Chapter 1

Intervals

The three great realms of calculus are differential calculus, integral calculus, and the theory of infinite series. Central to each of these is the notion of *limit*: derivatives, integrals, and infinite series can be defined as limits of appropriate objects. Before entering these realms, however, one must have some background. There are the basic facts about the algebra, the order properties, and the topology of the real line $\mathbb{R}$. One must know about functions and the various ways in which they can be combined (addition, composition, and so on). And, finally, one needs some basic facts about continuous functions, in particular the *intermediate value theorem* and the *extreme value theorem*.

Much of this elementary material appears in the appendices. The remainder (the topology of $\mathbb{R}$, continuity, and the *intermediate* and *extreme value theorems*) occupies the first six chapters. After these have been completed, we proceed with limits and the differential calculus of functions of a single variable.

As you will recall from beginning calculus, we are accustomed to calling certain intervals "open" (for example, $(0, 1)$ and $(-\infty, 3)$ are open intervals) and other intervals "closed" (for example, $[0, 1]$ and $[1, \infty)$ are closed intervals). In this chapter and the next we investigate the meaning of the terms open and closed. These concepts turn out to be rather more important than one might at first expect. It will become clear, after our discussion of such matters as continuity, connectedness, and compactness in subsequent chapters, just how important they really are.

1.1. Distance and neighborhoods

1.1.1. Definition. If x and a are real numbers, the DISTANCE BETWEEN x and a, which we denote by $d(x, a)$, is defined to be $|x - a|$.

1.1.2. Example. There are exactly two real numbers whose distance from the number 3 is 7.

Proof. We are looking for all numbers $x \in \mathbb{R}$ such that $d(x,3) = 7$. In other words we want solutions to the equation $|x-3| = 7$. There are two such solutions. If $x-3 \geq 0$, then $|x-3| = x-3$; so $x = 10$ satisfies the equation. On the other hand, if $x-3 < 0$, then $|x-3| = -(x-3) = 3-x$, in which case $x = -4$ is a solution. □

1.1.3. Exercise. Find the set of all points on the real line that are within five units of the number -2.

1.1.4. Problem. Find the set of all real numbers whose distance from 4 is greater than 15.

1.1.5. Definition. Let a be a point in $\mathbb{R}$, and let $\epsilon > 0$. The open interval $(a-\epsilon, a+\epsilon)$ centered at a is called the ϵ-NEIGHBORHOOD of a and is denoted by $J_\epsilon(a)$. Notice that this neighborhood consists of all numbers x whose distance from a is less than ϵ; that is, such that $|x-a| < \epsilon$.

1.1.6. Example. The $\frac{1}{2}$-neighborhood of 3 is the open interval $\left(\frac{5}{2}, \frac{7}{2}\right)$.

Proof. We have $d(x,3) < \frac{1}{2}$ only if $|x-3| < \frac{1}{2}$. Solve this inequality to obtain $J_{\frac{1}{2}}(3) = \left(\frac{5}{2}, \frac{7}{2}\right)$. □

1.1.7. Example. The open interval $(1,4)$ is an ϵ-neighborhood of an appropriate point.

Proof. The midpoint of $(1,4)$ is the point $\frac{5}{2}$. The distance from this point to either end of the interval is $\frac{3}{2}$. Thus $(1,4) = J_{\frac{3}{2}}\left(\frac{5}{2}\right)$. □

1.1.8. Problem. Find, if possible, a number ϵ such that the ϵ-neighborhood of $\frac{1}{3}$ contains both $\frac{1}{4}$ and $\frac{1}{2}$ but does not contain $\frac{17}{30}$. If such a neighborhood does not exist, explain why.

1.1.9. Problem. Find, if possible, a number ϵ such that the ϵ-neighborhood of $\frac{1}{3}$ contains $\frac{11}{12}$ but does not contain either $\frac{1}{2}$ or $\frac{5}{8}$. If such a neighborhood does not exist, explain why.

1.1.10. Problem. Let $U = \left(\frac{1}{4}, \frac{2}{3}\right)$ and $V = \left(\frac{1}{2}, \frac{6}{5}\right)$. Write U and V as ϵ-neighborhoods of appropriate points. (That is, find numbers a and ϵ such that $U = J_\epsilon(a)$, and find numbers b and δ such that $V = J_\delta(b)$.) Also write the sets $U \cup V$ and $U \cap V$ as ϵ-neighborhoods of appropriate points.

1.1.11. Problem. Generalize your solution to the preceding problem to show that the union and the intersection of any two ϵ-neighborhoods that overlap is itself an ϵ-neighborhood of some point. *Hint.* Since ϵ-neighborhoods are open intervals of finite length, we can write the given neighborhoods as (a,b) and (c,d). There are really just two distinct cases. One neighborhood may contain the other; say, $a \leq c < d \leq b$. Or each may have points that are not in the other; say $a < c < b < d$. Deal with the two cases separately.

1.1.12. Proposition. *If $a \in \mathbb{R}$ and $0 < \delta \leq \epsilon$, then $J_\delta(a) \subseteq J_\epsilon(a)$.*

Proof. Exercise.

1.2. Interior of a set

1.2.1. Definition. Let $A \subseteq \mathbb{R}$. A point a is an INTERIOR POINT of A if some ϵ-neighborhood of a lies entirely in A. That is, a is an interior point of A if and only if there exists $\epsilon > 0$ such that $J_\epsilon(a) \subseteq A$. The set of all interior points of A is denoted by A° and is called the INTERIOR of A.

1.2.2. Example. Every point of the interval $(0, 1)$ is an interior point of that interval. Thus $(0, 1)^\circ = (0, 1)$.

Proof. Let a be an arbitrary point in $(0, 1)$. Choose ϵ to be the smaller of the two (positive) numbers a and $1 - a$. Then $J_\epsilon(a) = (a - \epsilon, a + \epsilon) \subseteq (0, 1)$ (because $\epsilon \leq a$ implies $a - \epsilon \geq 0$, and $\epsilon \leq 1 - a$ implies $a + \epsilon \leq 1$). □

1.2.3. Example. If $a < b$, then every point of the interval (a, b) is an interior point of the interval. Thus $(a, b)^\circ = (a, b)$.

Proof. Problem.

1.2.4. Example. The point 0 is not an interior point of the interval $[0, 1)$.

Proof. Argue by contradiction. Suppose 0 belongs to the interior of $[0, 1)$. Then for some $\epsilon > 0$, the interval $(-\epsilon, \epsilon) = J_\epsilon(0)$ is contained in $[0, 1)$. But this is impossible since the number $-\frac{1}{2}\epsilon$ belongs to $(-\epsilon, \epsilon)$ but not to $[0, 1)$. □

1.2.5. Example. Let $A = [a, b]$ where $a < b$. Then $A^\circ = (a, b)$.

Proof. Problem.

1.2.6. Example. Let $A = \{x \in \mathbb{R} \colon x^2 - x - 6 \geq 0\}$. Then $A^\circ \neq A$.

Proof. Exercise.

1.2.7. Problem. Let $A = \{x \in \mathbb{R} \colon x^3 - 2x^2 - 11x + 12 \leq 0\}$. Find A°.

1.2.8. Example. The interior of the set $\mathbb{Q}$ of rational numbers is empty.

Proof. No open interval contains only rational numbers. □

1.2.9. Proposition. *If A and B are sets of real numbers with $A \subseteq B$, then $A^\circ \subseteq B^\circ$.*

Proof. Let $a \in A^\circ$. Then there is an $\epsilon > 0$ such that $J_\epsilon(a) \subseteq A \subseteq B$. This shows that $a \in B^\circ$. □

1.2.10. Proposition. *If A is a set of real numbers, then $A^{\circ\circ} = A^\circ$.*

Proof. Problem.

1.2.11. Proposition. *If A and B are sets of real numbers, then*

$$(A \cap B)^\circ = A^\circ \cap B^\circ.$$

Proof. Exercise. *Hint.* Show separately that $(A \cap B)^\circ \subseteq A^\circ \cap B^\circ$ and that $A^\circ \cap B^\circ \subseteq (A \cap B)^\circ$.

1.2.12. Proposition. *If A and B are sets of real numbers, then*

$$(A \cup B)^\circ \supseteq A^\circ \cup B^\circ \,.$$

Proof. Exercise.

1.2.13. Example. Equality may fail in the preceding proposition.

Proof. Problem. *Hint.* See if you can find sets A and B in $\mathbb{R}$ both of which have empty interior but whose union is all of $\mathbb{R}$.

Chapter 2

Topology of the real line

It is clear from the definition of interior that the interior of a set is always contained in the set. Those sets for which the reverse inclusion also holds are called *open sets*.

2.1. Open subsets of $\mathbb{R}$

2.1.1. Definition. A subset U of $\mathbb{R}$ is OPEN if $U^\circ = U$. That is, a set is open if and only if every point of the set is an interior point of the set. If U is an open subset of $\mathbb{R}$, we write $U \overset{\circ}{\subseteq} \mathbb{R}$.

Notice, in particular, that the empty set is open. This is a consequence of the way implication is defined in Appendix D.2: the condition that each point of $\emptyset$ be an interior point is *vacuously satisfied* because there *are no points* in $\emptyset$. (One argues that *if* an element x belongs to the empty set, *then* it is an interior point of the set. The hypothesis is false; so the implication is true.) Also notice that $\mathbb{R}$ itself is an open subset of $\mathbb{R}$.

2.1.2. Example. Bounded open intervals are open sets. That is, if $a < b$, then the open interval (a, b) is an open set.

Proof. Example 1.2.3. □

2.1.3. Example. The interval $(0, \infty)$ is an open set.

Proof. Problem.

One way of seeing that a set is open is to verify that each of its points is an interior point of the set. That is what the definition says. Often it is easier to observe that the set can be written as a union of bounded open intervals. That this happens exactly when a set is open is the point of the next proposition.

2.1.4. Proposition. *A nonempty subset of $\mathbb{R}$ is open if and only if it is a union of bounded open intervals.*

Proof. Let $U \subseteq \mathbb{R}$. First, let us suppose U is a nonempty open subset of $\mathbb{R}$. Each point of U is then an interior point of U. So for each $x \in U$, we may choose a bounded open interval $J(x)$ centered at x that is entirely contained in U. Since $x \in J(x)$ for each $x \in U$, we see that

$$U = \bigcup_{x\in U} \{x\} \subseteq \bigcup_{x\in U} J(x). \tag{1}$$

On the other hand, since $J(x) \subseteq U$ for each $x \in U$, we have (see Proposition F.1.8)

$$\bigcup_{x\in U} J(x) \subseteq U. \tag{2}$$

Together (1) and (2) show that U is a union of bounded open intervals.

For the converse suppose $U = \bigcup \mathfrak{J}$, where $\mathfrak{J}$ is a family of open bounded intervals. Let x be an arbitrary point of U. We need only show that x is an interior point of U. To this end choose an interval $J \in \mathfrak{J}$ that contains x. Since J is a bounded open interval, we may write $J = (a, b)$ where $a < b$. Choose ϵ to be the smaller of the numbers $x - a$ and $b - x$. Then it is easy to see that $\epsilon > 0$ and that $x \in J_\epsilon(x) = (x - \epsilon, x + \epsilon) \subseteq (a, b)$. Thus x is an interior point of U. □

2.1.5. Example. Every interval of the form $(-\infty, a)$ is an open set. So is every interval of the form (a, ∞). (Notice that this and Example 2.1.2 give us the very comforting result that the things we are accustomed to calling open intervals are indeed open sets.)

Proof. Problem.

The study of calculus has two main ingredients: algebra and topology. Algebra deals with operations and their properties and with the resulting structure of groups, fields, vector spaces, algebras, and the like. Topology, on the other hand, is concerned with closeness, ϵ-neighborhoods, and open sets and with the associated structure of metric spaces and various kinds of topological spaces. Almost everything in calculus results from the interplay between algebra and topology.

2.1.6. Definition. The word “topology” has a technical meaning. A family $\mathfrak{T}$ of subsets of a set X is a TOPOLOGY on X if

(a) $\emptyset$ and X belong to $\mathfrak{T}$;
(b) if $\mathfrak{S} \subseteq \mathfrak{T}$ (that is, if $\mathfrak{S}$ is a subfamily of $\mathfrak{T}$), then $\bigcup \mathfrak{S} \in \mathfrak{T}$; and
(c) if $\mathfrak{S}$ is a *finite* subfamily of $\mathfrak{T}$, then $\bigcap \mathfrak{S} \in \mathfrak{T}$.

We can paraphrase this definition by saying that a family of subsets of X that contains both $\emptyset$ and X is a topology on X if it is closed under arbitrary unions and finite intersections. If this definition doesn't make any sense to you at first reading, don't fret. This kind of abstract definition, although easy enough to remember, is irritatingly difficult to understand. Staring at it doesn't help. It appears that a bewildering array of entirely different things might turn out to be topologies. And this is indeed the case. An understanding and appreciation of the definition come only gradually. You will notice as you advance through this material that many important concepts, such as continuity, compactness, and connectedness, are defined by (or characterized by) open sets. Thus, theorems that involve these ideas will rely on properties of open sets for their proofs. This is true not only in the present realm of the real line but in the much wider world of metric spaces, which we will shortly encounter in all their fascinating variety. You will notice that two properties of open sets are used over and over: that unions of open sets are open, and that finite intersections of open sets are open. Nothing else about open sets turns out to be of much importance. Gradually one comes to see that these two facts completely dominate the discussion of continuity, compactness, and so on. Ultimately, it becomes clear that nearly everything in the proofs goes through in situations where *only* these properties are available—that is, in topological spaces.

Our goal at the moment is quite modest: we show that the family of all open subsets of $\mathbb{R}$ is indeed a topology on $\mathbb{R}$.

2.1.7. Proposition. *Let $\mathfrak{S}$ be a family of open sets in $\mathbb{R}$. Then*

(a) *the union of $\mathfrak{S}$ is an open subset of $\mathbb{R}$; and*

(b) *if $\mathfrak{S}$ is finite, the intersection of $\mathfrak{S}$ is an open subset of $\mathbb{R}$.*

Proof. Exercise.

2.1.8. Example. The set $U = \{x \in \mathbb{R} \colon x < -2\} \cup \{x > 0 \colon x^2 - x - 6 < 0\}$ is an open subset of $\mathbb{R}$.

Proof. Problem.

2.1.9. Example. The set $\mathbb{R} \setminus \mathbb{N}$ is an open subset of $\mathbb{R}$.

Proof. Problem.

2.1.10. Example. The family $\mathfrak{T}$ of open subsets of $\mathbb{R}$ is not closed under arbitrary intersections. (That is, there exists a family $\mathfrak{S}$ of open subsets of $\mathbb{R}$ such that $\bigcap \mathfrak{S}$ is *not* open.)

Proof. Problem.

2.2. Closed subsets of $\mathbb{R}$

Next we will investigate the closed subsets of $\mathbb{R}$. These will turn out to be the complements of open sets. But initially we will approach them from a different perspective.

2.2.1. Definition. A point b in $\mathbb{R}$ is an ACCUMULATION POINT of a set $A \subseteq \mathbb{R}$ if every ϵ-neighborhood of b contains at least one point of A distinct from b. (We do *not* require that b belong to A, although, of course, it may.) The set of all accumulation points of A is called the DERIVED SET of A and is denoted by A'. The CLOSURE of A, denoted by $\overline{A}$, is $A \cup A'$.

2.2.2. Example. Let $A = \{1/n \colon n \in \mathbb{N}\}$. Then 0 is an accumulation point of A. Furthermore, $\overline{A} = \{0\} \cup A$.

Proof. Problem.

2.2.3. Example. Let A be $(0,1) \cup \{2\} \subseteq \mathbb{R}$. Then $A' = [0,1]$ and $\overline{A} = [0,1] \cup \{2\}$.

Proof. Problem.

2.2.4. Example. Every real number is an accumulation point of the set $\mathbb{Q}$ of rational numbers (since every open interval in $\mathbb{R}$ contains infinitely many rationals); so $\overline{\mathbb{Q}}$ is all of $\mathbb{R}$.

2.2.5. Exercise. Let $A = \mathbb{Q} \cap (0, \infty)$. Find A°, A', and $\overline{A}$.

2.2.6. Problem. Let $A = (0,1] \cup \big([2,3] \cap \mathbb{Q}\big)$. Find

(a) A°;
(b) $\overline{A}$;
(c) $\overline{A^\circ}$;
(d) $\big(\overline{A}\big)^\circ$;
(e) $\overline{A^c}$;
(f) $\big(\overline{A^c}\big)^\circ$;
(g) $\big(A^c\big)^\circ$; and
(h) $\overline{\big(A^c\big)^\circ}$.

2.2.7. Example. Let A be a nonempty subset of $\mathbb{R}$. If A is bounded above, then $\sup A$ belongs to the closure of A. Similarly, if A is bounded below, then $\inf A$ belongs to $\overline{A}$.

Proof. Problem.

2.2.8. Problem. Starting with a set A, what is the greatest number of *different* sets you can get by applying successively the operations of closure, interior, and complement? Apply them as many times as you wish and in any order. For example, starting with the empty set doesn't produce much. We get only $\emptyset$ and $\mathbb{R}$. If we start with the closed interval $[0,1]$, we get four sets: $[0,1]$, $(0,1)$, $(-\infty, 0] \cup [1, \infty)$, and $(-\infty, 0) \cup (1, \infty)$. By making a more cunning choice of A, how many different sets can you get?

2.2.9. Proposition. *Let $A \subseteq \mathbb{R}$. Then*

(a) $\big(A^\circ\big)^c = \overline{A^c}$*; and*
(b) $\big(A^c\big)^\circ = \big(\overline{A}\big)^c$.

Proof. Exercise. *Hint.* Part (b) is a very easy consequence of (a).

2.2.10. Definition. A subset A of $\mathbb{R}$ is CLOSED if $\overline{A} = A$.

2.2.11. Example. Every closed interval $\big($that is, intervals of the form $[a,b]$ or $(-\infty, a]$ or $[a, \infty)$ or $(-\infty, \infty)\big)$ is closed.

Proof. Problem.

CAUTION. It is a common mistake to treat subsets of $\mathbb{R}$ as if they were doors or windows, and to conclude, for example, that a set is closed because it is not open, or that it cannot be closed because it is open. These "conclusions" are wrong! A subset of $\mathbb{R}$ may be open and not closed, or closed and not open, or both open and closed, or neither. For example, in $\mathbb{R}$,

(a) $(0, 1)$ is open but not closed;
(b) $[0, 1]$ is closed but not open;
(c) $\mathbb{R}$ is both open and closed; and
(d) $[0, 1)$ is neither open nor closed.

This is not to say, however, that there is no relationship between these properties. In the next proposition we discover that the correct relation has to do with complements.

2.2.12. Proposition. *A subset of* $\mathbb{R}$ *is open if and only if its complement is closed.*

Proof. Problem. *Hint.* Use Proposition 2.2.9.

2.2.13. Proposition. *The intersection of an arbitrary family of closed subsets of* $\mathbb{R}$ *is closed.*

Proof. Let $\mathfrak{A}$ be a family of closed subsets of $\mathbb{R}$. By *De Morgan's law* (see Proposition F.3.5) $\bigcap \mathfrak{A}$ is the complement of $\bigcup\{A^c \colon A \in \mathfrak{A}\}$. Since each set A^c is open (by Proposition 2.2.12), the union of $\{A^c \colon A \in \mathfrak{A}\}$ is open (by Proposition 2.1.7(a)); and its complement $\bigcap \mathfrak{A}$ is closed (Proposition 2.2.12 again). □

2.2.14. Proposition. *The union of a finite family of closed subsets of* $\mathbb{R}$ *is closed.*

Proof. Problem.

2.2.15. Problem. Give an example to show that the union of an arbitrary family of closed subsets of $\mathbb{R}$ need not be closed.

2.2.16. Definition. Let a be a real number. Any open subset of $\mathbb{R}$ that contains a is a NEIGHBORHOOD of a. Notice that an ϵ-neighborhood of a is a very special type of neighborhood: it is an interval, and it is symmetric about a. For most purposes the extra internal structure possessed by ϵ-neighborhoods is irrelevant to the matter at hand. To see that we can operate as easily with general neighborhoods as with ϵ-neighborhoods, do the next problem.

2.2.17. Problem. Let A be a subset of $\mathbb{R}$. Prove the following.

(a) A point a is an interior point of A if and only if some neighborhood of a lies entirely in A.
(b) A point b is an accumulation point of A if and only if every neighborhood of b contains at least one point of A distinct from b.

Chapter 8

Differentiation of real valued functions

Differential calculus is a highly geometric subject—a fact that is not always made entirely clear in elementary texts, where the study of derivatives as numbers often usurps the place of the fundamental notion of linear approximation. The contemporary French mathematician Jean Dieudonné comments on the problem in Chapter 8 of his magisterial multivolume treatise on the *Foundations of Modern Analysis* [**Die62**]:

> ... the fundamental idea of calculus [is] the "local" approximation of functions by *linear* functions. In the classical teaching of Calculus, this idea is immediately obscured by the accidental fact that, on a one-dimensional vector space, there is a one-to-one correspondence between linear forms and numbers, and therefore the derivative at a point is defined as a *number* instead of a *linear form*. This slavish subservience to the shibboleth of numerical interpretation at any cost becomes much worse when dealing with functions of several variables ...

The goal of this chapter is to display as vividly as possible the geometric underpinnings of the differential calculus. The emphasis is on "tangency" and "linear approximation", not on number.

8.1. The families $\mathfrak{O}$ and $\mathfrak{o}$

8.1.1. Notation. Let $a \in \mathbb{R}$. We denote by $\mathcal{F}_a$ the family of all real valued functions defined on a neighborhood of a. That is, f belongs to $\mathcal{F}_a$ if there exists an open set U such that $a \in U \subseteq \operatorname{dom} f$.

Notice that for each $a \in \mathbb{R}$, the set $\mathcal{F}_a$ is closed under addition and multiplication. (We define the sum of two functions f and g in $\mathcal{F}_a$ to be the function $f + g$

whose value at x is $f(x) + g(x)$ whenever x belongs to $\operatorname{dom} f \cap \operatorname{dom} g$. A similar convention holds for multiplication.)

Among the functions defined on a neighborhood of zero are two subfamilies of crucial importance; they are $\mathfrak{O}$ (the family of "*big-oh*" functions) and $\mathfrak{o}$ (the family of "*little-oh*" functions).

8.1.2. Definition. A function f in $\mathcal{F}_0$ belongs to $\mathfrak{O}$ if there exist numbers $c > 0$ and $\delta > 0$ such that

$$|f(x)| \le c\,|x|$$

whenever $|x| < \delta$.

A function f in $\mathcal{F}_0$ belongs to $\mathfrak{o}$ if for every $c > 0$ there exists $\delta > 0$ such that

$$|f(x)| \le c\,|x|$$

whenever $|x| < \delta$. Notice that f belongs to $\mathfrak{o}$ if and only if $f(0) = 0$ and

$$\lim_{h \to 0} \frac{|f(h)|}{|h|} = 0\,.$$

8.1.3. Example. Let $f(x) = \sqrt{|x|}$. Then f belongs to neither $\mathfrak{O}$ nor $\mathfrak{o}$. (A function belongs to $\mathfrak{O}$ only if in some neighborhood of the origin its graph lies between two lines of the form $y = cx$ and $y = -cx$.)

8.1.4. Example. Let $g(x) = |x|$. Then g belongs to $\mathfrak{O}$ but not to $\mathfrak{o}$.

8.1.5. Example. Let $h(x) = x^2$. Then h is a member of both $\mathfrak{O}$ and $\mathfrak{o}$.

Much of the elementary theory of differential calculus rests on a few simple properties of the families $\mathfrak{O}$ and $\mathfrak{o}$. These are given in Propositions 8.1.8–8.1.14.

8.1.6. Definition. A function $L\colon \mathbb{R} \to \mathbb{R}$ is LINEAR if

$$L(x + y) = L(x) + L(y)$$

and

$$L(cx) = cL(x)$$

for all x, y, $c \in \mathbb{R}$. The family of all linear functions from $\mathbb{R}$ into $\mathbb{R}$ will be denoted by $\mathfrak{L}$.

The collection of linear functions from $\mathbb{R}$ into $\mathbb{R}$ is not very impressive, as the next problem shows. When we get to spaces of higher dimension, the situation will become more interesting.

8.1.7. Example. A function $f\colon \mathbb{R} \to \mathbb{R}$ is linear if and only if its graph is a (nonvertical) line through the origin.

Proof. Problem.

CAUTION. Since linear functions must pass through the origin, straight lines are not in general graphs of linear functions.

8.1.8. Proposition. *Every member of $\mathfrak{o}$ belongs to $\mathfrak{O}$; so does every member of $\mathfrak{L}$. Every member of $\mathfrak{O}$ is continuous at 0.*

Proof. Obvious from the definitions. □

8.1.9. Proposition. *Other than the constant function zero, no linear function belongs to* $\mathfrak{o}$*.*

Proof. Exercise.

8.1.10. Proposition. *The family* $\mathfrak{O}$ *is closed under addition and multiplication by constants.*

Proof. Exercise.

8.1.11. Proposition. *The family* $\mathfrak{o}$ *is closed under addition and multiplication by constants.*

Proof. Problem.

The next two propositions say that the composite of a function in $\mathfrak{O}$ with one in $\mathfrak{o}$ (in either order) ends up in $\mathfrak{o}$.

8.1.12. Proposition. *If* $g \in \mathfrak{O}$ *and* $f \in \mathfrak{o}$*, then* $f \circ g \in \mathfrak{o}$*.*

Proof. Problem.

8.1.13. Proposition. *If* $g \in \mathfrak{o}$ *and* $f \in \mathfrak{O}$*, then* $f \circ g \in \mathfrak{o}$*.*

Proof. Exercise.

8.1.14. Proposition. *If* $\phi, f \in \mathfrak{O}$*, then* $\phi f \in \mathfrak{o}$*.*

Proof. Exercise.

Remark. The preceding propositions can be summarized rather concisely. (Notation: $\mathcal{C}_0$ is the set of all functions in $\mathcal{F}_0$ that are continuous at 0.)

$$
\begin{aligned}
&(1) \qquad \mathfrak{L} \cup \mathfrak{o} \subseteq \mathfrak{O} \subseteq \mathcal{C}_0\,. \\
&(2) \qquad \mathfrak{L} \cap \mathfrak{o} = 0\,. \\
&(3) \qquad \mathfrak{O} + \mathfrak{O} \subseteq \mathfrak{O}\,; \qquad \alpha\,\mathfrak{O} \subseteq \mathfrak{O}\,. \\
&(4) \qquad \mathfrak{o} + \mathfrak{o} \subseteq \mathfrak{o}\,; \qquad \alpha\,\mathfrak{o} \subseteq \mathfrak{o}\,. \\
&(5) \qquad \mathfrak{o} \circ \mathfrak{O} \subseteq \mathfrak{o}\,. \\
&(6) \qquad \mathfrak{O} \circ \mathfrak{o} \subseteq \mathfrak{o}\,. \\
&(7) \qquad \mathfrak{O} \cdot \mathfrak{O} \subseteq \mathfrak{o}\,.
\end{aligned}
$$

8.1.15. Problem. Show that $\mathfrak{O} \circ \mathfrak{O} \subseteq \mathfrak{O}$. That is, if $g \in \mathfrak{O}$ and $f \in \mathfrak{O}$, then $f \circ g \in \mathfrak{O}$. (As usual, the domain of $f \circ g$ is taken to be $\{x\colon g(x) \in \operatorname{dom} f\}$.)

8.2. Tangency

The fundamental idea of differential calculus is the local approximation of a "smooth" function by a translate of a linear one. Certainly, the expression "local approximation" could be taken to mean many different things. One sense of this expression that has stood the test of usefulness over time is *tangency*. Two functions are said to be tangent at zero if their difference lies in the family $\mathfrak{o}$. We can of course define tangency of functions at an arbitrary point (see Problem 8.2.12 below); but for our purposes, "tangency at 0" will suffice. All the facts we need to know concerning this relation turn out to be trivial consequences of the results we have just proved.

8.2.1. Definition. Two functions f and g in $\mathcal{F}_0$ are TANGENT AT ZERO, in which case we write $f \simeq g$, if $f - g \in \mathfrak{o}$.

8.2.2. Example. Let $f(x) = x$ and $g(x) = \sin x$. Then $f \simeq g$, since $f(0) = g(0) = 0$ and $\lim_{x\to 0} \frac{x - \sin x}{x} = \lim_{x\to 0}\left(1 - \frac{\sin x}{x}\right) = 0$.

8.2.3. Example. If $f(x) = x^2 - 4x - 1$ and $g(x) = \left(3x^2 + 4x - 1\right)^{-1}$, then $f \simeq g$.

Proof. Exercise.

8.2.4. Proposition. *The relation "tangency at zero" is an equivalence relation on $\mathcal{F}_0$.*

Proof. Exercise.

The next result shows that at most one linear function can be tangent at zero to a given function.

8.2.5. Proposition. *Let $S, T \in \mathfrak{L}$ and $f \in \mathcal{F}_0$. If $S \simeq f$ and $T \simeq f$, then $S = T$.*

Proof. Exercise.

8.2.6. Proposition. *If $f \simeq g$ and $j \simeq k$, then $f + j \simeq g + k$, and, furthermore, $\alpha f \simeq \alpha g$ for all $\alpha \in \mathbb{R}$.*

Proof. Problem.

Suppose that f and g are tangent at zero. Under what circumstances are $h \circ f$ and $h \circ g$ tangent at zero? And when are $f \circ j$ and $g \circ j$ tangent at zero? We prove next that sufficient conditions are h is linear and j belongs to $\mathfrak{O}$.

8.2.7. Proposition. *Let $f, g \in \mathcal{F}_0$ and $T \in \mathfrak{L}$. If $f \simeq g$, then $T \circ f \simeq T \circ g$.*

Proof. Problem.

8.2.8. Proposition. *Let $h \in \mathfrak{O}$ and $f, g \in \mathcal{F}_0$. If $f \simeq g$, then $f \circ h \simeq g \circ h$.*

Proof. Problem.

8.2.9. Example. Let $f(x) = 3x^2 - 2x + 3$ and $g(x) = \sqrt{-20x + 25} - 2$ for $x \leq 1$. Then $f \simeq g$.

Proof. Problem.

8.2.10. Problem. Let $f(x) = x^3 - 6x^2 + 7x$. Find a linear function $T\colon \mathbb{R} \to \mathbb{R}$ that is tangent to f at 0.

8.2.11. Problem. Let $f(x) = |x|$. Show that there is no linear function $T\colon \mathbb{R} \to \mathbb{R}$ that is tangent to f at 0.

8.2.12. Problem. Let $T_a\colon x \mapsto x + a$. The mapping T_a is called TRANSLATION BY a. Note that it is *not* linear (unless, of course, $a = 0$). We say that functions f and g in $\mathcal{F}_a$ are TANGENT AT a if the functions $f \circ T_a$ and $g \circ T_a$ are tangent at 0.

(a) Let $f(x) = 3x^2 + 10x + 13$ and $g(x) = \sqrt{-20x - 15}$. Show that f and g are tangent at -2.

(b) Develop a theory for the relationship "tangency at a" that generalizes our work on "tangency at 0".

8.2.13. Problem. Each of the following is an abbreviated version of a proposition. Formulate precisely and prove.

(a) $\mathcal{C}_0 + \mathfrak{O} \subseteq \mathcal{C}_0$.

(b) $\mathcal{C}_0 + \mathfrak{o} \subseteq \mathcal{C}_0$.

(c) $\mathfrak{O} + \mathfrak{o} \subseteq \mathfrak{O}$.

8.2.14. Problem. Suppose that $f \simeq g$. Then the following hold.

(a) If g is continuous at 0, so is f.

(b) If g belongs to $\mathfrak{O}$, so does f.

(c) If g belongs to $\mathfrak{o}$, so does f.

8.3. Linear approximation

One often hears that differentiation of a smooth function f at a point a in its domain is the process of finding the best "linear approximation" to f at a. This informal assertion is not quite correct. For example, as we know from beginning calculus, the tangent line at $x = 1$ to the curve $y = 4 + x^2$ is the line $y = 2x + 3$, which is not a linear function since it does not pass through the origin. To rectify this rather minor shortcoming, we first translate the graph of the function f so that the point $(a, f(a))$ goes to the origin and *then* find the best linear approximation at the origin. The operation of translation is carried out by a somewhat notorious acquaintance from beginning calculus, Δy. The source of its notoriety is two-fold: first, in many texts it is inadequately defined; and second, the notation Δy fails to alert the reader to the fact that under consideration is a function of *two* variables. We will be careful on both counts.

8.3.1. Definition. Let $f \in \mathcal{F}_a$. Define the function Δf_a by

$$\Delta f_a(h) := f(a + h) - f(a)$$

for all h such that $a + h$ is in the domain of f. Notice that since f is defined in a neighborhood of a, the function Δf_a is defined in a neighborhood of 0; that is, Δf_a belongs to $\mathcal{F}_0$. Notice also that $\Delta f_a(0) = 0$.

8.3.2. Problem. Let $f(x) = \cos x$ for $0 \le x \le 2\pi$.

(a) Sketch the graph of the function f.

(b) Sketch the graph of the function Δf_π.

8.3.3. Proposition. *If $f \in \mathcal{F}_a$ and $\alpha \in \mathbb{R}$, then*

$$\Delta(\alpha f)_a = \alpha\, \Delta f_a\,.$$

Proof. To show that two functions are equal, show that they agree at each point in their domain. Here

$$\begin{aligned}\Delta(\alpha f)_a(h) &= (\alpha f)(a+h) - (\alpha f)(a)\\ &= \alpha f(a+h) - \alpha f(a)\\ &= \alpha(f(a+h) - f(a))\\ &= \alpha\, \Delta f_a(h)\end{aligned}$$

for every h in the domain of Δf_a. □

8.3.4. Proposition. *If $f, g \in \mathcal{F}_a$, then*

$$\Delta(f+g)_a = \Delta f_a + \Delta g_a\,.$$

Proof. Exercise.

The last two propositions prefigure the fact that differentiation is a linear operator; the next result will lead to *Leibniz's rule* for differentiating products.

8.3.5. Proposition. *If $\phi, f \in \mathcal{F}_a$, then*

$$\Delta(\phi f)_a = \phi(a)\cdot\Delta f_a \;+\; \Delta\phi_a \cdot f(a) \;+\; \Delta\phi_a \cdot \Delta f_a\,.$$

Proof. Problem.

Finally, we present a result that prepares the way for the *chain rule*.

8.3.6. Proposition. *If $f \in \mathcal{F}_a$, $g \in \mathcal{F}_{f(a)}$, and $g \circ f \in \mathcal{F}_a$, then*

$$\Delta(g \circ f)_a = \Delta g_{f(a)} \circ \Delta f_a\,.$$

Proof. Exercise.

8.3.7. Proposition. *Let $A \subseteq \mathbb{R}$. A function $f\colon A \to \mathbb{R}$ is continuous at the point a in A if and only if Δf_a is continuous at 0.*

Proof. Problem.

8.3.8. Proposition. *If $f\colon U \to U_1$ is a bijection between subsets of $\mathbb{R}$, then for each a in U the function $\Delta f_a\colon U - a \to U_1 - f(a)$ is invertible and*

$$\left(\Delta f_a\right)^{-1} = \Delta\left(f^{-1}\right)_{f(a)}\,.$$

Proof. Problem.

8.4. Differentiability

We now have developed enough machinery to talk sensibly about *differentiating* real valued functions.

8.4.1. Definition. Let $f \in \mathcal{F}_a$. We say that f is DIFFERENTIABLE AT a if there exists a linear function that is tangent at 0 to Δf_a. If such a function exists, it is called THE DIFFERENTIAL OF f at a and is denoted by df_a. (Don't be put off by the slightly complicated notation; df_a is just a member of $\mathfrak{L}$ satisfying $df_a \simeq \Delta f_a$.) We denote by $\mathcal{D}_a$ the family of all functions in $\mathcal{F}_a$ that are differentiable at a.

The next proposition justifies the use of the definite article that modifies "differential" in the preceding paragraph.

8.4.2. Proposition. *Let $f \in \mathcal{F}_a$. If f is differentiable at a, then its differential is unique.* (*That is, there is at most one linear map tangent at* 0 *to* Δf_a.)

Proof. Proposition 8.2.5. □

8.4.3. Example. It is instructive to examine the relationship between the differential of f at a, which we defined in Definition 8.4.1, and the derivative of f at a as defined in beginning calculus. For $f \in \mathcal{F}_a$ to be differentiable at a it is necessary that there be a linear function $T\colon \mathbb{R} \to \mathbb{R}$ that is tangent at 0 to Δf_a. According to Example 8.1.7, there must exist a constant c such that $Tx = cx$ for all x in $\mathbb{R}$. For T to be tangent to Δf_a, it must be the case that

$$\Delta f_a - T \in \mathfrak{o}\,;$$

that is,

$$\lim_{h\to 0} \frac{\Delta f_a(h) - ch}{h} = 0\,.$$

Equivalently,

$$\lim_{h\to 0} \frac{f(a+h) - f(a)}{h} = \lim_{h\to 0} \frac{\Delta f_a(h)}{h} = c\,.$$

In other words, the function T, which is tangent to Δf_a at 0, must be a line through the origin whose slope is

$$\lim_{h\to 0} \frac{f(a+h) - f(a)}{h}\,.$$

This is, of course, the familiar "derivative of f at a" from beginning calculus. Thus for any real valued function f that is differentiable at a in $\mathbb{R}$,

$$df_a(h) = f'(a) \cdot h$$

for all $h \in \mathbb{R}$.

8.4.4. Problem. Explain carefully the quotation from Dieudonné given at the beginning of the chapter.

8.4.5. Example. Let $f(x) = 3x^2 - 7x + 5$ and $a = 2$. Then f is differentiable at a. (Sketch the graph of the differential df_a.)

Proof. Problem.

8.4.6. Example. Let $f(x) = \sin x$ and $a = \pi/3$. Then f is differentiable at a. (Sketch the graph of the differential df_a.)

Proof. Problem.

8.4.7. Proposition. *Let $T \in \mathfrak{L}$ and $a \in \mathbb{R}$. Then $dT_a = T$.*

Proof. Problem.

8.4.8. Proposition. *If $f \in \mathcal{D}_a$, then $\Delta f_a \in \mathfrak{O}$.*

Proof. Exercise.

8.4.9. Corollary. *Every function that is differentiable at a point is continuous there.*

Proof. Exercise.

8.4.10. Proposition. *If f is differentiable at a and $\alpha \in \mathbb{R}$, then αf is differentiable at a and*

$$d(\alpha f)_a = \alpha\, df_a\,.$$

Proof. Exercise.

8.4.11. Proposition. *If f and g are differentiable at a, then $f+g$ is differentiable at a and*

$$d(f+g)_a = df_a + dg_a\,.$$

Proof. Problem.

8.4.12. Proposition (Leibniz's rule)**.** *If $\phi, f \in \mathcal{D}_a$, then $\phi f \in \mathcal{D}_a$ and*

$$d(\phi f)_a = d\phi_a \cdot f(a) \;+\; \phi(a)\, df_a\,.$$

Proof. Exercise.

8.4.13. Theorem (The chain rule)**.** *If $f \in \mathcal{D}_a$ and $g \in \mathcal{D}_{f(a)}$, then $g \circ f \in \mathcal{D}_a$ and*

$$d(g \circ f)_a = dg_{f(a)} \circ df_a\,.$$

Proof. Exercise.

8.4.14. Problem (A problem set on functions from $\mathbb{R}$ into $\mathbb{R}$)**.** We are now in a position to derive the standard results, usually contained in the first semester of a beginning calculus course, concerning the differentiation of real valued functions of a single real variable. Having at our disposal the machinery developed earlier in this chapter, we may derive these results quite easily; and so the proof of each is a problem.

8.4.15. Definition. If $f \in \mathcal{D}_a$, the DERIVATIVE OF f AT a, denoted by $f'(a)$ or $Df(a)$, is defined to be $\lim_{h\to 0} \frac{f(a+h)-f(a)}{h}$. By Proposition 7.2.4 this is the same as $\lim_{x\to a} \frac{f(x)-f(a)}{x-a}$.

8.4.16. Proposition. *If $f \in \mathcal{D}_a$, then $Df(a) = df_a(1)$.*

Proof. Problem.

8.4.17. Proposition. *If f, $g \in \mathcal{D}_a$, then*

$$D(fg)(a) = Df(a) \cdot g(a) + f(a) \cdot Dg(a)\,.$$

Proof. Problem. *Hint.* Use *Leibniz's rule* (Proposition 8.4.12) and Proposition 8.4.16.

8.4.18. Example. Let $r(t) = \frac{1}{t}$ for all $t \neq 0$. Then r is differentiable and $Dr(t) = -\frac{1}{t^2}$ for all $t \neq 0$.

Proof. Problem.

8.4.19. Proposition. *If $f \in \mathcal{D}_a$ and $g \in \mathcal{D}_{f(a)}$, then $g \circ f \in \mathcal{D}_a$ and*

$$D(g \circ f)(a) = (Dg)(f(a)) \cdot Df(a)\,.$$

Proof. Problem.

8.4.20. Proposition. *If f, $g \in \mathcal{D}_a$ and $g(a) \neq 0$, then*

$$D\left(\frac{f}{g}\right)(a) = \frac{g(a)\,Df(a) - f(a)\,Dg(a)}{(g(a))^2}\,.$$

Proof. Problem. *Hint.* Write $\frac{f}{g}$ as $f \cdot (r \circ g)$ and use Propositions 8.4.16 and 8.4.19 and Example 8.4.18.

8.4.21. Proposition. *If $f \in \mathcal{D}_a$ and $Df(a) > 0$, then there exists $r > 0$ such that*

(a) *$f(x) > f(a)$ whenever $a < x < a + r$, and*
(b) *$f(x) < f(a)$ whenever $a - r < x < a$.*

Proof. Problem. *Hint.* Define $g(h) = h^{-1}\,\Delta f_a(h)$ if $h \neq 0$ and $g(0) = Df(a)$. Use Proposition 7.2.3 to show that g is continuous at 0. Then apply Proposition 3.3.21.

8.4.22. Proposition. *If $f \in \mathcal{D}_a$ and $Df(a) < 0$, then there exists $r > 0$ such that*

(a) *$f(x) < f(a)$ whenever $a < x < a + r$, and*
(b) *$f(x) > f(a)$ whenever $a - r < x < a$.*

Proof. Problem. *Hint.* Of course it is possible to obtain this result by doing Proposition 8.4.21 again with some inequalities reversed. That is the hard way.

8.4.23. Definition. Let $f\colon A \to \mathbb{R}$ where $A \subseteq \mathbb{R}$. The function f has a LOCAL (or RELATIVE) MAXIMUM at a point $a \in A$ if there exists $r > 0$ such that $f(a) \geq f(x)$ whenever $|x - a| < r$ and $x \in \operatorname{dom} f$. It has a LOCAL (or RELATIVE) MINIMUM at a point $a \in A$ if there exists $r > 0$ such that $f(a) \leq f(x)$ whenever $|x - a| < r$ and $x \in \operatorname{dom} f$.

Recall from Chapter 6 that $f\colon A \to \mathbb{R}$ is said to attain a MAXIMUM at a if $f(a) \geq f(x)$ for all $x \in \operatorname{dom} f$. This is often called a GLOBAL (or ABSOLUTE) MAXIMUM to help distinguish it from the local version just defined. It is clear that every global maximum is also a local maximum but not vice versa. (Of course a similar remark holds for minima.)

8.4.24. Proposition. *If $f \in \mathcal{D}_a$ and f has either a local maximum or a local minimum at a, then $Df(a) = 0$.*

Proof. Problem. *Hint.* Use Propositions 8.4.21 and 8.4.22.

8.4.25. Proposition (Rolle's theorem)**.** *Let $a < b$. If $f\colon [a,b] \to \mathbb{R}$ is continuous, if it is differentiable on (a,b), and if $f(a) = f(b)$, then there exists a point c in (a,b) such that $Df(c) = 0$.*

Proof. Problem. *Hint.* Argue by contradiction. Use the extreme value theorem (Theorem 6.3.3) and Proposition 8.4.24.

8.4.26. Theorem (Mean value theorem)**.** *Let $a < b$. If $f\colon [a,b] \to \mathbb{R}$ is continuous and if it is differentiable on (a,b), then there exists a point c in (a,b) such that*

$$Df(c) = \frac{f(b) - f(a)}{b - a}\,.$$

Proof. Problem. *Hint.* Let $y = g(x)$ be the equation of the line that passes through the points $(a, f(a))$ and $(b, f(b))$. Show that the function $f - g$ satisfies the hypotheses of Rolle's theorem (Proposition 8.4.25)

8.4.27. Proposition. *Let J be an open interval in $\mathbb{R}$. If $f\colon J \to \mathbb{R}$ is differentiable and $Df(x) = 0$ for every $x \in J$, then f is constant on J.*

Proof. Problem. *Hint.* Use the mean value theorem (Theorem 8.4.26).

27.3.3. Exercise. Let $f\colon \mathbb{R}^4 \to \mathbb{R}^3 \colon (w,x,y,z) \mapsto (wxz, x^2+2y^2+3z^2, wy\arctan z)$, let $a = (1,1,1,1)$, and let $v = (0,2,-3,1)$.

(a) Find $[df_a]$.

(b) Find $df_a(v)$.

27.3.4. Problem. Let $f\colon \mathbb{R}^2 \to \mathbb{R}^2 \colon x \mapsto ({x_1}^2 - {x_2}^2, 3x_1x_2)$, and let $a = (2,1)$.

(a) Find $[df_a]$.

(b) Use part (a) to find $df_a(-1,3)$.

27.3.5. Problem. Let $f\colon \mathbb{R}^3 \to \mathbb{R}^4 \colon (x,y,z) \mapsto (xy, y - z^2, 2xz, y+3z)$, let $a = (1,-2,3)$, and let $v = (2,1,-1)$.

(a) Find $[df_a]$.

(b) Use part (a) to calculate $D_vf(a)$.

27.3.6. Problem. Let $f\colon \mathbb{R}^2 \to \mathbb{R}^2 \colon (x,y) \mapsto (x^2y, 2xy^2)$, let $a = (2,-1)$, and let $u = (\frac{3}{5}, \frac{4}{5})$. Compute $D_uf(a)$ in three ways:

(a) Use the definition of directional derivative.

(b) Use Proposition 25.5.2.

(c) Use Proposition 25.5.9.

27.3.7. Problem. Suppose that $f \in \mathcal{D}_a(\mathbb{R}^3, \mathbb{R}^4)$ and that the Jacobian matrix of f at a is

$$\begin{bmatrix} b & c & e \\ g & h & i \\ j & k & l \\ m & n & p \end{bmatrix}.$$

Find $f_1(a)$, $f_2(a)$, $f_3(a)$, $\nabla f^1(a)$, $\nabla f^2(a)$, $\nabla f^3(a)$, and $\nabla f^4(a)$.

27.3.8. Problem. Let $f \in \mathcal{D}_a(\mathbb{R}^n, \mathbb{R}^m)$ and $v \in \mathbb{R}^n$. Show that

(a) $df_a(v) = \sum_{j=1}^m \langle \nabla f^j(a), v\rangle e^j$, and

(b) $\|df_a\| \le \sum_{j=1}^m \|\nabla f^j(a)\|$.

27.4. The chain rule

In some respects it is convenient for scientists to work with variables rather than functions. Variables denote the physical quantities in which a scientist is ultimately interested. (In thermodynamics, for example, T is temperature, P pressure, S entropy, and so on.) Functions usually have no such standard associations. Furthermore, a problem that deals with only a small number of variables may turn out to involve a dauntingly large number of functions if they are specified. The simplification provided by the use of variables may, however, be more apparent than real, and the price paid in increased ambiguity for their suggestiveness is often substantial. Below are a few examples of ambiguities produced by the combined effects of excessive reliance on variables, inadequate (if conventional) notation, and the unfortunate mannerism of using the same name for a function and a dependent variable ("Suppose $x = x(s,t)\dots$").

(A) If $z = f(x,y)$, what does $\frac{\partial}{\partial x}z(y,x)$ mean? (Perhaps $f_1(y,x)$? Possibly $f_2(y,x)$?)

(B) If $z = f(x,t)$ where $x = x(t)$, then what is $\frac{\partial z}{\partial t}$? (Is it $f_2(t)$? Perhaps the derivative of $t \mapsto f(x(t),t)$ is intended?)

(C) Let $f(x,y)$ be a function of two variables. Does the expression $z = f(tx,ty)$ have three partial derivatives $\frac{\partial z}{\partial x}$, $\frac{\partial z}{\partial y}$, and $\frac{\partial z}{\partial t}$? Do $\frac{\partial z}{\partial x}$ and $\frac{\partial f}{\partial x}$ mean the same thing?

(D) Let $w = w(x,y,t)$ where $x = x(s,t)$ and $y = y(s,t)$. A direct application of the chain rule (as stated in most beginning calculus texts) produces

$$\frac{\partial w}{\partial t} = \frac{\partial w}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial w}{\partial y}\frac{\partial y}{\partial t} + \frac{\partial w}{\partial t}\,.$$

Is this correct? Do the terms of the form $\frac{\partial w}{\partial t}$ cancel?

(E) Let $z = f(x,y) = g(r,\theta)$, where $x = r\cos\theta$ and $y = r\sin\theta$. Do $\frac{\partial z}{\partial r}$ and $\frac{\partial z}{\partial \theta}$ make sense? Do $\frac{\partial z}{\partial x}$ and $\frac{\partial z}{\partial y}$? How about z_1 and z_2? Are any of these equal?

(F) The formulas for changing polar to rectangular coordinates are $x = r\cos\theta$ and $y = r\sin\theta$. So if we compute the partial derivative of the variable r with respect to the variable x, we get

$$\frac{\partial r}{\partial x} = \frac{\partial}{\partial x}\sqrt{x^2+y^2} = \frac{x}{\sqrt{x^2+y^2}} = \frac{x}{r} = \cos\theta\,.$$

On the other hand, since $r = \dfrac{x}{\cos\theta}$, we use the chain rule to get

$$\frac{\partial r}{\partial x} = \frac{1}{\cos\theta} = \sec\theta\,.$$

Is something wrong here? What?

The principal goal of the present section is to provide a reliable formalism for dealing with partial derivatives of functions of several variables in such a way that questions like (A)–(F) can be avoided. The basic strategy is quite simple: when in doubt give names to the relevant functions (especially composite ones!), and then use the chain rule. Perhaps it should be remarked that one need not make a fetish of avoiding variables. Many problems stated in terms of variables can be solved quite simply without the intrusion of the names of functions; e.g., what is $\frac{\partial z}{\partial x}$ if $z = x^3y^2$? This section is intended as a guide for the perplexed. Although its techniques are often useful in dissipating confusion generated by inadequate notation, it is neither necessary nor even particularly convenient to apply them routinely to every problem that arises. Let us start by writing the chain rule for functions between Euclidean spaces in terms of partial derivatives. Suppose that $f \in \mathcal{D}_a(\mathbb{R}^p,\mathbb{R}^n)$ and $g \in \mathcal{D}_{f(a)}(\mathbb{R}^n,\mathbb{R}^m)$. Then according to Proposition 25.3.17

$$d(g\circ f)_a = dg_{f(a)} \circ df_a\,.$$

Replacing these linear transformations by their matrix representations and using Proposition 21.5.12, we obtain

$$\big[d(g \circ f)_a\big] = \big[dg_{f(a)}\big]\big[df_a\big]\,. \tag{47}$$

27.4.1. Proposition. *If $f \in \mathcal{D}_a(\mathbb{R}^p, \mathbb{R}^n)$ and $g \in \mathcal{D}_{f(a)}(\mathbb{R}^n, \mathbb{R}^m)$, then $\big[d(g\circ f)_a\big]$ is the $m \times p$ matrix whose entry in the jth row and kth column is $\sum_{i=1}^n g_i^j(f(a))f_k^i(a)$. That is,*

$$(g \circ f)_k^j(a) = \sum_{i=1}^n \big(g_i^j \circ f\big)(a) f_k^i(a)$$

for $1 \le j \le m$ and $1 \le k \le p$.

Proof. Multiply the two matrices on the right-hand side of (47) and use Proposition 27.3.2. □

It is occasionally useful to restate Proposition 27.4.1 in the following (clearly equivalent) way.

27.4.2. Corollary. *If $f \in \mathcal{D}_a(\mathbb{R}^p, \mathbb{R}^n)$ and $g \in \mathcal{D}_{f(a)}(\mathbb{R}^n, \mathbb{R}^m)$, then*

$$\big[d(g \circ f)_a\big] = \big[\langle \nabla g^j(f(a)), f_k(a)\rangle\big]_{j=1}^{m}\,{}_{k=1}^{p}\,.$$

27.4.3. Exercise. This is an exercise in translation of notation. Suppose $y = y(u, v, w, x)$ and $z = z(u, v, w, x)$ where $u = u(s, t)$, $v = v(s, t)$, $w = w(s, t)$, and $x = x(s, t)$. Show that (under suitable hypotheses)

$$\frac{\partial z}{\partial t} = \frac{\partial z}{\partial u}\frac{\partial u}{\partial t} + \frac{\partial z}{\partial v}\frac{\partial v}{\partial t} + \frac{\partial z}{\partial w}\frac{\partial w}{\partial t} + \frac{\partial z}{\partial x}\frac{\partial x}{\partial t}\,.$$

27.4.4. Problem. Suppose that the variables x, y, and z are differentiable functions of the variables α, β, γ, δ, and ϵ, that in turn depend in a differentiable fashion on the variables r, s, and t. As in Exercise 27.4.3 use Proposition 27.4.1 to write $\frac{\partial z}{\partial r}$ in terms of quantities such as $\frac{\partial z}{\partial \alpha}$, $\frac{\partial \delta}{\partial r}$, etc.

27.4.5. Exercise. Let $f(x, y, z) = (xy^2, 3x - z^2, xyz, x^2 + y^2, 4xz + 5)$, $g(s, t, u, v, w) = (s^2 + u^2 + v^2, s^2 v - 2tw^2)$, and $a = (1, 0, -1)$. Use the chain rule to find $\big[d(g \circ f)_a\big]$.

27.4.6. Problem. Let $f(x, y, z) = (x^3 y^2 \sin z, x^2 + y \cos z)$, $g(u, v) = (\sqrt{u}v, v^3)$, $k = g \circ f$, $a = (1, -2, \pi/2)$, and $h = (1, -1, 2)$. Use the chain rule to find $dk_a(h)$.

27.4.7. Problem. Let $f(x, y, z) = (x^2 y + y^2 z, xyz)$, $g(x, y) = (x^2 y, 3xy, x - 2y, x^2 + 3)$, and $a = (1, -1, 2)$. Use the chain rule to find $\big[d(g \circ f)_a\big]$.

We now consider a slightly more complicated problem. Suppose that $w = w(x, y, t)$ where $x = x(s, t)$ and $y = y(s, t)$ and that all the functions mentioned are differentiable. (This is problem (D) at the beginning of this section.) It is perhaps tempting to write

$$\begin{aligned} \frac{\partial w}{\partial t} &= \frac{\partial w}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial w}{\partial y}\frac{\partial y}{\partial t} + \frac{\partial w}{\partial t}\frac{\partial t}{\partial t} \\ &= \frac{\partial w}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial w}{\partial y}\frac{\partial y}{\partial t} + \frac{\partial w}{\partial t} \end{aligned} \tag{48}$$

(since $\frac{\partial t}{\partial t} = 1$). The trouble with this is that the $\frac{\partial w}{\partial t}$ on the left-hand side is not the same as the one on the right-hand side. The $\frac{\partial w}{\partial t}$ on the right-hand side refers only to the rate of change of w with respect to t insofar as t appears *explicitly* in the formula for w; the one on the left-hand side takes into account the fact that, in addition, w depends *implicitly* on t *via* the variables x and y. What to do? Use functions. Relate the variables by functions as follows.

$$\begin{array}{ccccc} & & x & & \\ s & \xrightarrow{\ f\ } & y & \xrightarrow{\ g\ } & w \\ t & & t & & \end{array} \tag{49}$$

Also let $h = g \circ f$. Notice that $f^3(s,t) = t$. Then according to the chain rule,

$$h_2 = \sum_{k=1}^{3} (g_k \circ f)\, f_2^k\,.$$

But $f_2^3 = 1$ (that is, $\frac{\partial t}{\partial t} = 1$). So

$$h_2 = (g_1 \circ f) f_2^1 + (g_2 \circ f) f_2^2 + g_3 \circ f. \tag{50}$$

The ambiguity of (48) has been eliminated in (50). The $\frac{\partial w}{\partial t}$ on the left-hand side is seen to be the derivative with respect to t of the composite $h = g \circ f$, whereas the $\frac{\partial w}{\partial t}$ on the right-hand side is just the derivative with respect to t of the function g.

One last point. Many scientific workers adamantly refuse to give names to functions. What do they do? Look back at diagram (49) and remove the names of the functions.

$$\begin{array}{ccccc} & & x & & \\ s & \longrightarrow & y & \longrightarrow & w \\ t & & t & & \end{array} \tag{51}$$

The problem is that the symbol "t" occurs twice. To specify differentiation of the composite function (our h) with respect to t, indicate that the "t" you are interested in is the one in the left column of (51). This may be done by listing everything else that appears in that column. That is, specify which variables are held constant. This specification conventionally appears as a subscript outside parentheses. Thus, the $\frac{\partial w}{\partial t}$ on the left-hand side of (48) (our h_2) is written as $\left(\frac{\partial w}{\partial t}\right)_s$ (and it is read "$\frac{\partial w}{\partial t}$ with s held constant"). Similarly, the $\frac{\partial w}{\partial t}$ on the right-hand side of (48)) (our g_3) involves differentiation with respect to t while x and y are fixed. So it is written $\left(\frac{\partial w}{\partial t}\right)_{x,y}$ (and it is read "$\frac{\partial w}{\partial t}$ with x and y held constant"). Thus, (48) becomes

$$\left(\frac{\partial w}{\partial t}\right)_s = \frac{\partial w}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial w}{\partial y}\frac{\partial y}{\partial t} + \left(\frac{\partial w}{\partial t}\right)_{x,y}. \tag{52}$$

It is not necessary to write, for example, an expression such as $\left(\frac{\partial w}{\partial x}\right)_{t,y}$ because there is no ambiguity; the symbol "x" occurs only once in (51). If you choose to use the convention just presented, it is best to use it only to avoid confusion; use it because you *must*, not because you *can*.

27.4.8. Exercise. Let $w = t^3 + 2yx^{-1}$ where $x = s^2 + t^2$ and $y = s \arctan t$. Use the chain rule to find $\left(\frac{\partial w}{\partial t}\right)_s$ at the point where $s = t = 1$.

We conclude this section with two more exercises on the use of the chain rule. Part of the difficulty here and in the problems at the end of the section is to interpret correctly what the problem says. The suggested solutions may seem longwinded, and they are. Nevertheless, these techniques prove valuable in situations complicated enough to be confusing. With practice, it is easy to do many of the indicated steps mentally.

27.4.9. Exercise. Show that if $z = xy + x\phi(yx^{-1})$, then $x\frac{\partial z}{\partial x} + y\frac{\partial z}{\partial y} = xy + z$. *Hint.* Start by restating the exercise in terms of functions. Add suitable hypotheses. In particular, suppose that $\phi\colon \mathbb{R} \to \mathbb{R}$ is differentiable. Let

$$j(x, y) = xy + x\phi(yx^{-1})$$

for x, $y \in \mathbb{R}$, $x \neq 0$. Then for each such x and y

$$xj_1(x, y) + yj_2(x, y) = xy + j(x, y)\,. \tag{53}$$

To prove this assertion proceed as follows.

(a) Let $g(x, y) = yx^{-1}$. Find $\left[dg_{(x,y)}\right]$.
(b) Find $\left[d(\phi \circ g)_{(x,y)}\right]$.
(c) Let $G(x, y) = \big(x, \phi(yx^{-1})\big)$. Use (b) to find $\left[dG_{(x,y)}\right]$.
(d) Let $m(x, y) = xy$. Find $\left[dm_{(x,y)}\right]$.
(e) Let $h(x, y) = x\phi(yx^{-1})$. Use (c) and (d) to find $\left[dh_{(x,y)}\right]$.
(f) Use (d) and (e) to find $\left[dj_{(x,y)}\right]$.
(g) Use (f) to prove (53).

27.4.10. Exercise. Show that if $f(u, v) = g(x, y)$ where f is a differentiable real valued function, $u = x^2 - y^2$, and $v = 2xy$, then

$$y\frac{\partial g}{\partial x} - x\frac{\partial g}{\partial y} = 2v\frac{\partial f}{\partial u} - 2u\frac{\partial f}{\partial v}. \tag{54}$$

Hint. The equations $u = x^2 - y^2$ and $v = 2xy$ give u and v in terms of x and y. Think of the function $h\colon (x, y) \mapsto (u, v)$ as a change of variables in $\mathbb{R}^2$. That is, define

$$h\colon \mathbb{R}^2 \to \mathbb{R}^2\colon (x, y) \mapsto (x^2 - y^2, 2xy)\,.$$

Then on the uv-plane (that is, the codomain of h) the function f is real valued and differentiable. The equation $f(u, v) = g(x, y)$ serves only to fix notation. It indicates that g is the composite function $f \circ h$. We may visualize the situation thus:

$$\begin{matrix} x \\ y \end{matrix} \xrightarrow{\ h\ } \begin{matrix} u \\ v \end{matrix} \xrightarrow{\ f\ } w \tag{55}$$

where $g = f \circ h$.

Now what are we trying to prove? The conclusion (54) is clear enough if we evaluate the partial derivatives at the right place. Recalling that we have defined h so that $u = h^1(x,y)$ and $v = h^2(x,y)$, we may write (54) in the following form:

$$yg_1(x,y) - xg_2(x,y) = 2h^2(x,y)f_1(h(x,y)) - 2h^1(x,y)f_2(h(x,y))\,. \tag{56}$$

Alternatively, we may write

$$\pi_2 g_1 - \pi_1 g_2 = 2h^2 f_1 - 2h^1 f_2,$$

where π_1 and π_2 are the usual coordinate projections. To verify (56), use the chain rule to find $\left[dg_{(x,y)}\right]$.

27.4.11. Problem. Let $w = \frac{1}{2}x^2y + \arctan(tx)$ where $x = t^2 - 3u^2$ and $y = 2tu$. Find $\left(\frac{\partial w}{\partial t}\right)_u$ when $t = 2$ and $u = -1$.

27.4.12. Problem. Let $z = \frac{1}{16}uw^2xy$ where $w = t^2 - u^2 + v^2$, $x = 2tu + tv$, and $y = 3uv$. Find $\left(\frac{\partial z}{\partial u}\right)_{t,v}$ when $t = 1$, $u = -1$, and $v = -2$.

27.4.13. Problem. If $z = f\left(\frac{x-y}{y}\right)$, then $x\frac{\partial z}{\partial x} + y\frac{\partial z}{\partial y} = 0$. State this precisely and prove it.

27.4.14. Problem. If ϕ is a differentiable function on an open subset of $\mathbb{R}^2$ and $w = \phi(u^2 - t^2, t^2 - u^2)$, then $t\frac{\partial w}{\partial u} + u\frac{\partial w}{\partial t} = 0$. *Hint.* Let $h(t,u) = (u^2 - t^2, t^2 - u^2)$ and $w = \psi(t,u)$ where $\psi = \phi \circ h$. Compute $\left[dh_{(t,u)}\right]$. Use the chain rule to find $\left[d\psi_{(t,u)}\right]$. Then simplify $t\psi_2(t,u) + u\psi_1(t,u)$.

27.4.15. Problem. If $f(u,v) = g(x,y)$ where f is a differentiable real valued function on $\mathbb{R}^2$ and if $u = x^3 + y^3$ and $v = xy$, then

$$x\frac{\partial g}{\partial x} + y\frac{\partial g}{\partial y} = 3u\frac{\partial f}{\partial u} + 2v\frac{\partial f}{\partial v}\,.$$

27.4.16. Problem. Let $f(x,y) = g(r,\theta)$ where (x,y) are Cartesian coordinates and (r,θ) are polar coordinates in the plane. Suppose that f is differentiable at all (x,y) in $\mathbb{R}^2$.

(a) Show that except at the origin

$$\frac{\partial f}{\partial x} = (\cos\theta)\frac{\partial g}{\partial r} - \frac{1}{r}(\sin\theta)\frac{\partial g}{\partial \theta}\,.$$

(b) Find a similar expression for $\frac{\partial f}{\partial y}$.

Hint. Recall that Cartesian and polar coordinates are related by $x = r\cos\theta$ and $y = r\sin\theta$.

27.4.17. Problem. Let n be a fixed positive integer. A function $f\colon \mathbb{R}^2 \to \mathbb{R}^2$ is HOMOGENEOUS OF DEGREE n if $f(tx,ty) = t^n f(x,y)$ for all t, x, $y \in \mathbb{R}$. If such a function f is differentiable, then

$$x\frac{\partial f}{\partial x} + y\frac{\partial f}{\partial y} = nf\,. \tag{57}$$

Hint. Try the following:

(a) Let $G(x,y,t) = (tx,ty)$. Find $\left[dG_{(t,x,y)}\right]$.

(b) Let $h = f \circ G$. Find $\left[dh_{(x,y,t)}\right]$.

(c) Let $H(x, y, t) = (t^n, f(x, y))$. Find $\left[dH_{(x,y,t)}\right]$.

(d) Let $k = m \circ H$ (where $m(u, v) = uv$). Find $\left[dk_{(x,y,t)}\right]$.

(e) By hypothesis $h = k$; so the answers to (b) and (d) must be the same. Use this fact to derive (57).

Quantifiers

Certainly, $2+2=4$ and $2+2=5$ are statements—one true, the other false. On the other hand the appearance of the variable x prevents the expression $x+2=5$ from being a statement. Such an expression we will call an OPEN SENTENCE; its truth is open to question since x is unidentified. There are three standard ways of converting open sentences into statements.

The first, and simplest, of these is to give the variable a particular value. If we "evaluate" the expression $x+2=5$ at $x=4$, we obtain the (false) statement $4+2=5$.

A second way of obtaining a statement from an expression involving a variable is UNIVERSAL QUANTIFICATION: we assert that the expression is true for all values of the variable. In the preceding example we get, "For all x, $x+2=5$". This is now a statement (and again false). The expression "for all x" (or equivalently, "for every x") is often denoted symbolically by $(\forall x)$. Thus the preceding sentence may be written $(\forall x)x+2=5$. (The parentheses are optional; they may be used in the interest of clarity.) We call $\forall$ a UNIVERSAL QUANTIFIER.

Frequently, there are several variables in an expression. They may all be universally quantified. For example

$$(\forall x)(\forall y)\, x^2-y^2=(x-y)(x+y) \tag{83}$$

is a (true) statement, which says that for every x and for every y the expression x^2-y^2 factors in the familiar way. The order of consecutive universal quantifiers is unimportant: the statement

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

says exactly the same thing as (83). For this reason the notation may be contracted slightly to read

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

A third way of obtaining a statement from an open sentence $P(x)$ is EXISTENTIAL QUANTIFICATION. Here we assert that $P(x)$ is true for *at least one* value of x.

This is often written "$(\exists x)$ such that $P(x)$" or more briefly "$(\exists x)P(x)$", and is read "there exists an x such that $P(x)$" or "$P(x)$ is true for some x". For example, if we existentially quantify the expression "$x + 2 = 5$", we obtain "$(\exists x)$ such that $x + 2 = 5$" (which happens to be true). We call $\exists$ an EXISTENTIAL QUANTIFIER.

As is true for universal quantifiers, the order of consecutive existential quantifiers is immaterial.

CAUTION. It is absolutely essential to realize that the order of an existential and a universal quantifier may *not* in general be reversed. For example,

$$(\exists x)(\forall y)\, x < y$$

says that there is a number x with the property that no matter how y is chosen, x is less than y; that is, there is a smallest real number. (This is, of course, false.) On the other hand

$$(\forall y)(\exists x)\, x < y$$

says that for every y we can find a number x smaller than y. (This is true: take x to be $y - 1$ for example.) *The importance of getting quantifiers in the right order cannot be overemphasized.*

There is one frequently used convention concerning quantifiers that should be mentioned. In the statement of definitions, propositions, theorems, etc., missing quantifiers are assumed to be universal; furthermore, they are assumed to be the innermost quantifiers.

A.1.1. Example. Let f be a real valued function defined on the real line $\mathbb{R}$. Many texts give the following definition. The function f is *continuous* at a point a in $\mathbb{R}$ if for every $\epsilon > 0$ there exists $\delta > 0$ such that

$$|f(x) - f(a)| < \epsilon \quad \text{whenever } |x - a| < \delta\,.$$

Here ϵ and δ are quantified; the function f and the point a are fixed for the discussion, so they do not require quantifiers. What about x? According to the convention just mentioned, x is universally quantified and that quantifier is the innermost one. Thus the definition reads, for every $\epsilon > 0$ there exists $\delta > 0$ such that for every x

$$|f(x) - f(a)| < \epsilon \quad \text{whenever } |x - a| < \delta\,.$$

A.1.2. Example. Sometimes all quantifiers are missing. In this case the preceding convention dictates that all variables are universally quantified. Thus

$$\text{Theorem.} \quad x^2 - y^2 = (x - y)(x + y)$$

is interpreted to mean

$$\text{Theorem.} \quad (\forall x)(\forall y)\, x^2 - y^2 = (x - y)(x + y)\,.$$

Sets

In this text everything is defined ultimately in terms of two primitive (that is, undefined) concepts: set and set membership. We assume that these are already familiar to the reader. In particular, it is assumed to be understood that distinct elements (or members, or points) can be regarded collectively as a single set (or family, or class, or collection). To indicate that x belongs to a set A (or that x is a member of A), we write $x \in A$; to indicate that it does not belong to A, we write $x \notin A$.

We specify a set by listing its members between braces (for instance, $\{1, 2, 3, 4, 5\}$ is the set of the first five natural numbers), by listing some of its members between braces with ellipses (three dots) indicating the missing members (e.g., $\{1, 2, 3, \dots\}$ is the set of all natural numbers), or by writing $\{x \colon P(x)\}$ where $P(x)$ is an open sentence that specifies what property the variable x must satisfy in order to be included in the set (e.g., $\{x \colon 0 \le x \le 1\}$ is the closed unit interval $[0, 1]$).

B.1.1. Problem. Let $\mathbb{N}$ be the set of natural numbers 1, 2, 3, ..., and let

$$S = \{x \colon x < 30 \text{ and } x = n^2 \text{ for some } n \in \mathbb{N}\}.$$

List all the elements of S.

B.1.2. Problem. Let $\mathbb{N}$ be the set of natural numbers 1, 2, 3, ..., and let

$$S = \{x \colon x = n + 2 \text{ for some } n \in \mathbb{N} \text{ such that } n < 6\}.$$

List all the elements of S.

B.1.3. Problem. Suppose that

$$S = \{x \colon x = n^2 + 2 \text{ for some } n \in \mathbb{N}\}$$

and that

$$T = \{3, 6, 11, 18, 27, 33, 38, 51\}.$$

(a) Find an element of S that is not in T.

(b) Find an element of T that is not in S.

B.1.4. Problem. Suppose that

$$S = \{x : x = n^2 + 2n \text{ for some } n \in \mathbb{N}\}$$

and that

$$T = \{x : x = 5n - 1 \text{ for some } n \in \mathbb{N}\}\,.$$

Find an element that belongs to both S and T.

CAUTION. How many elements are in the set $\{3, 7, 4, 7, 4, 1, 3, 1, 7\}$? Exactly four. Because

$$\{3, 7, 4, 7, 4, 1, 3, 1, 7\} = \{1, 3, 4, 7\} = \{4, 7, 3, 1\}.$$

Repetitions don't increase the size of a set. The order in which the elements are written does not matter.

It is prudent to give some thought to the possibility of sets having repeated elements. This occurs quite naturally in many contexts. The fundamental theorem of algebra, for example, says *every nonzero, single-variable, polynomial of degree n with complex coefficients has, counting multiplicity, exactly n complex roots.* Now let's look at an example: Consider the polynomial $p(x) = x^3 - x^2$. According to the preceding theorem, it is reasonable to write, "Let $S = \{r_1, r_2, r_3\}$ be the set of roots of p". However, when we factor p, we get $(x - 0)(x - 0)(x - 1)$, so that $r_1 = r_2 = 0$ and $r_3 = 1$. (A root r of a polynomial p has MULTIPLICITY m if the factor $(x - r)$ occurs exactly m times in the factorization of p.) So, despite its notation, the set S has only two elements, not three. There are only two *distinct* roots: the root 0 has multiplicity two, and the root 1 has multiplicity 1. To avoid consideration of uninteresting special cases, the following convention is frequently made.

B.1.5. Convention. When a set is *defined* by listing its elements, it is assumed that the elements are unique. For example, "Let $S = \{a, b, c\}$" is intended as a short version of, "Let $S = \{a, b, c\}$ be a set of three elements" or "Let $S = \{a, b, c\}$, where a, b, and c are distinct".

Since all of our subsequent work depends on the notions of set and of set membership, it is not altogether satisfactory to rely on intuition and shared understanding to provide a foundation for these crucial concepts. It is possible in fact to arrive at paradoxes using a naive approach to sets. (For example, ask yourself the question, "If S is the set of all sets that do not contain themselves as members, then does S belong to S?" If the answer is "yes", then it must be "no", and vice versa.) One satisfactory alternative to our intuitive approach to sets is axiomatic set theory. There are many ways of axiomatizing set theory to provide a secure foundation for subsequent mathematical development. Unfortunately, each of these ways turns out to be extremely intricate, and it is generally felt that the abstract complexities of axiomatic set theory do not serve well as a beginning to an introductory course in advanced calculus.

Most of the paradoxes inherent in an intuitive approach to sets have to do with sets that are too "large". For example, the set S mentioned in the preceding paragraph is enormous. Thus in what follows we will assume that in each situation all the mathematical objects we are then considering (sets, functions, etc.) belong to some appropriate "universal" set that is "small" enough to avoid set theoretic paradoxes. (Think of "universal" in terms of "universe of discourse", not "all-encompassing".) In many cases an appropriate universal set is clear from the context. Previously, we considered a statement

$$(\forall y)(\exists x)\, x < y\,.$$

The appearance of the symbol "$<$" suggests to most readers that x and y are real numbers. Thus the universal set from which the variables are chosen is the set $\mathbb{R}$ of all real numbers. When there is doubt that the universal set will be properly interpreted, it may be specified. In the example just mentioned, we might write

$$(\forall y \in \mathbb{R})(\exists x \in \mathbb{R})\, x < y.$$

This makes explicit the intended restriction that x and y be real numbers.

As another example recall that in Appendix A, we defined a real valued function f to be continuous at a point $a \in \mathbb{R}$ if

$$(\forall \epsilon > 0)(\exists \delta > 0) \text{ such that } (\forall x)\, |f(x) - f(a)| < \epsilon \text{ whenever } |x - a| < \delta.$$

Here the first two variables, ϵ and δ, are restricted to lie in the open interval $(0, \infty)$. Thus we might rewrite the definition as

$$\forall \epsilon \in (0,\infty)\ \exists \delta \in (0,\infty)$$
$$\text{such that } (\forall x \in \mathbb{R})\, |f(x) - f(a)| < \epsilon \text{ whenever } |x - a| < \delta.$$

The expressions $\forall x \in \mathbb{R}$, $\exists \delta \in (0, \infty)$, etc., are called RESTRICTED QUANTIFIERS.

B.1.6. Definition. Let S and T be sets. We say that S is a SUBSET of T and write $S \subseteq T$ (or $T \supseteq S$) if every member of S belongs to T. If $S \subseteq T$ we also say that S is CONTAINED IN T or that T CONTAINS S. Notice that the relation $\subseteq$ is REFLEXIVE (that is, $S \subseteq S$ for all S) and TRANSITIVE (that is, if $S \subseteq T$ and $T \subseteq U$, then $S \subseteq U$). It is also ANTISYMMETRIC (that is, if $S \subseteq T$ and $T \subseteq S$, then $S = T$). If we wish to claim that S is *not* a subset of T, we may write $S \not\subseteq T$. In this case there is at least one member of S that does not belong to T.

B.1.7. Example. Since every number in the closed interval $[0, 1]$ also belongs to the interval $[0, 5]$, it is correct to write $[0, 1] \subseteq [0, 5]$. Since the number π belongs to $[0, 5]$ but not to $[0, 1]$, we may also write $[0, 5] \not\subseteq [0, 1]$.

B.1.8. Definition. If $S \subseteq T$ but $S \neq T$, then we say that S is a PROPER SUBSET of T (or that S is PROPERLY CONTAINED IN T, or that T PROPERLY CONTAINS S) and write $S \subsetneqq T$.

B.1.9. Problem. Suppose that $S = \{x \colon x = 2n + 3 \text{ for some } n \in \mathbb{N}\}$ and that T is the set of all odd natural numbers 1, 3, 5, ...

(a) Is $S \subseteq T$? If not, find an element of S that does not belong to T.

(b) Is $T \subseteq S$? If not, find an element of T that does not belong to S.

B.1.10. Definition. The EMPTY SET (or NULL SET), which is denoted by $\emptyset$, is defined to be the set that has no elements. (Or, if you like, define it to be $\{x \colon x \neq x\}$.) It is regarded as a subset of every set, so that $\emptyset \subseteq S$ is always true. (*Note.* "$\emptyset$" is a letter of the Danish alphabet, not the Greek letter "phi".)

B.1.11. Definition. If S is a set, then the POWER SET of S, which we denote by $\mathfrak{P}(S)$, is the set of all subsets of S.

B.1.12. Example. Let $S = \{a, b, c\}$. (See Convention B.1.5.) Then the members of the power set of S are the empty set, the three one-element subsets, the three two-element subsets, and the set S itself. That is,

$$\mathfrak{P}(S) = \big\{\emptyset, \{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}, S\big\}\,.$$

B.1.13. Problem. In each of the words (a)–(d) below let S be the set of letters in the word. In each case find the number of members of S and the number of members of $\mathfrak{P}(S)$, the power set of S.

(a) lull

(b) appall

(c) attract

(d) calculus

CAUTION. In attempting to prove a theorem that has as a hypothesis "Let S be a set" do not include in your proof something like "Suppose $S = \{s_1, s_2, \dots, s_n\}$" or "Suppose $S = \{s_1, s_2, \dots\}$". In the first case you are tacitly assuming that S is finite and in the second that it is countable. Neither is justified by the hypothesis.

CAUTION. A single letter $\mathfrak{S}$ (an S in fraktur font) is an acceptable symbol in printed documents. Don't try to imitate it in handwritten work or on the blackboard. Use script letters instead.

Finally, a word on the use of the symbols $=$ and $:=$. In this text equality is used in the sense of identity. We write $x = y$ to indicate that x and y are two names for the same object. For example, $0.5 = 1/2 = 3/6 = 1/\sqrt{4}$ because 0.5, 1/2, 3/6, and $1/\sqrt{4}$ are different names for the same real number. You have probably encountered other uses of the term equality. In many high school geometry texts, for example, one finds statements to the effect that a triangle is isosceles if it has two equal sides (or two equal angles). What is meant of course is that a triangle is isosceles if it has two sides of *equal length* (or two angles of *equal angular measure*). We also make occasional use of the symbol $:=$ to indicate *equality by definition*. Thus when we write $a := b$ we are giving a new name a to an object b with which we are presumably already familiar.