Cartesian product $X \times X$. Conversely any subset R of $X \times X$ determines a corresponding relation $\sim$ on X, where elements u and v of X satisfy $u \sim v$ if and only if $(u, v) \in R$. There is thus a one-to-one correspondence between binary relations on a set X and subsets of $X \times X$.

Equivalence Relations

Let $\sim$ be a binary relation on a set S.

- The relation $\sim$ is *reflexive* on S if the following is true: $x \sim x$ for all elements x of S.
- The relation $\sim$ is *symmetric* on S if the following is true: if x and y are elements of S and if $x \sim y$ then $y \sim x$.
- The relation $\sim$ is *transitive* on S if the following is true: if x, y and z are elements of S and if if $x \sim y$ and $y \sim z$ then $x \sim z$.

Example. The relation $=$ (i.e., 'is equal to') is reflexive, symmetric and transitive on any set.

Example. The relation $<$ (i.e., 'is less than') is transitive on the set of real numbers but is neither reflexive nor symmetric.

Example. The relation $\leq$ (i.e., 'is less than of equal to') is reflexive and transitive on the set of real numbers but is not symmetric.

Example. The relation $\neq$ (i.e., 'is not equal to') is symmetric on the set of real numbers but is neither reflexive nor transitive.

Example. The relation 'has the same number of elements as' is reflexive, symmetric and transitive on any collection of finite sets.

Definition. An *equivalence relation* on a given set is a binary relation on that set which is reflexive, symmetric and transitive.

The relation of equality is an equivalence relation on any set.

The relation $<$ (i.e., 'is less than') is not an equivalence relation on the set of real numbers because it is neither reflexive nor symmetric.

Definition. Let $\sim$ be an equivalence relation on a set X. The *equivalence class* of x in X (with respect to the equivalence relation $\sim$) is the set C_x consisting of all elements of X that are related to x. Thus

$$C_x = \{y \in X : x \sim y\}.$$

Lemma 2.1. *Let $\sim$ be an equivalence relation on a set X, and, for each $x \in X$, let C_x denote equivalence class of x, defined by*

$$C_x = \{y \in X : x \sim y\}.$$

Then the following are true:

(i) *$x \in C_x$ for all $x \in X$;*

(ii) *$y \in C_x$ if and only if $C_x = C_y$;*

(iii) *if x and y are elements of X and if $C_x \cap C_y$ is non-empty, then $C_x = C_y$;*

(iv) *an element x of X belongs to exactly one equivalence class.*

Proof. The fact that $x \in C_x$ for all $x \in X$ follows immediately from the fact that any equivalence relation is required to be reflexive. This proves (i).

Suppose that $y \in C_x$. Then $x \sim y$. Also $y \sim x$, since any equivalence relation is transitive. If $z \in C_y$ then $x \sim y$ and $y \sim z$, and hence $x \sim z$, since any equivalence relation is transitive. It follows that if $z \in C_y$ then $z \in C_x$, and thus $C_y \subset C_x$. Similarly $C_x \subset C_y$. Thus if $y \in C_x$ then $C_x = C_y$. Conversely if $C_x = C_y$ then $y \in C_x$, since $y \in C_y$. This proves (ii).

Next note that if x and y are elements of X and if $C_x \cap C_y$ is non-empty, then there exists some element z of X such that $z \in C_x$ and $z \in C_y$. It follows from (ii) that $C_x = C_z$ and $C_y = C_z$, and therefore $C_x = C_y$. This proves (iii).

Finally (iv) is a consequence of (i) and (iii). ∎

Definition. Let X be a set. A *partition* of X is a collection of subsets of X with the property that every element of X belongs to exactly one of these subsets.

Let an equivalence relation be given on a set X. Then the collection of equivalence classes constitutes a partition of X. Conversely any partition of a set X determines an equivalence relation, where two elements of X are related if and only if they belong to the same subset in the partition.

Congruence

Let n be a positive integer. Two integers x and y are said to be *congruent* modulo n if $x - y$ is divisible by n. If x and y are congruent modulo n then we write $x \equiv y \bmod n$ to denote this fact; if x and y are not congruent modulo n then we write $x \not\equiv y \bmod n$.

Lemma 2.2. *Let n be a positive integer. Then the relation of congruence modulo n is an equivalence relation on the set $\mathbb{Z}$ of integers.*

Proof. Let x be an integer. Then $x \equiv x \bmod n$. Thus the relation of congruence modulo n is reflexive.

If x and y are integers, and if $x \equiv y \bmod n$ then $y \equiv x \bmod n$, for if $x - y$ is divisible by n then so is $y - x$. Thus the relation of congruence modulo n is symmetric.

Suppose that x, y and z are integers and that $x \equiv y \bmod n$ and $y \equiv z \bmod n$. Then $x - y$ and $y - z$ are divisible by n. But then $x - z$ is divisible by n, since $x - z = (x - y) + (y - z)$. It follows that $x \equiv z \bmod n$. Thus the relation of congruence modulo n is transitive.

The relation of congruence modulo n is reflexive, symmetric and transitive, and is thus an equivalence relation on the set of integers. ∎

Let n be a positive integer. The equivalence class of an integer x under the relation of congruence modulo n is referred to as the *congruence class* of x modulo n; this congruence class consists of all integers y for which $x - y$ is divisible by n. Every integer belongs to exactly one congruence class modulo n. The set of congruence classes of integers modulo n is denoted by $\mathbb{Z}_n$. It is easy to see that any integer is congruent to exactly one of the integers $0, 1, 2, \ldots, n-1$ modulo n. It follows that $\mathbb{Z}_n$ has n elements.

Let us denote by $[x]_n$ the congruence class of an integer x modulo a positive integer n. Then $[x]_n = [y]_n$ if and only if $x \equiv y \bmod n$.

Lemma 2.3. *Let n be a positive integer, and let x, y, u and v be integers. Suppose that $x \equiv u \bmod n$ and $y \equiv v \bmod n$. Then $x + y \equiv u + v \bmod n$ and $xy \equiv uv \bmod n$.*

Proof. The numbers $x - u$ and $y - v$ are divisible by n, since $x \equiv u \bmod n$ and $y \equiv v \bmod n$. It follows that $(x + y) - (u + v)$ is divisible by n, since

$$(x + y) - (u + v) = (x - u) + (y - v).$$

Also $xy - uv$ is divisible by n, since

$$xy - uv = (x - u)y + u(y - v).$$

Thus $x + y \equiv u + v \bmod n$ and $xy \equiv uv \bmod n$. ∎

Let n be a positive integer. Lemma 2.3 ensures that there are well-defined operations of addition and multiplication on the set $\mathbb{Z}_n$ of equivalence classes of integers modulo n. These operations are defined so that $[x]_n + [y]_n = [x + y]_n$ and $[x]_n[y]_n = [xy]_n$ for all integers x and y. Clearly these operations are commutative and associative, and multiplication on $\mathbb{Z}_n$ is distributive over addition.

When n is small the algebraic operations on $\mathbb{Z}_n$ can be tabulated. Here is the multiplation table for $\mathbb{Z}_6$:

$\times$	$[0]_6$	$[1]_6$	$[2]_6$	$[3]_6$	$[4]_6$	$[5]_6$
$[0]_6$	$[0]_6$	$[0]_6$	$[0]_6$	$[0]_6$	$[0]_6$	$[0]_6$
$[1]_6$	$[0]_6$	$[1]_6$	$[2]_6$	$[3]_6$	$[4]_6$	$[5]_6$
$[2]_6$	$[0]_6$	$[2]_6$	$[4]_6$	$[0]_6$	$[2]_6$	$[4]_6$
$[3]_6$	$[0]_6$	$[3]_6$	$[0]_6$	$[3]_6$	$[0]_6$	$[3]_6$
$[4]_6$	$[0]_6$	$[4]_6$	$[2]_6$	$[0]_6$	$[4]_6$	$[2]_6$
$[5]_6$	$[0]_6$	$[5]_6$	$[4]_6$	$[3]_6$	$[2]_6$	$[1]_6$

A congruence class in the body of such a table represents the sum or product of the congruence classes labelling the row and the column in which it occurs. For example, we can read off from the above table that $[2]_6 \times [5]_6 = [4]_6$.

3 Functions

Let X and Y be sets. A *function* $f: X \to Y$ from X to Y assigns to each element x of the set X exactly one element $f(x)$ of the set Y. The set X is the *domain* of the function, and the set Y is the *co-domain* of the function.

The notation $f: X \to Y$ is used to specify a function f whose domain is the set X and whose co-domain is the set Y.

A function is not fully specified unless its domain and co-domain are specified.

Example. Let us consider 'the function that sends x to $1/x^2$'. Note that $1/x^2$ is is not defined when $x = 0$. Therefore we cannot view this 'function' as a function on the set of real numbers. We can however take as the domain of the function the set $\mathbb{R} \setminus \{0\}$ of all non-zero real numbers. We thus obtain a function $f: \mathbb{R} \setminus \{0\} \to \mathbb{R}$ from the set $\mathbb{R} \setminus \{0\}$ of non-zero real numbers to the set $\mathbb{R}$ of real numbers, where $f(x) = 1/x^2$ for all non-zero real numbers x.

There is also a function $g: \mathbb{C} \setminus \{0\} \to \mathbb{C}$ from the set $\mathbb{C} \setminus \{0\}$ of non-zero complex numbers to the set $\mathbb{C}$ of complex numbers, where $g(z) = 1/z^2$ for all non-zero complex numbers z. The functions f and g have different domains, and are therefore considered to be different functions.

Note that there is no element x of the domain $\mathbb{R} \setminus \{0\}$ of $f: \mathbb{R} \setminus \{0\} \to \mathbb{R}$ for which $f(x) = 0$. Also $f(x) = f(-x)$ for all non-zero real numbers x. Thus, given an element y of the co-domain $\mathbb{R}$ of the function f, there need not be exactly one element x of the domain satisfying $f(x) = y$. There may not be any such elements x, as is the case when $y < 0$, or there may be more than one such element x, as is the case when $y > 0$.

Let X be a set. There is a function $i: X \to X$ from X to itself, where $i(x) = x$ for all $x \in X$. This function is referred to as the *identity map* of X.

Let $f: X \to Y$ be a function from a set X to a set Y. The *range* $f(X)$ of the function is defined to be the set $\{f(x) : x \in X\}$ of all elements of the co-domain Y that are of the form $f(x)$ for some element x of the domain. The *image* $f(A)$ of a subset A of X is defined to be the set $\{f(x) : x \in A\}$ of all elements of the co-domain Y that are of the form $f(x)$ for some element x of A. Also the *preimage* $f^{-1}(B)$ of a subset B of Y is defined to be the set $\{x \in X : f(x) \in B\}$ of all elements of x for which $f(x)$ is an element of B.

Note that the range of a function $f: X \to Y$ is the image $f(X)$ of the domain X of the function. Also $f^{-1}(Y) = X$.

Example. Let $f: \mathbb{R} \to \mathbb{R}$ be the function defined by $f(x) = x^2$ for all $x \in \mathbb{R}$. The range of f is the set $[0, +\infty)$ of non-negative real numbers. The image $f([1, 2])$ of the interval $[1, 2]$ is the interval $[1, 4]$. The preimage $f^{-1}([1, 4])$ of the interval $[1, 4]$ is the union $[-2, -1] \cup [1, 2]$ of the intervals $[1, 2]$ and $[-2, -1]$.

Let X, Y and Z be sets, and let $f: X \to Y$ and $g: Y \to Z$ be functions, where the domain Y of $g: Y \to Z$ is the co-domain of $f: X \to Y$. The composition function $g \circ f: X \to Z$ is defined by $(g \circ f)(x) = g(f(x))$ for all $x \in X$. Note that $g \circ f$ denotes the function 'f followed by g'.

Injective, Surjective and Bijective Functions

We now define *injective*, *surjective* and *bijective* functions:—

- a function $f: X \to Y$ is said to be *injective* (or *one-to-one*) if $f(u) \neq f(v)$ whenever u and v are elements of the domain X with $u \neq v$;

- a function $f: X \to Y$ is said to be *surjective* (or *onto*) if each element of the codomain of the function is the image $f(x)$ of at least one element x of the domain X;
- a function $f: X \to Y$ is said to be *bijective* (or is said to be a *one-to-one correspondence*) if it is both injective and surjective.

Injective, surjective and bijective functions are also referred to as *injections*, *surjections* and *bijections* respectively.

Note that a function $f: X \to Y$ is bijective if and only if, given any element y of the co-domain Y of the function, there exists exactly one element x of the domain X satisfying $f(x) = y$.

Example. Let $\mathbb{N}$ denote the set $\{1, 2, 3, 4, \ldots\}$ of positive integers. Let $f: \mathbb{N} \to \mathbb{N}$ be the function defined by $f(n) = n^2$ for all positive integers n. This function is injective, for if m and n are positive integers and if $m \neq n$ then $m^2 \neq n^2$. The function is not surjective, since there is no positive integer n satisfying $f(n) = 3$.

Example. Let $g: \mathbb{R} \to [0, +\infty)$ be the function from the set $\mathbb{R}$ of real numbers to the set $[0, +\infty)$ of non-negative real numbers that sends each real number x to x^2. This function is not injective, since $g(2) = g(-2) = 4$. It is surjective: for any non-negative real number y, there is a real number $\sqrt{y}$ satisfying $g(\sqrt{y}) = y$.

Example. Let $h: \mathbb{N} \to \mathbb{N}$ be the function from the set $\mathbb{N}$ of positive integers to itself defined by

$$h(n) = \begin{cases} n+1 & \text{if } n \text{ is odd;} \\ n-1 & \text{if } n \text{ is even.} \end{cases}$$

Thus $h(1) = 2$, $h(2) = 1$, $h(3) = 4$, $h(4) = 3$, etc. The function is injective. Indeed let m and n be positive integers with $m \neq n$. If m is odd and n is even then $h(m) \neq h(n)$, since $h(m)$ is even and $h(n)$ is odd. If m is even and n is odd then $h(m) \neq h(n)$, since $h(m)$ is odd and $h(n)$ is even. If m and n are both odd then $h(m) \neq h(n)$ since $h(m) = m+1$, $h(n) = n+1$ and $m+1 \neq n+1$. If m and n are both even then $h(m) \neq h(n)$ since $h(m) = m-1$, $h(n) = n-1$ and $m-1 \neq n-1$. We have thus verified that $h(m) \neq h(n)$ for all positive integers m and n satisfying $m \neq n$. Thus the function is injective.

Let n be a positive integer. If n is odd then $n = h(n+1)$. If n is even then $n = h(n-1)$. Thus the function is surjective.

The function $h: \mathbb{N} \to \mathbb{N}$ is therefore bijective.

Lemma 3.1. *Let X, Y and Z be sets, and let $f: X \to Y$ and $g: Y \to Z$ be functions.*

(i) *If $f: X \to Y$ and $g: Y \to Z$ are injective, then so is $g \circ f: X \to Z$.*

(ii) *If $f: X \to Y$ and $g: Y \to Z$ are surjective, then so is $g \circ f: X \to Z$.*

(iii) *If $f: X \to Y$ and $g: Y \to Z$ are bijective, then so is $g \circ f: X \to Z$.*

Proof. First suppose that $f: X \to Y$ and $g: Y \to Z$ are injective. We must prove that $g \circ f: X \to Z$ is injective. Let u and v be elements of X with $u \neq v$. Then $f(u) \neq f(v)$, since $f: X \to Y$ is injective. But then $g(f(u)) \neq g(f(v))$, since $g: Y \to Z$ is injective. It follows that $g \circ f: X \to Z$ is injective. This proves (i).

Next suppose that $f: X \to Y$ and $g: Y \to Z$ are surjective. We must prove that $g \circ f: X \to Z$ is surjective. Let z be an element of Z. Then there exists $y \in Y$ satisfying $g(y) = z$, since $g: Y \to Z$ is surjective. Then there exists $x \in X$ satisfying $f(x) = y$, since $f: X \to Y$ is surjective. But then $g(f(x) = z$. It follows that $g \circ f: X \to Z$ is surjective. This proves (ii).

Clearly (iii) follows from (i) and (ii). ∎

Inverse Functions

Definition. Let X and Y be sets, and let $f: X \to Y$ be a function from X to Y. A function $g: Y \to X$ from Y to X is said to be the *inverse* of $f: X \to Y$ if $g(f(x)) = x$ for all $x \in X$ and $f(g(y)) = y$ for all $y \in Y$.

We denote by $f^{-1}: Y \to X$ the inverse of a function $f: X \to Y$, provided that such an inverse exists.

Example. Consider the function $f: [1, 2] \to [1, 4]$, where $f(x) = x^2$ for all $x \in [1, 2]$. The inverse of this function is the function $g: [1, 4] \to [1, 2]$, where $g(y) = \sqrt{y}$ for all $y \in [1, 4]$.

Example. Consider the function $h: \mathbb{R} \to \mathbb{R}$, where $h(x) = x^2$ for all real numbers x. This function does not have an well-defined inverse. Indeed no function $k: \mathbb{R} \to \mathbb{R}$ has the property that $y = h(k(y))$ for all real numbers y, since this identity clearly cannot be satisfied when $y < 0$.

Lemma 3.2. *Let X and Y be sets. A function $f: X \to Y$ has a well-defined inverse if and only if it is a bijection. Moreover the inverse of a bijection is itself a bijection.*

Proof. Let $f: X \to Y$ be a function which has a well-defined inverse $f^{-1}: Y \to X$. Let u and v be elements of X. Then $u = f^{-1}(f(u))$ and $v = f^{-1}(f(v))$. Thus if $u \neq v$ then $f(u) \neq f(v)$. The function $f: X \to Y$ is therefore injective. The function $f: X \to Y$ is also surjective, since $y = f(f^{-1}(y))$ for all $y \in Y$. We have thus shown that if a function $f: X \to Y$ has a well defined inverse then it is both injective and surjective, and is thus a bijection.

Conversely suppose that $f: X \to Y$ is a bijection. Then, given any element y of Y, there exists exactly one element x of X satisfying $f(x) = y$. We therefore define $f^{-1}(y)$ to be the unique element x of X satisfying $f(x) = y$. Clearly $f(f^{-1}(y)) = y$ for all $y \in Y$. Thus $f \circ f^{-1}$ is the identity map of Y. We must also show that $f^{-1} \circ f$ is the identity map of X. Let x be an element of X. Then $f(f^{-1}(f(x))) = f(x)$, since $f \circ f^{-1}$ is the identity map of Y. But $f: X \to Y$ is injective. It follows that $f^{-1}(f(x)) = x$, since the elements x and $f^{-1}(f(x))$ are mapped by f to the same element of the set Y. We have thus shown that if the function $f: X \to Y$ is a bijection then it has a well-inverse.

If $g: Y \to X$ is the inverse of a bijection $f: X \to Y$ then f is the inverse of g, and therefore $g: Y \to X$ must be a bijection. ∎

4 Permutations

A *permutation* of a set S is a bijective function $p: S \to S$ from S to itself.

The *identity* permutation of a set S is the permutation that fixes every element of S.

Permutations of a finite set S are conveniently represented in a two row form

$$\begin{pmatrix} x_1 & x_2 & \dots & x_n \\ p(x_1) & p(x_2) & \dots & p(x_n) \end{pmatrix},$$

where $x_1, x_2, \dots, x_n$ are the elements of the set S and $p(x_1), p(x_2), \dots, p(x_n)$ are the images of these elements under the permutation p being represented. Thus for example

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}$$

represents the permutation of the set $\{1, 2, 3\}$ that sends 1 to 2, sends 2 to 3, and sends 3 to 1.

Example. There are two permutations of a set $\{a, b\}$ with two elements. These are the identity permutation $\begin{pmatrix} a & b \\ a & b \end{pmatrix}$ and the transposition $\begin{pmatrix} a & b \\ b & a \end{pmatrix}$ that interchanges the elements a and b. Indeed if p is a permutation of $\{a, b\}$ and if $p(a) = a$, then $p(b) \neq a$ (since a must be the image of exactly one element of the set), hence $p(b) = b$ and thus p is the identity permutation. Similarly if $p(a) = b$ then $p(b) \neq b$ and therefore $p(b) = a$, and thus p is the transposition that interchanges the elements a and b.

Example. There are six permutations of a set $\{a, b, c\}$ with three elements. These are

$$\begin{pmatrix} a & b & c \\ a & b & c \end{pmatrix}, \quad \begin{pmatrix} a & b & c \\ a & c & b \end{pmatrix},$$

$$\begin{pmatrix} a & b & c \\ b & a & c \end{pmatrix}, \quad \begin{pmatrix} a & b & c \\ b & c & a \end{pmatrix},$$

$$\begin{pmatrix} a & b & c \\ c & a & b \end{pmatrix}, \quad \begin{pmatrix} a & b & c \\ c & b & a \end{pmatrix}.$$

Indeed it is easy to see that the first two include all permutations that send a to itself, the next two include all permutations that send a to b, and the remaining two include all permutations that send a to c.

Lemma 4.1. *Let S be a set. Then the composition $q \circ p$ of any two permutations of S is itself a permutation of S. Also any permutation p of S has a well-defined inverse p^{-1}.*

Proof. The composition of two bijections is a bijection (Lemma 3.1). Therefore the composition of two permutations is a permutation.

A function is a bijection if and only if it has a well-defined inverse (Lemma 3.2). It follows that any permutation has a well-defined inverse. ∎

Let p, q and r be permutations of a set S. Then $(r \circ q) \circ p = r \circ (q \circ p)$, since

$$(p \circ (q \circ r))(x) = (p(q \circ r)(x)) = p(q(r(x)) = (p \circ q)(r(x)) = ((p \circ q) \circ r)(x)$$

for any element x of S. Thus composition of permutations is associative.

Let S be a set, and let $a_1, a_2, \ldots, a_n$ be distinct elements of S. We denote by $(a_1\, a_2 \cdots a_n)$ the permutation of S that sends a_i to a_{i+1} for $i = 1, 2, \ldots, n-1$, sends a_n to a_1, and fixes all other elements of S. Such a permutation is called a *cycle* of order n, or *n-cycle*. A cycle of length 2 is also called a *transposition*.

(Note that evaluating a composition of cycles, we shall compose them from right to left, in accordance with standard practice when composing functions.)

Example. We list all permutations of a set $\{a, b, c, d\}$ with exactly four elements. There is the identity permutation that fixes every element of the set. There are six transpositions. These are $(a\, b)$, $(a\, c)$, $(a\, d)$, $(b\, c)$, $(b\, d)$ and $(c\, d)$, where

$$(a\, b) = \begin{pmatrix} a & b & c & d \\ b & a & c & d \end{pmatrix}, \quad (a\, c) = \begin{pmatrix} a & b & c & d \\ c & b & a & d \end{pmatrix}, \text{ etc.}$$

There are eight 3-cycles. These are

$$(b\, c\, d), \quad (b\, d\, c), \quad (a\, c\, d), \quad (a\, d\, c),$$

$$(a\,b\,d), \quad (a\,d\,b), \quad (a\,b\,c), \quad (a\,c\,b),$$

where

$$(b\,c\,d) = \begin{pmatrix} a & b & c & d \\ a & c & d & b \end{pmatrix}, \quad (b\,d\,c) = \begin{pmatrix} a & b & c & d \\ a & d & b & c \end{pmatrix}, \text{ etc.}$$

There are six 4-cycles. These are

$$(a\,b\,c\,d), \quad (a\,b\,d\,c), \quad (a\,c\,b\,d),$$

$$(a\,c\,d\,b), \quad (a\,d\,b\,c), \quad (a\,d\,c\,b).$$

We now show that there are three more permutations of the set. Let p be a permutation of $\{a, b, c, d\}$. Now if p fixes any element of the set then it must permute the remaining three elements amongst themselves and must therefore be the identity permutation, a transposition or a 3-cycle. (All six permutations of a set with three elements are of this type.) Also if $p(a) = u$ and $p(u) = v$, where a, u and v are all distinct, then it is not difficult to verify that p must be either a 3-cycle or a 4-cycle. Thus if p is a permutation of the set which is not the identity permutation and is not a cycle, and if $p(a) = u$, then $p(u) = a$, and p must also transpose the other two elements of the set. It follows that there are three permutations of the set $\{a, b, c, d\}$ that are not the identity permutation and are not cycles. These are

$$\begin{pmatrix} a & b & c & d \\ b & a & d & c \end{pmatrix}, \quad \begin{pmatrix} a & b & c & d \\ c & d & a & b \end{pmatrix}, \quad \begin{pmatrix} a & b & c & d \\ d & c & b & a \end{pmatrix}.$$

Note that

$$\begin{pmatrix} a & b & c & d \\ b & a & d & c \end{pmatrix} = (a\,b) \circ (c\,d);$$
$$\begin{pmatrix} a & b & c & d \\ c & d & a & b \end{pmatrix} = (a\,c) \circ (b\,d);$$
$$\begin{pmatrix} a & b & c & d \\ d & c & b & a \end{pmatrix} = (a\,d) \circ (b\,c).$$

We have now found all 24 permutations of a set with four elements: the identity permutation, six transpositions, eight 3-cycles, six 4-cycles, and three further permutations.

Two cycles $(a_1\,a_2\,\cdots\,a_m)$ and $(b_1\,b_2\,\cdots\,b_n)$ are said to be *disjoint* when the elements $a_1, a_2, \ldots, a_m$ and $b_1, b_2, \ldots, b_n$ are distinct (i.e., no pair of these elements coincide).

It is easy to see that if $(a_1\,a_2\,\cdots\,a_m)$ and $(b_1\,b_2\,\cdots\,b_n)$ are disjoint cycles then

$$(a_1\,a_2\,\cdots\,a_m) \circ (b_1\,b_2\,\cdots\,b_n) = (b_1\,b_2\,\cdots\,b_n) \circ (a_1\,a_2\,\cdots\,a_m).$$

Proposition 4.2. *Any permutation of a finite set S is the identity permutation, a cycle, or a composition of two or more disjoint cycles.*

Proof. We prove the result by induction on the number of elements in the set S (using the Principle of Complete Induction). The result is trivially true if S has only one element, since in this case the only permutation of S is the identity permutation. Suppose that the result is known to be true for all permutations of sets with fewer than k elements. We show that the result then holds for all permutations of sets with k elements.

Let S be a set with k elements and let p be a permutation of S. Choose an element a_1 of S, and let elements $a_2, a_3, a_4, \ldots$ of S be defined by the requirement that $p(a_i) = a_{i+1}$ for all positive integers i.

Let n be the largest positive integer for which the elements $a_1, a_2, \ldots, a_n$ of S are distinct. We claim that $p(a_n) = a_1$.

Now the choice of n ensures that the elements $a_1, a_2, \ldots, a_n, a_{n+1}$ are not distinct. Therefore $a_{n+1} = a_j$ for some positive integer j between 1 and n. If j were greater than one then we would have $a_j = p(a_{j-1})$ and $a_j = p(a_n)$, which is impossible since if p is a permutation of S then exactly one element of S must be sent to a_j by p. Therefore $j = 1$, and thus $p(a_n) = a_1$. Let $\sigma_1 = (a_1\, a_2 \cdots a_n)$.

Let T be the set $S \setminus \{a_1, a_2, \ldots, a_n\}$ consisting of all elements of S other than $a_1, a_2, \ldots, a_n$. Now $a_1 = p(a_n)$, and $a_i = p(a_{i-1})$ for $i = 2, 3, \ldots, n$. Thus if $x \in T$ then $p(x) \neq a_i$ for $i = 1, 2, \ldots, n$ (since the function $p\colon S \to S$ is injective), and therefore $p(x) \in T$. We can therefore define a function $q\colon T \to T$, where $q(x) = p(x)$ for all $x \in T$. This function has a well-defined inverse $q^{-1}\colon T \to T$ where $q^{-1}(x) = p^{-1}(x)$ for all $x \in T$. It follows that $q\colon T \to T$ is a permutation of T. The induction hypothesis ensures that this permutation is the identity permutation of T, or is a cycle, or can be expressed as a composition of two or more disjoint cyles. These cycles extend to permutations of S that fix the elements $a_1, a_2, \ldots, a_n$, and these permutations of S are also cycles. It follows that either $p = \sigma_1$ (and q is the identity permutation of T), or else $p = \sigma_1 \circ \sigma_2 \ldots \circ \sigma_m$, where $\sigma_2, \sigma_3, \ldots, \sigma_m$ are disjoint cycles of S that fix $a_1, a_2, \ldots, a_n$ and correspond to cycles of T. Thus if the result holds for permutations of sets with fewer than k elements, then it holds for permutations of sets with k elements. It follows by induction on k that the result holds for permutations of finite sets. ∎

Recall that a *transposition* is a permutation $(a\, b)$ of a set S that interchanges two elements a and b of S and fixes the remaining elements.

Lemma 4.3. *Every permutation of a finite set with more than one element can be expressed as a finite composition of transpositions.*

Proof. Each cycle can be expressed as a composition of transpositions. Indeed if $a_1, a_2, \ldots, a_n$ are distinct elements of a finite set S then

$$(a_1\, a_2 \cdots a_n) = (a_1\, a_2) \circ (a_2\, a_3) \circ \cdots \circ (a_{n-1}\, a_n).$$

It follows from Proposition 4.2 that a permutation of S that is not the identity permutation can be expressed as a finite composition of transpositions. Moreover the identity permutation of S can be expressed as the composition of any transposition with itself, provided that S has more than one element. The result follows. ∎

Theorem 4.4. *A permutation of a finite set cannot be expressed in one way as a composition of an odd number of transpositions and in another way as a composition of an even number of transpositions.*

Proof. Let n be an integer greater than one. Given any permutation p of the set $\{1, 2, \ldots, n\}$ we define $\epsilon(p) = (-1)^{m(p)}$, where $m(p)$ is the number of ordered pairs (i, j) of integers between 1 and n satisfying $i < j$ and $p(i) > p(j)$.

Let p and q be permutations of the set $\{1, 2, \ldots, n\}$. We claim that $\epsilon(p \circ q) = \epsilon(p)\epsilon(q)$. Define numbers a, b, c and d as follows:—

- let a be the number of ordered pairs (i, j) of integers between 1 and n satisfying $i < j$, $q(i) < q(j)$ and $p(q(i)) < p(q(j))$;
- let b be the number of ordered pairs (i, j) of integers between 1 and n satisfying $i < j$, $q(i) < q(j)$ and $p(q(i)) > p(q(j))$;

- let c be the number of ordered pairs (i, j) of integers between 1 and n satisfying $i < j$, $q(i) > q(j)$ and $p(q(i)) < p(q(j))$;
- let d be the number of pairs (i, j) of integers between 1 and n satisfying $i < j$, $q(i) > q(j)$ and $p(q(i)) > p(q(j))$.

Clearly $m(q) = c + d$ and $m(p \circ q) = b + d$. Now $m(p)$ can be described as the number of subsets of $\{1, 2, \ldots, n\}$ with exactly two elements where the order of these elements is reversed under the permutation p. Moreover each such subset is of the form $\{q(i), q(j)\}$ where i and j are uniquely determined integers between 1 and n satisfying $i < j$. It follows that $m(p)$ is equal to the number of ordered pairs (i, j) of integers between 1 and n satisfying $i < j$ for which either $q(i) < q(j)$ and $p(q(i)) > p(q(j))$ or else $q(i) > q(j)$ and $p(q(i)) < p(q(j))$. Thus $m(p) = b + c$. It follows that

$$\epsilon(p)\epsilon(q) = (-1)^{b+2c+d} = (-1)^{b+d} = \epsilon(p \circ q).$$

Now let t be a transposition $(k\,l)$, where $k < l$. Then the set of all ordered pairs (i, j) satisfying $i < j$ and $t(i) > t(j)$ consists of the pair (k, l) together with all pairs (k, j) and (j, l) with $k < j < l$. It follows that $m(t) = 1 + 2(l - k - 1)$, and thus $\epsilon(t) = -1$. It follows that if a permutation p of $\{1, 2, \ldots, n\}$ can be expressed a composition of g transpositions then $(-1)^g = \epsilon(p)$. Therefore p cannot be expressed in one way as a composition of an odd number of transpositions and in another way as a composition of an even number of transpositions. This result holds also for permutations of any set with n elements, for if these elements are labelled as $x_1, x_2, \ldots, x_n$ then a permutation of the set sends each element x_i to $x_{p(i)}$, where p is a corresponding permutation of the set $\{1, 2, \ldots, n\}$. ∎

A permutation of a finite set is said to be *even* if it is expressible as the composition of an even number of transpositions. A permutation of a finite set is said to be *odd* if it is expressible as the composition of an odd number of transpositions.

Any permutation of a finite set is expressible as a composition of transpositions (Lemma 4.3) and must therefore be either even or odd. However Theorem 4.4 ensures that a permutation of a finite set cannot be both even and odd.

Lemma 4.5. *An n-cycle is even if n is odd, and is odd if n is even.*

Proof. An n-cycle $(a_1, a_2, \ldots, a_n)$ is expressible as a composition of $n - 1$ transpositions, since

$$(a_1\, a_2\, \cdots\, a_n) = (a_1\, a_2) \circ (a_2\, a_3) \circ \cdots \circ (a_{n-1}\, a_n).$$

Thus an n-cycle is even if $n - 1$ is even, and is odd if $n - 1$ is odd. ∎

Example. Let us classify the permutations of a set $\{a, b, c, d\}$ of 4 elements into even and odd permutations. The identity permutation is even. The six transpositions are all odd. The eight 3-cycles are all even. The six 4-cycles are all odd. The three remaining permutations $(a\,b) \circ (c\,d)$, $(a\,c) \circ (b\,d)$ and $(a\,d) \circ (b\,c)$ are all even. Note that there are 12 even permutations and 12 odd permutations of a set with 4 elements.

5 Countable and Uncountable Sets

Definition. A set A is said to be *countable* if there exists an injection from the set A to the set $\mathbb{N}$ of positive integers. A set which is not countable is said to be *uncountable*.

(A countable set is thus one which may be put in one-to-one correspondence with some subset of the set $\mathbb{N}$ of positive integers.)

Example. Any subset A of the set $\mathbb{N}$ of positive integers is countable: the required injection is the 'inclusion map' $i\colon A \to \mathbb{N}$ defined by $i(x) = x$ for all $x \in A$.

Example. The set $\mathbb{Z}$ of integers is countable: an injection $f\colon \mathbb{Z} \to \mathbb{N}$ from $\mathbb{Z}$ to $\mathbb{N}$ is given by the formula

$$f(n) = \begin{cases} 2n & \text{if } n > 0; \\ 1 - 2n & \text{if } n \leq 0. \end{cases}$$

Example. Any subset of a countable set is countable. Indeed if A is a subset of a countable set B, and if $f\colon B \to \mathbb{N}$ is an injection mapping B into the set $\mathbb{N}$ of positive integers then the restriction $f|A\colon A \to \mathbb{N}$ is also an injection, where $(f|A)(x) = f(x)$ for all $x \in A$. In particular, any subset of the set $\mathbb{Z}$ of integers is countable.

Example. The union of two countable sets is countable. Indeed let A and B be countable sets, and let $f\colon A \to \mathbb{N}$ and $g\colon B \to \mathbb{N}$ be injections. We can obtain an injection $h\colon A \cup B \to \mathbb{N}$ by defining

$$h(x) = \begin{cases} 2f(x) & \text{if } x \in A; \\ 2g(x) - 1 & \text{if } x \in B \setminus A. \end{cases}$$

Note that the function h maps all elements of A to even numbers, and maps all elements of $B \setminus A$ to odd numbers. Thus if x and y are elements of $A \cup B$ and if $h(x) = h(y)$ then either $h(x)$ is even, in which case x and y are both elements of A, or else $h(x)$ is odd, in which case x and y are both elements of $B \setminus A$. In either case it follows that $x = y$, since the functions $f\colon A \to \mathbb{N}$ and $g\colon B \to \mathbb{N}$ are injective. Therefore the function $h\colon A \cup B \to \mathbb{N}$ is indeed injective, and hence the union $A \cup B$ of the countable sets A and B is itself countable.

Example. Any finite union of countable sets is countable. Indeed suppose that $A = A_1 \cup A_2 \cup A_3 \cup \cdots \cup A_n$, where $A_1, A_2, A_3, \ldots, A_n$ are countable sets. Then there exist injections $f_j\colon A_j \to \mathbb{N}$ for $j = 1, 2, 3, \ldots, n$. Let $f\colon A \to \mathbb{N}$ be defined by $f(x) = k(x) + nf_{k(x)}(x)$, where $k(x)$ is the smallest value of j for which $x \in A_j$. It is easy to verify that $f\colon A \to \mathbb{N}$ is injective. Thus A is countable.

Example. The Cartesian product $\mathbb{N} \times \mathbb{N}$ is countable. An injection $f\colon \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ is given by

$$f(m, n) = \tfrac{1}{2}(m + n - 2)(m + n - 1) + m$$

for each ordered pair (m, n) of positive integers m and n. The values of $f(m, n)$ are given for small values of m and n in the following table, in which rows are labelled by m and columns by n:

	1	2	3	4	5	6	...
1	1	2	4	7	11	16	...
2	3	5	8	12	17		
3	6	9	13	18			
4	10	14	19				
5	15	20					
6	21						
⋮	⋮						

Lemma 5.1. *Let $A_1, A_2, A_3, \ldots$ be an infinite sequence of countable sets. Then the union $\bigcup_{n=1}^{\infty} A_n$ of the sets in the sequence is a countable set.*

Proof. For each positive integer n there exists an injection $g_n: A_n \to \mathbb{N}$, since A_n is countable. Let $A = \bigcup_{n=1}^{\infty} A_n$ be the union of the sets $A_1, A_2, A_3, \ldots$. We obtain a function $h: A \to \mathbb{N} \times \mathbb{N}$ by defining $h(x) = (k(x), g_{k(x)}(x))$, where $k(x)$ is defined to be the smallest positive integer n for which $x \in A_n$.

We now show that $h: A \to \mathbb{N} \times \mathbb{N}$ is injective. Let x and y be elements of A for which $h(x) = h(y)$, and let $n = k(x)$. Then the elements x and y both belong to A_n, since $n = k(x) = k(y)$. Moreover $g_n(x) = g_n(y)$. But $g_n: A \to \mathbb{N}$ is injective. Therefore $x = y$. We conclude that $h: A \to \mathbb{N} \times \mathbb{N}$ is indeed injective.

Now $\mathbb{N} \times \mathbb{N}$ is countable. Indeed the function $f: \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ defined by

$$f(m, n) = \tfrac{1}{2}(m + n - 2)(m + n - 1) + m$$

is injective (see the example above). The function $f \circ h: A \to \mathbb{N}$ is then an injection from A to $\mathbb{N}$, since it is the composition of two injections. Thus the union A of the sets $A_1, A_2, \ldots, A_n$ is countable, as required. ∎

Lemma 5.2. *A finite Cartesian product of countable sets is countable.*

Proof. First we show that the Cartesian product $A \times B$ of two countable sets A and B is countable. Now there exist injections $g: A \to \mathbb{N}$ and $h: B \to \mathbb{N}$ mapping the sets A and B into the set $\mathbb{N}$ of positive integers, since A and B are countable. Moreover the Cartesian product $\mathbb{N} \times \mathbb{N}$ is countable, and thus there exists an injection $f: \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ from $\mathbb{N} \times \mathbb{N}$ to $\mathbb{N}$. Let $k: A \times B \to \mathbb{N}$ be the function defined by $k(x, y) = f(g(x), h(y))$ for all $x \in A$ and $y \in B$. Suppose that $k(x, y) = k(u, v)$ where $x, u \in A$ and $y, v \in B$. Then $(g(x), h(y)) = (g(u), h(v))$, since the function $f: \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ is injective, and therefore $g(x) = g(u)$ and $h(y) = h(v)$. But then $x = u$ and $y = v$, since the functions $f: A \to \mathbb{N}$ and $g: B \to \mathbb{N}$ are injective, and thus $(x, y) = (u, v)$. Thus the function $k: A \times B \to \mathbb{N}$ is indeed injective. This shows that the Cartesian product $A \times B$ of two countable sets is countable.

Suppose that $n > 2$. Then a Cartesian product $A_1 \times A_2 \times \cdots \times A_{n-1} \times A_n$ of n sets can be regarded as the Cartesian product of the sets $A_1 \times A_2 \times \cdots \times A_{n-1}$ and A_n; this follows on representing an n-tuple $(x_1, x_2, \ldots, x_n)$ in the n-fold Cartesian product as the ordered pair $((x_1, x_2, \ldots, x_{n-1}), x_n)$, where the first component of the ordered pair belongs to the Cartesian product of the first $n-1$ sets. A straightforward induction on n now shows that a Cartesian product of a finite number of countable sets is countable, as required. ∎

Lemma 5.3. *A non-empty set A is countable if and only if there exists a surjection from the set $\mathbb{N}$ of positive integers to A.*

Proof. Suppose that there exists a surjection $g: \mathbb{N} \to A$. Then, for each element x of A there exists at least one positive integer n satisfying $g(n) = x$. We can therefore define a function $f: A \to \mathbb{N}$ which sends each element x of A to the smallest positive integer n for which $g(n) = x$. Clearly $g(f(x)) = x$ for all $x \in A$. Thus if x and y are elements of A and if $f(x) = f(y)$ then $x = g(f(x)) = g(f(y)) = y$. It follows that the function $f: A \to \mathbb{N}$ is injective, and thus the set A is countable.

Conversely suppose that the non-empty set A is countable. Then there exists an injection $f: A \to \mathbb{N}$ from A to the set $\mathbb{N}$ of positive integers. Choose an element w of A. Now, for each element n of $f(A)$ there exists a unique element x of A for which $f(x) = n$, since $f: A \to \mathbb{N}$ is an injection. It follows that

there is a function $g: \mathbb{N} \to A$ defined such that $g(n)$ is the unique element of A satisfying $f(g(n)) = n$ for each positive integer n that belongs to to $f(A)$, and $g(n) = w$ for each positive integer n that does not belong to $f(A)$. Moreover $g(f(x)) = x$ for all elements x of A, and therefore $g: \mathbb{N} \to A$ is surjective, as required. ∎

The Set of Real Numbers is Uncountable

Any non-negative real number less than one can be represented by means of a decimal expansion

$$0.d_1\, d_2\, d_3\, d_4\, d_5\, d_6\, d_7 \cdots$$

where the digits $d_1, d_2, d_3, \ldots$ are integers between zero and nine. We say that this decimal expansion has recurring nines if there is some positive integer m such that $d_i = 9$ for all $i \geq 9$. If a number can be represented by a terminating decimal expansion (in which all but finitely many of the digits are zero), then it can also be represented as a non-terminating decimal expansion with recurring nines. However every non-negative real number less than one has a unique decimal expansion without recurring nines.

Theorem 5.4. *The set of non-negative real numbers less than one is uncountable.*

Proof. Let A be the set of all non-negative real number less than 1. We prove that that A is uncountable, using the method of *reductio ad absurdum* (also known as *proof by contradiction*). We show that if the hypothesis that A is countable were adopted then this would lead to a contradiction.

Thus suppose that the set A were countable. Then there would exist a surjection $f: \mathbb{N} \to A$ (Lemma 5.3. We shall show that this would emply the existence of an element y of A that was not in the image $f(A)$ of this surjection. But the existence of such an element y is clearly impossible. Thus the hypothesis that A is countable leads to the required contradiction.

Supposing that a surjection $f: \mathbb{N} \to A$ were to exist. For each positive integer i let

$$f(i) = 0.d_{i1}\, d_{i2}\, d_{i3}\, d_{i4}\, d_{i5}\, d_{i6}\, d_{i7} \cdots$$

be the decimal expansion of $f(i)$ without recurring nines. This decimal expansion is uniquely determined. Let

$$e_j = \begin{cases} d_{jj} + 1 & \text{if } 0 \leq d_{jj} \leq 7, \\ 0 & \text{if } d_{jj} = 8 \text{ or } 9, \end{cases}$$

and let y be the element of A with decimal expansion

$$y = 0.e_1\, e_2\, e_3\, e_4\, e_5\, e_6\, e_7 \cdots.$$

Note that this decimal expansion does not involve the digit nine. Now $e_i \neq d_{ii}$ for each positive integer i. The uniqueness of decimal expansions without recurring nines would then ensure that $y \neq f(i)$ for each positive integer i, since the decimal expansions of these numbers would differ in the ith digit. Thus y would be an element of A that was not in the image of the surjection $f: \mathbb{N} \to A$. But $f(\mathbb{N}) = A$, so that we have arrived at the required contradiction. Since the hypothesis that the set A is countable leads to a contradiction, it must be the case that the set A is uncountable. ∎

Corollary 5.5. *The set of all real numbers is uncountable.*

Proof. Any subset of a countable set is countable. Therefore any set with an uncountable subset must itself be uncountable. In particular the set of real numbers must be uncountable. ∎

Problems

1. Use the Principle of Mathematical Induction to show that $\sum_{i=1}^{n} i^2 = \frac{1}{6}n(n+1)(2n+1)$ for all positive integers n.

2. Let $X = \{2, 4, 6, 8, 10\}$ and $Y = \{4, 5, 6, 7\}$. What are $X \cup Y$, $X \cap Y$, $X \setminus Y$ and $Y \setminus X$?

3. Let $\mathbb{N}$ denote the set $\{1, 2, 3, \ldots\}$ of positive integers, and let $\mathbb{Z}$ denote the set of all (positive and negative) integers. Determine which of the following functions are injective and which are surjective:

   (a) the function $f\colon \mathbb{N} \to \mathbb{N}$ defined by $f(n) = n + 1$ for all positive integers n;

   (b) the function $g\colon \mathbb{Z} \to \mathbb{Z}$ defined by $g(n) = n + 1$ for all integers n;

   (c) the function $h\colon \{1, 2, 3\} \to \{1, 2, 3\}$ defined by $h(1) = 2$, $h(2) = 3$ and $h(3) = 3$;

   (d) the function $k\colon \{1, 2, 3\} \to \{1, 2, 3\}$ defined by $k(1) = 2$, $k(2) = 3$ and $k(3) = 1$;

   (e) the function $l\colon \mathbb{Z} \to \mathbb{Z}$ defined by $l(n) = n^3$ for all integers n.

4. Give an example consisting of sets X, Y and Z and functions $f\colon X \to Y$ and $g\colon Y \to Z$ such that $g \circ f\colon X \to Z$ is bijective though neither $f\colon X \to Y$ nor $g\colon Y \to Z$ is bijective.

5. Let X and Y be non-empty sets. Prove that a function $f\colon X \to Y$ is surjective if and only if there exists a function $g\colon Y \to X$ such that $f \circ g$ is the identity map of Y.

6. Let X and Y be non-empty sets. Prove that a function $f\colon X \to Y$ is injective if and only if there exists a function $h\colon Y \to X$ such that $h \circ f$ is the identity map of X.

7. Calculate the following composition of permutations of the set $\{a, b, c, d, e\}$:

$$\begin{pmatrix} a & b & c & d & e \\ b & e & c & a & d \end{pmatrix} \circ \begin{pmatrix} a & b & c & d & e \\ a & c & e & b & d \end{pmatrix}$$

8. Calculate the permutation of $\{a, b, c, d, e\}$ resulting from the following composition of cycles:

$$(a\,b\,c) \circ (a\,d\,e) \circ (b\,e) \circ (a\,d\,e).$$

   [Remember that permutations (including cycles) are composed from right to left.]

9. Determine by inspection whether the following composition of cycles is an even or odd permutation of the set $\{a, b, c, d, e, f\}$:

$$(a\,b) \circ (d\,f\,a\,c) \circ (e\,b\,d) \circ (a\,e\,f\,d)$$

10. Let A and B be non-empty sets, and let $f\colon A \to B$ be a surjection. Suppose that A is countable. Explain why B is countable.

11. Let A be a countable set with infinitely many elements, and let $i\colon A \to \mathbb{N}$ be an injection. For each positive integer n, let $f(n)$ be the unique element a of A with the property that $i(a)$ is the nth smallest element of $i(A)$. Verify that $f\colon \mathbb{N} \to A$ is a bijection from $\mathbb{N}$ to A.