# 1

# THINGS FAMILIAR AND LESS FAMILIAR



## 1. A FEW PRELIMINARY REMARKS

For many readers this book will be their first contact with abstract mathematics. The subject to be discussed is usually called "abstract algebra," but the difficulties that the reader may encounter are not so much due to the "algebra" part as they are to the "abstract" part.

On seeing some area of abstract mathematics for the first time, be it in analysis, topology, or what-not, there seems to be a common reaction for the novice. This can best be described by a feeling of being adrift, of not having something solid to hang on to. This is not too surprising, for while many of the ideas are fundamentally quite simple, they are subtle and seem to elude one's grasp the first time around. One way to mitigate this feeling of limbo, or asking oneself "What is the point of all this?," is to take the concept at hand and see what it says in particular cases. In other words, the best road to good understanding of the notions introduced is to look at examples. This is true in all of mathematics, but it is particularly true for the subject matter of abstract algebra.

Can one, with a few strokes, quickly describe the essence, purpose, and background for the material we shall study? Let's give it a try.

We start with some collection of objects $S$ and endow this collection with an algebraic structure by assuming that we can combine, in one or several ways (usually two), elements of this set $S$ to obtain, once more, elements of this set $S$. These ways of combining elements of $S$ we call *operations* on $S$.

Then we try to condition or regulate the nature of $S$ by imposing certain rules on how these operations behave on $S$. These rules are usually called the *axioms* defining the particular structure on $S$. These axioms are for us to define, but the choice made comes, historically in mathematics, from noticing that there are many concrete mathematical systems that satisfy these rules or axioms. We shall study some of the basic axiomatic algebraic systems in this book, namely *groups*, *rings*, and *fields*.

Of course, one could try many sets of axioms to define new structures. What would we require of such a structure? Certainly we would want that the axioms be consistent, that is, that we should not be led to some nonsensical contradiction computing within the framework of the allowable things the axioms permit us to do. But that is not enough. We can easily set up such algebraic structures by imposing a set of rules on a set $S$ that lead to a pathological or weird system. Furthermore, there may be very few examples of something obeying the rules we have laid down.

Time has shown that certain structures defined by "axioms" play an important role in mathematics (and other areas as well) and that certain others are of no interest. The ones we mentioned earlier, namely groups, rings, and fields, have stood the test of time.

A word about the use of "axioms." In everyday language "axiom" means a self-evident truth. But we are not using everyday language; we are dealing with mathematics. An axiom is not a universal truth—but one of several rules spelling out a given mathematical structure. The axiom is true in the system we are studying because we have forced it to be true by hypothesis. It is a license, in the particular structure, to do certain things.

We return to something we said earlier about the reaction that many students have on their first encounter with this kind of algebra, namely a lack of feeling that the material is something they can get their teeth into. Do not be discouraged if the initial exposure leaves you in a bit of a fog. Stick with it, try to understand what a given concept says, and most importantly, look at particular, concrete examples of the concept under discussion.

## PROBLEMS

**1.** Let $S$ be a set having an operation $*$ which assigns an element $a * b$ of $S$ for any $a, b \in S$. Let us assume that the following two rules hold:

1. If $a, b$ are any objects in $S$, then $a * b = a$.
2. If $a, b$ are any objects in $S$, then $a * b = b * a$.

Show that $S$ can have at most one object.

**2.** Let $S$ be the set of all integers $0, \pm 1, \pm 2, \ldots, \pm n, \ldots$. For $a, b$ in $S$ define $*$ by $a * b = a - b$. Verify the following:

**(a)** $a * b \neq b * a$ unless $a = b$.

**(b)** $(a * b) * c \neq a * (b * c)$ in general. Under what conditions on $a, b, c$ is $(a * b) * c = a * (b * c)$?

**(c)** The integer 0 has the property that $a * 0 = a$ for every $a$ in $S$.

**(d)** For $a$ in $S$, $a * a = 0$.

**3.** Let $S$ consist of the two objects $\square$ and $\triangle$. We define the operation $*$ on $S$ by subjecting $\square$ and $\triangle$ to the following conditions:

1. $\square * \triangle = \triangle = \triangle * \square$.
2. $\square * \square = \square$.
3. $\triangle * \triangle = \square$.

Verify by explicit calculation that if $a, b, c$ are any elements of $S$ (i.e., $a, b$ and $c$ can be any of $\square$ or $\triangle$), then:

**(a)** $a * b$ is in $S$.

**(b)** $(a * b) * c = a * (b * c)$.

**(c)** $a * b = b * a$.

**(d)** There is a particular $a$ in $S$ such that $a * b = b * a = b$ for *all* $b$ in $S$.

**(e)** Given $b$ in $S$, then $b * b = a$, where $a$ is the particular element in Part (d).

## 2. SET THEORY

With the changes in the mathematics curriculum in the schools in the United States, many college students have had some exposure to set theory. This introduction to set theory in the schools usually includes the elementary notions and operations with sets. Going on the assumption that many readers will have some acquaintance with set theory, we shall give a rapid survey of those parts of set theory that we shall need in what follows.

First, however, we need some notation. To avoid the endless repetition of certain phrases, we introduce a shorthand for these phrases. Let $S$ be a collection of objects; the objects of $S$ we call the *elements* of $S$. To denote that a given element, $a$, is an element of $S$, we write $a \in S$—this is read "$a$ is an element of $S$." To denote the contrary, namely that an object $a$ is *not* an element of $S$, we write $a \notin S$. So, for instance, if $S$ denotes the set of all positive integers $1, 2, 3, \ldots, n, \ldots$, then $165 \subset S$, whereas $-13 \notin S$.

We often want to know or prove that given two sets $S$ and $T$, one of these is a part of the other. We say that $S$ is a *subset* of $T$, which we write $S \subset T$ (read "$S$ is contained in $T$") if every element of $S$ is an element of $T$.

In terms of the notation we now have: $S \subset T$ if $s \subset S$ implies that $s \in T$. We can also denote this by writing $T \supset S$, read "$T$ contains $S$." (This does not exclude the possibility that $S = T$, that is, that $S$ and $T$ have exactly the same elements.) Thus, if $T$ is the set of all positive integers and $S$ is the set of all positive even integers, then $S \subset T$, and $S$ is a subset of $T$. In the definition given above, $S \supset S$ for any set $S$; that is, $S$ is always a subset of itself.

We shall frequently need to show that two sets $S$ and $T$, defined perhaps in distinct ways, are equal, that is, they consist of the same set of elements. The usual strategy for proving this is to show that both $S \subset T$ and $T \subset S$. For instance, if $S$ is the set of all positive integers having 6 as a factor and $T$ is the set of all positive integers having both 2 and 3 as factors, then $S = T$. (Prove!)

The need also arises for a very peculiar set, namely one having no elements. This set is called the *null* or *empty* set and is denoted by $\emptyset$. It has the property that it is a subset of *any* set $S$.

Let $A$, $B$ be subsets of a given set $S$. We now introduce methods of constructing other subsets of $S$ from $A$ and $B$. The first of these is the *union* of $A$ and $B$, written $A \cup B$, which is defined: $A \cup B$ is that subset of $S$ consisting of those elements of $S$ that are elements of $A$ *or* are elements of $B$. The "or" we have just used is somewhat different in meaning from the ordinary usage of the word. Here we mean that an element $e$ is in $A \cup B$ if it is in $A$, or is in $B$, or is in *both*. The "or" is not meant to exclude the possibility that both things are true. Consequently, for instance, $A \cup A = A$.

If $A = \{1, 2, 3\}$ and $B = \{2, 4, 6, 10\}$, then $A \cup B = \{1, 2, 3, 4, 6, 10\}$.

We now proceed to our second way of constructing new sets from old. Again let $A$ and $B$ be subsets of a set $S$; by the *intersection* of $A$ and $B$, written $A \cap B$, we shall mean the subset of $S$ consisting of those elements that are both in $A$ *and* in $B$. Thus, in the example above, $A \cap B = \{2\}$. It should be clear from the definitions involved that $A \cap B \subset A$ and $A \cap B \subset B$. Particular examples of intersections that hold universally are: $A \cap A = A$, $A \cap S = A$, $A \cap \emptyset = \emptyset$.

This is an opportune moment to introduce a notational device that will be used time after time. Given a set $S$, we shall often be called on to describe the subset $A$ of $S$, whose elements satisfy a certain property $P$. We shall write this as $A = \{s \in S \mid s \text{ satisfies } P\}$. For instance, if $A$, $B$ are subsets of $S$, then $A \cup B = \{s \in S \mid s \in A \text{ or } s \subset B\}$ while $A \cap B = \{s \subset S \mid s \in A \text{ and } s \in B\}$.

Although the notions of union and intersection of subsets of $S$ have been defined for two subsets, it is clear how one can define the union and intersection of any number of subsets.

We now introduce a third operation we can perform on sets, the *difference* of two sets. If $A$, $B$ are subsets of $S$, we define $A - B = \{a \in A \mid a \notin B\}$.

So if $A$ is the set of all positive integers and $B$ is the set of all even integers, then $A - B$ is the set of all positive odd integers. In the particular case when $A$ is a subset of $S$, the difference $S - A$ is called the *complement* of $A$ in $S$ and is written $A'$.

We represent these three operations pictorially. If $A$ is Ⓐ and $B$ is Ⓑ, then

1. $A \cup B =$  is the shaded area.

2. $A \cap B =$  is the shaded area.

3. $A - B =$  is the shaded area.

4. $B - A =$  is the shaded area.

Note the relation among the three operations, namely the equality $A \cup B = (A \cap B) \cup (A - B) \cup (B - A)$. As an illustration of how one goes about proving the equality of sets constructed by such set-theoretic constructions, we prove this latter alleged equality. We first show that $(A \cap B) \cup (A - B) \cup (B - A) \subset A \cup B$; this part is easy for, by definition, $A \cap B \subset A$, $A - B \subset A$, and $B - A \subset B$, hence

$$(A \cap B) \cup (A - B) \cup (B - A) \subset A \cup A \cup B = A \cup B.$$

Now for the other direction, namely that $A \cup B \subset (A \cap B) \cup (A - B) \cup (B - A)$. Given $u \in A \cup B$, if $u \in A$ and $u \in B$, then $u \in A \cap B$, so it is certainly in $(A \cap B) \cup (A - B) \cup (B - A)$. On the other hand, if $u \in A$ but $u \notin B$, then, by the very definition of $A - B$, $u \in A - B$, so again it is certainly in $(A \cap B) \cup (A - B) \cup (B - A)$. Finally, if $u \in B$ but $u \notin A$, then $u \in B - A$, so again it is in $(A \cap B) \cup (A - B) \cup (B - A)$. We have thus covered all the possibilities and have shown that $A \cup B \subset (A \cap B) \cup (A - B) \cup (B - A)$. Having the two opposite containing relations of $A \cup B$ and $(A \cap B) \cup (A - B) \cup (B - A)$, we obtain the desired equality of these two sets.

We close this brief review of set theory with yet another construction we can carry out on sets. This is the *Cartesian product* defined for the two sets $A$, $B$ by $A \times B = \{(a, b) \mid a \in A, b \in B\}$, where we *declare* the ordered pair $(a, b)$ to be equal to the ordered pair $(a_1, b_1)$ if and only if $a = a_1$ and $b = b_1$. Here, too, we need not restrict ourselves to two sets; for instance, we

can define, for sets $A$, $B$, $C$, their Cartesian product as the set of ordered triples $(a, b, c)$, where $a \in A$, $b \in B$, $c \in C$ and where equality of two ordered triples is defined component-wise.

## PROBLEMS

### Easier Problems

**1.** Describe the following sets verbally.
- **(a)** $S = \{\text{Mercury, Venus, Earth, . . . , Pluto}\}$.
- **(b)** $S = \{\text{Alabama, Alaska, . . . , Wyoming}\}$.

**2.** Describe the following sets verbally.
- **(a)** $S = \{2, 4, 6, 8, \ldots\}$.
- **(b)** $S = \{2, 4, 8, 16, 32, \ldots\}$.
- **(c)** $S = \{1, 4, 9, 16, 25, 36, \ldots\}$.

**3.** If $A$ is the set of all residents of the United States, $B$ the set of all Canadian citizens, and $C$ the set of all women in the world, describe the sets $A \cap B \cap C$, $A - B$, $A - C$, $C - A$ verbally.

**4.** If $A = \{1, 4, 7, a\}$ and $B = \{3, 4, 9, 11\}$ and you have been told that $A \cap B = \{4, 9\}$, what must $a$ be?

**5.** If $A \subset B$ and $B \subset C$, prove that $A \subset C$.

**6.** If $A \subset B$, prove that $A \cup C \subset B \cup C$ for any set $C$.

**7.** Show that $A \cup B = B \cup A$ and $A \cap B = B \cap A$.

**8.** Prove that $(A - B) \cup (B - A) = (A \cup B) - (A \cap B)$. What does this look like pictorially?

**9.** Prove that $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.

**10.** Prove that $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$.

**11.** Write down all the subsets of $S = \{1, 2, 3, 4\}$.

### Middle-Level Problems

***12.** If $C$ is a subset of $S$, let $C'$ denote the complement of $C$ in $S$. Prove the *De Morgan Rules* for subsets $A$, $B$ of $S$, namely:
- **(a)** $(A \cap B)' = A' \cup B'$.
- **(b)** $(A \cup B)' = A' \cap B'$.

***13.** Let S be a set. For any two subsets of S we define

$$A + B = (A - B) \cup (B - A) \quad \text{and} \quad A \cdot B = A \cap B.$$

Prove that:

**(a)** $A + B = B + A$.

**(b)** $A + \emptyset = A$.

**(c)** $A \cdot A = A$.

**(d)** $A + A = \emptyset$.

**(e)** $A + (B + C) = (A + B) + C$.

**(f)** If $A + B = A + C$, then $B = C$.

**(g)** $A \cdot (B + C) = A \cdot B + A \cdot C$.

***14.** If $C$ is a finite set, let $m(C)$ denote the number of elements in $C$. If $A$, $B$ are finite sets, prove that

$$m(A \cup B) = m(A) + m(B) - m(A \cap B).$$

**15.** For three finite sets $A$, $B$, $C$ find a formula for $m(A \cup B \cup C)$. (**Hint:** First consider $D = B \cup C$ and use the result of Problem 14.)

**16.** Take a shot at finding $m(A_1 \cup A_2 \cup \cdots \cup A_n)$ for $n$ finite sets $A_1, A_2, \ldots, A_n$.

**17.** Use the result of Problem 14 to show that if 80% of all Americans have gone to high school and 70% of all Americans read a daily newspaper, then *at least* 50% of Americans have both gone to high school and read a daily newspaper.

**18.** A public opinion poll shows that 93% of the population agreed with the government on the first decision, 84% on the second, and 74% on the third, for three decisions made by the government. At least what percentage of the population agreed with the government on all three decisions? (**Hint:** Use the results of Problem 15.)

**19.** In his book *A Tangled Tale*, Lewis Carroll proposed the following riddle about a group of disabled veterans: "Say that 70% have lost an eye, 75% an ear, 80% an arm, 85% a leg. What percentage, *at least*, must have lost all four?" Solve Lewis Carroll's problem.

***20.** Show, for finite sets $A$, $B$, that $m(A \times B) = m(A)m(B)$.

**21.** If $S$ is a set having five elements:

**(a)** How many subsets does $S$ have?

**(b)** How many subsets having four elements does $S$ have?

**(c)** How many subsets having two elements does $S$ have?

**Harder Problems**

**22. (a)** Show that a set having $n$ elements has $2^n$ subsets.

**(b)** If $0 < m < n$, how many subsets are there that have exactly $m$ elements?

## 3. MAPPINGS

One of the truly universal concepts that runs through almost every phase of mathematics is that of a *function* or *mapping* from one set to another. One could safely say that there is no part of mathematics where the notion does not arise or play a central role. The definition of a function from one set to another can be given in a formal way in terms of a subset of the Cartesian product of these sets. Instead, here, we shall give an informal and admittedly nonrigorous definition of a mapping (function) from one set to another.

Let $S$, $T$ be sets; a *function* or *mapping* $f$ from $S$ to $T$ is a *rule* that assigns to *each* element $s \in S$ a *unique* element $t \in T$. Let's explain a little more thoroughly what this means. If $s$ is a given element of $S$, then there is *only one* element $t$ in $T$ that is associated to $s$ by the mapping. As $s$ varies over $S$, $t$ varies over $T$ (in a manner depending on $s$). Note that by the definition given, the following is *not* a mapping. Let $S$ be the set of all people in the world and $T$ the set of all countries in the world. Let $f$ be the rule that assigns to every person his or her country of citizenship. Then $f$ is not a mapping from $S$ to $T$. Why not? Because there are people in the world that enjoy a dual citizenship; for such people there would not be a *unique* country of citizenship. Thus, if Mary Jones is both an English and French citizen, $f$ would not make sense, as a mapping, when applied to Mary Jones. On the other hand, the rule $f: \mathbb{R} \to \mathbb{R}$, where $\mathbb{R}$ is the set of real numbers, defined by $f(a) = a^2$ for $a \in \mathbb{R}$, is a perfectly good function from $\mathbb{R}$ to $\mathbb{R}$. It should be noted that $f(-2) = (-2)^2 = 4 = f(2)$, and $f(-a) = f(a)$ for all $a \in \mathbb{R}$.

We denote that $f$ is a mapping from $S$ to $T$ by $f: S \to T$ and for the $t \in T$ mentioned above we write $t = f(s)$; we call $t$ the *image* of $s$ under $f$.

The concept is hardly a new one for any of us. Since grade school we have constantly encountered mappings and functions, often in the form of formulas. But mappings need not be restricted to sets of numbers. As we see below, they can occur in any area.

**Examples**

1. Let $S$ = {all men who have ever lived} and $T$ = {all women who have ever lived}. Define $f: S \to T$ by $f(s)$ = mother of $s$. Therefore, $f$(John F. Kennedy) = Rose Kennedy, and according to our definition, Rose Kennedy is the image under $f$ of John F. Kennedy.

2. Let $S$ = {all legally employed citizens of the United States} and $T$ = {positive integers}. Define, for $s \in S$, $f(s)$ by $f(s)$ = Social Security Number of $s$. (For the purpose of this text, let us assume that all legally employed citizens of the United States have a Social Security Number.) Then $f$ defines a mapping from $S$ to $T$.

3. Let $S$ be the set of all objects for sale in a grocery store and let $T$ = {all real numbers}. Define $f: S \to T$ by $f(s)$ = price of $s$. This defines a mapping from $S$ to $T$.

4. Let $S$ be the set of all integers and let $T = S$. Define $f: S \to T$ by $f(m) = 2m$ for any integer $m$. Thus the image of 6 under this mapping, $f(6)$, is given by $f(6) = 2 \cdot 6 = 12$, while that of $-3$, $f(-3)$, is given by $f(-3) = 2(-3) = -6$. If $s_1, s_2 \in S$ are in $S$ and $f(s_1) = f(s_2)$, what can you say about $s_1$ and $s_2$?

5. Let $S = T$ be the set of all real numbers; define $f: S \to T$ by $f(s) = s^2$. Does every element of $T$ come up as an image of some $s \in S$? If not, how would you describe the set of all images $\{f(s) \mid s \in S\}$? When is $f(s_1) = f(s_2)$?

6. Let $S = T$ be the set of all real numbers; define $f: S \to T$ by $f(s) = s^3$. This is a function from $S$ to $T$. What can you say about $\{f(s) \mid s \in S\}$? When is $f(s_1) = f(s_2)$?

7. Let $T$ be any nonempty set and let $S = T \times T$, the Cartesian product of $T$ with itself. Define $f: T \times T \to T$ by $f(t_1, t_2) = t_1$. This mapping from $T \times T$ to $T$ is called the *projection* of $T \times T$ onto its first component.

8. Let $S$ be the set of all positive integers and let $T$ be the set of all positive rational numbers. Define $f: S \times S \to T$ by $f(m, n) = m/n$. This defines a mapping from $S \times S$ to $T$. Note that $f(1, 2) = \frac{1}{2}$ while $f(3, 6) = \frac{3}{6} = \frac{1}{2} = f(1, 2)$, although $(1, 2) \neq (3, 6)$. Describe the subset of $S \times S$ consisting of those $(a, b)$ such that $f(a, b) = \frac{1}{2}$.

The mappings to be defined in Examples 9 and 10 are mappings that occur for any nonempty sets and play a special role.

9. Let $S$, $T$ be nonempty sets, and let $t_0$ be a fixed element of $T$. Define $f: S \to T$ by $f(s) = t_0$ for every $s \in S$; $f$ is called a *constant* function from $S$ to $T$.

10. Let $S$ be any nonempty set and define $i: S \to S$ by $i(s) = s$ for every $s \in S$. We call this function of $S$ to itself the *identity function* (or *identity mapping*) on S. We may, at times, denote it by $i_S$ (and later in the book, by $e$).

Now that we have the notion of a mapping we need some way of identifying when two mappings from one set to another are equal. This is not God given; it is for us to decide how to declare $f = g$ where $f: S \to T$ and $g: S \to T$. What is more natural than to define this equality via the actions of $f$ and $g$ on the elements of $S$? More precisely, we declare that $f = g$ if and only if $f(s) = g(s)$ for *every* $s \in S$. If $S$ is the set of all real numbers and $f$ is defined on $S$ by $f(s) = s^2 + 2s + 1$, while $g$ is defined on $S$ by $g(s) = (s + 1)^2$, our definition of the equality of $f$ and $g$ is merely a statement of the familiar identity $(s + 1)^2 = s^2 + 2s + 1$.

Having made the definition of equality of two mappings, we now want to single out certain types of mappings by the way they behave.

**Definition.** The mapping $f: S \to T$ is *onto* or *surjective* if every $t \in T$ is the image under $f$ of some $s \in S$; that is, if and only if, given $t \in T$, there exists an $s \in S$ such that $t = f(s)$.

In the examples we gave earlier, in Example 1 the mapping is not onto, since not every woman that ever lived was the mother of a male child. Similarly, in Example 2 the mapping is not onto, for not every positive integer is the Social Security Number of some U.S. citizen. The mapping in Example 4 fails to be onto because not every integer is even; and in Example 5, again, the mapping is not onto, for the number $-1$, for instance, is not the square of any real number. However, the mapping in Example 6 is onto because every real number has a unique real cube root. The reader can decide whether or not the given mappings are onto in the other examples.

If we define $f(S) = \{f(s) \in T \mid s \in S\}$, another way of saying that the mapping $f: S \to T$ is onto is by saying that $f(S) = T$.

Another specific type of mapping plays an important and particular role in what follows.

**Definition.** A mapping $f: S \to T$ is said to be *one-to-one* (written 1-1) or *injective* if for $s_1 \neq s_2$ in $S$, $f(s_1) \neq f(s_2)$ in $T$. Equivalently, $f$ is 1-1 if $f(s_1) = f(s_2)$ implies that $s_1 = s_2$.

In other words, a mapping is 1-1 if it takes distinct objects into distinct images. In the examples of mappings we gave earlier, the mapping of Example 1 is not 1-1, since two brothers would have the same mother. However in Example 2 the mapping is 1-1 because distinct U.S. citizens have distinct Social Security numbers (provided that there is no goof-up in Washington, which is unlikely). The reader should check if the various other examples of mappings are 1-1.

Given a mapping $f: S \to T$ and a subset $A \subset T$, we may want to look at $B = \{s \in S \mid f(s) \in A\}$; we use the notation $f^{-1}(A)$ for this set $B$, and call $f^{-1}(A)$ the *inverse image of A under f*. Of particular interest is $f^{-1}(t)$, the inverse image of the subset $\{t\}$ of $T$ consisting of the element $t \in T$ alone. If the inverse image of $\{t\}$ consists of only one element, say $s \in S$, we could try to define $f^{-1}(t)$ by defining $f^{-1}(t) = s$. As we note below, this need not be a mapping from $T$ to $S$, but is so if $f$ is 1-1 and onto. We shall use the same notation $f^{-1}$ in cases of both subsets and elements. This $f^{-1}$ does *not* in general define a mapping from $T$ to $S$ for several reasons. First, if $f$ is not onto, then

there is some $t$ in $T$ which is not the image of any element $s$, so $f^{-1}(t) = \emptyset$. Second, if $f$ is not 1-1, then for some $t \in T$ there are at least two distinct $s_1 \neq s_2$ in $S$ such that $f(s_1) = t = f(s_2)$. So $f^{-1}(t)$ is *not* a unique element of $S$—something we require in our definition of mapping. However, if $f$ is both 1-1 and onto $T$, then $f^{-1}$ indeed defines a mapping of $T$ onto $S$. (Verify!) This brings us to a very important class of mappings.

**Definition.** The mapping $f: S \to T$ is said to be a 1-1 *correspondence* or *bijection* if $f$ is both 1-1 and onto.

Now that we have the notion of a mapping and have singled out various types of mappings, we might very well ask: "Good and well, but what can we do with them?" As we shall see in a moment, we can introduce an operation of combining mappings in certain circumstances.

Consider the situation $g: S \to T$ and $f: T \to U$. Given an element $s \in S$, then $g$ sends it into the element $g(s)$ in $T$; so $g(s)$ is ripe for being acted on by $f$. Thus we get an element $f(g(s)) \in U$. We claim that this procedure provides us with a mapping from $S$ to $U$. (Verify!) We define this more formally in the

**Definition.** If $g: S \to T$ and $f: T \to U$, then the *composition* (or *product*), denoted by $f \circ g$, is the mapping $f \circ g: S \to U$ defined by $(f \circ g)(s) = f(g(s))$ for every $s \in S$.

Note that to compose the two mappings $f$ and $g$—that is, for $f \circ g$ to have any sense—the *terminal set*, $T$, for the mapping $g$ *must be* the *initial set* for the mapping $f$. *One special time when we can always compose any two mappings is when $S = T = U$, that is, when we map $S$ into itself.* Although special, this case is of the utmost importance.

We verify a few properties of this composition of mappings.

**Lemma 1.3.1.** If $h: S \to T$, $g: T \to U$, and $f: U \to V$, then $f \circ (g \circ h) = (f \circ g) \circ h$.

*Proof.* How shall we go about proving this lemma? To verify that two mappings are equal, we merely must check that they do the same thing to every element. Note first of all that both $f \circ (g \circ h)$ and $(f \circ g) \circ h$ define mappings from $S$ to $V$, so it makes sense to speak about their possible equality.

Our task, then, is to show that for every $s \in S$, $(f \circ (g \circ h))(s) = ((f \circ g) \circ h)(s)$. We apply the definition of composition to see that

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

Unraveling

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

we do indeed see that

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

for every $s \in S$. Consequently, by definition, $f \circ (g \circ h) = (f \circ g) \circ h$. □

(The symbol □ will always indicate that the proof has been completed.)

This equality is described by saying that mappings, under composition, satisfy the *associative law*. Because of the equality involved there is really no need for parentheses, so we write $f \circ (g \circ h)$ as $f \circ g \circ h$.

**Lemma 1.3.2.** If $g : S \to T$ and $f : T \to U$ are both 1-1, then $f \circ g : S \to U$ is also 1-1.

*Proof.* Let us suppose that $(f \circ g)(s_1) = (f \circ g)(s_2)$; thus, by definition, $f(g(s_1)) = f(g(s_2))$. Since $f$ is 1-1, we get from this that $g(s_1) = g(s_2)$; however, $g$ is also 1-1, thus $s_1 = s_2$ follows. Since $(f \circ g)(s_1) = (f \circ g)(s_2)$ forces $s_1 = s_2$, the mapping $f \circ g$ is 1-1. □

We leave the proof of the next Remark to the reader.

**Remark.** If $g : S \to T$ and $f : T \to U$ are both onto, then $f \circ g : S \to U$ is also onto.

An immediate consequence of combining the Remark and Lemma 1.3.2 is to obtain

**Lemma 1.3.3.** If $g : S \to T$ and $f : T \to U$ are both bijections, then $f \circ g : S \to U$ is also a bijection.

If $f$ is a 1-1 correspondence of $S$ onto $T$, then the "object" $f^{-1} : T \to S$ defined earlier can easily be shown to be a 1-1 mapping of $T$ onto $S$. In this case it is called the *inverse* of $f$. In this situation we have

**Lemma 1.3.4.** If $f : S \to T$ is a bijection, then $f \circ f^{-1} = i_T$ and $f^{-1} \circ f = i_S$, where $i_S$ and $i_T$ are the identity mappings of $S$ and $T$, respectively.

*Proof.* We verify one of these. If $t \in T$, then $(f \circ f^{-1})(t) = f(f^{-1}(t))$. But what is $f^{-1}(t)$? By definition, $f^{-1}(t)$ is that element $s_0 \in S$ such that

$t = f(s_0)$. So $f(f^{-1}(t)) = f(s_0) = t$. In other words, $(f \circ f^{-1})(t) = t$ for every $t \in T$; hence $f \circ f^{-1} = i_T$, the identity mapping on $T$. □

We leave the last result of this section for the reader to prove.

**Lemma 1.3.5.** If $f: S \to T$ and $i_T$ is the identity mapping of $T$ onto itself and $i_S$ is that of $S$ onto itself, then $i_T \circ f = f$ and $f \circ i_S = f$.

## PROBLEMS

### Easier Problems

1. For the given sets $S$, $T$ determine if a mapping $f: S \to T$ is clearly and unambiguously defined; if not, say why not.
   - **(a)** $S$ = set of all women, $T$ = set of all men, $f(s)$ = husband of $s$.
   - **(b)** $S$ = set of positive integers, $T = S$, $f(s) = s - 1$.
   - **(c)** $S$ = set of positive integers, $T$ = set of nonnegative integers, $f(s) = s - 1$.
   - **(d)** $S$ = set of nonnegative integers, $T = S$, $f(s) = s - 1$.
   - **(e)** $S$ = set of all integers, $T = S$, $f(s) = s - 1$.
   - **(f)** $S$ = set of all real numbers, $T = S$, $f(s) = \sqrt{s}$.
   - **(g)** $S$ = set of all positive real numbers, $T = S$, $f(s) = \sqrt{s}$.
2. In those parts of Problem 1 where $f$ does define a function, determine if it is 1-1, onto, or both.

*3. If $f$ is a 1-1 mapping of $S$ onto $T$, prove that $f^{-1}$ is a 1-1 mapping of $T$ onto $S$.

*4. If $f$ is a 1-1 mapping of $S$ onto $T$, prove that $f^{-1} \circ f = i_S$.

5. Give a proof of the Remark after Lemma 1.3.2.

*6. If $f: S \to T$ is onto and $g: T \to U$ and $h: T \to U$ are such that $g \circ f = h \circ f$, prove that $g = h$.

*7. If $g: S \to T$, $h: S \to T$, and if $f: T \to U$ is 1-1, show that if $f \circ g = f \circ h$, then $g = h$.

8. Let $S$ be the set of all integers and $T = \{1, -1\}$; $f: S \to T$ is defined by $f(s) = 1$ if $s$ is even, $f(s) = -1$ if $s$ is odd.
   - **(a)** Does this define a function from $S$ to $T$?
   - **(b)** Show that $f(s_1 + s_2) = f(s_1)f(s_2)$. What does this say about the integers?
   - **(c)** Is $f(s_1 s_2) = f(s_1)f(s_2)$ also true?

**9.** Let $S$ be the set of all real numbers. Define $f: S \to S$ by $f(s) = s^2$, and $g: S \to S$ by $g(s) = s + 1$.
   **(a)** Find $f \circ g$.
   **(b)** Find $g \circ f$.
   **(c)** Is $f \circ g = g \circ f$?

**10.** Let $S$ be the set of all real numbers and for $a, b \in S$, where $a \neq 0$, define $f_{a,b}(s) = as + b$.
   **(a)** Show that $f_{a,b} \circ f_{c,d} = f_{u,v}$ for some real $u, v$. Give explicit values for $u, v$ in terms of $a, b, c$, and $d$.
   **(b)** Is $f_{a,b} \circ f_{c,d} = f_{c,d} \circ f_{a,b}$ always?
   **(c)** Find all $f_{a,b}$ such that $f_{a,b} \circ f_{1,1} = f_{1,1} \circ f_{a,b}$.
   **(d)** Show that $f_{a,b}^{-1}$ exists and find its form.

**11.** Let $S$ be the set of all positive integers. Define $f: S \to S$ by $f(1) = 2$, $f(2) = 3$, $f(3) = 1$, and $f(s) = s$ for any other $s \in S$. Show that $f \circ f \circ f = i_S$. What is $f^{-1}$ in this case?

**Middle-Level Problems**

**12.** Let $S$ be the set of nonnegative rational numbers, that is, $S = \{m/n \mid m, n$ nonnegative integers, $n \neq 0\}$, and let $T$ be the set of all integers.
   **(a)** Does $f: S \to T$ defined by $f(m/n) = 2^m 3^n$ define a legitimate function from $S$ to $T$?
   **(b)** If not, how could you modify the definition of $f$ so as to get a legitimate function?

**13.** Let $S$ be the set of all positive integers of the form $2^m 3^n$, where $m > 0$, $n > 0$, and let $T$ be the set of all rational numbers. Define $f: S \to T$ by $f(2^m 3^n) = m/n$. Prove that $f$ defines a function from $S$ to $T$. (On what properties of the integers does this depend?)

**14.** Let $f: S \to S$, where $S$ is the set of all integers, be defined by $f(s) = as + b$, where $a, b$ are integers. Find the necessary and sufficient conditions on $a, b$ in order that $f \circ f = i_S$.

**15.** Find all $f$ of the form given in Problem 14 such that $f \circ f \circ f = i_S$.

**16.** If $f$ is a 1-1 mapping of $S$ onto itself, show that $(f^{-1})^{-1} = f$.

**17.** If $S$ is a finite set having $m > 0$ elements, how many mappings are there of $S$ into itself?

**18.** In Problem 17, how many 1-1 mappings are there of $S$ into itself?

**19.** Let $S$ be the set of all real numbers, and define $f: S \to S$ by $f(s) = s^2 + as + b$, where $a, b$ are fixed real numbers. Prove that for no values at $a, b$ can $f$ be onto or 1-1.

**20.** Let $S$ be the set of all positive real numbers. For positive reals $a, c$ and nonnegative reals $b, d$, is it ever possible that the mapping $f: S \to S$ defined by $f(s) = (as + b)/(cs + d)$ satisfies $f \circ f = i_S$? Find all such $a, b, c, d$ that do the trick.

**21.** Let $S$ be the set of all rational numbers and let $f_{a,b}: S \to S$ be defined by $f_{a,b}(s) = as + b$, where $a \neq 0$, $b$ are rational numbers. Find all $f_{c,d}$ of this form satisfying $f_{c,d} \circ f_{a,b} = f_{a,b} \circ f_{c,d}$ for *every* $f_{a,b}$.

**22.** Let $S$ be the set of all integers and $a, b, c$ rational numbers. Define $f: S \to S$ by $f(s) = as^2 + bs + c$. Find necessary and sufficient conditions on $a, b, c$, so that $f$ defines a mapping on $S$ [**Note:** $a, b, c$ need not be integers; for example, $f(s) = \frac{1}{2}s(s + 1) = \frac{1}{2}s^2 + \frac{1}{2}s$ *does* always give us an integer for integral $s$.]

**Harder Problems**

**23.** Let $S$ be the set of all integers of the form $2^m3^n$, $m \geq 0$, $n \geq 0$, and let $T$ be the set of all positive integers. Show that there is a 1-1 correspondence of $S$ onto $T$.

**24.** Prove that there is a 1-1 correspondence of the set of all positive integers onto the set of all positive rational numbers.

**25.** Let $S$ be the set of all real numbers and $T$ the set of all positive reals. Find a 1-1 mapping $f$ of $S$ onto $T$ such that $f(s_1 + s_2) = f(s_1)f(s_2)$ for all $s_1, s_2 \in S$.

**26.** For the $f$ in Problem 25, find $f^{-1}$ explicitly.

**27.** If $f, g$ are mappings of $S$ into $S$ and $f \circ g$ is a *constant* function, then
**(a)** What can you say about $f$ if $g$ is onto?
**(b)** What can you say about $g$ if $f$ is 1-1?

**28.** If $S$ is a finite set and $f$ is a mapping of $S$ *onto* itself, show that $f$ must be 1-1.

**29.** If $S$ is a finite set and $f$ is a 1-1 mapping of $S$ into itself, show that $f$ must be surjective.

**30.** If $S$ is a finite set and $f$ is a 1-1 mapping of $S$, show that for some integer $n > 0$,

$$\underbrace{f \circ f \circ f \circ \cdots \circ f}_{n \text{ times}} = i_S.$$

**31.** If $S$ has $m$ elements in Problem 30, find an $n > 0$ (in terms of $m$) that works simultaneously for all 1-1 mappings of $S$ into itself.

## 4. $A(S)$ (THE SET OF 1-1 MAPPINGS OF $S$ ONTO ITSELF)

We focus our attention in this section on particularly nice mappings of a nonempty set, $S$, into itself. Namely, we shall consider the set, $A(S)$, of all 1-1 mappings of $S$ onto itself. Although most of the concern in the book will be in the case in which $S$ is a finite set, we do not restrict ourselves to that situation here.

When $S$ has a finite number of elements, say $n$, then $A(S)$ has a special name. It is called the *symmetric group of degree n* and is often denoted by $S_n$. Its elements are called *permutations* of $S$. If we are interested in the structure of $S_n$, it really does not matter much what our underlying set $S$ is. So, you can think of $S$ as being the set $\{1, \ldots, n\}$. Chapter 3 will be devoted to a study, in some depth, of $S_n$. In the investigation of finite groups, $S_n$ plays a central role.

There are many properties of the set $A(S)$ on which we could concentrate. We have chosen to develop those aspects here which will motivate the notion of a group and which will give the reader some experience, and feeling for, working in a group-theoretic framework. Groups will be discussed in Chapter 2.

We begin with a result that is really a compendium of some of the results obtained in Section 3.

**Lemma 1.4.1.** $A(S)$ satisfies the following:

(a) $f, g \in A(S)$ implies that $f \circ g \in A(S)$.

(b) $f, g, h \in A(S)$ implies that $(f \circ g) \circ h = f \circ (g \circ h)$.

(c) There exists an element—the identity mapping $i$—such that $f \circ i = i \circ f = f$ for every $f \in A(S)$.

(d) Given $f \in A(S)$, there exists a $g \in A(S)$ $(g = f^{-1})$ such that $f \circ g = g \circ f = i$.

*Proof.* All these things were done in Section 3, either in the text material or in the problems. We leave it to the reader to find the relevant part of Section 3 that will verify each of the statements (a) through (d). □

We should now like to know how many elements there are in $A(S)$ when $S$ is a finite set having $n$ elements. To do so, we first make a slight digression.

Suppose that you can do a certain thing in $r$ different ways and a second independent thing in $s$ different ways. In how many distinct ways can you do both things together? The best way of finding out is to picture this in

a concrete context. Suppose that there are $r$ highways running from Chicago to Detroit and $s$ highways running from Detroit to Ann Arbor. In how many ways can you go first to Detroit, then to Ann Arbor? Clearly, for every road you take from Chicago to Detroit you have $s$ ways of continuing on to Ann Arbor. You can start your trip from Chicago in $r$ distinct ways, hence you can complete it in

$$\underbrace{s + s + s + \cdots + s}_{r \text{ times}} = rs$$

different ways.

It is fairly clear that we can extend this from doing two independent things to doing $m$ independent ones, for an integer $m \geq 2$. If we can do the first things in $r_1$ distinct ways, the second in $r_2$ ways, . . . , the $m$th in $r_m$ distinct ways, then we can do all these together in $r_1 r_2 \ldots r_m$ different ways.

Let's recall something many of us have already seen:

**Definition.** If $n$ is a positive integer, then $n!$ (read "*n factorial*") is defined by $n! = 1 \cdot 2 \cdot 3 \cdots n$.

**Lemma 1.4.2.** If $S$ has $n$ elements, then $A(S)$ has $n!$ elements.

*Proof.* Let $f \in A(S)$, where $S = \{x_1, x_2, \ldots, x_n\}$. How many choices does $f$ have as a place to send $x_1$? Clearly $n$, for we can send $x_1$ under $f$ to *any* element of $S$. But now $f$ is *not* free to send $x_2$ anywhere, since $f$ is 1-1, we must have $f(x_1) \neq f(x_2)$. So we can send $x_2$ anywhere except onto $f(x_1)$. Hence $f$ can send $x_2$ into $n - 1$ different images. Continuing this way, we see that $f$ can send $x_i$ into $n - (i - 1)$ different images. Hence the number of such $f$'s is $n(n-1)(n-2)\cdots 1 = n!$ □

**Example**

The number $n!$ gets very large quickly. To be able to see the picture in its entirety, we look at the special case $n = 3$, where $n!$ is still quite small.

Consider $A(S) = S_3$, where $S$ consists of the three elements $x_1, x_2, x_3$. We list all the elements of $S_3$, writing out each mapping explicitly by what it does to each of $x_1, x_2, x_3$.

1. $i: x_1 \to x_1, x_2 \to x_2, x_3 \to x_3$.
2. $f: x_1 \to x_2, x_2 \to x_3, x_3 \to x_1$.
3. $g: x_1 \to x_2, x_2 \to x_1, x_3 \to x_3$.
4. $g \circ f: x_1 \to x_1, x_2 \to x_3, x_3 \to x_2$. (Verify!)

5. $f \circ g : x_1 \to x_3, x_2 \to x_2, x_3 \to x_1$. (Verify!)
6. $f \circ f : x_1 \to x_3, x_2 \to x_1, x_3 \to x_2$. (Verify!)

Since we have listed here six different elements of $S_3$, and $S_3$ has only six elements, we have a complete list of all the elements of $S_3$. What does this list tell us? To begin with, we note that $f \circ g \neq g \circ f$, so one familiar rule of the kind of arithmetic we have been used to is violated. Since $g \in S_3$ and $g \in S_3$, we must have $g \circ g$ also in $S_3$. What is it? If we calculate $g \circ g$, we easily get $g \circ g = i$. Similarly, we get

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

Note also that $f \circ (f \circ f) = i$, hence $f^{-1} = f \circ f$. Finally, we leave it to the reader to show that $g \circ f = f^{-1} \circ g$.

It is a little cumbersome to write this product in $A(S)$ using the $\circ$. *From now on we shall drop it and write $f \circ g$ merely as $fg$.* Also, we shall start using the shorthand of exponents, to avoid expressions like $f \circ f \circ f \circ \cdots \circ f$. We define, for $f \in A(S)$, $f^0 = i$, $f^2 = f \circ f = ff$, and so on. For negative exponents $n$ we define $f^{-n}$ by $f^{-n} = (f^{-1})^n$, where $n$ is a positive integer. The usual rules of exponents prevail, namely $f^r f^s = f^{r+s}$ and $(f^r)^s = f^{rs}$. We leave these as exercises—somewhat tedious ones at that—for the reader.

**Example**

Do not jump to conclusions that all familiar properties of exponents go over. For instance, in the example of the $f, g \in S_3$ defined above, we claim that $(fg)^2 \neq f^2 g^2$. To see this, we note that

$$fg : x_1 \to x_3, x_2 \to x_2, x_3 \to x_1,$$

so that $(fg)^2 : x_1 \to x_1, x_2 \to x_2, x_3 \to x_3$, that is, $(fg)^2 = i$. On the other hand, $f^2 \neq i$ and $g^2 = i$, hence $f^2 g^2 = f^2 \neq i$, whence $(fg)^2 \neq f^2 g^2$ in this case.

However, some other familiar properties do go over. For instance, if $f, g, h$ are in $A(S)$ and $fg = fh$, then $g = h$. Why? Because, from $fg = fh$ we have $f^{-1}(fg) = f^{-1}(fh)$, therefore, $g = ig = (f^{-1}f)g = f^{-1}(fg) = f^{-1}(fh) = (f^{-1}f)h = ih = h$. Similarly, $gf = hf$ implies that $g = h$. So we can cancel an element in such an equation provided that we *do not change sides*. In $S_3$ our $f, g$ satisfy $gf = f^{-1}g$, but since $f \neq f^{-1}$ we *cannot cancel* the $g$ here.

## PROBLEMS

Recall that $fg$ stands for $f \circ g$ and, also, what $f^n$ means. $S$, without subscripts, will be a nonempty set.

**Easier Problems**

1. If $s_1 \neq s_2$ are in $S$, show that there is an $f \in A(S)$ such that $f(s_1) = s_2$.
2. If $s_1 \in S$, let $H = \{f \in A(S) \mid f(s_1) = s_1\}$. Show that:
   **(a)** $i \in H$.
   **(b)** If $f, g \in H$, then $fg \in H$.
   **(c)** If $f \in H$, then $f^{-1} \in H$.
3. Suppose that $s_1 \neq s_2$ are in $S$ and $f(s_1) = s_2$, where $f \in A(S)$. Then if $H$ is as in Problem 2 and $K = \{g \in A(S) \mid g(s_2) = s_2\}$, show that:
   **(a)** If $g \in K$, then $f^{-1}gf \in H$.
   **(b)** If $h \in H$, then there is some $g \in K$ such that $h = f^{-1}gf$.
4. If $f, g, h \in A(S)$, show that $(f^{-1}gf)(f^{-1}hf) = f^{-1}(gh)f$. What can you say about $(f^{-1}gf)^n$?
5. If $f, g \in A(S)$ and $fg = gf$, show that:
   **(a)** $(fg)^2 = f^2g^2$.
   **(b)** $(fg)^{-1} = f^{-1}g^{-1}$.
6. Push the result of Problem 5, for the same $f$ and $g$, to show that $(fg)^m = f^mg^m$ for all integers $m$.
7. Verify the rules of exponents, namely $f^rf^s = f^{r+s}$ and $(f^r)^s = f^{rs}$ for $f \in A(S)$ and positive integers $r, s$.
8. If $f, g \in A(S)$ and $(fg)^2 = f^2g^2$, prove that $fg = gf$.
9. If $S = \{x_1, x_2, x_3, x_4\}$, let $f, g \in S_4$ be defined by

$$f: x_1 \to x_2, x_2 \to x_3, x_3 \to x_4, x_4 \to x_1,$$

   and

$$g: x_1 \to x_2, x_2 \to x_1, x_3 \to x_3, x_4 \to x_4.$$

   Calculate:
   **(a)** $f^2, f^3, f^4$.
   **(b)** $g^2, g^3$.
   **(c)** $fg$.
   **(d)** $gf$.
   **(e)** $(fg)^3, (gf)^3$.
   **(f)** $f^{-1}, g^{-1}$
10. If $f \in S_3$, show that $f^6 = i$.
11. Can you find a positive integer $m$ such that $f^m = i$ for *all* $f \in S_4$?

**Middle-Level Problems**

*12. If $f \in S_n$, show that there is some positive integer $k$, depending on $f$, such that $f^k = i$. (**Hint:** Consider the positive powers of $f$.)

*13. Show that there is a positive integer $t$ such that $f^t = i$ for *all* $f \in S_n$.

14. If $m < n$, show that there is a 1-1 mapping $F: S_m \to S_n$ such that $F(fg) = F(f)F(g)$ for all $f, g \in S_m$.

15. If $S$ has three or more elements, show that we can find $f, g \in A(S)$ such that $fg \neq gf$.

16. Let $S$ be an infinite set and let $M \subset A(S)$ be the set of all elements $f \in A(S)$ such that $f(s) \neq s$ for at most a finite number of $s \in S$. Prove that:
   **(a)** $f, g \in M$ implies that $fg \in M$.
   **(b)** $f \in M$ implies that $f^{-1} \in M$.

17. For the situation in Problem 16, show, if $f \in A(S)$, that $f^{-1}Mf = \{f^{-1}gf \mid g \in M\}$ must equal $M$.

18. Let $S \supset T$ and consider the subset $U(T) = \{f \in A(S) \mid f(t) \in T$ for every $t \in T\}$. Show that:
   **(a)** $i \in U(T)$.
   **(b)** $f, g \in U(T)$ implies that $fg \in U(T)$.

19. If the $S$ in Problem 18 has $n$ elements and $T$ has $m$ elements, how many elements are there in $U(T)$? Show that there is a mapping $F: U(T) \to S_m$ such that $F(fg) = F(f)F(g)$ for $f, g \in U(T)$ and $F$ is onto $S_m$.

20. If $m < n$, can $F$ in Problem 19 ever be 1-1? If so, when?

21. In $S_n$ show that the mapping $f$ defined by

$$f: x_1 \to x_2, x_2 \to x_3, x_3 \to x_4, \cdots, x_{n-1} \to x_n, x_n \to x_1$$

[i.e., $f(x_i) = x_{i+1}$ if $i < n$, $f(x_n) = x_1$] can be written as $f = g_1 g_2 \cdots g_{n-1}$ where each $g_i \in S_n$ *interchanges exactly two elements of* $S = \{x_1, \ldots, x_n\}$, leaving the other elements fixed in $S$.

**Harder Problems**

22. If $f \in S_n$, show that $f = h_1 h_2 \cdots h_m$ for some $h_j \in S_n$ such that $h_j^2 = i$.

*23. Call an element in $S_n$ a *transposition* if it interchanges two elements, leaving the others fixed. Show that any element in $S_n$ is a product of transpositions. (This sharpens the result of Problem 22.)

24. If $n$ is at least 3, show that for some $f$ in $S_n$, $f$ cannot be expressed in the form $f = g^3$ for any $g$ in $S_n$.

**25.** If $f \in S_n$ is such that $f \neq i$ but $f^3 = i$, show that we can number the elements of $S$ in such a way that $f(x_1) = x_2$, $f(x_2) = x_3$, $f(x_3) = x_1$, $f(x_4) = x_5$, $f(x_5) = x_6$, $f(x_6) = x_4$, $\ldots$, $f(x_{3k+1}) = x_{3k+2}$, $f(x_{3k+2}) = x_{3k+3}$, $f(x_{3k+3}) = x_{3k+1}$ for some $k$, and, for all the other $x_t \in S$, $f(x_t) = x_t$.

**26.** View a fixed shuffle of a deck of 52 cards as a 1-1 mapping of the deck onto itself. Show that repeating this fixed shuffle a finite (positive) number of times will bring the deck back to its original order.

***27.** If $f \in A(S)$, call, for $s \in S$, the *orbit* of $s$ (relative to $f$) the set $O(s) = \{f^j(s) \mid \text{all integers } j\}$. Show that if $s, t \in S$, then either $O(s) \cap O(t) = \emptyset$ or $O(s) = O(t)$.

**28.** If $S = \{x_1, x_2, \ldots, x_{12}\}$ and $f \in S_{12}$ is defined by $f(x_i) = x_{i+1}$ if $i = 1, 2, \ldots, 11$ and $f(x_{12}) = x_1$, find the orbits of all the elements of $S$ (relative to $f$).

**29.** If $f \in A(S)$ satisfies $f^3 = i$, show that the orbit of any element of $S$ has one or three elements.

***30.** Recall that a *prime number* is an integer $p > 1$ such that $p$ cannot be factored as a product of smaller positive integers. If $f \in A(S)$ satisfies $f^p = i$, what can you say about the size of the orbits of the elements of $S$ relative to $f$? What property of the prime numbers are you using to get your answer?

**31.** Prove that if $S$ has more than two elements, then the only elements $f_0$ in $A(S)$ such that $f_0 f = f f_0$ for all $f \in A(S)$ must satisfy $f_0 = i$.

***32.** We say that $g \in A(S)$ *commutes* with $f \in A(S)$ if $fg = gf$. Find all the elements in $A(S)$ that commute with $f: S \to S$ defined by $f(x_1) = x_2$, $f(x_2) = x_1$, and $f(s) = s$ if $s \neq x_1, x_2$.

**33.** In $S_n$ show that the only elements commuting with $f$ defined by $f(x_i) = x_{i+1}$ if $i < n$, $f(x_n) = x_1$, are the powers of $f$, namely $i = f^0, f, f^2, \ldots, f^{n-1}$.

**34.** For $f \in A(S)$, let $C(f) = \{g \in A(S) \mid fg = gf\}$. Prove that:

**(a)** $g, h \in C(f)$ implies that $gh \in C(f)$.

**(b)** $g \in C(f)$ implies that $g^{-1} \in C(f)$.

**(c)** $C(f)$ is not empty.

## 5. THE INTEGERS

The mathematical set most familiar to everybody is that of the positive integers $1, 2, \ldots$, which we shall often call $\mathbb{N}$. Equally familiar is the set, $\mathbb{Z}$, of all integers—positive, negative, and zero. Because of this acquaintance with $\mathbb{Z}$, we shall give here a rather sketchy survey of the properties of $\mathbb{Z}$ that we shall use often in the ensuing material. Most of these properties are well known to all of us; a few are less well known.

The basic assumption we make about the set of integers is the

**Well-Ordering Principle.** Any nonempty set of nonnegative integers has a smallest member.

More formally, what this principle states is that given a nonempty set $V$ of nonnegative integers, there is an element $v_0 \in V$ such that $v_0 \leq v$ for every $v \in V$. This principle will serve as the foundation for our ensuing discussion of the integers.

The first application we make of it is to show something we all know and have taken for granted, namely that we can divide one integer by another to get a remainder that is smaller. This is known as *Euclid's Algorithm*. We give it a more formal statement and a proof based on well-ordering.

**Theorem 1.5.1 (Euclid's Algorithm).** If $m$ and $n$ are integers with $n > 0$, then there exist integers $q$ and $r$, with $0 \leq r < n$, such that $m = qn + r$.

*Proof.* Let $W$ be the set of $m - tn$, where $t$ runs through all the integers, i.e., $W = \{m - tn \mid t \in \mathbb{Z}\}$. Note that $W$ contains some nonnegative integers, for if $t$ is large enough and negative, then $m - tn > 0$. Let $V = \{v \in W \mid v \geq 0\}$; by the well-ordering principle $V$ has a smallest element, $r$. Since $r \in V$, $r \geq 0$, and $r = m - qn$ for some $q$ (for that is the form of all elements in $W \supset V$). We claim that $r < n$. If not, $r = m - qn \geq n$, hence $m - (q + 1)n \geq 0$. But this puts $m - (q + 1)n$ in $V$, yet $m - (q + 1)n < r$, contradicting the minimal nature of $r$ in $V$. With this, Euclid's Algorithm is proved. □

Euclid's Algorithm will have a host of consequences for us, especially about the notion of divisibility. Since we are speaking about the integers, *be it understood that all letters used in this section will be integers*. This will save a lot of repetition of certain phrases.

**Definition.** Given integers $m \neq 0$ and $n$ we say that $m$ *divides* $n$, written as $m \mid n$, if $n = cm$ for some integer $c$.

Thus, for instance, $2 \mid 14$, $(-7) \mid 14$, $4 \mid (-16)$. If $m \mid n$, we call $m$ a *divisor* or *factor* of $n$, and $n$ a *multiple* of $m$. To indicate that $m$ is not a divisor of $n$, we write $m \nmid n$; so, for instance, $3 \nmid 5$.

The basic elementary properties of divisibility are laid out in

**Lemma 1.5.2.** The following are true:

(a) $1 \mid n$ for all $n$.

(b) If $m \neq 0$, then $m \mid 0$.

(c) If $m \mid n$ and $n \mid q$, then $m \mid q$.
(d) If $m \mid n$ and $m \mid q$, then $m \mid (un + vq)$ for all $u, v$.
(e) If $m \mid 1$, then $m = 1$ or $m = -1$.
(f) If $m \mid n$ and $n \mid m$, then $m = \pm n$.

*Proof.* The proofs of all these parts are easy, following immediately from the definition of $m \mid n$. We leave all but Part (d) as exercises but prove Part (d) here to give the flavor of how such proofs go.

So suppose that $m \mid n$ and $m \mid q$. Then $n = cm$ and $q = dm$ for some $c$ and $d$. Therefore, $un + vq = u(cm) + v(dm) = (uc + vd)m$. Thus, from the definition, $m \mid (un + vq)$. □

Having the concept of a divisor of an integer, we now want to introduce that of the *greatest common divisor* of two (or more) integers. Simply enough, this should be the largest possible integer that is a divisor of both integers in question. However, we want to avoid using the size of an integer for reasons that may become clear much later when we talk about rings. So we make the definition in what may seem as a strange way.

**Definition.** Given $a, b$ (not both 0), then their *greatest common divisor* $c$ is defined by:

(a) $c > 0$.
(b) $c \mid a$ and $c \mid b$.
(c) If $d \mid a$ and $d \mid b$, then $d \mid c$.

We write this $c$ as $c = (a, b)$.

In other words, the greatest common divisor of $a$ and $b$ is the positive number $c$ which divides $a$ and $b$ and is divisible by every $d$ which divides $a$ and $b$.

Defining something does not guarantee its existence. So it is incumbent on us to prove that $(a, b)$ exists, and is, in fact, unique. The proof actually shows more, namely that $(a, b)$ is a nice combination of $a$ and $b$. This combination is not unique; for instance,

$$(24, 9) = 3 = 3 \cdot 9 + (-1)24 = (-5)9 + 2 \cdot 24.$$

**Theorem 1.5.3.** If $a, b$ are not both 0, then their *greatest common divisor* $c = (a, b)$ exists, is unique, and, moreover, $c = m_0 a + n_0 b$ for some suitable $m_0$ and $n_0$.

*Proof.* Since not both $a$ and $b$ are 0, the set $A = \{ma + nb \mid m, n \in \mathbb{Z}\}$ has nonzero elements. If $x \in A$ and $x < 0$, then $-x$ is also in $A$ and $-x > 0$, for if $x = m_1a + n_1b$, then $-x = (-m_1)a + (-n_1)b$, so is in $A$. Thus $A$ has positive elements; hence, by the well-ordering principle there is a smallest positive element, $c$, in $A$. Since $c \in A$, by the form of the elements of $A$ we know that $c = m_0a + n_0b$ for some $m_0, n_0$.

We claim that $c$ is our required greatest common divisor. First note that if $d \mid a$ and $d \mid b$, then $d \mid (m_0a + n_0b)$ by Part (d) of Lemma 1.5.2, that is, $d \mid c$. So, to verify that $c$ is our desired element, we need only show that $c \mid a$ and $c \mid b$.

By Euclid's Algorithm, $a = qc + r$, where $0 \leq r < c$, that is, $a = q(m_0a + n_0b) + r$. Therefore, $r = -qn_0b + (1 - qm_0)a$. *So $r$ is in $A$.* But $r < c$ and is in $A$, so by the choice of $c$, *$r$ cannot be* positive. Hence $r = 0$; in other words, $a = qc$ and so $c \mid a$. Similarly, $c \mid b$.

For the uniqueness of $c$, if $t > 0$ also satisfied $t \mid a$, $t \mid b$ and $d \mid t$ for all $d$ such that $d \mid a$ and $d \mid b$, we would have $t \mid c$ and $c \mid t$. By Part (f) of Lemma 1.5.2 we get that $t = c$ (since both are positive). □

Let's look at an explicit example, namely $a = 24$, $b = 9$. By direct examination we know that $(24, 9) = 3$; note that $3 = 3 \cdot 9 + (-1)24$. What is $(-24, 9)$?

How is this done for positive numbers $a$ and $b$ which may be quite large? If $b > a$, interchange $a$ and $b$ so that $a > b > 0$. Then we can find $(a, b)$ by

1. observing that $(a, b) = (b, r)$ where $a = qb + r$ with $0 \leq r < b$ (Why?);
2. finding $(b, r)$, which now is easier since one of the numbers is smaller than before.

So, for example, we have

$$(100, 28) = (28, 16) \qquad \text{since } 100 = 3(28) + 16$$
$$(28, 16) = (16, 12) \qquad \text{since } 28 = 1(16) + 12$$
$$(16, 12) = (12, 4) \qquad \text{since } 16 = 1(12) + 4$$

This gives us

$$(100, 28) = (12, 4) = 4.$$

It is possible to find the actual values of $m_0$ and $n_0$ such that

$$4 = m_0 \cdot 100 + n_0 \cdot 28$$

by going backwards through the calculations made to find 4:

$$\begin{aligned} &\text{Since } 16 = 1(12) - 4, && 4 = 16 + (-1)12 \\ &\text{Since } 28 = 1(16) + 12, && 12 = 28 + (-1)16 \\ &\text{Since } 100 = 3(28) + 16, && 16 = 100 + (-3)28 \end{aligned}$$

But then

$$\begin{aligned} 4 &= 16 + (-1)12 = 16 + (-1)(28 + (-1)16) \\ &= (-1)28 + (2)16 = (-1)28 + (2)(100 + (-3)28) \\ &= (2)100 + (-7)28 \end{aligned}$$

so that $m_0 = 2$ and $n_0 = -7$.

This shows how Euclid's Algorithm can be used to compute $(a, b)$ for any positive integers $a$ and $b$.

We shall include some exercises at the end of this section on other properties of $(a, b)$.

We come to the very important

**Definition.** We say that $a$ and $b$ are *relatively prime* if $(a, b) = 1$.

So the integers $a$ and $b$ are relatively prime if they have no nontrivial common factor. An immediate corollary to Theorem 1.5.3 is

**Theorem 1.5.4.** The integers $a$ and $b$ are relatively prime if and only if $1 = ma + nb$ for suitable integers $m$ and $n$.

Theorem 1.5.4 has an immediate consequence

**Theorem 1.5.5.** If $a$ and $b$ are relatively prime and $a \mid bc$, then $a \mid c$.

*Proof.* By Theorem 1.5.4, $ma + nb = 1$ for some $m$ and $n$, hence $(ma + nb)c = c$, that is, $mac + nbc = c$. By assumption, $a \mid bc$ and by observation $a \mid mac$, hence $a \mid (mac + nbc)$ and so $a \mid c$. □

**Corollary.** If $b$ and $c$ are both relatively prime to $a$, then $bc$ is also relatively prime to $a$.

*Proof.* We pick up the proof of Theorem 1.5.5 at $mac + nbc = c$. If $d = (a, bc)$, then $d \mid a$ and $d \mid bc$, hence $d \mid (mac + nbc) = c$. Since $d \mid a$ and $d \mid c$

and $(a, c) = 1$, we get that $d = 1$. Since $1 = d = (a, bc)$, we have that $bc$ is relatively prime to $a$. □

We now single out an ultra-important class of positive integers, which we met before in Problem 30, Section 4.

**Definition.** A *prime number*, or a *prime*, is an integer $p > 1$, such that for any integer $a$ either $p \mid a$ or $p$ is relatively prime to $a$.

*This definition coincides with the usual one, namely that we cannot factor $p$ nontrivially.* For if $p$ is a prime as defined above and $p = ab$ where $1 \leq a < p$, then $(a, p) = a$ (Why?) and $p$ does not divide $a$ since $p > a$. It follows that $a = 1$, so $p = b$. On the other hand, if $p$ is a prime in the sense that it cannot be factored nontrivially, and if $a$ is an integer not relatively prime to $p$, then $(a, b)$ is not 1 and it divides $a$ and $p$. But then $(a, b)$ equals $p$, by our hypothesis, so $p$ divides $a$.

Another result coming out of Theorem 1.5.5 is

**Theorem 1.5.6.** If $p$ is a prime and $p \mid (a_1 a_2 \cdots a_n)$, then $p \mid a_i$ for some $i$ with $1 \leq i \leq n$.

*Proof.* If $p \mid a_1$, there is nothing to prove. Suppose that $p \nmid a_1$; then $p$ and $a_1$ are relatively prime. But $p \mid a_1(a_2 \cdots a_n)$, hence by Theorem 1.5.5, $p \mid a_2 \cdots a_n$. Repeat the argument just given on $a_2$, and continue. □

The primes play a very special role in the set of integers larger than 1 in that every integer $n > 1$ is either a prime or is the product of primes. We shall show this in the next theorem. In the theorem after the next we shall show that there is a uniqueness about the way $n > 1$ factors into prime factors. The proofs of both these results lean heavily on the well-ordering principle.

**Theorem 1.5.7.** If $n > 1$, then either $n$ is a prime or $n$ is the product of primes.

*Proof.* Suppose that the theorem is false. Then there must be an intger $m > 1$ for which the theorem fails. Therefore, the set $M$ for which the theorem fails is nonempty, so, by the well-ordering principle, $M$ has a least element $m$. Clearly, since $m \in M$, $m$ cannot be a prime, thus $m = ab$, where $1 < a < m$ and $1 < b < m$. Because $a < m$ and $b < m$ and $m$ is the least element in $M$, we cannot have $a \in M$ or $b \in M$. Since $a \notin M$, $b \notin M$, by the definition of $M$ the theorem must be true for both $a$ and $b$. Thus $a$ and $b$ are

primes or the product of primes; from $m = ab$ we get that $m$ is a product of primes. This puts $m$ outside of $M$, contradicting that $m \in M$. This proves the theorem. □

We asserted above that there is a certain uniqueness about the decomposition of an integer into primes. We make this precise now. To avoid trivialities of the kind $6 = 2 \cdot 3 = 3 \cdot 2$ (so, in a sense, 6 has two factorizations into the primes 2 and 3), we shall state the theorem in a particular way.

**Theorem 1.5.8.** Given $n > 1$, then there is one and only one way to write $n$ in the form $n = p_1^{a_1} p_2^{a_2} \cdots p_k^{a_k}$, where $p_1 < p_2 < \cdots < p_k$ are primes and the exponents $a_1, a_2, \ldots, a_k$ are all positive.

*Proof.* We start as we did above by assuming that the theorem is false, so there is a least integer $m > 1$ for which it is false. This *m must have two distinct factorizations* as $m = p_1^{a_1} p_2^{a_2} \cdots p_k^{a_k} = q_1^{b_1} q_2^{b_2} \cdots q_\ell^{b_\ell}$ where $p_1 < p_2 < \cdots < p_k$, $q_1 < q_2 < \cdots < q_\ell$ are primes and where the exponents $a_1, \ldots, a_k$ and $b_1, \ldots, b_\ell$ are all positive. Since $p_1 \mid p_1^{a_1} \cdots p_k^{a_k} = q_1^{b_1} \cdots q_\ell^{b_\ell}$, by Theorem 1.5.6 $p_1 \mid q_i^{b_i}$ for some $i$; hence, again by Theorem 1.5.6, $p_1 \mid q_i$, hence $p_1 = q_i$. By the same token $q_1 = p_j$ for some $j$; thus $p_1 \leq p_j = q_i \leq q_1 = p_1$. This gives us that $p_1 = q_1$. Now since $m/p_1 < m$, $m/p_1$ has the unique factorization property. But $m/p_1 = p_1^{a_1 - 1} p_2^{a_2} \cdots p_k^{a_k} = p_1^{b_1 - 1} q_2^{b_2} \cdots q_\ell^{b_\ell}$ and since $m/p_1$ can be factored in one and only one way in this form, we easily get $k = \ell$, $p_2 = q_2, \ldots, p_k = q_k$, $a_1 - 1 = b_1 - 1$, $a_2 = b_2, \ldots, a_k = b_k$. So we see that the primes and their exponents arising in the factorization of $m$ are unique. This contradicts the lack of such uniqueness for $m$, and so proves the theorem. □

What these last two theorems tell us is that we can build up the integers from the primes in a very precise and well-defined manner. One would expect from this that there should be many—that is, an infinity—of primes. This old result goes back to Euclid; in fact, the argument we shall give is due to Euclid.

**Theorem 1.5.9.** There is an infinite number of primes.

*Proof.* If the result were false, we could enumerate *all* the primes in $p_1, p_2, \ldots, p_k$. Consider the integer $q = 1 + p_1 p_2 \cdots p_k$. Since $q > p_i$ for every $i = 1, 2, \ldots, k$, $q$ cannot be a prime. Since $p_i \nmid q$, for we get a remainder of 1 on dividing $q$ by $p_i$, $q$ is not divisible by any of $p_1, \ldots, p_k$. So $q$ is not a prime nor is it divisible by any prime. This violates Theorem 1.5.7, thereby proving the theorem. □

Results much sharper than Theorem 1.5.9 exist about how many primes there are up to a given point. The famous prime number theorem states that for large $n$ the number of primes less than or equal to $n$ is "more or less" $n/\log_e n$, where this "more or less" is precisely described. There are many open questions about the prime numbers.

## PROBLEMS

### Easier Problems

**1.** Find $(a, b)$ and express $(a, b)$ as $ma + nb$ for:
- **(a)** (116, 84).
- **(b)** (85, 65).
- **(c)** (72, 26).
- **(d)** (72, 25).

**2.** Prove all the parts of Lemma 1.5.2, except part (d).

**3.** Show that $(ma, mb) = m(a, b)$ if $m > 0$.

**4.** Show that if $a \mid m$ and $b \mid m$ and $(a, b) = 1$, then $(ab) \mid m$.

**5.** Factor the following into primes.
- **(a)** 36.
- **(b)** 120.
- **(c)** 720.
- **(d)** 5040.

**6.** If $m = p_1^{a_1} \cdots p_k^{a_k}$ and $n = p_1^{b_1} \cdots p_k^{b_k}$, where $p_1, \ldots, p_k$ are distinct primes and $a_1, \ldots, a_k$ are nonnegative and $b_1, \ldots, b_k$ are nonnegative, express $(m, n)$ as $p_1^{c_1} \cdots p_k^{c_k}$ by describing the $c$'s in terms of the $a$'s and $b$'s.

***7.** Define the *least common multiple* of positive integers $m$ and $n$ to be the smallest positive integer $v$ such that both $m \mid v$ and $n \mid v$.
- **(a)** Show that $v = mn/(m, n)$.
- **(b)** In terms of the factorization of $m$ and $n$ given in Problem 6, what is $v$?

**8.** Find the least common multiple of the pairs given in Problem 1.

**9.** If $m, n > 0$ are two integers, show that we can find integers $u, v$ with $-n/2 < v \leq n/2$ such that $m = un + v$.

**10.** To check that a given integer $n > 1$ is a prime, prove that it is enough to show that $n$ is not divisible by any prime $p$ with $p \leq \sqrt{n}$.

**11.** Check if the following are prime.
- **(a)** 301.
- **(b)** 1001.
- **(c)** 473.

**12.** Starting with 2, 3, 5, 7, . . . , construct the positive integers $1 + 2 \cdot 3$, $1 + 2 \cdot 3 \cdot 5$, $1 + 2 \cdot 3 \cdot 5 \cdot 7$, . . . . Do you always get a prime number this way?

**Middle-Level Problems**

**13.** If $p$ is an odd prime, show that $p$ is of the form:
- **(a)** $4n + 1$ or $4n + 3$ for some $n$.
- **(b)** $6n + 1$ or $6n + 5$ for some $n$.

**14.** Adapt the proof of Theorem 1.5.9 to prove:
- **(a)** There is an infinite number of primes of the form $4n + 3$.
- **(b)** There is an infinite number of primes of the form $6n + 5$.

**15.** Show that no integer $u = 4n + 3$ can be written as $u = a^2 + b^2$, where $a, b$ are integers.

**16.** If $T$ is an infinite subset of $\mathbb{N}$, the set of all positive integers, show that there is a 1-1 mapping of $T$ *onto* $\mathbb{N}$.

**17.** If $p$ is a prime, prove that one cannot find nonzero integers $a$ and $b$ such that $a^2 = pb^2$. (This shows that $\sqrt{p}$ is *irrational*.)

## 6. MATHEMATICAL INDUCTION

If we look back at Section 5, we see that at several places –for instance, in the proof of Theorem 1.5.6—we say "argue as above and continue." This is not very satisfactory as a means of nailing down an argument. What is clear is that we need some technique of avoiding such phrases when we want to prove a proposition about *all* the positive integers. This is provided for us by the *Principle of Mathematical Induction*; in fact, this will be the usual method that we shall use for proving theorems about all the positive integers.

**Theorem 1.6.1.** Let $P(n)$ be a statement about the positive integers such that:

(a) $P(1)$ is true.

(b) If $P(k)$ happens to be true for some integer $k \geq 1$, then $P(k + 1)$ is also true.

Then $P(n)$ is true for all $n \geq 1$.

*Proof.* Actually, the arguments given in proving Theorems 1.5.7 and 1.5.8 are a prototype of the argument we give here.

Suppose that the theorem is false; then, by well-ordering, there is a least integer $m > 1$ for which $P(m)$ is not true. Since $P(1)$ is true, $m \neq 1$, hence $m > 1$. Now $1 \leq m - 1 < m$, so by the choice of $m$, $P(m - 1)$ must be valid. But then by the *inductive hypothesis* [Part (b)] we must have that $P(m)$ is true. This contradicts that $P(m)$ is not true. Thus there can be no integer for which $P$ is not true, and so the theorem is proved. □

We illustrate how to use induction with some rather diverse examples.

**Examples**

1. Suppose that $n$ tennis balls are put in a straight line, touching each other. Then we claim that these balls make $n - 1$ contacts.

*Proof.* If $n = 2$, the matter is clear. If for $k$ balls we have $k - 1$ contacts, then adding one ball (on a line) adds one contact. So $k + 1$ balls would have $k$ contacts. So if $P(n)$ is what is stated above about the tennis balls, we see that if $P(k)$ happens to be true, then so is $P(k + 1)$. Thus, by the theorem, $P(n)$ is true for all $n \geq 1$. □

2. If $p$ is a prime and $p \mid a_1 a_2 \cdots a_n$, then $p \mid a_i$ for some $1 \leq i \leq n$.

*Proof.* Let $P(n)$ be the statement in Example 2. Then $P(1)$ is true, for if $p \mid a_1$, it certainly divides $a_i$ for some $1 \leq i \leq 1$.

Suppose we know that $P(k)$ is true, and that $p \mid a_1 a_2 \cdots a_k a_{k+1}$. Thus, by Theorem 1.5.6, since $p \mid (a_1 a_2 \cdots a_k)a_{k+1}$, either $p \mid a_{k+1}$ (a desired conclusion) or $p \mid a_1 \cdots a_k$. In this second possibility, since $P(k)$ is true we have that $p \mid a_i$ for some $1 \leq i \leq k$. Combining both possibilities, we get that $p \mid a_j$ for some $1 \leq j \leq k + 1$. So Part (b) of Theorem 1.6.1 holds; hence $P(n)$ is true for all $n \geq 1$. □

3. For $n \geq 1$, $1 + 2 + \cdots + n = \frac{1}{2}n(n + 1)$.

*Proof.* If $P(n)$ is the proposition that $1 + 2 + \cdots + n = \frac{1}{2}n(n + 1)$, then $P(1)$ is certainly true, for $1 = \frac{1}{2}(1 + 1)$. If $P(k)$ should be true, this means that

$$1 + 2 + \cdots + k = \tfrac{1}{2}k(k + 1).$$

The question is: Is $P(k + 1)$ then also true, that is, is $1 + 2 + \cdots + k + (k + 1) = \frac{1}{2}(k + 1)((k + 1) + 1)$? Now $1 + 2 + \cdots + k + (k + 1) = (1 + 2 + \cdots k) + (k + 1) = \frac{1}{2}k(k + 1) + (k + 1)$, since $P(k)$ is valid. But $\frac{1}{2}k(k + 1) + (k + 1) = \frac{1}{2}(k(k + 1) + 2(k + 1)) = \frac{1}{2}(k + 1)(k + 2)$, which assures us that $P(k + 1)$ is true. Thus the proposition $1 + 2 + \cdots + n = \frac{1}{2}n(n + 1)$ is true for all $n \geq 1$. □

We must emphasize one point here: Mathematical induction is *not* a method for finding results about integers; it is a means of verifying a result. We could, by other means, find the formula given above for $1 + 2 + \cdots + n$.

Part (b) of Theorem 1.6.1 is usually called the *induction step*.

In the problems we shall give some other versions of the principle of induction.

## PROBLEMS

### Easier Problems

**1.** Prove that $1^2 + 2^2 + 3^2 + \cdots + n^2 = \frac{1}{6}n(n+1)(2n+1)$ by induction.

**2.** Prove that $1^3 + 2^3 + \cdots + n^3 = \frac{1}{4}n^2(n+1)^2$ by induction.

**3.** Prove that a set having $n \geq 2$ elements has $\frac{1}{2}n(n-1)$ subsets having exactly two elements.

**4.** Prove that a set having $n \geq 3$ elements has $n(n-1)(n-2)/3!$ subsets having exactly three elements.

**5.** If $n \geq 4$ and $S$ is a set having $n$ elements, guess (from Problems 3 and 4) how many subsets having exactly 4 elements there are in $S$. Then verify your guess using mathematical induction.

***6.** Complete the proof of Theorem 1.5.6, replacing the last sentence by an induction argument.

**7.** If $a \neq 1$, prove that $1 + a + a^2 + \cdots + a^n = (a^{n+1} - 1)/(a - 1)$ by induction.

**8.** By induction, show that

$$\frac{1}{1 \cdot 2} + \frac{1}{2 \cdot 3} + \cdots + \frac{1}{n(n+1)} = \frac{n}{n+1}.$$

***9.** Suppose that $P(n)$ is a proposition about positive integers $n$ such that $P(n_0)$ is valid, and if $P(k)$ is true, so must $P(k+1)$ be. What can you say about $P(n)$? Prove your statement.

***10.** Let $P(n)$ be a proposition about integers $n$ such that $P(1)$ is true and such that if $P(j)$ is true for all positive integers $j < k$, then $P(k)$ is true. Prove that $P(n)$ is true for all positive integers $n$.

### Middle-Level Problems

**11.** Give an example of a proposition that is *not* true for any positive integer, yet for which the induction step [Part (b) of Theorem 1.6.1] holds.

**12.** Prove by induction that a set having $n$ elements has exactly $2^n$ subsets.

**13.** Prove by induction on $n$ that $n^3 - n$ is always divisible by 3.

**14.** Using induction on $n$, generalize the result in Problem 13 to: If $p$ is a prime number, then $n^p - n$ is always divisible by $p$. (**Hint:** The binomial theorem.)

**15.** Prove by induction that for a set having $n$ elements the number of 1 1 mappings of this set onto itself is $n!$.

## 7. COMPLEX NUMBERS

We all know something about the integers, rational numbers, and real numbers· indeed, this assumption has been made for some of the text material and many of the problems have referred to these numbers. Unfortunately, the complex numbers and their properties are much less known to present-day college students. At one time the complex numbers were a part of the high school curriculum and the early college one. This is no longer the case. So we shall do a rapid development of this very important mathematical set.

The set of *complex numbers*, $\mathbb{C}$, is the set of all $a - bi$, where $a$, $b$ are real and where we *declare*:

1. $a - bi = c + di$, for $a, b, c, d$ real, if and only if $a = c$ and $b = d$.
2. $(a + bi) \pm (c + di) = (a \pm c) + (b \pm d)i$.
3. $(a + bi)(c + di) = (ac - bd) + (ad + bc)i$.

This last property—multiplication—can best be remembered by using $i^2 = -1$ and multiplying out formally with this relation in mind.

For the complex number $z = a + bi$, $a$ is called the *real part* of $z$ and $b$ the *imaginary part* of $z$. If $a$ is 0, we call $z$ *purely imaginary*.

We shall write $0 + 0i$ as 0 and $a + 0i$ as $a$. Note that $z + 0 = z$, $z1 = z$ for any complex number $z$.

Given $z = a + bi$, there is a complex number related to $z$, which we write as $\overline{z}$, defined by $\overline{z} = a - bi$. This complex number, $\overline{z}$, is called the *complex conjugate* of $z$. Taking the complex conjugate gives us a mapping of $\mathbb{C}$ onto itself. We claim

**Lemma 1.7.1.** If $z, w \in \mathbb{C}$, then:

(a) $\overline{(\overline{z})} = z$.
(b) $\overline{(z + w)} = \overline{z} + \overline{w}$.
(c) $\overline{(zw)} = \overline{z}\,\overline{w}$
(d) $z\overline{z}$ is real and nonnegative and is, in fact, positive if $z \neq 0$.

(e) $z + \overline{z}$ is twice the real part of $z$.

(f) $z - \overline{z}$ is twice the imaginary part of $z$ times $i$.

*Proof.* Most of the parts of this lemma are straightforward and merely involve using the definition of complex conjugate. We do verify Parts (c) and (d).

Suppose that $z = a + bi$, $w = c + di$, where $a, b, c, d$ are real. So $zw = (ac - bd) - (ad + bc)i$, hence

$$\overline{(zw)} = \overline{(ac - bd)} + \overline{(ad + bc)i} = (ac - bd) - (ad + bc)i.$$

On the other hand, $\overline{z} = a - bi$ and $\overline{w} = c - di$, hence, by the definition of the product in $\mathbb{C}$, $\overline{z}\,\overline{w} = (ac - bd) - (ad + bc)\,i$. Comparing this with the result that we obtained for $\overline{(zw)}$, we see that indeed $\overline{(zw)} = \overline{z}\,\overline{w}$. This verifies Part (c).

We go next to the proof of Part (d). Suppose that $z = a - bi \neq 0$, then $\overline{z} = a - bi$ and $z\overline{z} = a^2 + b^2$. Since $a, b$ are real and not both 0, $a^2 + b^2$ is real and positive, as asserted in Part (d). □

The proof of Part (d) of Lemma 1.7.1 shows that if $z = a + bi \neq 0$, then $z\overline{z} = a^2 + b^2 \neq 0$ and $z(\overline{z}/(a^2 + b^2)) = 1$, so

$$\frac{\overline{z}}{a^2 + b^2} = \frac{a}{a^2 + b^2} - \left(\frac{b}{a^2 + b^2}\right)i$$

acts like the inverse $1/z$ of $z$. This allows us to carry out division in $\mathbb{C}$, staying in $\mathbb{C}$ while doing so.

We now list a few properties of $\mathbb{C}$.

**Lemma 1.7.2.** $\mathbb{C}$ behaves under its sum and product according to the following: If $u, v, w \in \mathbb{C}$, then

(a) $u + v = v + u$.

(b) $(u + v) + w = u + (v + w)$.

(c) $uv = vu$.

(d) $(uv)w = u(vw)$.

(e) $u \neq 0$ implies that $u^{-1} = 1/u$ exists in $\mathbb{C}$ such that $uu^{-1} = 1$.

*Proof.* We leave the proofs of these various parts to the reader. □

These properties of $\mathbb{C}$ make of $\mathbb{C}$ what we shall call a *field*, which we shall study in much greater depth later in the book. What the lemma says is that we are allowed to calculate in $\mathbb{C}$ more or less as we did with real numbers. However, $\mathbb{C}$ has a much richer structure than the set of real numbers.

We now introduce a "size" function on $\mathbb{C}$.

**Definition.** If $z = a + bi \in \mathbb{C}$, then the *absolute value* of $z$, written as $|z|$, is defined by $|z| = \sqrt{z\overline{z}} = \sqrt{a^2 + b^2}$.

We shall see, in a few moments, what this last definition means geometrically. In the meantime we prove

**Lemma 1.7.3.** If $u, v \in \mathbb{C}$, then $|uv| = |u|\,|v|$.

*Proof.* By definition, $|u| = \sqrt{u\overline{u}}$ and $|v| = \sqrt{v\overline{v}}$. Now

$$\begin{aligned} |uv| &= \sqrt{(uv)\,\overline{(uv)}} = \sqrt{(uv)\,(\overline{u}\,\overline{v})} \qquad \text{(by Part (c) of Lemma 1.7.1)} \\ &= \sqrt{(u\overline{u})(v\overline{v})} \qquad \text{(by Lemma 1.7.2)} \\ &= \sqrt{u\overline{u}}\,\sqrt{v\overline{v}} = |u|\,|v|. \square \end{aligned}$$

Another way of verifying this lemma is to write $u = a + bi$, $v = c + di$, $uv = (ac - bd) + (ad + bc)i$ and to note the identity

$$(ac - bd)^2 + (ad + bc)^2 = (a^2 + b^2)(c^2 + d^2).$$

Note several small points about conjugates. If $z \in \mathbb{C}$, then $z$ is real if and only if $\overline{z} = z$, and $z$ is purely imaginary if and only if $\overline{z} = -z$. If $z, w \in \mathbb{C}$, then

$$\overline{(z\overline{w} + \overline{z}w)} = z\,\overline{\overline{w}} + \overline{\overline{z}}\,\overline{w} = \overline{z}w + z\overline{w},$$

so $z\overline{w} + \overline{z}w$ is real. We want to get an upper bound for $|z\overline{w} + \overline{z}w|$; this will come up in the proof of Theorem 1.7.5 below.

But first we must digress for a moment to obtain a statement about quadratic expressions.

**Lemma 1.7.4.** Let $a, b, c$ be real, with $a > 0$. If $a\alpha^2 - b\alpha + c \geq 0$ for every real $\alpha$, then $b^2 - 4ac \leq 0$.

*Proof.* Consider the quadratic expression for $\alpha = -b/2a$. We get $a(-b/2a)^2 + b(-b/2a) + c \geq 0$. Simplifying this, we obtain that $(4ac - b^2)/4a \geq 0$, and since $a > 0$, we end up with $4ac - b^2 \geq 0$, and so $b^2 - 4ac \leq 0$. $\square$

We use this result immediately to prove the important

**Theorem 1.7.5 (Triangle Inequality).** For $z, w \in \mathbb{C}$, $|z + w| \leq |z| + |w|$.

*Proof.* If $z = 0$, there is nothing to prove, so we may assume that $z \neq 0$; thus $z\overline{z} > 0$. Now, for $\alpha$ real,

$$0 \leq |\alpha z + w|^2 = (\alpha z + w)\,\overline{(\alpha z + w)} = (\alpha z + w)(\alpha \bar{z} + \bar{w})$$
$$= \alpha^2 z\bar{z} + \alpha(z\bar{w} + \bar{z}w) + w\bar{w}.$$

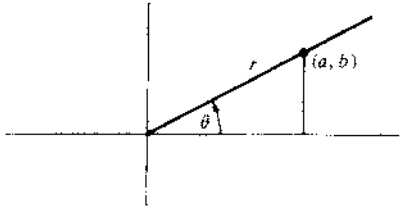
If $a = z\bar{z} > 0, b = z\bar{w} + \bar{z}w, c = w\bar{w}$, then Lemma 1.7.4 tells us that $b^2 - 4ac = (z\bar{w} + \bar{z}w)^2 - 4(z\bar{z})(w\bar{w}) \leq 0$, hence $(z\bar{w} + \bar{z}w)^2 \leq 4(z\bar{z})(w\bar{w}) = 4|z|^2|w|^2$. Therefore, $z\bar{w} + \bar{z}w \leq 2|z|\,|w|$.

For $\alpha = 1$ above,

$$|z + w|^2 = z\bar{z} + w\bar{w} + z\bar{w} + \bar{z}w = |z|^2 + |w|^2 + z\bar{w} + \bar{z}w$$
$$\leq |z|^2 + |w|^2 + 2|z|\,|w|$$

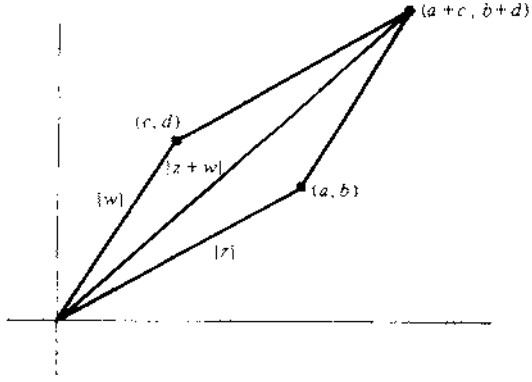
from the result above. In other words, $|z + w|^2 \leq (|z| + |w|)^2$; taking square roots we get the desired result, $|z + w| \leq |z| + |w|$. □

Why is this result called the triangle inequality? The reason will be clear once we view the complex numbers geometrically. Represent the complex number $z = a - bi$ as the point having coordinates $(a, b)$ in the $x$-$y$ plane.




The distance $r$ of this point from the origin is $\sqrt{a^2 + b^2}$, in other words, $|z|$. The angle $\theta$ is called the *argument* of $z$ and, as we see, $\tan\theta = b/a$. Also, $a = r\cos\theta, b = r\sin\theta$; therefore, $z = a + bi = r(\cos\theta + i\sin\theta)$. This representation of $z$ is called its *polar form*.

Given $z = a + bi$, $w = c + di$, then their sum is $z + w = (a + c) + (b + d)i$. Geometrically, we have the picture:



The statement $|z + w| \leq |z| + |w|$ merely reflects the fact that in a triangle one side is of smaller length than the sum of the lengths of the other two sides; thus the term *triangle inequality*.

The complex numbers that come up in the polar form $\cos\theta + i\sin\theta$ are very interesting numbers indeed. Specifically,

$$|\cos\theta + i\sin\theta| = \sqrt{\cos^2\theta + \sin^2\theta} = \sqrt{1} = 1,$$

so they give us many complex numbers of absolute value 1. In truth they give us *all* the complex numbers of absolute value 1; to see this just go back and look at the polar form of such a number.

Let's recall two basic identities from trigonometry, $\cos(\theta + \psi) = \cos\theta\cos\psi - \sin\theta\sin\psi$ and $\sin(\theta + \psi) = \sin\theta\cos\psi + \cos\theta\sin\psi$. Therefore, if $z = r(\cos\theta + i\sin\theta)$ and $w = s(\cos\psi + i\sin\psi)$, then

$$\begin{aligned} zw &= r(\cos\theta + i\sin\theta)\cdot s(\cos\psi + i\sin\psi) \\ &= rs(\cos\theta\cos\psi - \sin\theta\sin\psi) + i\,rs(\sin\theta\cos\psi + \cos\theta\sin\psi) \\ &= rs[\cos(\theta + \psi) + i\sin(\theta + \psi)]. \end{aligned}$$

Thus, in multiplying two complex numbers, the argument of the product is the sum of the arguments of the factors.

This has another very interesting consequence.

**Theorem 1.7.6 (De Moivre's Theorem).** For any integer $n \geq 1$, $[r(\cos\theta + i\sin\theta)]^n = r^n[\cos(n\theta) + i\sin(n\theta)]$.

*Proof.* We proceed by induction on $n$. If $n = 1$, the statement is obviously true. Assume then that for some $k$, $[r(\cos\theta + i\sin\theta)]^k = r^k[\cos k\theta + i\sin k\theta]$. Thus

$$\begin{aligned} [r(\cos\theta + i\sin\theta)]^{k+1} &= [r(\cos\theta + i\sin\theta)]^k \cdot r(\cos\theta + i\sin\theta) \\ &= r^k(\cos k\theta + i\sin k\theta)\cdot r(\cos\theta + i\sin\theta) \\ &= r^{k+1}[\cos(k+1)\theta + i\sin(k+1)\theta] \end{aligned}$$

by the result of the paragraph above. This completes the induction step; hence the result is true for all integers $n \geq 1$. □

In the problems we shall see that De Moivre's Theorem is true for all integers $m$; in fact, it is true even if $m$ is rational.

Consider the following special case:

$$\theta_n = \cos\frac{2\pi}{n} + i\sin\frac{2\pi}{n}, \qquad \text{where } n \geq 1 \text{ is an integer.}$$

By De Moivre's Theorem,

$$\left(\cos\left(\frac{2\pi}{n}\right) + i\sin\left(\frac{2\pi}{n}\right)\right)^n$$
$$= \cos\left(n\left(\frac{2\pi}{n}\right)\right) + i\sin\left(n\left(\frac{2\pi}{n}\right)\right)$$
$$= \cos 2\pi + i\sin 2\pi = 1.$$

So $\theta_n^n = 1$; you can verify that $\theta_n^m \neq 1$ if $0 < m < n$. This property of $\theta_n$ makes it one of the *primitive nth roots of unity*, which will be encountered in Problem 26.

## PROBLEMS

### Easier Problems

**1.** Multiply.
- **(a)** $(6 - 7i)(8 + i)$.
- **(b)** $(\frac{2}{3} + \frac{2}{2}i)(\frac{2}{3} - \frac{2}{2}i)$.
- **(c)** $(6 + 7i)(8 - i)$.

**2.** Express $z^{-1}$ in the form $z^{-1} = a - bi$ for:
- **(a)** $z = 6 + 8i$.
- **(b)** $z = 6 - 8i$.
- **(c)** $z = -\frac{1}{\sqrt{2}} + \frac{1}{\sqrt{2}}i$.

***3.** Show that $(\overline{z})^{-1} = (\overline{z^{-1}})$.

**4.** Find $(\cos\theta + i\sin\theta)^{-1}$.

**5.** Verify parts *a*, *b*, *e*, *f* of Lemma 1.7.1.

***6.** Show that $z$ is real if and only if $\overline{z} = z$, and is purely imaginary if and only if $\overline{z} = -z$.

**7.** Verify the commutative law of multiplication $zw = wz$ in $\mathbb{C}$.

**8.** Show that for $z \neq 0$, $|z^{-1}| = 1/|z|$.

**9.** Find:
- **(a)** $|6 - 4i|$.
- **(b)** $|\frac{1}{2} + \frac{2}{3}i|$.
- **(c)** $\left|\frac{1}{\sqrt{2}} + \frac{1}{\sqrt{2}}i\right|$.

**10.** Show that $|\overline{z}| = |z|$.

**11.** Find the polar form for

**(a)** $z = \frac{\sqrt{2}}{2} + \frac{1}{\sqrt{2}}i$.

**(b)** $z = 4i$.

**(c)** $z = \frac{6}{\sqrt{2}} + \frac{6}{\sqrt{2}}i$.

**(d)** $z = -\frac{13}{2} + \frac{39}{2\sqrt{3}}i$.

**12.** Prove that $(\cos(\frac{1}{2}\theta) - i\sin(\frac{1}{2}\theta))^2 = \cos\theta + i\sin\theta$.

**13.** By direct multiplication show that $(\frac{1}{2} + \frac{1}{2}\sqrt{3}\,i)^3 = -1$.

**Middle-Level Problems**

**14.** Show that $(\cos\theta + i\sin\theta)^m = \cos(m\theta) + i\sin(m\theta)$ for *all* integers $m$.

**15.** Show that $(\cos\theta + i\sin\theta)^r = \cos(r\theta) + i\sin(r\theta)$ for all rational numbers $r$.

**16.** If $z \in \mathbb{C}$ and $n \geq 1$ is any positive integer, show that there are $n$ distinct complex numbers $w$ such that $z = w^n$.

**17.** Find the necessary and sufficient condition on $k$ such that:

$$\left(\cos\left(\frac{2\pi k}{n}\right) + i\sin\left(\frac{2\pi k}{n}\right)\right)^n = 1 \quad \text{and}$$

$$\left(\cos\left(\frac{2\pi k}{n}\right) + i\sin\left(\frac{2\pi k}{n}\right)\right)^m \neq 1 \quad \text{if } 0 < m < n.$$

**18.** Viewing the $x$ $y$ plane as the set of all complex numbers $x + iy$, show that multiplication by $i$ induces a 90° rotation of the $x$ $y$ plane in a counterclockwise direction.

**19.** In Problem 18, interpret geometrically what multiplication by the complex number $a + bi$ does to the $x$-$y$ plane.

***20.** Prove that $|z + w|^2 + |z - w|^2 = 2(|z|^2 + |w|^2)$.

**21.** Consider the set $A = \{a + bi \mid a, b \in \mathbb{Z}\}$. Prove that there is a 1-1 correspondence of $A$ onto $\mathbb{N}$. ($A$ is called the set of *Gaussian integers*.)

**22.** If $a$ is a (complex) root of the polynomial

$$x^n + \alpha_1 x^{n-1} + \cdots + \alpha_{n-1}x + \alpha_n,$$

where the $\alpha_i$ are real, show that $\bar{a}$ must also be a root. [$r$ is a root of a polynomial $p(x)$ if $p(r) = 0$.]

**Harder Problems**

**23.** Find the necessary and sufficient conditions on $z$ and $w$ in order that $|z + w| = |z| + |w|$.

**24.** Find the necessary and sufficient conditions on $z_1, \cdots, z_k$ in order that $|z_1 + \cdots + z_k| = |z_1| + \cdots + |z_k|$.

***25.** The complex number $\theta$ is said to have *order* $n \geq 1$ if $\theta^n = 1$ and $\theta^m \neq 1$ for $0 < m < n$. Show that if $\theta$ has order $n$ and $\theta^k = 1$, where $k > 0$, then $n \mid k$.

***26.** Find all complex numbers $\theta$ having order $n$. (These are the *primitive* $n$th roots of unity.)