PART I

BACKGROUND MATERIAL

CHAPTER 1

SETS AND FUNCTIONS

This first chapter introduces the notation, terminology and basic concepts needed for what lies ahead. We review some basic facts about sets in general, about sets of numbers, and about functions. We also take the opportunity to introduce some elementary functions of several (mainly two or three) variables that will be used in several examples in later chapters.

1.1 SETS IN GENERAL

Why should we begin our discussion with sets and not with numbers? After all, most of the sets we deal with are sets of numbers. Furthermore, the mathematical concept of number is older than that of set and is probably more intuitive.

Even though both concepts seem to be primitive (we shall not define either), sets are, in fact, more fundamental than numbers and can be used to generate number systems. Today, most mathematics is based on a solid set theoretic foundation, too lengthy to

Analysis in Vector Spaces.
By M. A. Akcoglu, P. F. A. Bartha and D. M. Ha

present here. Instead, we will confine ourselves to the elements of "naive set theory," doing little more than reviewing some standard notation and presenting a few basic facts about sets.

Definition 1.1.1 Terminology for sets. The following are basic notations and terms involving sets and set relations.

1. A *set* is a collection of objects. If A is a set, then the objects in A are the *elements* or *members* or *points* of A. The notation $x \in A$ means that x is a member of A and $x \notin A$ means that x is not a member of A.

2. The set having no elements is the *empty set*, denoted by $\emptyset$. A nonempty set is a set that contains at least one element.

3. Let A and B be sets. Then A is a *subset of* B if $x \in A$ implies $x \in B$. The notation $A \subset B$ means that A is a subset of B and $A \not\subset B$ means that A is not a subset of B. Thus, $A \not\subset B$ if and only if there is an $x \in A$ such that $x \notin B$. Hence, it follows that $\emptyset \subset B$ for every set B, since $\emptyset$ has no elements, and hence no element x for which $x \notin B$. We say that A is a *proper subset* of B if $A \subset B$ and there is an $x \in B$ such that $x \notin A$.

4. Let X be a set. Then the *power set of* X, denoted by $\mathcal{P}(X)$, is the set of all subsets of X.

5. If $A \subset B$ and $B \subset A$, then we write $A = B$.

6. Let S be a set. For each $x \in S$, let $P(x)$ be a statement about x that is either true or false. We define

$$\{\, x \in S \mid P(x) \,\}$$

to be the subset of S consisting of those x in S for which $P(x)$ is true. When S is clear from the context, this set is also expressed as $\{\, x \mid P(x) \,\}$.

Example 1.1.2 Let $A = \{1, 5, \{1, 5\}\}$, $B = \{1, 5, \emptyset\}$.

1. The members of A are $1, 5$ and $\{1, 5\}$. The members of B are $1, 5$ and $\emptyset$.

2. $A \not\subset B$ because $\{1, 5\} \in A$ but $\{1, 5\} \notin B$. Also, $B \not\subset A$ because $\emptyset \in B$ but $\emptyset \notin A$.

3. $\{1, 5\}$ is a proper subset of both A and B.

4. We have

$$\mathcal{P}(A) = \{\emptyset, \{1\}, \{5\}, \{\{1, 5\}\}, \{1, 5\}, \{1, \{1, 5\}\}, \{5, \{1, 5\}\}, \{1, 5, \{1, 5\}\}\}.$$

Note that $\{1,5\}$ is an element in A so that $\{\{1,5\}\}$ is a subset of A. Thus, $\{\{1,5\}\}$ is an element in $\mathcal{P}(A)$ and $\{\{1,5\}\} \neq \{1,5\}$. Note that A has 3 elements and $\mathcal{P}(A)$ has $8 = 2^3$ elements. In general, whenever X is a finite set with n elements, $\mathcal{P}(X)$ has 2^n elements. We can see this by observing that every subset of X is formed by considering each element of X and either including it or omitting it. $\triangle$

Definition 1.1.3 Operations on sets. Let A, B be sets.

1. We define

$$\begin{aligned} A \cup B &= \{ x \mid x \in A \text{ or } x \in B \} \\ A \cap B &= \{ x \mid x \in A \text{ and } x \in B \} . \end{aligned}$$

   The sets $A \cup B$ and $A \cap B$ are called, respectively, the *union of A and B* and the *intersection of A and B*.

2. If $A \cap B = \emptyset$, then A and B are *disjoint* sets. If $A \cap B \neq \emptyset$, then we say that A and B *intersect*, or A *intersects* B, or B *intersects* A. The sets in a collection of sets are called *pairwise disjoint* if any two (different) sets in this collection are disjoint.

3. The *complement of B in A*, denoted by $A \setminus B$, is defined by

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

   When all sets $A, B, \ldots$ under discussion are subsets of a fixed set S, we write A^c for $S \setminus A$ and B^c for $S \setminus B$. If S is left implicit, then $A^c = S \setminus A$ is called the *complement of A* rather than the complement of A in S.

4. The *symmetric difference* of A and B is defined by

$$A \triangle B = (A \setminus B) \cup (B \setminus A).$$

5. The *Cartesian product of A and B*, denoted by $A \times B$, is defined by

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

   An element (a,b) of $A \times B$ is an *ordered pair*. Note that $A \times B \neq B \times A$ unless $A = B$. Similarly, $A \times B \times C = \{(a,b,c) \mid a \in A, b \in B, c \in C\}$ consists of ordered 3-tuples. The Cartesian product of any finite number of sets is defined in a similar way.

Basic properties of set operations are summarized in the following lemma.

Lemma 1.1.4 *The following are true for all subsets A, B and C of a set X.*

(1) $A \cup B = B \cup A$, $A \cap B = B \cap A$.

(2) $(A \cup B) \cup C = A \cup (B \cup C)$, $(A \cap B) \cap C = A \cap (B \cap C)$.

(3) $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.

(4) $(A \cup B)^c = A^c \cap B^c$.

(5) $(A \cap B)^c = A^c \cup B^c$.

In parts (4) and (5), the complements are with respect to X.

Proof. We will prove part (3) to illustrate the general method; the other assertions are of comparable difficulty or even easier. They are left as exercises.

Let $x \in A \cap (B \cup C)$. Then $x \in A$ and $x \in B \cup C$. Thus, either $x \in B$ or $x \in C$. If $x \in B$, then $x \in A \cap B$. If $x \in C$, then $x \in A \cap C$. Hence, $x \in A \cap B$ or $x \in A \cap C$. That is, $x \in (A \cap B) \cup (A \cap C)$. Thus,

$$A \cap (B \cup C) \quad \subset \quad (A \cap B) \cup (A \cap C) .$$

Conversely, let $y \in (A \cap B) \cup (A \cap C)$. Then either $y \in A \cap B$ or $y \in A \cap C$. If $y \in A \cap B$, then of course, $y \in A \cap (B \cup C)$. Similarly, if $y \in A \cap C$, then $y \in A \cap (B \cup C)$. Thus, in both cases, $y \in A \cap (B \cup C)$. Hence,

$$(A \cap B) \cup (A \cap C) \quad \subset \quad A \cap (B \cup C).$$

These two inclusions imply that $(A \cap B) \cup (A \cap C) = A \cap (B \cup C)$. □

Definition 1.1.5 General unions and intersection. Let $\mathcal{G}$ be a collection of sets. Hence $\mathcal{G}$ is a set whose elements G are also sets. The union and intersection of the sets in this collection are defined in an obvious way as

$$\begin{aligned} \bigcup_{G \in \mathcal{G}} G &= \{ x \mid \text{there is a } G \in \mathcal{G} \text{ such that } x \in G \} , \\ \bigcap_{G \in \mathcal{G}} G &= \{ x \mid x \in G \text{ for all } G \in \mathcal{G} \} . \end{aligned}$$

Additional concepts and notation for dealing with collections of sets will be introduced in the examples below.

Relations and Equivalences

Definition 1.1.6 Relations. Let A and B be sets. A *relation* between the elements of A and the elements of B is a subset R of $A \times B$. If $(a, b) \in R$, then one says that R is *satisfied by a and b* or that *a is related to b by R.* One may omit explicit mention of R if this relation is understood from the context. There is an all-important class of relations: *functions*. They are introduced in Section 1.3. Another important class of relations is formed by *equivalences*, introduced below in Definition 1.1.9.

Example 1.1.7 Let A be a set and let $B = \mathcal{P}(A)$ be the power set of A. Then

$$R = \{ (a, S) \in A \times \mathcal{P}(A) \mid a \in S \}$$

defines a relation between the elements of A and the subsets of A. This relation is satisfied by a and S if and only if $a \in S$. △

Example 1.1.8 Let A be a set. Then

$$R = \{ (U, V) \in \mathcal{P}(A) \times \mathcal{P}(A) \mid U \subset V \}$$

defines a relation R between elements of $\mathcal{P}(A)$. An ordered pair (U, V) of subsets of A satisfies this relation if and only if $U \subset V$. △

Definition 1.1.9 Equivalences. Let A be a set. Let $E \subset A \times A$. Then E is called an *equivalence relation on A* (or an equivalence on A, or simply an equivalence when A is clear from the context) if it has the following properties.

Reflexivity If $a \in A$, then $(a, a) \in E$.

Symmetry If $(a, b) \in E$, then $(b, a) \in E$.

Transitivity If $(a, b) \in E$ and $(b, c) \in E$, then $(a, c) \in E$.

If $E \subset A \times A$ is an equivalence, then one usually writes $a \sim b$ to indicate that $(a, b) \in E$. With this notation the properties above are expressed as follows. Let $a, b, c \in A$. Then (1) $a \sim a$, (2) if $a \sim b$ then $b \sim a$, and (3) if $a \sim b$ and $b \sim c$ then $a \sim c$. If $a \sim b$, then one also says that a and b are *equivalent* (with respect to the underlying equivalence).

Example 1.1.10 Let A be a set and let $C \subset A$. Define a relation $E \subset A \times A$ such that, for any two points a and b in A, $(a, b) \in E$ if and only if both a and b are in C or both a and b are not in C. Hence let

$$E = \{ (a, b) \in A \times A \quad \text{Either } [\, a \in C \text{ and } b \in C \,] \text{ or } [\, a \notin C \text{ and } b \notin C \,] \} .$$

It is easy to check that E is an equivalence . $\triangle$

Definition 1.1.11 Equivalence classes. Let $\sim$ be an equivalence on A. Let $p \in A$. Then

$$[p] = \{ a \in A \mid a \sim p \}$$

is called *the equivalence class of* p. A subset P of A is called *an equivalence class* if there is a $p \in A$ such that $P = [p]$.

Theorem 1.1.12 *Let $\sim$ be an equivalence relation on A.*

(1) Let x, y in A. Then $[x] = [y]$ if and only if $x \sim y$.

(2) Two different equivalence classes are disjoint.

(3) The union of all equivalence classes is A.

Proof. Assume that $x \sim y$. If $a \in [x]$, then $a \sim x$. By transitivity, $a \sim y$. Thus, $[x] \subset [y]$. Similarly, $[y] \subset [x]$. Hence, $[x] = [y]$. Conversely, if $[x] = [y]$, then of course, $x \in [y]$ and $x \sim y$. This proves (1).

Now assume that $[x] \cap [y] \neq \emptyset$. We will show that $[x] = [y]$. Let $a \in [x] \cap [y]$. Then $a \sim x$ and $a \sim y$. Hence, by symmetry and transitivity of $\sim$, we have $x \sim y$. Thus, by part (1), $[x] = [y]$. Hence two different equivalence classes can not intersect. This proves (2).

Finally, $a \in [a]$ for all $a \in A$. Hence, every element of A belongs to an equivalence class. This proves (3). $\square$

Definition 1.1.13 Complete set of representatives. Let A be a set with an equivalence. Let P be an equivalence class. Any point $p \in P$ is called a *representative* for P. A subset R of A is called a *complete set of representatives* if each equivalence class has exactly one point in R as its representative.

Example 1.1.14 Let A be a set and let C be a subset of A. Let $\sim$ be the equivalence defined in Example 1.1.10. Then C and $A \setminus C$ are the only two equivalence classes. If both C and $A \setminus C$ are nonempty, then any two-point set consisting one point from C and one point from $A \setminus C$ is a complete set of representatives. $\triangle$

Example 1.1.15 Let X and Y be sets and let $Z = X \times Y$. Define a relation $\sim$ on Z as follows: $(x, y) \sim (x, y')$ for all $x \in X$ and all y, y' in Y. We verify easily that this relation is an equivalence on Z. Let $b \in Y$ be a fixed point in Y. It is easy to see that $X \times \{b\}$ is a complete set of representatives for this equivalence. $\triangle$

Problems

1.1 Give an example of a family of sets such that any two sets in the family intersect (that is, they have nonempty intersection) but the intersection of all the sets in this family is empty.

1.2 Let $\mathcal{A}$ be a collection of subsets of a set X. Show that

$$(\bigcup_{A\in\mathcal{A}} A)^c = \bigcap_{A\in\mathcal{A}} A^c \text{ and } (\bigcap_{A\in\mathcal{A}} A)^c = \bigcup_{A\in\mathcal{A}} A^c.$$

Hence, $(A \cup B)^c = A^c \cap B^c$ and $(A \cap B)^c = A^c \cup B^c$.

1.3 If $\mathcal{A}$ is a collection of subsets of a set X and if $B \subset X$, then show that

$$\begin{aligned}\left(\bigcup_{A\in\mathcal{A}} A\right) \cap B &= \bigcup_{A\in\mathcal{A}} (A \cap B) \text{ and} \\ \left(\bigcap_{A\in\mathcal{A}} A\right) \cup B &= \bigcap_{A\in\mathcal{A}} (A \cup B).\end{aligned}$$

1.4 Show that $A\triangle B = (A \cup B) \setminus (A \cap B)$. Deduce that $A\triangle B = \emptyset$ if and only if $A = B$ and $A\triangle B = X$ if and only if $B = A^c$.

1.5 Show that $A\triangle B \subset (A\triangle C) \cup (C\triangle B)$ for any three sets. Give an example to show that in general the inclusion is a proper inclusion.

1.6 A collection of subsets of a set X is called an *algebra of sets* if it satisfies the following three conditions:

1. $X \in \mathcal{A}$;
2. if $A \in \mathcal{A}$, then also $A^c \in \mathcal{A}$;
3. if $A, B \in \mathcal{A}$, then also $A \cup B \in \mathcal{A}$.

Show that if $\mathcal{A}$ is an algebra of sets and if $A, B \in \mathcal{A}$, then $A \cap B$, $A \setminus B$, and $A\triangle B$ are also in $\mathcal{A}$.

1.7 A collection of nonempty subsets of a set X is called a *partition* of X if the sets in this collection are pairwise disjoint and if their union is X. Show that any partition is the family of equivalence classes with respect to an equivalence on X.

1.8 Let $\mathcal{A}$ be a partition of X and $\mathcal{B}$ a partition of Y. Show that the family

$$\mathcal{C} = \{ A \times B \mid A \in \mathcal{A},\ B \in \mathcal{B} \}$$

is a partition of $X \times Y$.

1.2 SETS OF NUMBERS

In this course we will deal mainly with sets of numbers. We shall assume that the set of *natural numbers*, $\mathbb{N}$, the set of *integers*, $\mathbb{Z}$, the set of *nonnegative integers*, $\mathbb{Z}^+$, and the set of *rational numbers*, $\mathbb{Q}$, are familiar. These sets are

$$\begin{aligned} \mathbb{N} &= \{1, 2, 3, \ldots\}, \\ \mathbb{Z} &= \{\ldots, -2, -1, 0, 1, 2, \ldots\}, \\ \mathbb{Z}^- &= \{0, 1, 2, \ldots\}, \\ \mathbb{Q} &= \{\, a/b \mid a, b \in \mathbb{Z}, b \neq 0 \,\}. \end{aligned}$$

We shall assume that the reader is acquainted with the addition and multiplication operations and the order relations on these sets. Another familiar set of numbers is $\mathbb{R}$, the set of *real numbers*. Note that $\mathbb{N} \subset \mathbb{Z}^+ \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R}$. We will discuss $\mathbb{R}$ in some detail in Chapter 2. However, we shall assume that the basic properties of the real numbers are familiar. These properties are used below to give other examples of sets of numbers.

Intervals in $\mathbb{R}$

Example 1.2.1 Intervals. Four basic types of unbounded interval are defined in terms of a number $p \in \mathbb{R}$. These intervals are, in standard notation,

$$\begin{aligned} [p, \infty) &= \{\, r \in \mathbb{R} \mid p \leq r \,\}, & (p, \infty) &= \{\, r \in \mathbb{R} \mid p < r \,\}, \\ (-\infty, p\,] &= \{\, r \in \mathbb{R} \mid r \leq p \,\}, & (-\infty, p\,) &= \{\, r \in \mathbb{R} \mid r < p \,\}. \end{aligned}$$

Intersections of these intervals give other types of intervals. For example, again in standard notation,

$$[\,a, b\,) = (-\infty, b\,) \cap [\,a, \infty) = \{\, t \in \mathbb{R} \mid a \leq t < b \,\}.$$

Here a and b are two fixed numbers in $\mathbb{R}$. Note that $[\,a, b\,) = \emptyset$ if $b \leq a$. $\triangle$

Example 1.2.2 Collections of intervals. Let $r > 0$ be a fixed number. For each $a \in \mathbb{R}$, let $I_a = [a - r, a + r)$. Then $\mathfrak{I} = \{\, I_a \mid a \in \mathbb{R} \,\}$ is a collection of intervals. Denote this collection as $\{I_a\}_{a \in \mathbb{R}}$ and the union and intersection of the intervals in this collection as $\bigcup_{a \in \mathbb{R}} I_a$ and $\bigcap_{a \in \mathfrak{I}} I_a$. Obviously,

$$\bigcup\nolimits_{a \in \mathbb{R}} I_a = \mathbb{R} \quad \text{and} \quad \bigcap\nolimits_{a \in \mathfrak{I}} I_a = \emptyset.$$

We can also consider subcollections of this collection. For example, $\{I_a\}_{0 \leq a \leq 1}$ is such a subcollection. We see easily that

$$\bigcup\nolimits_{0 \leq a \leq 1} I_a = [-r, 1 + r) \quad \text{and} \quad \bigcap\nolimits_{0 \leq a \leq 1} I_a = [1 - r, r).$$

Hence $\bigcap_{0<a\leq 1} I_a = \emptyset$ if $1 - r \geq r$, that is, if $r \leq 1/2$. $\triangle$

Examples 1.2.3 Lines and half-planes in $\mathbb{R}^2$. We consider $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$ as the usual xy-plane. It consists of all ordered pairs (x, y), with $x, y \in \mathbb{R}$. A line in $\mathbb{R}^2$ is a set of the form $L = \{ (x, y) \in \mathbb{R}^2 \mid Ax + By + C = 0 \}$, where A, B, and C are three fixed numbers in $\mathbb{R}$ and at least one of A or B is nonzero. A line divides $\mathbb{R}^2$ into three pairwise disjoint sets

$$
\begin{aligned}
L &= \{ (x, y) \in \mathbb{R}^2 \mid Ax + By + C = 0 \}, \\
H_1 &= \{ (x, y) \in \mathbb{R}^2 \mid Ax + By + C > 0 \}, \\
H_2 &= \{ (x, y) \in \mathbb{R}^2 \mid Ax + By + C < 0 \}.
\end{aligned}
$$

Here H_1 and H_2 are the two half-planes bounded by the line L. We may refer to them as the *lower* and *upper* half-planes or as the *left-hand* and *right-hand* half-planes, depending on the position of L. Finally, one may also refer to the equation $Ax + By + C = 0$ as a *line*. $\triangle$

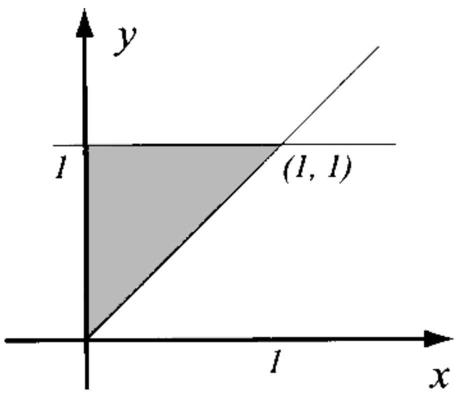

Figure 1.1. Triangle in Example 1.2.4.

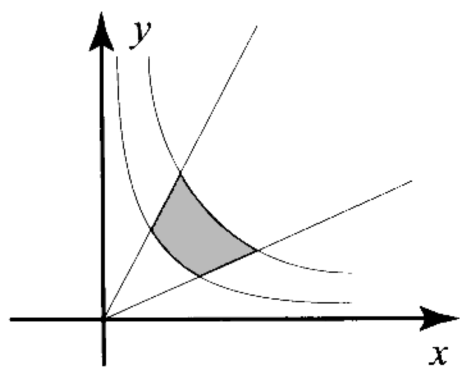

Figure 1.2. Region in Example 1.2.5.

Example 1.2.4 Let $R \subset \mathbb{R}^2$ be the triangle in the xy-plane bounded by the lines $x = 0$, $y = 1$, and $x = y$ (see Figure 1.1). Specify R by a set of inequalities.

Solution. We see that R is the intersection of the following three half-planes. (1) The right-hand side of $x = 0$, (2) the lower part of $y = 1$, and (3) the upper part of $x = y$. Hence $(x, y) \in R$ if and only if $x \geq 0$ and $y \leq 1$ and $y \geq x$. We can express this more concisely as

$$(x, y) \in R \text{ if and only if } 0 \leq x \leq y \leq 1.$$

Here we have assumed that R contains the inside, the edges, and the vertices of this triangle. The inside of R without the vertices and without the edges corresponds to the relation $0 < x < y < 1$. $\triangle$

Example 1.2.5 Let R be the region in the xy-plane that lies in the first quadrant (that is, $x \geq 0$ and $y \geq 0$) and is between the hyperbolas $xy = 1$, $xy = 2$, and bounded by the lines $2y = x$, $y = 2x$ (see Figure 1.2). Specify R by a set of inequalities.

Solution. The region in the first quadrant that is between the hyperbolas $xy = 1$ and $xy = 2$ is specified by the conditions that $x > 0$ and $1 \leq xy \leq 2$. Similarly, the region in the first quadrant that is between the lines $2y = x$ and $y = 2x$ is specified by the conditions that $x > 0$ and $1/2 \leq y/x \leq 1$. Hence we see that $(x, y) \in R$ if and only if

$$1 \leq xy \leq 2 \text{ and } 1/2 \leq y/x \leq 2 \text{ and } x > 0. \quad \triangle$$

Remark 1.2.6 Note that in the last two examples we have used undefined (yet intuitive) terms like *bounded by* and *upper part*. In such cases, the final formal statements in terms of inequalities may be considered as the definition of these intuitive terms.

Discs in $\mathbb{R}^2$

Example 1.2.7 Collections of discs. Let $r > 0$. If (a, b) is a point in the xy-plane (i.e., a point in $\mathbb{R}^2$), then let

$$D_r(a, b) = \left\{ (x, y) \in \mathbb{R}^2 \mid (x - a)^2 - (y - b)^2 < r^2 \right\}$$

be the (open) disc of radius r about the point (a, b). This notation is convenient because we shall often need to refer to open discs, just as we often refer to intervals in $\mathbb{R}$. We see easily that

$$\bigcup_{(a, b) \in \mathbb{R}^2} D_r(a, b) = \mathbb{R}^2 \text{ and } \bigcap_{(a, b) \in \mathbb{R}^2} D_r(a, b) = \emptyset.$$

It may require some work to identify

$$U(G, r) = \bigcup_{(a, b) \in G} D_r(a, b) \text{ and } I(G, r) = \bigcap_{(a, b) \in G} D_r(a, b)$$

for various regions G in the xy-plane. If

$$G = \left\{ (a, b) \in \mathbb{R}^2 \mid a^2 + b^2 = 1 \right\}$$

is the unit circle about the origin, for example, then

$$U(G, r) = \begin{cases} \left\{ (x, y) \mid (1 - r)^2 < (x^2 + y^2) < (1 + r)^2 \right\} & \text{if } r \leq 1 \\ \left\{ (x, y) \mid (x^2 + y^2) < (1 + r)^2 \right\} & \text{if } r > 1, \end{cases}$$

$$I(G, r) = \begin{cases} \emptyset & \text{if } r \leq 1 \\ \left\{ (x, y) \mid (x^2 + y^2) < (r - 1)^2 \right\} & \text{if } r > 1. \end{cases}$$

In obtaining these results, it is helpful to keep in mind the geometric interpretation of the above sets. For example, $U(G, r)$ is the set of all points in the plane that have a distance less than r to a point in G. $\triangle$

The Induction Principle

There is an "obvious" property of $\mathbb{N}$ called the *well-ordering property*. Since this property does not follow formally from the rules of the order relation, it is stated as an axiom for $\mathbb{N}$.

Axiom 1.2.8 The well-ordering axiom. Any nonempty subset of $\mathbb{N}$ contains a smallest element. More explicitly, if $T \subset \mathbb{N}$ and if T is not empty, then there is an $n \in T$ such that $n \leq m$ for all $m \in T$.

The well-ordering of $\mathbb{N}$ implies the induction principle for $\mathbb{N}$.

Theorem 1.2.9 Induction principle. *Let $S \subset \mathbb{N}$. Suppose that $1 \in S$ and that $k + 1 \in S$ whenever $k \in S$. Then $S = \mathbb{N}$.*

Proof. Assume, on the contrary, that $S \neq \mathbb{N}$. Let $T = \mathbb{N} \setminus S$. Then T is a nonempty subset of $\mathbb{N}$. Hence, by the well-ordering axiom, there is a smallest element a in T. Since $a \notin S$ and $1 \in S$, we must have $a > 1$. Thus, $a - 1 \in \mathbb{N}$. Since a is the smallest element in T, we must have $a - 1 \notin T$. Hence, $a - 1 \in S$. By property 2, we have $a = (a - 1) + 1 \in S$, a contradiction. $\square$

Example 1.2.10 Let $S \subset \mathbb{Z}$. Assume that $a \in S$ and that $k + 1 \in S$ whenever $k \in S$. Show that $\{ a + k \mid k \in \mathbb{N} \} \subset S$.

Solution. Set $T = \{ k \in \mathbb{N} \mid a + k \in S \}$. We will show that $T = \mathbb{N}$. First, since $a \in S$, we have $a + 1 \in S$. Hence, $1 \in T$. Assume that $k \in T$. Then $a + k \in S$. Hence, by assumption, $a + (k + 1) = (a + k) + 1 \in S$. Thus, $k + 1 \in T$. By the induction principle, $T = \mathbb{N}$. $\triangle$

The following example shows how the induction principle gives us the familiar method of proof by induction. The idea is to prove that some result holds for $n = 1$ (the *base case*) and to show that if it holds for n, then it holds for $n + 1$ (the *inductive step*).

Example 1.2.11 Show that $1 + 2 + \cdots + n = n(n + 1)/2$ for all $n \in \mathbb{N}$.

Solution. Let $S = \{ n \in \mathbb{N} \mid 1 + 2 + \cdots + n = n(n+1)/2 \}$. Then $1 \in S$ (the base case). For the inductive step, suppose that $n \in S$. Then

$$\begin{aligned} 1 + 2 + \cdots + n + (n+1) &= (1 + 2 + \cdots + n) + (n+1) \\ &= (n(n+1)/2) + (n+1) \\ &= (n+1)(n+2)/2 \end{aligned}$$

shows that $(n+1) \in S$. Hence $S = \mathbb{N}$ by the induction principle 1.2.9. $\triangle$

Remarks 1.2.12 Limitations of the induction principle. The induction principle is useful in stating certain arguments in a clear and concise way. But this principle may not be very helpful in obtaining new results. For example, the principle will not help you to guess the result $1 + \cdots + n = n(n+1)/2$. To obtain this result, you need other methods.

Definition 1.2.13 Binomial coefficients. Let $r \in \mathbb{R}$ and $k \in \mathbb{N}$. Define

$$\binom{r}{0} = 1, \quad \binom{r}{1} = r, \quad \text{and} \quad \binom{r}{k+1} = \frac{r(r-1)\cdots(r-k)}{(k+1)!}.$$

These are the general binomial coefficients. These expressions will be used in examples and later in the discussion of multilinear functions. If $r \in \mathbb{N}$, then $\binom{r}{k}$ is the number of ways to select k objects from a collection of r objects.

Problems

1.9 Express the set

$$C = \left\{ x \in \mathbb{R} \;\middle|\; 0 < x^2 - 5x + 4 \le 10 \right\} \subset \mathbb{R}$$

in terms of intervals.

1.10 There are four bounded regions in the xy-plane bounded by the lines $y = x$ and $y = 2x + 1$, and by the ellipse $x^2 + 4y^2 = 16$. One of these regions is

$$\left\{ (x, y) \in \mathbb{R}^2 \;\middle|\; x^2 + 4y^2 \le 16 \text{ and } y \ge x \text{ and } y \ge 2x + 1 \right\}.$$

Express the other three regions similarly.

1.11 Let $G = \left\{ (x, y) \in \mathbb{R}^2 \mid -1 \le x \le 1, \ -1 \le y \le 1 \right\}$. Consider the sets $U(G, 1)$ and $I(G, 1)$ defined in Example 1.2.7. Express these sets in simpler terms.

1.12 Let $(a_1, b_1), (a_2, b_2)$ be in $U(C, r)$. Consider $\mathbb{R}^2$ as the xy-plane and use the customary vectorial notations. A subset C of the xy-plane is called *convex* if whenever C contains two points $(a_1,\, b_1)$ and $(a_2,\, b_2)$, then C also contains all the points on the line segment joining these points. Let C be the set of all points $(x,\, y)$ that can be expressed as

$$(x,\, y) = p\,(1,\, 0) + q\,(2,\, 0) + r\,(0,\, 2) + s\,(-1,\, -1)\,,$$

where p, q, r and s are all nonnegative and $p + q + r + s = 1$. Show that C is a convex set and describe it in simpler terms.

1.13 Define a relation $D \subset \mathbb{Z} \times \mathbb{Z}$ as

$$D = \{\, (a,\, b) \in \mathbb{Z} \times \mathbb{Z} \mid \text{There is a } k \in \mathbb{Z} \text{ such that } a = kb \,\}\,.$$

This relation is called *divisibility*: $(a,\, b) \in D$ just in case a is *divisible by b*. Show that divisibility is reflexive and transitive but not symmetric.

1.14 Let $p \in \mathbb{N}$. Define a relation $C_p \subset \mathbb{Z} \times \mathbb{Z}$ by

$$C_p = \{\, (a,\, b) \in \mathbb{Z} \times \mathbb{Z} \mid (a - b,\, p) \in D \,\}\,,$$

where D is the divisibility relation defined in Problem 1.13. This relation among the integers is called *congruence modulo p*.

1. Show that congruence modulo p is an equivalence on $\mathbb{Z}$.
2. Show that there are exactly p equivalence classes for this equivalence.
3. Show that the set of integers

$$R = \{\, 1,\, 2,\, \ldots,\, p \,\}$$

   is a complete set of representatives for congruence modulo p.

4. What is the equivalence class represented by $k \in \mathbb{Z}$?

1.15 Define a relation $C_1 \subset \mathbb{R} \times \mathbb{R}$ by the condition that $(r,\, s) \in C_1$ if and only if $(r - s) \in \mathbb{Z}$. This relation among the real numbers is called *congruence modulo* 1.

1. Show that congruence modulo 1 is an equivalence on $\mathbb{R}$.
2. Show that the interval

$$R = [0,\, 1) = \{\, t \in \mathbb{R} \mid 0 \le t < 1 \,\}$$

is a complete set of representatives for congruence modulo 1.

3. What is the equivalence class represented by $t \in R$?

Note. Let $p \in \mathbb{N}$, $p \geq 2$. Congruence modulo p can be defined on $\mathbb{R}$, but this is not customary.

1.16 Define a relation among the points $(x, y) \in \mathbb{R}^2$ as follows. A point (x, y) is related to a point (x', y') if and only if $x^2 + y^2 = x'^2 + y'^2$. Show that this is an equivalence. What are the equivalence classes? What is a complete set of representatives? Is the x-axis

$$\{ (x, 0) \in \mathbb{R}^2 \mid x = 0 \}$$

a complete set of representatives? Why or why not?

1.17 Define a relation among the points $(x, y) \in \mathbb{R}^2$ as follows. A point (x, y) is related to a point (x', y') if and only if $xy = x'y'$. Show that this is an equivalence. What are the equivalence classes? What is the equivalence class containing the origin $(0, 0)$? What is a complete set of representatives? Is the line

$$\{ (x, y) \in \mathbb{R}^2 \mid x = y \}$$

a complete set of representatives? Why or why not?

1.18 Let C be a convex set in the xy-plane (see Problem 1.12) and let $r > 0$. Show that the sets $U(C, r)$ and $I(C, r)$ are also convex, with the notation of Example 1.2.7. (Hint for the convexity of $I(C, r)$: show that the intersection of any family of convex sets is convex.)

1.19 Let $n \in \mathbb{N}$. Show by induction that $4^n - 3n - 1$ is divisible by 9. (Divisibility is defined in Problem 1.13.)

1.20 Let $r \in \mathbb{R}$ and $n \in \mathbb{Z}^+$.

1. Show that

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

2. Use the induction principle to show that for all integers $n \geq 0$,

$$\sum_{k=0}^{n} \binom{r+k}{k} = \binom{r+n+1}{n}.$$

1.21 (*Binomial Theorem*) Let $a, b \in \mathbb{R}$ and $n \in \mathbb{Z}^+$. Use the induction principle to show that

$$(a + b)^n = \sum_{k=0}^{n} \begin{pmatrix} n \\ k \end{pmatrix} a^{n-k} b^k .$$

1.22 Let $r, s \in \mathbb{R}$ and $n \in \mathbb{Z}^+$. Show that

$$\sum_{k=0}^{n} \begin{pmatrix} r \\ k \end{pmatrix} \begin{pmatrix} s \\ n-k \end{pmatrix} = \begin{pmatrix} r+s \\ n \end{pmatrix} .$$

1.23 Let $n \in \mathbb{N}$. Show that $\sum_{k=1}^{n} k^2 = (1/6)n(n + 1)(2n + 1)$.

1.3 FUNCTIONS

The concept of a function is one of the most important ideas in mathematics. Functions are certainly of paramount importance in analysis. A function from a set X to a set Y is a special type of relation between the elements of X and Y. A subset of $X \times Y$ that defines a function is called a *graph*.

Definition 1.3.1 Graphs. Let X and Y be nonempty sets. A subset Γ of $X \times Y$ is called a *graph* if, whenever (x, y) and (x, y') are both in Γ, then $y = y'$. Hence Γ is a graph if for each $x \in X$ there is at most one point $(x, y) \in \Gamma$. *The domain of* Γ, denoted by $\operatorname{Dom}\Gamma$, is defined by

$$\operatorname{Dom}\Gamma = \{\, x \in X \mid \text{there is a } y \in Y \text{ such that } (x, y) \in \Gamma \,\} .$$

Also, *the domain space of* Γ is X and *the range space of* Γ is Y.

Definition 1.3.2 Functions. Let $\Gamma \subset X \times Y$ be a graph and let $D = \operatorname{Dom}\Gamma$. The relation defined by this graph Γ is called a *function* $f : D \to Y$ from D to Y. It is customary to denote graphs and functions by different symbols. What distinguishes a function f from a general relation is the condition that each $x \in D$ is related to a *unique* point $y \in Y$ by f. One calls y *the value of* f *at* x, or *the image of* x *under* f, and one writes this as $y = f(x)$. If $x \in D$, then one also says that f *is defined at* x. The domain D of Γ is also called *the domain* of f and denoted as $\operatorname{Dom} f$. *The domain space of* f is X and *the range space of* f is Y.

Remark 1.3.3 Roles of X and Y. Note that the sets X and Y are not uniquely determined by the function f. In fact, X can be any set containing the domain of f, and Y can be any set containing all the values of f. They do determine, however, the nature of the function under consideration. If $X = Y = \mathbb{R}$, for example, then we are dealing with a real-valued function defined on a set of real numbers. Also, the sets X and Y are important for defining bijections in Definition 1.3.20 below.

Remarks 1.3.4 Role of the graph. The formal definition of a function given above in Definition 1.3.2 is satisfactory but rarely used in the actual statement of a function. One usually defines $f(x)$ for a general point x by an explicit rule or computational formula. Nevertheless, the graphical definition of a function is an important idea that will have applications later.

Remarks 1.3.5 Notation for functions. When a function is defined by an explicit rule, then one denotes this function by $y = f(x)$ or by $f(x)$. This notation is not strictly correct: $y = f(x)$ is a point in Y rather than the function $f : D \to Y$ itself. Nevertheless, $y = f(x)$ is convenient notation which causes no confusion in practice. When a function is given as $y = f(x)$, $D = \operatorname{Dom} f$ is understood to be the set of x for which $f(x)$ is defined.

We shall sometimes express a function as $y = y(x)$. This notation indicates that we are dealing with a function the points of whose domain space are denoted by x and points in the range space by y. It eliminates the unnecessary symbol f. Still, the most common notation in practice is $y = f(x)$.

Definition 1.3.6 Restrictions of a function. Let $f : D \to Y$ be a function and $A \subset D$. The *restriction of f to A* is a new function that has the value $f(x)$ for $x \in A$ but is undefined if $x \notin A$. In general, it is not necessary to use a different notation for the restricted function. This is understood from the context.

Definition 1.3.7 Identity functions. Let X be a set. Define a function $I_X : X \to X$ as $I_X(x) = x$ for all $x \in X$. It is called the *identity function* on X, or simply the identity on X. We also write I instead of I_X if X is understood from the context.

Definition 1.3.8 Sequences. Let $\mathbb{K}$ be a subset of $\mathbb{Z}$ such that $\mathbb{K}$ is bounded below but not above. A function defined on $\mathbb{K}$ is called a *sequence*. One usually takes $\mathbb{K}$ as $\mathbb{N}$ or an unbounded subset of $\mathbb{N}$. The range space of a sequence can be any nonempty set Y. By a *sequence in Y*, we mean a sequence with the range space Y. $\mathbb{K}$ is called an *index set*. The value of a sequence $a : \mathbb{K} \to Y$ at $k \in \mathbb{K}$ is called the kth term of the sequence and is denoted by a_k. The sequence itself may be denoted as $a : \mathbb{K} \to Y$, as a_k, $k \in \mathbb{K}$, as $\{a_k\}$, or simply as a_k if the domain $\mathbb{K}$ is understood. This last notation is not strictly correct, but it is convenient to use when the meaning is clear from the context. A sequence is usually given by a formula involving n. Such a sequence is defined for all $n \in \mathbb{N}$ for which this formula is meaningful. For example, $a_n = (n - 5)$ defines a sequence $a : \mathbb{N} \to \mathbb{Z}$ and $b_n = 1/(n - 5)$ defines a sequence $b : \mathbb{K} \to \mathbb{Q}$, where $\mathbb{K} = \mathbb{N} \setminus \{5\}$.

Examples 1.3.9 Graphs in $\mathbb{R} \times \mathbb{R} = \mathbb{R}^2$. The graphs introduced in Definition 1.3.1 above become ordinary graphs in the plane when $X = Y = \mathbb{R}$.

1. Let $G = \{ (x, x^2) \in \mathbb{R}^2 \mid x \in \mathbb{R} \}$. Then G is a graph and $\operatorname{Dom} G = \mathbb{R}$. The function defined by G is $y = x^2$. As noted above in Remarks 1.3.4, one defines this function directly by the formula $y = x^2$, without mentioning the graph G. In this case, G is a parabola.

2. Let $G' = \{ (x, y) \in \mathbb{R}^2 \mid x = y^2 \}$. Then G' is once again a parabola, but G' is not a graph. If $x > 0$, then both $(x, x^{1/2})$ and $(x, -x^{1/2})$ are in G' and $x^{1/2} \neq -x^{1/2}$. The uniqueness requirement of Definition 1.3.1 is not satisfied.

   Note that G' is obtained from G by interchanging x and y. As x and y do not appear symmetrically in the definition of a graph, it turns out that G is a graph but G' is not.

3. Let $G_1 = \{ (x, y) \in \mathbb{R}^2 \mid x = y^2,\ y \geq 0 \}$. G_1 is the upper branch of the parabola G' above. G_1 is a graph and $\operatorname{Dom} G = [0, \infty)$, the positive part of the x-axis. The function g_1 defined by G_1 is $y = x^{1/2}$. Note that this formula is meaningful only if $x \geq 0$. Hence $D = [0, \infty)$ is the domain of g_1. The lower branch G_2 of G' is another graph. The function g_2 defined by G_2 is $y = -x^{1/2}$. Also, $\operatorname{Dom} g_2 = [0, \infty)$.

4. Let $H = \{ (x, y) \in \mathbb{R}^2 \mid xy = 1 \}$ be a hyperbola. We see that H is a graph. In fact, if (x, y) and (x, y') are both in H, then $xy = xy' = 1$ implies that $x \neq 0$ and $y = y'$. We see easily that the domain of H is

$$D = \operatorname{Dom} H = (\mathbb{R} \setminus \{0\}) = (-\infty, 0) \cup (0, \infty).$$

   The function defined by H is $y = 1/x$, $x \neq 0$. $\triangle$

Examples 1.3.10 More general functions. In this course we deal mainly with functions for which the domain space is $\mathbb{R}^m$ and the range space is $\mathbb{R}^n$, where $m, n \in \mathbb{N}$. An efficient way to work with these functions in specific examples is to define them directly by a set of formulas.

1. Let $m = 2$, $n = 1$. Identify the domain space $\mathbb{R}^2$ with the xy-plane and the range space $\mathbb{R}$ with the z-axis. A formula $z = f(x, y)$ gives us a real-valued function defined on a subset D of the xy-plane. For example, let $f(x, y) = xy/(x^2 + y^2)$. Then the domain of f is the set D of all $(x, y) \in \mathbb{R}^2$ for which the expression $(xy)/(x^2 + y^2)$ is meaningful. We see that D contains all points in $\mathbb{R}^2$ except the origin $(0, 0)$ of $\mathbb{R}^2$. The graph Γ of f is a subset of $\mathbb{R}^2 \times \mathbb{R} = \mathbb{R}^3$. More explicitly,

$$\Gamma = \{ (x, y, z) \in \mathbb{R}^3 \mid z = xy/(x^2 + y^2) \text{ and } x^2 + y^2 \neq 0 \}.$$

Geometrically, Γ is a surface in $\mathbb{R}^3$.

2. Let $m = n = 2$. Identify the domain space $\mathbb{R}^2$ with the xy-plane and the range space $\mathbb{R}^2$ with the uv-plane. A set of formulas

$$u = u(x, y), \quad v = v(x, y)$$

gives us a function defined on a subset of the xy-plane and taking values in the uv-plane. For example

$$u = x^2 + y^2, \quad v = y/x$$

is such a set of formulas. The domain of this function is the set

$$D = \left\{ (x, y) \mid (x, y) \in \mathbb{R}^2, \ x \neq 0 \right\}.$$

Hence one obtains D by removing the y-axis from the xy-plane. $\triangle$

Remark 1.3.11 Coordinate changes. Certain functions for which both the domain and range spaces are $\mathbb{R}^n$ are called *coordinate changes* in $\mathbb{R}^n$. Coordinate changes are considered in later chapters in some detail. Here we provide a few examples of functions $\mathbb{R}^n \to \mathbb{R}^n$ that are used as coordinate changes. $\triangle$

Example 1.3.12 Polar coordinates. Polar coordinates are given as

$$x = x(r, \theta) = r\cos\theta, \quad y = y(r, \theta) = r\sin\theta.$$

This is a coordinate change in $\mathbb{R}^2$. Both the domain and range spaces for this function are $\mathbb{R}^2$. The domain space $\mathbb{R}^2$ is identified with the $r\theta$-plane and the range space $\mathbb{R}^2$ with the xy-plane. This function takes the point $(r, \theta) \in \mathbb{R}^2$ in the $r\theta$-plane to the point $(r\cos\theta, r\sin\theta) \in \mathbb{R}^2$ in the xy-plane. In the present context, $r \in \mathbb{R}$ and $\theta \in \mathbb{R}$ are two real numbers and $(r, \theta) \in \mathbb{R}^2$ is an ordered pair of real numbers. We see that $(r\cos\theta, r\sin\theta) \in \mathbb{R}^2$ is defined for all $(r, \theta) \in \mathbb{R}^2$. Hence the domain of this function is also $\mathbb{R}^2$. $\triangle$

Example 1.3.13 Cylindrical coordinates. These coordinates are given as

$$x = r\cos\theta, \quad y = r\sin\theta, \quad z = \zeta.$$

This is a coordinate change in $\mathbb{R}^3$. The domain space $\mathbb{R}^3$ is identified with the $r\theta\zeta$-space and the range space $\mathbb{R}^3$ with the xyz-space. Actually, the standard notation for ζ is also z. Hence the domain space is the $r\theta z$-space and the range space is the xyz-space. The domain of this function is also $\mathbb{R}^3$, since

$$(r\cos\theta, r\sin\theta, \zeta) \in \mathbb{R}^3$$

is defined for all $(r, \theta, \zeta) \in \mathbb{R}^3$. $\triangle$

Example 1.3.14 Spherical coordinates. These coordinates are given as

$$x = \rho \sin\varphi \cos\theta, \quad y = \rho \sin\varphi \sin\theta, \quad z = \rho \cos\varphi.$$

This is a coordinate change in $\mathbb{R}^3$. The domain space $\mathbb{R}^3$ is identified with the $\rho\varphi\theta$-space and the range space $\mathbb{R}^3$ with the xyz-space. The domain of this function is also $\mathbb{R}^3$, since

$$(\rho \sin\varphi \cos\theta, \ \rho \sin\varphi \sin\theta, \ \rho \cos\varphi) \in \mathbb{R}^3$$

is defined for all $(\rho, \varphi, \theta) \in \mathbb{R}^3$. $\triangle$

Images Under Functions

Definition 1.3.15 Images of sets. Let $f : D \to Y$ be a function with domain space X and range space Y.

1. **Direct images under a function.** Let $U \subset X$. Then the *direct image of U* under f is defined as the set

$$f(U) = \{\, y \in Y \mid y = f(x), \ \ x \in U \cap D \,\}.$$

   Hence $f(U)$ is the set of all values $f(x)$, where $x \in U \cap D$. Note that $f(U) = \emptyset$ if and only if $U \cap D = \emptyset$. Also note that $f(U) = f(U \cap D)$ for any set $U \subset X$.

2. **The range of a function.** The direct image of the domain space is called the *range of f* and denoted as Range f. Hence

$$\text{Range}\, f = f(X) = f(D).$$

   The range of f is the set of all points $y \in Y$ in the range space that are the images of points $x \in D$.

3. **Inverse images under a function.** Let $V \subset Y$. Then the inverse image of V under f is the set

$$f^{-1}(V) = \{\, x \in X \mid f(x) \in V \,\} = \{\, x \in D \mid f(x) \in V \,\}.$$

   Here the first equality is the definition of $f^{-1}(V)$ as the set of all points $x \in X$ which have images $f(x)$ in V. Since $f(x)$ exists only for $x \in D$, we need to consider only points $x \in D$. This is expressed by the second equality.

Example 1.3.16 Images under polar coordinates. Let $f : \mathbb{R}^2 \to \mathbb{R}^2$ be the polar coordinates defined in Example 1.3.12. The domain space is represented by the

$r\theta$-plane and the range space by the xy-plane. The value of f at the point (r, θ) in the domain space is the point

$$(x, y) = f(r, \theta) = (r\cos\theta, r\sin\theta)$$

in the range space. Note that

$$\begin{aligned} f(r, \theta) &= f(r, \theta + 2k\pi) \\ &= f(-r, \theta + (2k+1)\pi) \text{ for all } k \in \mathbb{Z}. \end{aligned}$$

Hence, if an inverse image $f^{-1}(V)$ contains a point (r, θ), then it also contains all the points of the form

$$(r, \theta + 2k\pi) \text{ and } (-r, \theta - (2k+1)\pi), \text{ where } k \in \mathbb{Z}.$$

Vertical lines in the $r\theta$-plane correspond to the constant values of r. We see that the direct images of these lines in the $r\theta$-plane are concentric circles about the origin in the xy-plane. The images of the horizontal lines in the $r\theta$-plane are lines passing through the origin in the xy-plane. Let C be a circle of radius a about the origin in the xy-plane. Then the inverse image $f^{-1}(C)$ of this circle consists of two vertical lines $r = \pm a$ in the $r\theta$-plane. Let L be a line in the xy-plane passing through the origin and making an angle of φ with the positive x-axis. Then the inverse image $f^{-1}(L)$ of this lines consists of infinitely many horizontal lines in the $r\theta$-plane given as $\theta = \varphi + k\pi$, $k \in \mathbb{Z}$.

Figure 1.3 shows a part of the inverse image of the shaded region in the xy-plane. This region is bounded by two circles about the origin and two lines passing through the origin. Its inverse image under f is the union of infinitely many rectangles in the $r\theta$-plane. Figure 1.3 shows four of these rectangles. The direct image of each of these rectangles is the same shaded region in the xy-plane. $\triangle$

This last example shows that a function may have the same value at many different points. Functions for which this does not happen are important. They are called *one-to-one* functions.

Definition 1.3.17 One-to-one functions. Let f be a function with $D = \operatorname{Dom} f$. Let $A \subset D$. Then f is said to be *one-to-one* (or *injective*) on A if $f(x_1) \neq f(x_2)$ whenever $x_1, x_2 \in A$ and $x_1 \neq x_2$. Equivalently, f is one-to-one on A if $x_1 = x_2$ whenever $x_1, x_2 \in A$ and $f(x_1) = f(x_2)$.

Theorem 1.3.18 *Let f be a function with $D = Dom\, f$. Let $A \subset D$ and $B = f(A)$. Then f is one-to-one on A if and only if there is a function*

$$g : B \to A \quad \textit{such that } g(f(x)) = x \textit{ for all } x \in A.$$

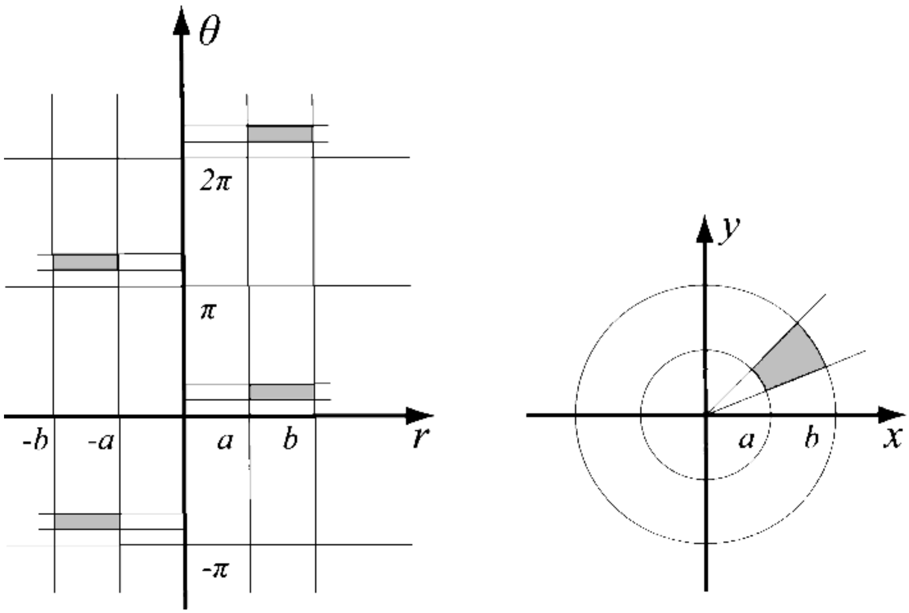

Figure 1.3. Images under polar coordinates.

Also, if such a function g exists, then it is unique and $f(g(y)) = y$ for each $y \in B$.

Proof. Assume that f is one-to-one on A. Let $B = f(A)$. Then for each $y \in B$, there is a unique $x \in A$ such that $f(x) = y$. Hence, we can define a function $g : B \to A$ by letting $x = g(y)$ whenever $f(x) = y$. Hence, $x = g(y) = g(f(x))$ for all $x \in A$.

Conversely, assume the existence of g. If $x_1, x_2 \in A$ and $f(x_1) = f(x_2)$, then

$$x_1 = g(f(x_1)) = g(f(x_2)) = x_2.$$

Hence f is one-to-one on A.

Finally, if g exists, then it is unique. In fact, the previous argument shows that if g exists, then for each $b \in B$ there is a unique $a \in A$ such that $b = f(a)$ and $g(b) = g(f(a)) = a$. If h is another function on B such that $h(f(x)) = x$ for all $x \in A$, then $h(b) = h(f(a)) = a = g(b)$. Hence $h = g$. Finally, $g(y) = x$ and $f(x) = y$ show that $f(g(y)) = y$. ◻

Definition 1.3.19 Inverse functions. Let f be a function with $D = \operatorname{Dom} f$. Let $A \subset D$ and $B = f(A)$. A function defined on B is called an *inverse function of*

f *on* A if $g(f(x)) = x$ for all $x \in A$. Theorem 1.3.18 above shows that f has an inverse function on A if and only if f is one-to-one on A. Also, if an inverse function g exists, then it is unique. Its value at $y \in B$ is the unique solution of the equation $y = f(x)$. Finally, the same theorem also shows that if g is the inverse of f on A, then f is the inverse of g on $B = f(A)$.

Definition 1.3.20 Invertible functions. A function $f : A \to B$ is called an *invertible function*, or a *bijection*, or a *one-to-one correspondence* between A and B if it has an inverse function $g : B \to A$. Theorem 1.3.18 shows that $f : A \to B$ is an invertible function between A and B if and only if f is one-to-one on A and $B = f(A)$.

Definition 1.3.21 One-to-one and onto functions. When a function $f : A \to B$ is said to be invertible, it is understood that it is invertible between A and B. Hence, in the case of an invertible function f, the sets A and B in the notation $f : A \to B$ become important. The invertibility of $f : A \to B$ means that f is one-to-one on A and that $B = f(A)$. An invertible function $f : A \to B$ is also called a *one-to-one and onto function*. Here the sets A and B are again important. It is understood that f is one-to-one on A and that it maps A *onto* $B = f(A)$.

Example 1.3.22 Let $y = f(x) = (2x - 1)/(3x - 1)$. This function is defined for all $x \neq -1/3$. Hence $D = \text{Dom}\, f = (\mathbb{R} \setminus \{-1/3\})$. The equation

$$y = f(x) = \frac{2x - 1}{3x + 1}$$

is uniquely solved as

$$x = g(y) = \frac{1 + y}{2 - 3y}$$

for each $y \neq 2/3$. Hence f is one-to-one on D, and its inverse on D is g. Also, $\text{Dom}\, g = (\mathbb{R} \setminus \{2/3\})$. $\triangle$

Example 1.3.23 Let $y = f(x) = x^2 + x + 1$. Then $f(x)$ is defined for all $x \in \mathbb{R}$. Hence $D = \text{Dom}\, f = \mathbb{R}$. The solution of the equation $x^2 + x + 1 = y$ is given by the formula

$$x = -(1/2)(1 \pm (4y - 3)^{1/2}).$$

This formula defines x if and only if $4y - 3 \geq 0$, that is, if and only if $y \geq 3/4$. Hence $f(D) = f(\mathbb{R}) = [3/4, \infty)$. Finally, the equation $x^2 + x + 1 = y$ has two solutions for each $y > 3/4$. These two solutions are symmetrical with respect to the point $x = -1/2$. There is only one solution in $A = [-1/2, \infty)$ and also only one solution in $A' = (-\infty, -1/2]$. Hence f is one-to-one on A and also one-to-one on A'. The inverse of f on A is

$$g(y) = (1/2)(-1 + (4y - 3)^{1/2})$$

and the inverse of f on A' is

$$g'(y) = (1/2)(-1 - (4y - 3)^{1/2}).$$

Note that $f(A) = f(A') = f(\mathbb{R}) = [3/4, \infty)$. There are (infinitely many) other sets $C \subset \mathbb{R}$ such that f is one-to-one on C and such that $f(C) = f(\mathbb{R})$. For example, $C = (-\infty, -1) \cup [-1/2, 0]$ is such a set. $\triangle$

Example 1.3.24 Polar coordinates were discussed in Examples 1.3.12 and 1.3.16. They are defined as a function $f : \mathbb{R}^2 \to \mathbb{R}^2$ such that

$$f(r, \theta) = (x, y), \text{ where } x = r\cos\theta \text{ and } y = r\sin\theta.$$

As observed in 1.3.16, f is not one-to-one on $\mathbb{R}^2 = \operatorname{Dom} f$. Hence f does not have an inverse function on $\mathbb{R}^2$. But there are many sets $A \subset \operatorname{Dom} f$ such that f is one-to-one on A. For example any function is one-to-one on a singleton set. The important point is to find a set $A \subset \mathbb{R}^2 = \operatorname{Dom} f$ such that f is one-to-one on A and such that $f(A) = f(\mathbb{R}^2)$. There are many such sets. Let, for example, $\alpha \in \mathbb{R}$ be a fixed number and define

$$\begin{aligned} A_\alpha &= \left\{ (r, \theta) \in \mathbb{R}^2 \mid 0 < r,\ \alpha \le \theta < \alpha + 2\pi \right\} \cup \{(0, 0)\}, \\ A'_\alpha &= \left\{ (r, \theta) \in \mathbb{R}^2 \mid r < 0,\ \alpha \le \theta < \alpha + 2\pi \right\} \cup \{(0, 0)\}. \end{aligned}$$

It is easy to check that f is one-to-one on each of these sets and also that

$$f(A_\alpha) = f(A'_\alpha) = f(\mathbb{R}^2) = \mathbb{R}^2 \text{ for each } \alpha \in \mathbb{R}.$$

Note that f is not one-to-one, for example, on

$$C_0 = \left\{ (r, \theta) \in \mathbb{R}^2 \mid 0 \le r,\ 0 \le \theta < 2\pi \right\}.$$

In fact, in this case f maps all the points on the vertical segment

$$S_0 = \left\{ (r, \theta) \in \mathbb{R}^2 \mid r = 0,\ 0 \le \theta < 2\pi \right\} \subset C_0$$

to $(0, 0)$. This is a triviality, and it is ignored in many cases. One defines the polar coordinates of (x, y) as (r, θ) such that $0 \le r$, $\alpha \le \theta < \alpha + 2\pi$, and such that $x = r\cos\theta$ and $y = r\sin\theta$. Here α is a fixed specific number like $\alpha = 0$ or $\alpha = -\pi$, depending on the problem. Everyone knows that this does not determine the θ value for $(x, y) = (0, 0)$. But everyone also knows that this is not an important point in most cases. $\triangle$

Composition of Functions

Definition 1.3.25 Composition of functions. If f and g are two functions, then the *composition* $g \circ f$ is defined by $(g \circ f)(x) = g(f(x))$. The domain of $g \circ f$ is specified by this definition in an obvious way. It is the set of all x for which $g(f(x))$ is defined. Let $F = \text{Dom}\, f$ and $G = \text{Dom}\, g$. We see that

$$\text{Dom}\,(g \circ f) = F \cap f^{-1}(G) = f^{-1}(G).$$

In fact, F is the set of all x for which $f(x)$ is defined and

$$f^{-1}(G) = \{\, x \in F \mid f(x) \in G \,\} \subset F$$

is the set of all $x \in F$ for which $f(x)$ is in the domain of g. Hence $g(f(x))$ is defined if and only if $x \in f^{-1}(G)$.

The composition of more than two functions is defined similarly. If f, g, and h are three functions, for example, then $h \circ g \circ f$ is defined as

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

The domain of $h \circ g \circ f$ is the set of all x such that $h(g(f(x)))$ is defined. If

$$F = \text{Dom}\, f\,,\ G = \text{Dom}\, g\,,\ \text{and } H = \text{Dom}\, h,$$

then we see that

$$\text{Dom}\,(h \circ g \circ f) = f^{-1}(g^{-1}(H)).$$

In fact, $x \in \text{Dom}\,(h \circ g \circ f)$ if and only if $g(f(x)) \in H$. This happens if and only if $f(x) \in g^{-1}(H)$, which is equivalent to $x \in f^{-1}(g^{-1}(H))$.

Lemma 1.3.26 Images under compositions. *Let f and g be two functions. Then*

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

for any set E. Also, if $A \subset \text{Dom}\,(g \circ f)$, then

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

Proof. For the first part, note that

$$\begin{aligned} x \in (g \circ f)^{-1}(E) &\iff (g \circ f)(x) \in E \iff g(f(x)) \in E \\ &\iff f(x) \in g^{-1}(E) \iff x \in f^{-1}(g^{-1}(E)). \end{aligned}$$

Also,

$$\begin{aligned} (g \circ f)(A) &= \{\, (g \circ f)(x) \mid x \in A \,\} = \{\, g(f(x)) \mid x \in A \,\} \\ &= \{\, g(y) \mid y \in f(A) \,\} = g(f(A)). \quad \square \end{aligned}$$

Remarks 1.3.27 Compositions and inverses. Let X be a set. The identity function $I_X : X \to X$ on X was defined in Definition 1.3.7 as $I_X(x) = x$ for all $x \in X$. Let f be a function with $A = \operatorname{Dom} f$ and $B = f(A)$. Let g be a function with $B = \operatorname{Dom} g$. Then g is the inverse of f on A if and only if $g \circ f = I_A$. In this case also, $f \circ g = I_B$. These remarks follow directly from Theorem 1.3.18.

Problems

1.24 Let $f : D \to Y$ be a function with the domain $D \subset X$ and the domain space X. Let $\mathcal{E}$ be a collection of subsets of X. Show that

$$f\left(\bigcup_{E\in\mathcal{E}} E\right) = \bigcup_{E\in\mathcal{E}} f(E) \text{ and } f\left(\bigcap_{E\in\mathcal{E}} E\right) \subset \bigcap_{E\in\mathcal{E}} f(E).$$

Give an example to show that $f(\bigcap_{E\in\mathcal{E}} E) \neq \bigcap_{E\in\mathcal{E}} f(E)$ is possible.

1.25 Let $f : D \to Y$ be a function with the range space Y. Let $\mathcal{F}$ be a collection of subsets of Y. Show that

$$f^{-1}\left(\bigcup_{F\in\mathcal{F}} F\right) = \bigcup_{F\in\mathcal{F}} f^{-1}(F) \text{ and } f^{-1}\left(\bigcap_{F\in\mathcal{F}} F\right) = \bigcap_{F\in\mathcal{F}} f^{-1}(F).$$

1.26 Let $f : D \to Y$ be a function with the domain $D \subset X$, the domain space X, the range space Y, and the range $R = f(X) = f(D) \subset Y$.

1. Show that $f(f^{-1}(B)) = B \cap R$ for all $B \subset Y$.
2. Show that $A \cap D \subset f^{-1}(f(A))$ for all $A \subset X$.
3. Give examples to show that $f^{-1}(f(A)) \neq A \cap D$ is possible.
4. Show that if f is one-to-one on D, then $f^{-1}(f(A)) = A \cap D$.

1.27 Let $f(x) = x^2 - 6x - 7$. What is the range $f(D)$ of f? Is f one-to-one on its domain? If not, find two different sets $P, Q \subset \mathbb{R}$ such that f is one-to-one on P and one-to-one on Q and such that $f(P) = f(Q) = f(\mathbb{R})$. Find the inverse function of f on P and the inverse function of f on Q.

1.28 Let $f : D \to Y$ be a function with the domain $D \subset X$, the domain space X, the range space Y, and the range $R = f(X) = f(D) \subset Y$. Define a relation on D by the condition that $a \in D$ is related to $b \in D$ if and only if $f(a) = f(b)$.

1. Show that this is an equivalence in D.

2. Let $P \subset D$ be an equivalence class and let $V \subset Y$ be any subset of Y. Show that either $P \subset f^{-1}(V)$ or $P \cap f^{-1}(V) = \emptyset$.

3. Let $A \subset D$ be a complete set of representatives for this equivalence. Show that f is one-to-one on A and $R = f(A)$.

1.29 Let $f : \mathbb{R}^2 \to \mathbb{R}^2$ be the polar coordinates defined by

$$(x, y) = f(r, \theta) = (r\cos\theta, r\sin\theta) .$$

Find $f^{-1}(A)$ for

$$A = \left\{ (x, y) \in \mathbb{R}^2 \;\middle|\; 1 \le (x^2 + y^2) \le 4 \text{ and } 0 \le (y/x) \le 1 \right\} .$$

1.30 Define the function f from the xy-plane to the uv-plane by

$$(u, v) = f(x, y) = (3x + 2y, 6x + 4y).$$

1. What is the domain $D \subset \mathbb{R}^2$ of f?

2. What is the range $R = f(\mathbb{R}^2) = f(D) \subset \mathbb{R}^2$ of f?

3. Let $a, b \in \mathbb{R}$, and let L_a be the line $u = a$ and M_b the line $v = b$ in the uv-plane. What are the inverse images $f^{-1}(L_a)$ and $f^{-1}(M_b)$?

4. Let $(a, b) \in R$. What is $f^{-1}(\{ (a, b) \})$?

5. Find some examples of $A \subset D$ such that f is one-to-one on A and such that $f(A) = f(D)$.

1.31 Define the function f from the xy-plane to the uv-plane by

$$(u, v) = f(x, y) = (3x + 2y, 6x - 4y).$$

Repeat the parts of Problem 1.30 for this example.

1.32 Define the function f from the xy-plane to the uv-plane by

$$(u, v) = f(x, y) = (xy, y/x).$$

Repeat the parts of Problem 1.30 for this example.

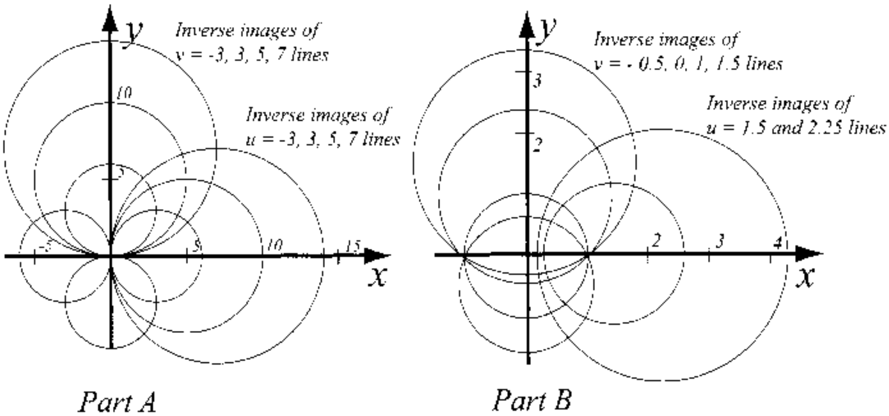

Figure 1.4. Hints for Problems 1.33 and 1.34.

1.33 Define the function f from the xy-plane to the uv-plane by

$$(u, v) = f(x, y) = ((x^2 + y^2)/(2x), (x^2 + y^2)/(2y)).$$

Repeat the parts of Problem 1.30 for this example. (*Hint.* In Part A of Figure 1.4 we see the inverse images of the lines $u = -3, 3, 5, 7$ and the lines $v = -3, 3, 5, 7$.)

1.34 Define the function f from the xy-plane to the uv-plane by

$$(u, v) = f(x, y) = ((x^2 + y^2 + 1)/(2x), (x^2 + y^2 - 1)/(2y)).$$

Repeat the parts of Problem 1.30 for this example. (*Hint.* In Part B of Figure 1.4 we see the inverse images of the lines $u = 1.5, 2.25$ and the lines $v = -0.5, 0, 1, 1.5$.)

1.35 Define the function f from the xy-plane to the uv-plane by

$$(u, v) = f(x, y) = (p(x, y) + q(x, y), p(x, y) - q(x, y)),$$

where $p(x, y) = ((x + 1)^2 + y^2)^{1/2}$ and $q(x, y) = ((x - 1)^2 + y^2)^{1/2}$. Repeat the parts of Problem 1.30 for this example.

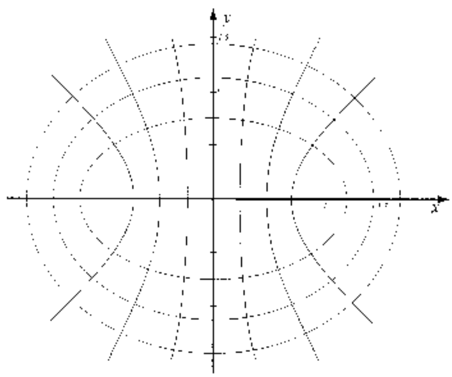

Figure 1.5. Hint for Problem 1.35.