CHAPTER 0

Foundations

This text discusses the theory of finite–dimensional vector spaces in sufficient detail to enable the reader to understand and solve most linear algebra problems in mathematics and physics likely to be encountered outside of specialized research. In other words, we treat the general theory of determinants and matrices along with their relationship to linear transformations. Our approach will generally be rather abstract since we feel that most readers already have a reasonable amount of experience in visualizing vectors in three dimensions. Furthermore, we will not discuss any analytic geometry. Those readers who wish to learn something about this subject are referred to the books listed in the bibliography.

In this chapter, we briefly go through some elementary concepts from analysis dealing with numbers and functions. While most readers will probably be familiar with this material, it is worth summarizing the basic definitions that we will be using throughout this text, and thus ensure that everyone is on an equal footing to begin with. This has the additional advantage in that it also makes this text virtually self–contained and all the more useful for self–study. The reader should feel free to skim this chapter now, and return to certain sections if and when the need arises.

0.1 SETS

For our purposes, it suffices to assume that the concept of a set is intuitively clear, as is the notion of the set of integers. In other words, a **set** is a collection of objects, each of which is called a **point** or an **element** of the set. For example, the set of integers consists of the numbers 0, ±1, ±2, . . . and will be denoted by $\mathbb{Z}$. Furthermore, the set $\mathbb{Z}^+$ consisting of the numbers 1, 2, . . . will be called the set of **positive integers**, while the collection 0, 1, 2, . . . is called the set of **natural numbers** (or **nonnegative** integers). If m and $n \neq 0$ are integers, then the set of all numbers of the form m/n is called the set of **rational numbers**, and will be denoted by $\mathbb{Q}$. We shall shortly show that there exist real numbers not of this form. The most important sets of numbers that we shall be concerned with are the set $\mathbb{R}$ of real numbers and the set $\mathbb{C}$ of complex numbers (both of these sets will be discussed below).

If S and T are sets, then S is said to be a **subset** of T if every element of S is also an element of T, i.e., $x \in S$ implies $x \in T$. If in addition $S \neq T$, then S is said to be a **proper** subset of T. To denote the fact that S is a subset of T, we write $S \subset T$ (or sometimes $T \supset S$ in which case T is said to be a **superset** of S). Note that if $S \subset T$ and $T \subset S$, then $S = T$. This fact will be extremely useful in many proofs. The set containing no elements at all is called the **empty set** and will be denoted by $\varnothing$.

Next, consider the set of all elements which are members of T but not members of S. This defines the set denoted by $T - S$ and called the **complement** of S in T. (Many authors denote this set by $T \setminus S$, but we shall not use this notation.) In other words, $x \in T - S$ means that $x \in T$ but $x \notin S$. If (as is usually the case) the set T is understood and $S \subset T$, then we write the complement of S as S^c.

Example 0.1 Let us prove the useful fact that if $A, B \subset X$ with $A^c \subset B$, then it is true that $B^c \subset A$. To show this, we simply note that $x \in B^c$ implies $x \notin B$, which then implies $x \notin A^c$, and hence $x \in A$. This observation is quite useful in proving many identities involving sets. ∥

Now let $S_1, S_2, \ldots$ be a collection of sets. (Such a collection is called a **family** of sets.) For simplicity we write this collection as $\{S_i\}$, $i \in I$. The set I is called an **index set**, and is most frequently taken to be the set $\mathbb{Z}^+$. The **union** $\cup_{i \in I} S_i$ of the collection $\{S_i\}$ is the set of all elements that are members of at least one of the S_i. Since the index set is usually understood, we will simply write this as $\cup S_i$. In other words, we write

$$\cup S_i = \{x: x \in S_i \text{ for at least one } i \in I\} .$$

This notation will be used throughout this text, and is to be read as "the set of all x such that x is an element of S_i for at least one $i \in I$." Similarly, the **intersection** $\cap S_i$ of the S_i is given by

$$\cap S_i = \{x: x \in S_i \text{ for all } i \in I\} .$$

For example, if $S, T \subset X$, then $S - T = S \cap T^c$ where $T^c = X - T$. Furthermore, two sets S_1 and S_2 are said to be **disjoint** if $S_1 \cap S_2 = \varnothing$.

We now use these concepts to prove the extremely useful "**De Morgan Formulas**."

Theorem 0.1 Let $\{S_i\}$ be a family of subsets of some set T. Then

(a) $\cup S_i^c = (\cap S_i)^c$

(b) $\cap S_i^c = (\cup S_i)^c$

Proof (a) $x \in \cup S_i^c$ if and only if x is an element of some S_i^c, hence if and only if x is not an element of some S_i, hence if and only if x is not an element of $\cap S_i$, and therefore if and only if $x \in (\cap S_i)^c$.

(b) $x \in \cap S_i^c$ if and only if x is an element of every S_i^c, hence if and only if x is not an element of any S_i, and therefore if and only if $x \in (\cup S_i)^c$. ∎

While this may seem like a rather technical result, it is in fact directly useful not only in mathematics, but also in many engineering fields such as digital electronics where it may be used to simplify various logic circuits.

Finally, if $S_1, S_2, \ldots, S_n$ is a collection of sets, we may form the (ordered) set of all n–tuples $(x_1, \ldots, x_n)$ where each $x_i \in S_i$. This very useful set is denoted by $S_1 \times \cdots \times S_n$ and called the **Cartesian product** of the S_i.

Example 0.2 Probably the most common example of the Cartesian product is the plane $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$. Each point $\vec{x} \in \mathbb{R}^2$ has **coordinates** (x, y) where $x, y \in \mathbb{R}$. In order to facilitate the generalization to $\mathbb{R}^n$, we will generally write $\vec{x} = (x_1, x_2)$ or $\vec{x} = (x^1, x^2)$. This latter notation is used extensively in more advanced topics such as tensor analysis, and there is usually no confusion between writing the components of $\vec{x}$ as superscripts and their being interpreted as exponents (see Chapter 11). //

Exercises

1. Let A, B and C be sets. Prove that
   (a) $(A - B) \cap C = (A \cap C) - (B \cap C)$.
   (b) $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.
   (c) $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$.
   (d) $(A - B) - C = A - (B \cup C)$.
   (e) $A - (B \cup C) = (A - B) \cap (A - C)$.

2. The **symmetric difference** $A \Delta B$ of two sets A and B is defined by

$$A \Delta B = (A - B) \cup (B - A).$$

   Show that
   (a) $A \Delta B = (A \cup B) - (A \cap B) = B \Delta A$.
   (b) $A \cap (B \Delta C) = (A \cap B) \Delta (A \cap C)$.
   (c) $A \cup B = (A \Delta B) \Delta (A \cap B)$.
   (d) $A - B = A \Delta (A \cap B)$.

3. Let $\mathcal{R}$ be a nonempty collection of sets with the property that $A, B \in \mathcal{R}$ implies that both $A \cap B$ and $A \Delta B$ are also in $\mathcal{R}$. Show that $\mathcal{R}$ must contain the empty set, $A \cup B$ and $A - B$. (The collection $\mathcal{R}$ is called a **ring of sets**, and is of fundamental importance in measure theory and Lebesgue integration.)

0.2 MAPPINGS

Given two sets S and T, a **mapping** or **function** f from S to T is a rule which assigns a *unique* element $y \in T$ to each element $x \in S$. Symbolically, we write this mapping as $f: S \to T$ or $f: x \mapsto f(x)$ (this use of the colon should not be confused with its usage meaning "such that"). The set S is called the **domain** of f and T is called the **range** of f. Each point $f(x) \in T$ is called the **image** of x under f (or the **value** of f at x), and the collection $\{f(x) \in T: x \in S\}$ of all such image points is called the **image** of f. In general, whenever a new mapping is given, we must check to see that it is in fact **well-defined**. In other words, we must verify that $x = y$ implies $f(x) = f(y)$. We will use this requirement several times throughout the text.

If $A \subset S$, the set $\{f(x): x \in A\}$ is called the **image** of A under f and is denoted by f(A). If f is a mapping from S to T and $A \subset S$, then the **restriction** of f to A, denoted by $f|A$ (or sometimes f_A), is the function from A to T defined by $f|A: x \in A \mapsto f(x) \in T$. If $x' \in T$, then any element $x \in S$ such that $f(x) = x'$ is called an **inverse image** of x' (this is sometimes also called a **preimage** of x'). Note that in general there may be more than one inverse

image for any particular $x' \in T$. Similarly, if $A' \subset T$, then the inverse image of A' is the subset of S given by $\{x \in S: f(x) \in A'\}$. We will denote the inverse image of A' by $f^{-1}(A')$.

Let f be a mapping from S to T. Note that every element of T need not necessarily be the image of some element of S. However, if $f(S) = T$, then f is said to be **onto** or **surjective**. In other words, f is surjective if given any $x' \in T$ there exists $x \in S$ such that $f(x) = x'$. In addition, f is said to be **one-to-one** or **injective** if $x \neq y$ implies that $f(x) \neq f(y)$. An alternative characterization is to say that f is injective if $f(x) = f(y)$ implies that $x = y$.

If f is both injective and surjective, then f is said to be **bijective**. In this case, given any $x' \in T$ there exists a *unique* $x \in S$ such that $x' = f(x)$. If f is bijective, then we may define the **inverse mapping** $f^{-1}: T \to S$ in the following way. For any $x' \in T$, we let $f^{-1}(x')$ be that (unique) element $x \in S$ such that $f(x) = x'$.

Example 0.3 Consider the function $f: \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^2$. This mapping is clearly not surjective since $f(x) \geq 0$ for any $x \in \mathbb{R}$. Furthermore, it is also not injective. Indeed, it is clear that $2 \neq -2$ but $f(2) = f(-2) = 4$. Note also that both the domain and range of f are the whole set $\mathbb{R}$, but that the image of f is just the subset of all nonnegative real numbers (i.e., the set of all $x \in \mathbb{R}$ with $x \geq 0$).

On the other hand, it is easy to see that the mapping $g: \mathbb{R} \to \mathbb{R}$ defined by $g(x) = ax + b$ for any $a, b \in \mathbb{R}$ (with $a \neq 0$) is a bijection. In this case the inverse mapping is simply given by $g^{-1}(x') = (x' - b)/a$. ∥

Example 0.4 If f is a mapping defined on the collections $\{A_i\}$ and $\{B_i\}$ of sets, then we claim that

$$f(\cup A_i) = \cup f(A_i)$$

and

$$f^{-1}(\cup B_i) = \cup f^{-1}(B_i) \ .$$

To prove these relationships we proceed in our usual manner. Thus we have $x' \in f(\cup A_i)$ if and only if $x' = f(x)$ for some $x \in \cup A_i$, hence if and only if x' is in some $f(A_i)$, and therefore if and only if $x' \in \cup f(A_i)$. This proves the first statement. As to the second statement, we have $x \in f^{-1}(\cup B_i)$ if and only if $f(x) \in \cup B_i$, hence if and only if $f(x)$ is in some B_i, hence if and only if x is in some $f^{-1}(B_i)$, and therefore if and only if $x \in \cup f^{-1}(B_i)$.

Several similar relationships that will be referred to again are given in the exercises. ∥

Now consider the sets S, T and U along with the mappings f: S → T and g: T → U. We define the **composite mapping** (sometimes also called the **product**) g ∘ f: S → U by

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

for all x ∈ S. In general, f ∘ g ≠ g ∘ f, and we say that the composition of two functions is not **commutative**. However, if we also have a mapping h: U → V, then for any x ∈ S we have

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

This means that

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

and hence the composition of mappings is **associative**.

As a particular case of the composition of mappings, note that if f: S → T is a bijection and f(x) = x′ ∈ T where x ∈ S, then

$$(f \circ f^{-1})(x') = f(f^{-1}(x')) = f(x) = x'$$

and

$$(f^{-1} \circ f)(x) = f^{-1}(f(x)) = f^{-1}(x') = x \ .$$

If we write $f \circ f^{-1} = I_T$, then the mapping I_T has the property that $I_T(x') = x'$ for every x′ ∈ T. We call I_T the **identity mapping** on T. Similarly, the composition mapping $f^{-1} \circ f = I_S$ is called the identity mapping on S. In the particular case that S = T, then $f \circ f^{-1} = f^{-1} \circ f = I$ is also called the identity mapping.

An extremely important result follows by noting that (even if S ≠ T)

$$\begin{aligned}(f^{-1} \circ g^{-1})(g \circ f)(x) &= (f^{-1} \circ g^{-1})(g(f(x))) = f^{-1}(g^{-1}(g(f(x)))) \\ &= f^{-1}(f(x)) = x\end{aligned}$$

Since it is also easy to see that $(g \circ f)(f^{-1} \circ g^{-1})(x') = x'$, we have shown that

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

Exercises

1. Let f be a mapping of sets. For each of the following, state any conditions on f that may be required (e.g., surjective or injective), and then prove the statement:
   (a) $A_1 \subset A_2$ implies $f(A_1) \subset f(A_2)$.
   (b) $f(A)^c \subset f(A^c)$ is true if and only if f is surjective.
   (c) $f(\cap A_i) \subset \cap f(A_i)$.
   (d) $B_1 \subset B_2$ implies $f^{-1}(B_1) \subset f^{-1}(B_2)$.
   (e) $f^{-1}(\cap B_i) = \cap f^{-1}(B_i)$.
   (f) $f^{-1}(B^c) = f^{-1}(B)^c$.

2. Given a nonempty set A, we define the **identity mapping** i_A: $A \to A$ by $i_A(a) = a$ for every $a \in A$. Let f: $A \to A$ be any mapping.
   (a) Show that $f \circ i_A = i_A \circ f = f$.
   (b) If f is bijective (so that f^{-1} exists), show that $f \circ f^{-1} = f^{-1} \circ f = i_A$.
   (c) Let f be a bijection, and suppose that g is any other mapping with the property that $g \circ f = f \circ g = i_A$. Show that $g = f^{-1}$.

0.3 ORDERINGS AND EQUIVALENCE RELATIONS

Given any two sets S and T, a subset R of $S \times T$ is said to be a **relation** between S and T. If $R \subset S \times T$ and $(x, y) \in R$, then it is common to write xRy to show that x and y are "R-related." In particular, consider the relation symbolized by $\leq$ and defined as having the following properties on a set S:

(a) $x \leq x$ (reflexivity);
(b) $x \leq y$ and $y \leq x$ implies $x = y$ for all $x, y \in S$ (antisymmetry);
(c) $x \leq y$ and $y \leq z$ implies $x \leq z$ for all $x, y, z \in S$ (transitivity).

Any relation on a non-empty set S having these three properties is said to be a **partial ordering**, and S is said to be a **partially ordered set**. We will sometimes write $y \geq x$ instead of the equivalent notation $x \leq y$. The reason for including the qualifying term "partial" in this definition is shown in our next example.

Example 0.5 Let S be any set, and let $\mathcal{P}(S)$ be the collection of all subsets of S (this is sometimes called the **power set** of S). If A, B and C are subsets of S, then clearly $A \subset A$ so that (a) is satisfied; $A \subset B$ and $B \subset A$ implies $A = B$ then satisfies (b); and $A \subset B$ and $B \subset C$ implies $A \subset C$ satisfies (c). Therefore

$\subset$ defines a partial ordering on $\mathcal{P}(S)$, and the subsets of S are said to be **ordered by inclusion**. Note however, that if $A \subset S$ and $B \subset S$ but $A \not\subset B$ and $B \not\subset A$, then there is no relation between A and B, and we say that A and B are not **comparable**. //

The terminology used in this example is easily generalized as follows. If S is any partially ordered set and $x, y \in S$, then we say that x and y are **comparable** if either $x \leq y$ or $y \leq x$.

If, in addition to properties (a) – (c), a relation R also has the property that any two elements are comparable, then R is said to be a **total ordering**. In other words, a total ordering also has the property that

(d) either $x \leq y$ or $y \leq x$ for all $x, y \in S$.

Let S be a set partially ordered by $\leq$ and suppose $A \subset S$. It should be clear that A may be considered to be a partially ordered set by defining $a \leq b$ for all $a, b \in A$ if $a \leq b$ where a and b are considered to be elements of S. (This is similar to the restriction of a mapping.) We then say that A has a partial ordering $\leq$ **induced** by the ordering on S. If A is *totally* ordered by the ordering induced by $\leq$, then A is frequently called a **chain** in S.

Let A be a non-empty subset of a partially ordered set S. An element $x \in S$ is called an **upper bound** for A if $a \leq x$ *for all* $a \in A$. If it so happens that x is an element of A, then x is said to be a **largest element** of A. Similarly, $y \in S$ is called a **lower bound** for A if $y \leq a$ *for all* $a \in A$, and y is a **smallest element** of A if $y \in A$. If A has an upper (lower) bound, then we say that A is **bounded above** (**below**). Note that largest and smallest elements need not be unique.

Suppose that A is bounded above by $\alpha \in S$, and in addition, suppose that for any other upper bound x of A we have $\alpha \leq x$. Then we say that α is a **least upper bound** (or **supremum**) of A, and we write $\alpha = \text{lub } A = \sup A$. As expected, if A is bounded below by $\beta \in S$, and if $y \leq \beta$ for all other lower bounds $y \in S$, then β is called a **greatest lower bound** (or **infimum**), and we write $\beta = \text{glb } A = \inf A$. In other words, if it exists, the least upper (greatest lower) bound for A is a smallest (largest) element of the set of all upper (lower) bounds for A.

From property (b) above and the definitions of inf and sup, we see that if they exist, the least upper bound and the greatest lower bound are unique. (For example, if β and β' are both greatest lower bounds, then $\beta \leq \beta'$ and $\beta' \leq \beta$ implies that $\beta = \beta'$.) Hence it is meaningful to talk about *the* least upper bound and *the* greatest lower bound.

Let S be a partially ordered set, and suppose $A \subset S$. An element $\alpha \in A$ is said to be **maximal in A** if for any element $a \in A$ with $\alpha \leq a$, we have $a = \alpha$. In other words, no element of A other than α itself is greater than or equal to

α. Similarly, an element $\beta \in A$ is said to be **minimal in A** if for any $b \in A$ with $b \leq \beta$, we have $b = \beta$. Note that a maximal element may not be a largest element (since two elements of a partially ordered set need not be comparable), and there may be many maximal elements in A.

We now state Zorn's lemma, one of the most fundamental results in set theory, and hence in all of mathematics. While the reader can hardly be expected to appreciate the significance of this lemma at the present time, it is in fact extremely powerful.

Zorn's Lemma Let S be a partially ordered set in which every chain has an upper bound. Then S contains a maximal element.

It can be shown (see any book on set theory) that Zorn's lemma is logically equivalent to the **axiom of choice**, which states that given any non-empty family of non-empty disjoint sets, a set can be formed which contains precisely one element taken from each set in the family. Although this seems like a rather obvious statement, it is important to realize that either the axiom of choice or some other statement equivalent to it must be postulated in the formulation of the theory of sets, and thus Zorn's lemma is not really provable in the usual sense. In other words, Zorn's lemma is frequently taken as an axiom of set theory. However, it is an indispensable part of some of what follows although we shall have little occasion to refer to it directly.

Up to this point, we have only talked about one type of relation, the partial ordering. We now consider another equally important relation. Let S be any set. A relation $\approx$ on S is said to be an **equivalence relation** if it has the following properties for all $x, y, z \in S$:

(a) $x \approx x$ for all $x \in S$ (reflexivity);
(b) $x \approx y$ implies $y \approx x$ (symmetry);
(c) $x \approx y$ and $y \approx z$ implies $x \approx z$ for all $x, y, z \in S$ (transitivity).

Note that only (b) differs from the defining relations for a partial ordering.

A **partition** of a set S is a family $\{S_i\}$ of non-empty subsets of S such that $\cup S_i = S$ and $S_i \cap S_j \neq \varnothing$ implies $S_i = S_j$. Suppose $x \in S$ and let $\approx$ be an equivalence relation on S. The subset of S defined by $[x] = \{y: y \approx x\}$ is called the **equivalence class** of x. The most important property of equivalence relations is contained in the following theorem.

Theorem 0.2 The family of all distinct equivalence classes of a set S forms a partition of S. (This is called the partition induced by $\approx$.) Moreover, given any partition of S, there is an equivalence relation on S that induces this partition.

Proof Let $\approx$ be an equivalence relation on a set S, and let x be any element of S. Since $x \approx x$, it is obvious that $x \in [x]$. Thus each element of S lies in at least one non-empty equivalence class. We now show that any two equivalence classes are either disjoint or are identical. Let $[x_1]$ and $[x_2]$ be two equivalence classes, and let y be a member of both classes. In other words, $y \approx x_1$ and $y \approx x_2$. Now choose any $z \in [x_1]$ so that $z \approx x_1$. But this means that $z \approx x_1 \approx y \approx x_2$ so that any element of $[x_1]$ is also an element of $[x_2]$, and hence $[x_1] \subset [x_2]$. Had we chosen $z \in [x_2]$ we would have found that $[x_2] \subset [x_1]$. Therefore $[x_1] = [x_2]$, and we have shown that if two equivalence classes have any element in common, then they must in fact be identical.

Let $\{S_i\}$ be any partition of S. We define an equivalence relation on S by letting $x \approx y$ if $x, y \in S_i$ for any $x, y \in S$. It should be clear that this does indeed satisfy the three conditions for an equivalence relation, and that this equivalence relation induces the partition $\{S_i\}$. ∎

As we will see in the next chapter, this theorem has a direct analogue in the theory of groups.

Exercises

1. Let $\mathbb{Z}^+$ denote the set of positive integers. We write $m|n$ to denote the fact that m divides n, i.e., $n = km$ for some $k \in \mathbb{Z}^+$.

   (a) Show that $|$ defines a partial ordering on $\mathbb{Z}^+$.

   (b) Does $\mathbb{Z}^+$ contain either a maximal or minimal element relative to this partial ordering?

   (c) Prove that any subset of $\mathbb{Z}^+$ containing exactly two elements has a greatest lower bound and a least upper bound.

   (d) For each of the following subsets of $\mathbb{Z}^+$, determine whether or not it is a chain in $\mathbb{Z}^+$, find a maximal and minimal element, an upper and lower bound, and a least upper bound:

   (i) {1, 2, 4, 6, 8}.
   (ii) {1, 2, 3, 4, 5}.
   (iii) {3, 6, 9, 12, 15, 18}.
   (iv) {4, 8, 16, 32, 64, 128}.

2. Define a relation $\approx$ on $\mathbb{R}$ by requiring that $a \approx b$ if $|a| = |b|$. Show that this defines an equivalence relation on $\mathbb{R}$.

3. For any $a, b \in \mathbb{R}$, let $a \sim b$ mean $ab > 0$. Does $\sim$ define an equivalence relation? What happens if we use $ab \geq 0$ instead of $ab > 0$?

0.4 CARDINALITY AND THE REAL NUMBER SYSTEM

We all have an intuitive sense of what it means to say that two finite sets have the same number of elements, but our intuition leads us astray when we come to consider infinite sets. For example, there are as many perfect squares (1, 4, 9, 16, etc.) among the positive integers as there are positive integers. That this is true can easily be seen by writing each positive integer paired with its square:

$$\begin{array}{cccccc} 1, & 2, & 3, & 4, & \ldots \\ 1^2, & 2^2, & 3^2, & 4^2, & \ldots \end{array}$$

While it seems that the perfect squares are only sparsely placed throughout the integers, we have in fact constructed a bijection of all positive integers with all of the perfect squares of integers, and we are forced to conclude that in this sense they both have the "same number of elements."

In general, two sets S and T are said to have the same **cardinality**, or to possess the same number of elements, if there exists a bijection from S to T. A set S is **finite** if it has the same cardinality as either $\varnothing$ or the set $\{1, 2, \ldots, n\}$ for some positive integer n; otherwise, S is said to be **infinite**. However, there are varying degrees of "infinity." A set S is **countable** if it has the same cardinality as a subset of the set $\mathbb{Z}^+$ of positive integers. If this is not the case, then we say that S is **uncountable**. Any infinite set which is numerically equivalent to (i.e., has the same cardinality as) $\mathbb{Z}^+$ is said to be **countably infinite**. We therefore say that a set is **countable** if it is countably infinite or if it is non-empty and finite.

It is somewhat surprising (as was first discovered by Cantor) that the set $\mathbb{Q}^+$ of all positive rational numbers is in fact countable. The elements of $\mathbb{Q}^+$ can not be listed in order of increasing size because there is no smallest such number, and between any two rational numbers there are infinitely many others (see Theorem 0.4 below). To show that $\mathbb{Q}^+$ is countable, we shall construct a bijection from $\mathbb{Z}^+$ to $\mathbb{Q}^+$.

To do this, we first consider all positive rationals whose numerator and denominator add up to 2. In this case we have only $1/1 = 1$. Next we list those positive rationals whose numerator and denominator add up to 3. If we agree

to always list our rationals with numerators in increasing order, then we have 1/2 and 2/1 = 2. Those rationals whose numerator and denominator add up to 4 are then given by 1/3, 2/2 = 1, 3/1 = 3. Going on to 5 we obtain 1/4, 2/3, 3/2, 4/1 = 4. For 6 we have 1/5, 2/4 = 1/2, 3/3 = 1, 4/2 = 2, 5/1 = 5. Continuing with this procedure, we list together all of our rationals, omitting any number already listed. This gives us the sequence

$$1, 1/2, 2, 1/3, 3, 1/4, 2/3, 3/2, 4, 1/5, 5, \ldots$$

which contains each positive rational number exactly once, and provides our desired bijection.

We have constructed several countably infinite sets of real numbers, and it is natural to wonder whether there are in fact any uncountably infinite sets. It was another of Cantor's discoveries that the set $\mathbb{R}$ of all real numbers is actually uncountable. To prove this, let us assume that we have listed (in some manner similar to that used for the set $\mathbb{Q}^+$) all the real numbers in decimal form. What we shall do is construct a decimal $.d_1d_2d_3\cdots$ that is not on our list, thus showing that the list can not be complete. Consider only the portion of the numbers on our list to the right of the decimal point, and look at the first number on the list. If the first digit after the decimal point of the first number is a 1, we let $d_1 = 2$; otherwise we let $d_1 = 1$. No matter how we choose the remaining d's, our number will be different from the first on our list. Now look at the second digit after the decimal point of the second number on our list. Again, if this second digit is a 1, we let $d_2 = 2$; otherwise we let $d_2 = 1$. We have now constructed a number that differs from the first two numbers on our list. Continuing in this manner, we construct a decimal $.d_1d_2d_3\cdots$ that differs from every other number on our list, contradicting the assumption that all real numbers can be listed, and proving that $\mathbb{R}$ is actually uncountable.

Since it follows from what we showed above that the set $\mathbb{Q}$ of all rational numbers on the real line is countable, and since we just proved that the set $\mathbb{R}$ is uncountable, it follows that a set of **irrational** numbers must exist and be uncountably infinite.

From now on we will assume that the reader understands what is meant by the real number system, and we proceed to investigate some of its most useful properties. A complete axiomatic treatment that justifies what we already know would take us too far afield, and the interested reader is referred to, e.g., Rudin (1976).

Let S be any ordered set, and let $A \subset S$ be non-empty and bounded above. We say that S has the **least upper bound property** if sup A exists in S. In the special case of $S = \mathbb{R}$, we have the following extremely important axiom.

Archimedean Axiom Every non-empty set of real numbers which has an upper (lower) bound has a least upper bound (greatest lower bound).

The usefulness of this axiom is demonstrated in the next rather obvious though important result, sometimes called the **Archimedean property** of the real number line.

Theorem 0.3 Let $a, b \in \mathbb{R}$ and suppose $a > 0$. Then there exists $n \in \mathbb{Z}^+$ such that $na > b$.

Proof Let S be the set of all real numbers of the form na where n is a positive integer. If the theorem were false, then b would be an upper bound for S. But by the Archimedean axiom, S has a least upper bound $\alpha = \sup S$. Since $a > 0$, we have $\alpha - a < \alpha$ and $\alpha - a$ can not be an upper bound of S (by definition of α). Therefore, there exists an $m \in \mathbb{Z}^+$ such that $ma \in S$ and $\alpha - a < ma$. But then $\alpha < (m + 1)a \in S$ which contradicts the fact that $\alpha = \sup S$. ∎

One of the most useful facts about the real line is that the set $\mathbb{Q}$ of all rational numbers is **dense** in $\mathbb{R}$. By this we mean that given any two distinct real numbers, we can always find a rational number between them. This means that any real number may be approximated to an arbitrary degree of accuracy by a rational number. It is worth proving this using Theorem 0.3.

Theorem 0.4 Suppose $x, y \in \mathbb{R}$ and assume that $x < y$. Then there exists a rational number $p \in \mathbb{Q}$ such that $x < p < y$.

Proof Since $x < y$ we have $y - x > 0$. In Theorem 0.3, choose $a = y - x$ and $b = 1$ so there exists $n \in \mathbb{Z}^+$ such that $n(y - x) > 1$, or alternatively,

$$1 + nx < ny .$$

Applying Theorem 0.3 again, we let $a = 1$ and both $b = nx$ and $b = -nx$ to find integers $m_1, m_2 \in \mathbb{Z}^+$ such that $m_1 > nx$ and $m_2 > -nx$. Rewriting the second of these as $-m_2 < nx$, we combine the two inequalities to obtain

$$-m_2 < nx < m_1$$

so that nx lies between two integers. But if nx lies between two integers, it must lie between two consecutive integers $m - 1$ and m for some $m \in \mathbb{Z}$ where $-m_2 \leq m \leq m_1$. Thus $m - 1 \leq nx < m$ implies that $m \leq 1 + nx$ and $nx < m$. We therefore obtain

$$nx < m \leq 1 + nx < ny$$

or, equivalently (since $n \neq 0$), $x < m/n < y$. ∎

Corollary Suppose $x, y \in \mathbb{R}$ and assume that $x < y$. Then there exist integers $m \in \mathbb{Z}$ and $k \geq 0$ such that $x < m/2^k < y$.

Proof Simply note that the proof of Theorem 0.4 could be carried through if we choose an integer $k \geq 0$ so that $2^k(y - x) > 1$, and replace n by 2^k throughout. ∎

In addition to the real number system $\mathbb{R}$ we have been discussing, it is convenient to introduce the **extended real number system** as follows. To the real number system $\mathbb{R}$, we adjoin the symbols $+\infty$ and $-\infty$ which are *defined* to have the property that $-\infty < x < +\infty$ *for all* $x \in \mathbb{R}$. This is of great notational convenience. We stress however, that neither $+\infty$ or $-\infty$ are considered to be elements of $\mathbb{R}$.

Suppose A is a non-empty set of real numbers. We have already defined sup A in the case where A has an upper bound. If A is non-empty and has no upper bound, then we say that $\sup A = +\infty$, and if $A = \emptyset$, then $\sup A = -\infty$. Similarly, if $A \neq \emptyset$ and has no lower bound, then $\inf A = -\infty$, and if $A = \emptyset$, then $\inf A = +\infty$.

Suppose $a, b \in \mathbb{R}$ with $a \leq b$. Then the **closed interval** [a, b] from a to b is the subset of $\mathbb{R}$ defined by

$$[a, b] = \{x \in \mathbb{R}: a \leq x \leq b\} .$$

Similarly, the **open interval** (a, b) is defined to be the subset

$$(a, b) = \{x \in \mathbb{R}: a < x < b\} .$$

We may also define the **open-closed** and **closed-open** intervals in the obvious way. The **infinity symbols** $\pm\infty$ thus allow us to talk about intervals of the form $(-\infty, b]$, $[a, +\infty)$ and $(-\infty, +\infty)$.

Another property of the sup that will be needed later on is contained in the following theorem. By way of notation, we define $\mathbb{R}^+$ to be the set of all real numbers > 0, and $\overline{\mathbb{R}}^+ = \mathbb{R}^+ \cup \{0\}$ to be the set of all real numbers ≥ 0.

Theorem 0.5 Let A and B be non-empty bounded sets of real numbers, and define the sets

$$A + B = \{x + y: x \in A \text{ and } y \in B\}$$

and

$$AB = \{xy: x \in A \text{ and } y \in B\} .$$

Then

(a) For all $A, B \subset \mathbb{R}$ we have $\sup (A + B) = \sup A + \sup B$.

(b) For all $A, B \subset \overline{\mathbb{R}}^+$ we have $\sup (AB) \leq (\sup A)(\sup B)$.

Proof (a) Let $\alpha = \sup A$, $\beta = \sup B$, and suppose $x + y \in A + B$. Then

$$x + y \leq \alpha + y \leq \alpha + \beta$$

so that $\alpha + \beta$ is an upper bound for $A + B$. Now note that given $\varepsilon > 0$, there exists $x \in A$ such that $\alpha - \varepsilon/2 < x$ (or else α would not be the least upper bound). Similarly, there exists $y \in B$ such that $\beta - \varepsilon/2 < y$. Then $\alpha + \beta - \varepsilon < x + y$ so that $\alpha + \beta$ must be the least upper bound for $A + B$.

(b) If $x \in A \subset \overline{\mathbb{R}}^+$ we must have $x \leq \sup A$, and if $y \in B \subset \overline{\mathbb{R}}^+$ we have $y \leq \sup B$. Hence $xy \leq (\sup A)(\sup B)$ for all $xy \in AB$, and therefore $A \neq \varnothing$ and $B \neq \varnothing$ implies

$$\sup (AB) \leq (\sup A)(\sup B) .$$

The reader should verify that strict equality holds if $A \subset \mathbb{R}^+$ and $B \subset \mathbb{R}^+$. ∎

The last topic in our treatment of real numbers that we wish to discuss is the absolute value. Note that if $x \in \mathbb{R}$ and $x^2 = a$, then we also have $(-x)^2 = a$. We *define* $\sqrt{a}$, for $a \geq 0$, to be the unique *positive* number x such that $x^2 = a$, and we call x the **square root** of a.

Suppose $x, y \geq 0$ and let $x^2 = a$ and $y^2 = b$. Then $x = \sqrt{a}$, $y = \sqrt{b}$ and we have $(\sqrt{a}\sqrt{b})^2 = (xy)^2 = x^2 y^2 = ab$ which implies that

$$\sqrt{ab} = \sqrt{a}\sqrt{b} .$$

For any $a \in \mathbb{R}$, we define its **absolute value** $|a|$ by $|a| = \sqrt{a^2}$. It then follows that $|-a| = |a|$, and hence

$$|a| = \begin{cases} a \text{ if } a \geq 0 \\ -a \text{ if } a < 0 \end{cases}$$

This clearly implies that

$$a \leq |a| .$$

In addition, if $a, b \geq 0$ and $a \leq b$, then we have $(\sqrt{a})^2 = a \leq b = (\sqrt{b})^2$ so that $\sqrt{a} \leq \sqrt{b}$.

The absolute value has two other useful properties. First, we note that

$$|ab| = \sqrt{(ab)^2} = \sqrt{a^2 b^2} = \sqrt{a^2}\sqrt{b^2} = |a||b|.$$

Second, we see that

$$\begin{aligned} |a+b|^2 &= (a+b)^2 \\ &= a^2 + b^2 + 2ab \\ &\le |a|^2 + |b|^2 + 2|ab| \\ &= |a|^2 + |b|^2 + 2|a||b| \\ &= (|a| + |b|)^2 \end{aligned}$$

and therefore

$$|a + b| \le |a| + |b| .$$

Using these results, many other useful relationships may be obtained. For example, $|a| = |a + b - b| \le |a + b| + |-b| = |a + b| + |b|$ so that

$$|a| - |b| \le |a + b| .$$

Others are to be found in the exercises.

Example 0.6 Let us show that if $\varepsilon > 0$, then $|x| < \varepsilon$ if and only if $-\varepsilon < x < \varepsilon$. Indeed, we see that if $x > 0$, then $|x| = x < \varepsilon$, and if $x < 0$, then $|x| = -x < \varepsilon$ which implies $-\varepsilon < x < 0$ (we again use the fact that $a < b$ implies $-b < -a$). Combining these results shows that $|x| < \varepsilon$ implies $-\varepsilon < x < \varepsilon$. We leave it to the reader to reverse the argument and complete the proof.

A particular case of this result that will be of use later on comes from letting $x = a - b$. We then see that $|a - b| < \varepsilon$ if and only if $-\varepsilon < a - b < \varepsilon$. Rearranging, this may be written in the form $b - \varepsilon < a < b + \varepsilon$. The reader should draw a picture of this relationship. //

Exercises

1. Prove that if A and B are countable sets, then $A \times B$ is countable.

2. (a) A real number x is said to be **algebraic** (over the rationals) if it satisfies some polynomial equation of positive degree with rational coefficients:

$$x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 = 0 \ .$$

Given the fact (which we will prove in Chapter 6) that each polynomial equation has only finitely many roots, show that the set of all algebraic numbers is countable.
(b) We say that a real number x is **transcendental** if it is not algebraic (the most common transcendental numbers are π and e). Using the fact that the reals are uncountable, show that the set of all transcendental numbers is also uncountable.

3. If $a, b \geq 0$, show that $\sqrt{ab} \leq (a + b)/2$.

4. For any $a, b \in \mathbb{R}$, show that:
   (a) $| |a| - |b| | \leq |a + b|$.
   (b) $| |a| - |b| | \leq |a - b|$.

5. (a) If $A \subset \mathbb{R}$ is nonempty and bounded below, show $\sup(-A) = -\inf A$.
   (b) If $A \subset \mathbb{R}$ is nonempty and bounded above, show $\inf(-A) = -\sup A$.

0.5 INDUCTION

Another important concept in the theory of sets is called "well-ordering." In particular, we say that a totally ordered set S is **well-ordered** if *every* nonempty subset A of S has a smallest element. For example, consider the set S of all rational numbers in the interval [0, 1]. It is clear that 0 is the smallest element of S, but the subset of S consisting of all rational numbers > 0 has no smallest element (this is a consequence of Theorem 0.4).

For our purposes, it is an (apparently obvious) axiom that every nonempty set of natural numbers has a smallest element. In other words, the natural numbers are well-ordered. The usefulness of this axiom is that it allows us to prove an important property called **induction**.

Theorem 0.6 Assume that for all $n \in \mathbb{Z}^+$ we are given an assertion A(n), and assume it can be shown that:
(a) A(1) is true;
(b) If A(n) is true, then A(n + 1) is true.
Then A(n) is true for all $n \in \mathbb{Z}^+$.

Proof If we let S be that subset of $\mathbb{Z}^+$ for which A(n) is not true, then we must show that $S = \varnothing$. According to our well-ordering axiom, if $S \neq \varnothing$ then S contains a least element which we denote by N. By assumption (a), we must

have $N \neq 1$ and hence $N > 1$. Since N is a least element, $N - 1 \notin S$ so that $A(N-1)$ must be true. But then (b) implies that $A(N)$ must be true which contradicts the definition of N. ∎

Example 0.7 Let $n > 0$ be an integer. We define **n factorial**, written n!, to be the number

$$n! = n(n-1)(n-2)\cdots(2)(1)$$

with 0! defined to be 1. The **binomial coefficient** $\binom{n}{k}$ is defined by

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

where n and k are nonnegative integers. We leave it to the reader (see Exercise 0.6.1) to show that

$$\binom{n}{k} = \binom{n}{n-k}$$

and

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

What we wish to prove is the **binomial theorem**:

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

We proceed by induction as follows. For n = 1, we have

$$\binom{1}{0} x^0 y^1 + \binom{1}{1} x^1 y^0 = (x+y)^1$$

so that the assertion is true for n = 1. We now assume the theorem holds for n, and proceed to show that it also holds for n + 1. We have

$$\begin{aligned} (x+y)^{n+1} &= (x+y)(x+y)^n = (x+y)\left[\sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}\right] \\ &= \sum_{k=0}^{n} \binom{n}{k} x^{k+1} y^{n-k} + \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k+1} \end{aligned} \qquad (*)$$

By relabelling the summation index, we see that for any function f with domain equal to {0, 1, . . . , n} we have

$$\sum_{k=0}^{n} f(k) = f(0) + f(1) + \cdots + f(n) = \sum_{k=1}^{n+1} f(k-1).$$

We use this fact in the first sum in (*), and separate out the k = 0 term in the second to obtain

$$(x+y)^{n+1} = \sum_{k=1}^{n+1} \binom{n}{k-1} x^k y^{n-k+1} + y^{n+1} + \sum_{k=1}^{n} \binom{n}{k} x^{k+1} y^{n-k+1}.$$

We now separate out the k = n + 1 term from the first sum in this expression and group terms to find

$$\begin{aligned}(x+y)^{n+1} &= x^{n+1} + y^{n+1} + \sum_{k=1}^{n} \left[\binom{n}{k-1} + \binom{n}{k} \right] x^k y^{n-k+1} \\ &= x^{n+1} + y^{n+1} + \sum_{k=1}^{n} \binom{n+1}{k} x^k y^{n+1-k} \\ &= \sum_{k=0}^{n+1} \binom{n+1}{k} x^k y^{n+1-k}\end{aligned}$$

as was to be shown. //

0.6 COMPLEX NUMBERS

At this time we wish to formally define the complex number system $\mathbb{C}$, although most readers should already be familiar with its basic properties. The motivation for the introduction of such numbers comes from the desire to solve equations such as $x^2 + 1 = 0$ which leads to the square root of a negative number. We may proceed by manipulating square roots of negative numbers as if they were square roots of positive numbers. However, a consequence of this is that on the one hand, $(\sqrt{-1})^2 = -1$, while on the other hand

$$(\sqrt{-1})^2 = \sqrt{-1}\sqrt{-1} = \sqrt{(-1)(-1)} = \sqrt{+1} = 1 .$$

In order to avoid paradoxical manipulations of this type, the symbol i was introduced by Euler (in 1779) with the defining property that $i^2 = -1$. Then, if $a > 0$, we have $\sqrt{-a} = i\sqrt{a}$. Using this notation, a **complex number** $z \in \mathbb{C}$ is a

number of the form $z = x + iy$ where $x \in \mathbb{R}$ is called the **real part** of z (written Re z), and $y \in \mathbb{R}$ is called the **imaginary part** of z (written Im z).

Two complex numbers $x + iy$ and $u + iv$ are said to be equal if $x = u$ and $y = v$. Algebraic operations in $\mathbb{C}$ are *defined* as follows:

Addition: $(x + iy) + (u + iv) = (x + u) + i(y + v)$.
Subtraction: $(x + iy) - (u + iv) = (x - u) + i(y - v)$.
Multiplication: $(x + iy)(u + iv) = (xu - yv) + i(xv + yu)$.
Division: $(x + iy)/(u + iv) = (x + iy)(u - iv)/(u + iv)(u - iv)$
$= [(xu + yv) + i(yu - vx)]/(u^2 + v^2)$.

It should be clear that the results for multiplication and division may be obtained by formally multiplying out the terms and using the fact that $i^2 = -1$.

The **complex conjugate** z^* of a complex number $z = x + iy$ is defined to be the complex number $z^* = x - iy$. Note that if $z, w \in \mathbb{C}$ we have

$$\begin{aligned}(z+w)^* &= z^* + w^* \\ (zw)^* &= z^* w^* \\ z + z^* &= 2\,\mathrm{Re}\, z \\ z - z^* &= 2i\,\mathrm{Im}\, z\end{aligned}$$

The **absolute value** (or **modulus**) $|z|$ of a complex number $z = x + iy$ is defined to be the real number

$$|z| = \sqrt{x^2 + y^2} = \left(zz^*\right)^{1/2}.$$

By analogy to the similar result for real numbers, if $z, w \in \mathbb{C}$ then (using the fact that $z = x + iy$ implies $\mathrm{Re}\, z = x \le \sqrt{x^2+y^2} = |z|$)

$$\begin{aligned}|z+w|^2 &= (z+w)(z+w)^* \\ &= zz^* + zw^* + z^* w + ww^* \\ &= |z|^2 + 2\,\mathrm{Re}(zw^*) + |w|^2 \\ &\le |z|^2 + 2|zw^*| + |w|^2 \\ &= |z|^2 + 2|z||w| + |w|^2 \\ &= (|z| + |w|)^2\end{aligned}$$

and hence taking the square root of both sides yields

$$|z + w| \le |z| + |w| .$$

Let the sum $z_1 + \cdots + z_n$ be denoted by $\sum_{i=1}^{n} z_i$. The following theorem is known as **Schwartz's inequality**.

Theorem 0.7 Let $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$ be complex numbers. Then

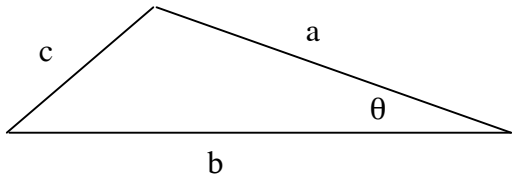
$$\left|\sum_{j=1}^{n} a_j b_j^*\right| \le \left(\sum_{j=1}^{n} |a_j|^2\right)\left(\sum_{j=1}^{n} |b_j|^2\right)$$

Proof Write (suppressing the limits on the sum) $A = \sum_j |a_j|^2$, $B = \sum_j |b_j|^2$ and $C = \sum_j a_j b_j^*$. If $B = 0$, then $b_j = 0$ for all $j = 1, \ldots, n$ and there is nothing to prove, so we assume that $B \neq 0$. We then have

$$\begin{aligned}
0 &\le \textstyle\sum_i |Ba_i - Cb_i|^2 \\
&= \textstyle\sum_i (Ba_i - Cb_i)(Ba_i^* - Cb_i^*) \\
&= B^2 \textstyle\sum_i |a_i|^2 - BC^* \sum_i a_i b_i^* - BC \sum_i a_i^* b_i + |C|^2 \sum_i |b_i|^2 \\
&= B^2 A - B|C|^2 - B|C|^2 + |C|^2 B \\
&= B(AB - |C|^2).
\end{aligned}$$

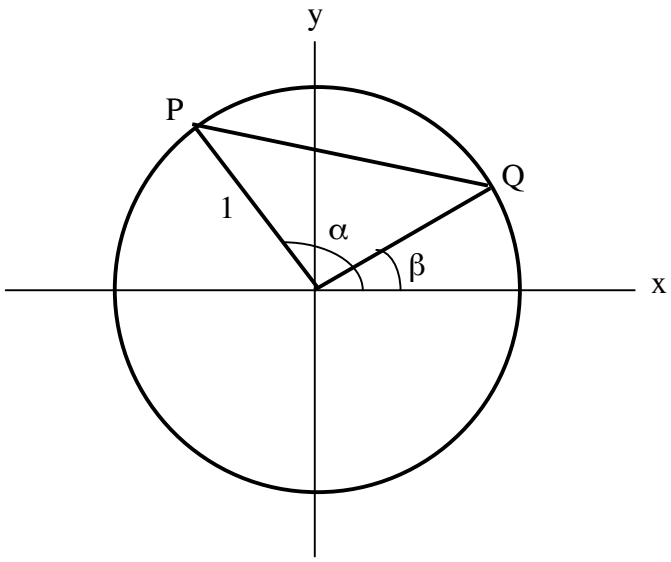
But $B \ge 0$ so that $AB - |C|^2 \ge 0$ and hence $|C|^2 \le AB$. ∎

It is worth our going through some additional elementary properties of complex numbers that will be needed on occasion throughout this text. Purely for the sake of logical consistency, let us first prove some basic trigonometric relationships. Our starting point will be the so-called "law of cosines" which states that $c^2 = a^2 + b^2 - 2ab\cos\theta$ (see the figure below).



A special case of this occurs when $\theta = \pi/2$, in which case we obtain the famous Pythagorean theorem $a^2 + b^2 = c^2$. (While most readers should already be familiar with these results, we prove them in Section 2.4.)

Now consider a triangle inscribed in a *unit* circle as shown below:



The point P has coordinates $(x_P, y_P) = (\cos\alpha, \sin\alpha)$, and Q has coordinates $(x_Q, y_Q) = (\cos\beta, \sin\beta)$. Applying the Pythagorean theorem to the right triangle with hypotenuse defined by the points P and Q (and noting $x_Q^2 + y_Q^2 = x_P^2 + y_P^2 = 1$), we see that the square of the distance between the points P and Q is given by

$$\begin{aligned}(PQ)^2 &= (x_Q - x_P)^2 + (y_Q - y_P)^2 \\ &= (x_Q{}^2 + y_Q{}^2) + (x_P{}^2 + y_P{}^2) - 2(x_P x_Q + y_P y_Q) \\ &= 2 - 2(\cos\alpha\cos\beta + \sin\alpha\sin\beta).\end{aligned}$$

On the other hand, we can apply the law of cosines to obtain the distance PQ, in which case we find that $(PQ)^2 = 2 - 2\cos(\alpha - \beta)$. Equating these expressions yields the basic result

$$\cos(\alpha - \beta) = \cos\alpha\cos\beta + \sin\alpha\sin\beta \ .$$

Replacing β by $-\beta$ we obtain

$$\cos(\alpha + \beta) = \cos\alpha\cos\beta - \sin\alpha\sin\beta \ .$$

If we let $\alpha = \pi/2$, then we have $\cos(\pi/2 - \beta) = \sin\beta$, and if we now replace β by $\pi/2 - \beta$, we find that $\cos\beta = \sin(\pi/2 - \beta)$. Finally, we can use these last results to obtain formulas for $\sin(\alpha \pm \beta)$. In particular, we replace β by $\alpha + \beta$ to obtain

$$\begin{aligned}\sin(\alpha+\beta) &= \cos(\pi/2-(\alpha+\beta)) \\ &= \cos(\pi/2-\alpha-\beta) \\ &= \cos(\pi/2-\alpha)\cos\beta+\sin(\pi/2-\alpha)\sin\beta \\ &= \sin\alpha\cos\beta+\cos\alpha\sin\beta\end{aligned}$$

Again, replacing β by $-\beta$ yields

$$\sin(\alpha-\beta) = \sin\alpha\cos\beta-\cos\alpha\sin\beta \ .$$

(The reader may already know that these results are simple to derive using the Euler formula $\exp(\pm i\theta) = \cos\theta \pm i\sin\theta$ which follows from the Taylor series expansions of exp x, sin x and cos x, along with the definition $i^2 = -1$.)

It is often of great use to think of a complex number $z = x + iy$ as a point in the xy-plane. If we define

$$r = |z| = \sqrt{x^2+y^2}$$

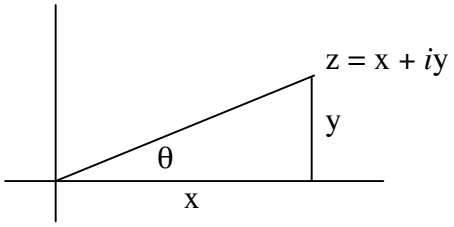
and

$$\tan\theta = y/x$$

then a complex number may also be written in the form

$$z = x+iy = r(\cos\theta+i\sin\theta) = r\exp(i\theta)$$

(see the figure below).



Given two complex numbers

$$z_1 = r_1(\cos\theta_1+i\sin\theta_1)$$

and

$$z_2 = r_2(\cos\theta_2+i\sin\theta_2)$$

we can use the trigonometric addition formulas derived above to show that

$$z_1z_2 = r_1r_2[\cos(\theta_1+\theta_2)+i\sin(\theta_1+\theta_2)] \ .$$

In fact, by induction it should be clear that this can be generalized to (see Exercise 0.6.5)

$$z_1 z_2 \cdots z_n = r_1 r_2 \cdots r_n[\cos(\theta_1 + \theta_2 + \cdots + \theta_n) + i\sin(\theta_1 + \theta_2 + \cdots + \theta_n)] \ .$$

In the particular case where $z_1 = \cdots = z_n$, we find that

$$z^n = r^n(\cos n\theta + i\sin n\theta) \ .$$

This is often called **De Moivre's theorem**.

One of the main uses of this theorem is as follows. Let w be a complex number, and let $z = w^n$ (where n is a positive integer). We say that w is an n*th* **root** of z, and we write this as $w = z^{1/n}$. Next, we observe from De Moivre's theorem that writing $z = r(\cos \theta + i\sin \theta)$ and $w = s(\cos \phi + i\sin \phi)$ yields (assuming that $z \neq 0$)

$$r(\cos \theta + i\sin \theta) = s^n(\cos n\phi + i\sin n\phi) \ .$$

But $\cos \theta = \cos(\theta \pm 2k\pi)$ for $k = 0, \pm1, \pm2, \ldots$, and therefore $r = s^n$ and $n\phi = \theta \pm 2k\pi$. (This follows from the fact that if $z_1 = x_1 + iy_1 = r_1(\cos \theta_1 + i\sin \theta_1)$ and $z_2 = x_2 + iy_2 = r_2(\cos \theta_2 + i\sin \theta_2)$, then $z_1 = z_2$ implies $x_1 = x_2$ and $y_1 = y_2$ so that $r_1 = r_2$, and hence $\theta_1 = \theta_2$.) Then s is the real positive nth root of r, and $\phi = \theta/n \pm 2k\pi/n$. Since this expression for ϕ is the same if any two integers k differ by a multiple of n, we see that there are precisely n distinct solutions of $z = w^n$ (when $z \neq 0$), and these are given by

$$w = r^{1/n}[\cos(\theta + 2k\pi)/n + i\sin(\theta + 2k\pi)/n]$$

where $k = 0, 1, \ldots, n - 1$.

Exercises

1. (a) Show

$$\binom{n}{k} = \binom{n}{n-k}.$$

(b) Show

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

2. Prove by induction the formula $1 + 2 + \cdots + n = n(n + 1)/2$.

3. Prove by induction the formula

$$1 + x + x^2 + \cdots + x^{n-1} = \frac{1 - x^n}{1 - x}$$

where x is any real number $\neq 1$.

4. Prove by induction that for any complex numbers $z_1, \ldots, z_n$ we have:
(a)

$$\left| \sum_{i=1}^{n} z_i \right| \leq \sum_{i=1}^{n} |z_i|$$

(b) $|z_1 z_2 \cdots z_n| = |z_1||z_2| \cdots |z_n|$.

5. Prove by induction that for any complex numbers $z_1, \ldots, z_n$ we have

$$z_1 z_2 \cdots z_n = r_1 r_2 \cdots r_n[\cos(\theta_1 + \theta_2 + \cdots + \theta_n) + i \sin(\theta_1 + \theta_2 + \cdots + \theta_n)]$$

where $z_j = r_j \exp(i\theta_j)$.

0.7 ADDITIONAL PROPERTIES OF THE INTEGERS

The material of this section will be generalized in Sections 6.1 and 6.2 to the theory of polynomials. However, it will also be directly useful to us in our discussion of finite fields in Section 6.6. Most of this section should be familiar to the reader from very elementary courses.

Our first topic is the division of an arbitrary integer $a \in \mathbb{Z}$ by a positive integer $b \in \mathbb{Z}^+$. For example, we can divide 11 by 4 to obtain $11 = 2 \cdot 4 + 3$. As another example, -7 divided by 2 yields $-7 = -4 \cdot 2 + 1$. Note that each of these examples may be written in the form $a = qb + r$ where $q \in \mathbb{Z}$ and $0 \leq r < b$. The number q is called the **quotient**, and the number r is called the **remainder** in the division of a by b. In the particular case that $r = 0$, we say that b **divides** a and we write this as $b|a$. If $r \neq 0$, then b does not divide a, and this is written as $b \nmid a$. If an integer $p \in \mathbb{Z}^+$ is not divisible by any positive integer other than 1 and p itself, then p is said to be **prime**.

It is probably worth pointing out the elementary fact that if $a|b$ and $a|c$, then $a|(mb + nc)$ for any $m, n \in \mathbb{Z}$. This is because $a|b$ implies $b = q_1 a$, and $a|c$ implies $c = q_2 a$. Thus $mb + nc = (mq_1 + nq_2)a$ so that $a|(mb + nc)$.

Theorem 0.8 (**Division Algorithm**) If $a \in \mathbb{Z}$ and $b \in \mathbb{Z}^+$, then there exist unique integers q and r such that $a = qb + r$ where $0 \leq r < b$.

Proof Define $S = \{a - nb \geq 0: n \in \mathbb{Z}\}$. In other words, S consists of all non-negative integers of the form $a - bn$. It is easy to see that $S \neq \varnothing$. Indeed, if $a \geq 0$ we simply choose $n = 0$ so that $a \in S$, and if $a < 0$ we choose $n = a$ so that $a - ba = a(1 - b) \in S$ (since $a < 0$ and $1 - b \leq 0$). Since S is a nonempty subset of the natural numbers, we may apply the well-ordering property of the natural numbers to conclude that S contains a least element $r \geq 0$. If we let q be the value of n corresponding to this r, then we have $a - qb = r$ or $a = qb + r$ where $0 \leq r$. We must show that $r < b$. To see this, suppose that $r \geq b$. Then

$$a - (q + 1)b = a - qb - b = r - b \geq 0$$

so that $a - (q + 1)b \in S$. But $b > 0$ so that

$$a - (q + 1)b = (a - qb) - b < a - qb = r$$

which contradicts the definition of r as the least element of S. Hence $r < b$.

To prove uniqueness, we suppose that we may write $a = q_1 b + r_1$ and $a = q_2 b + r_2$ where $0 \leq r_1 < b$ and $0 \leq r_2 < b$. Equating these two formulas yields $q_1 b + r_1 = q_2 b + r_2$ or $(q_1 - q_2)b = r_2 - r_1$, and therefore $b|(r_2 - r_1)$. Using the fact that $0 \leq r_1 < b$ and $0 \leq r_2 < b$, we see that $r_2 - r_1 < b - r_1 \leq b$. Similarly we have $r_1 - r_2 < b - r_2 \leq b$ or $-b < r_2 - r_1$. This means that $-b < r_2 - r_1 < b$. Therefore $r_2 - r_1$ is a multiple of b that lies strictly between $-b$ and b, and thus we must have $r_2 - r_1 = 0$. Then $(q_1 - q_2)b = 0$ with $b \neq 0$, and hence $q_1 - q_2 = 0$ also. This shows that $r_1 = r_2$ and $q_1 = q_2$ which completes the proof of uniqueness. ∎

Suppose we are given two integers $a, b \in \mathbb{Z}$ where we assume that a and b are not both zero. We say that an integer $d \in \mathbb{Z}^+$ is the **greatest common divisor** of a and b if $d|a$ and $d|b$, and if c is any other integer with the property that $c|a$ and $c|b$, then $c|d$. We denote the greatest common divisor of a and b by $\gcd\{a, b\}$. Our next theorem shows that the gcd always exists and is unique. Furthermore, the method of proof shows us how to actually compute the gcd.

Theorem 0.9 (**Euclidean Algorithm**) If a, b $\in \mathbb{Z}$ are not both zero, then there exists a unique positive integer d $\in \mathbb{Z}^+$ such that

(a) d|a and d|b.

(b) If c $\in \mathbb{Z}$ is such that c|a and c|b, then c|d.

Proof First assume b > 0. Applying the division algorithm, there exist unique integers q_1 and r_1 such that

$$a = q_1 b + r_1 \quad \text{with } 0 \le r_1 < b \ .$$

If $r_1 = 0$, then b|a and we may take d = b to satisfy both parts of the theorem. If $r_1 \neq 0$, then we apply the division algorithm again to b and r_1, obtaining

$$b = q_2 r_1 + r_2 \quad \text{with } 0 \le r_2 < r_1 \ .$$

Continuing this procedure, we obtain a sequence of nonzero remainders $r_1, r_2, \ldots, r_k$ where

$$\begin{aligned} a &= q_1 b + r_1 && \text{with} \quad 0 \le r_1 < b \\ b &= q_2 r_1 + r_2 && \text{with} \quad 0 \le r_2 < r_1 \\ r_1 &= q_3 r_2 + r_3 && \text{with} \quad 0 \le r_3 < r_2 \\ &\vdots \\ r_{k-2} &= q_k r_{k-1} + r_k && \text{with} \quad 0 \le r_k < r_{k-1} \\ r_{k-1} &= q_{k+1} r_k \end{aligned} \qquad (*)$$

That this process must terminate with a zero remainder as shown is due to the fact that each remainder is a nonnegative integer with $r_1 > r_2 > \cdots$. We have denoted the last nonzero remainder by r_k.

We now claim that $d = r_k$. Since $r_{k-1} = q_{k+1} r_k$, we have $r_k | r_{k-1}$. Then, because $r_{k-2} = q_k r_{k-1} + r_k$ and $r_k | r_k$ and $r_k | r_{k-1}$, we have $r_k | r_{k-2}$. Continuing in this manner, we see that $r_k | r_{k-1}$, $r_k | r_{k-2}$, . . . , $r_k | r_1$, r_k|b and r_k|a. This shows that r_k is a common divisor of a and b. To show that r_k is in fact the greatest common divisor, we first note that if c|a and c|b, then $c | r_1$ because $r_1 = a - q_1 b$. But now we see in the same way that $c | r_2$, and working our way through the above set of equations we eventually arrive at $c | r_k$. Thus r_k is a gcd as claimed.

If b < 0, we repeat the above process with a and −b rather than a and b. Since b and −b have the same divisors, it follows that a gcd of {a, −b} will be a gcd of {a, b} (note we have not yet shown the uniqueness of the gcd). If b = 0, then we can simply let d = |a| to satisfy both statements in the theorem.

As to uniqueness of the gcd, suppose we have integers d_1 and d_2 that satisfy both statements of the theorem. Then applying part (b) to both d_1 and d_2,

we must have $d_1|d_2$ and $d_2|d_1$. But both d_1 and d_2 are positive, and hence $d_1 = d_2$. ∎

Corollary If $d = \gcd\{a, b\}$ where a and b are not both zero, then $d = am + bn$ for some $m, n \in \mathbb{Z}$.

Proof Referring to equations (*) in the proof of Theorem 0.9, we note that the equation for r_{k-2} may be solved for r_k to obtain $r_k = r_{k-2} - r_{k-1} q_k$. Next, the equation $r_{k-3} = q_{k-1}r_{k-2} + r_{k-1}$ may be solved for r_{k-1}, and this is then substituted into the previous equation to obtain $r_k = r_{k-2}(1 + q_{k-1}q_k) - r_{k-3} q_k$. Working our way up equations (*), we next eliminate r_{k-2} to obtain r_k in terms of r_{k-3} and r_{k-4}. Continuing in this manner, we eventually obtain r_k in terms of b and a. ∎

If $a, b \in \mathbb{Z}$ and $\gcd\{a, b\} = 1$, then we say that a and b are **relatively prime** (or sometimes **coprime**). The last result on integers that we wish to prove is the result that if p is prime and $p|ab$ (where $a, b \in \mathbb{Z}$), then either $p|a$ or $p|b$.

Theorem 0.10 (a) Suppose $a, b, c \in \mathbb{Z}$ where $a|bc$ and a and b are relatively prime. Then $a|c$.

(b) If p is prime and $a_1, \ldots, a_n \in \mathbb{Z}$ with $p|a_1 \cdots a_n$, then $p|a_i$ for some $i = 1, \ldots, n$.

Proof (a) By the corollary to Theorem 0.9 we have $\gcd\{a, b\} = 1 = am + bn$ for some $m, n \in \mathbb{Z}$. Multiplying this equation by c we obtain $c = amc + bnc$. But $a|bc$ by hypothesis so clearly $a|bnc$. Since it is also obvious that $a|amc$, we see that $a|c$.

(b) We proceed by induction on n, the case $n = 1$ being trivial. We therefore assume that $n > 1$ and $p|a_1 \cdots a_n$. If $p|a_1 \cdots a_{n-1}$, then $p|a_i$ for some $i = 1, \ldots, n - 1$ by our induction hypothesis. On the other hand, if $p \nmid a_1 \cdots a_{n-1}$ then $\gcd\{p, a_1, \ldots, a_{n-1}\} = 1$ since p is prime. We then apply part (a) with $a = p$, $b = a_1 \cdots a_{n-1}$ and $c = a_n$ to conclude that $p|a_n$. ∎

Exercises

1. Find the gcd of the following sets of integers:
   (a) {6, 14}.
   (b) {−75, 105}.
   (c) {14, 21, 35}.

2. Find the gcd of each set and write it in terms of the given integers:
   (a) $\{1001, 33\}$.
   (b) $\{-90, 1386\}$.
   (c) $\{-2860, -2310\}$.

3. Suppose p is prime and $p \nmid a$ where $a \in \mathbb{Z}$. Prove that a and p are relatively prime.

4. Prove that if $\gcd\{a, b\} = 1$ and $c|a$, then $\gcd\{b, c\} = 1$.

5. If $a, b \in \mathbb{Z}^+$, then $m \in \mathbb{Z}^+$ is called the **least common multiple** (abbreviated lcm) if $a|m$ and $b|m$, and if $c \in \mathbb{Z}$ is such that $a|c$ and $b|c$, then $m|c$. Suppose $a = p_1^{s_1} \cdots p_k^{s_k}$ and $b = p_1^{t_1} \cdots p_k^{t_k}$ where $p_1, \ldots, p_k$ are distinct primes and each s_i and t_i are ≥ 0.
   (a) Prove that $a|b$ if and only if $s_i \leq t_i$ for all $i = 1, \ldots, k$.
   (b) For each $i = 1, \ldots, k$ let $u_i = \min\{s_i, t_i\}$ and $v_i = \max\{s_i, t_i\}$. Prove that $\gcd\{a, b\} = p_1^{u_1} \cdots p_k^{u_k}$ and $\text{lcm}\{a, b\} = p_1^{v_1} \cdots p_k^{v_k}$.

6. Prove the **Fundamental Theorem of Arithmetic**: Every integer > 1 can be written as a unique (except for order) product of primes. Here is an outline of the proof:
   (a) Let $S = \{a \in \mathbb{Z}: a > 1$ and a can not be written as a product of primes.$\}$ (In particular, note that S contains no primes.) Show that $S = \varnothing$ by assuming the contrary and using the well-ordered property of the natural numbers.
   (b) To prove uniqueness, assume that $n > 1$ is an integer that has two different expansions as $n = p_1 \cdots p_s = q_1 \cdots q_t$ where all the p_i and q_j are prime. Show that $p_1|q_j$ for some $j = 1, \ldots, t$ and hence that $p_1 = q_j$. Thus p_1 and q_j can be canceled from each side of the equation. Continue this argument to cancel one p_i with one q_j, and then finally concluding that $s = t$.