CHAPTER 4 GRAPHS

Mention the real numbers to a mathematician and the image of a straight line will probably form in her mind, quite involuntarily. And most likely she will neither banish nor too eagerly embrace this mental picture of the real numbers. "Geometric intuition" will allow her to interpret statements about numbers in terms of this picture, and may even suggest methods of proving them. Although the properties of the real numbers which were studied in Part I are not greatly illuminated by a geometric picture, such an interpretation will be a great aid in Part II.

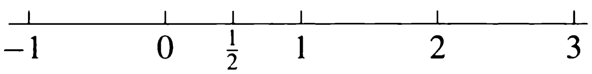


FIGURE 1

You are probably already familiar with the conventional method of considering the straight line as a picture of the real numbers, i.e., of associating to each real number a point on a line. To do this (Figure 1) we pick, arbitrarily, a point which we label 0, and a point to the right, which we label 1. The point twice as far to the right is labeled 2, the point the same distance from 0 to 1, but to the left of 0, is labeled -1, etc. With this arrangement, if $a < b$, then the point corresponding to a lies to the left of the point corresponding to b. We can also draw rational numbers, such as $\frac{1}{2}$, in the obvious way. It is usually taken for granted that the irrational numbers also somehow fit into this scheme, so that every real number can be drawn as a point on the line. We will not make too much fuss about justifying this assumption, since this method of "drawing" numbers is intended solely as a method of picturing certain abstract ideas, and our proofs will never rely on these pictures (although we will frequently use a picture to suggest or help explain a proof). Because this geometric picture plays such a prominent, albeit inessential role, geometric terminology is frequently employed when speaking of numbers—thus a number is sometimes called a *point*, and $\mathbf{R}$ is often called the *real line*.

The number $|a-b|$ has a simple interpretation in terms of this geometric picture: it is the distance between a and b, the length of the line segment which has a as one end point and b as the other. This means, to choose an example whose frequent occurrence justifies special consideration, that the set of numbers x which satisfy $|x - a| < \varepsilon$ may be pictured as the collection of points whose distance from a is less than ε. This set of points is the "interval" from $a - \varepsilon$ to $a + \varepsilon$, which may also be described as the points corresponding to numbers x with $a - \varepsilon < x < a + \varepsilon$ (Figure 2).

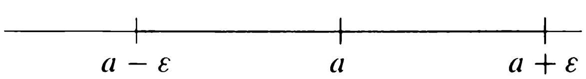


FIGURE 2

Sets of numbers which correspond to intervals arise so frequently that it is desirable to have special names for them. The set $\{x : a < x < b\}$ is denoted by (a, b) and called the **open interval** from a to b. This notation naturally creates some ambiguity, since (a, b) is also used to denote a pair of numbers, but in context it is always clear (or can easily be made clear) whether one is talking about a pair or an interval. Note that if $a \geq b$, then $(a, b) = \emptyset$, the set with no elements; in practice, however, it is almost always assumed (explicitly if one has been careful, and implicitly otherwise), that whenever an interval (a, b) is mentioned, the number a is less than b.

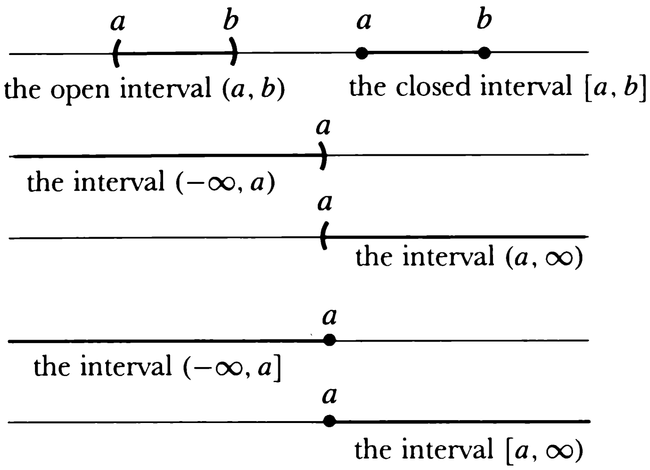


FIGURE 3

The set $\{x : a \leq x \leq b\}$ is denoted by $[a, b]$ and is called the **closed interval** from a to b. This symbol is usually reserved for the case $a < b$, but it is sometimes used for $a = b$, also. The usual pictures for the intervals (a, b) and $[a, b]$ are shown in Figure 3; since no reasonably accurate picture could ever indicate the difference between the two intervals, various conventions have been adopted. Figure 3 also shows certain "infinite" intervals. The set $\{x : x > a\}$ is denoted by (a, ∞), while the set $\{x : x \geq a\}$ is denoted by $[a, \infty)$; the sets $(-\infty, a)$ and $(-\infty, a]$ are defined similarly. At this point a standard warning must be issued: the symbols ∞ and $-\infty$, though usually read "infinity" and "minus infinity," are *purely* suggestive; there is no number "∞" which satisfies $\infty \geq a$ for all numbers a. While the symbols ∞ and $-\infty$ will appear in many contexts, it is always necessary to define these uses in ways that refer only to numbers. The set **R** of all real numbers is also considered to be an "interval," and is sometimes denoted by $(-\infty, \infty)$.

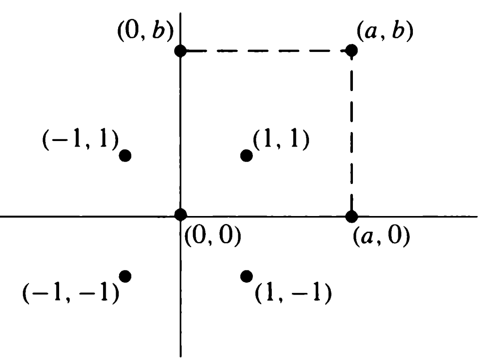


FIGURE 4

Of even greater interest to us than the method of drawing numbers is a method of drawing pairs of numbers. This procedure, probably also familiar to you, requires a "coordinate system," two straight lines intersecting at right angles. To distinguish these straight lines, we call one the *horizontal axis*, and one the *vertical axis*. (More prosaic terminology, such as the "first" and "second" axes, is probably preferable from a logical point of view, but most people hold their books, or at least their blackboards, in the same way, so that "horizontal" and "vertical" are more descriptive.) Each of the two axes could be labeled with real numbers, but we can also label points on the horizontal axis with pairs $(a, 0)$ and points on the vertical axis with pairs $(0, b)$, so that the intersection of the two axes, the "origin" of the coordinate system, is labeled $(0, 0)$. Any pair (a, b) can now be drawn as in Figure 4, lying at the vertex of the rectangle whose other three vertices are labeled $(0, 0)$, $(a, 0)$, and $(0, b)$. The numbers a and b are called the *first* and *second coordinates*, respectively, of the point determined in this way.

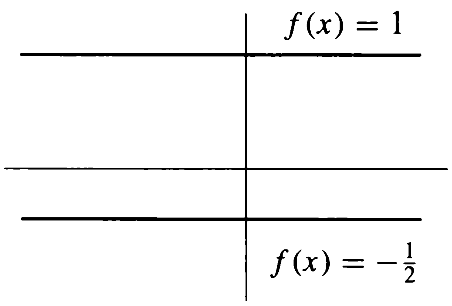


FIGURE 5

Our real concern, let us recall, is a method of drawing functions. Since a function is just a collection of pairs of numbers, we can draw a function by drawing each of the pairs in the function. The drawing obtained in this way is called the **graph** of the function. In other words, the graph of f contains all the points corresponding to pairs $(x, f(x))$. Since most functions contain infinitely many pairs, drawing the graph promises to be a laborious undertaking, but, in fact, many functions have graphs which are quite easy to draw.

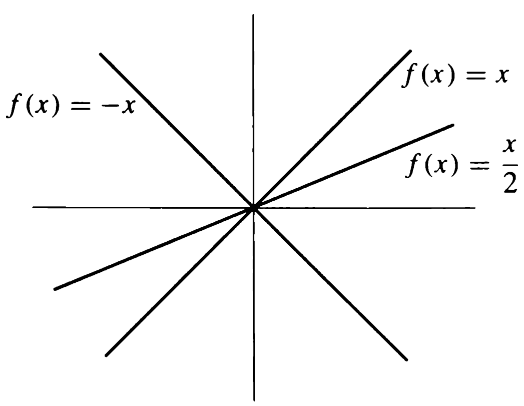


FIGURE 6

Not surprisingly, the simplest functions of all, the constant functions $f(x) = c$, have the simplest graphs. It is easy to see that the graph of the function $f(x) = c$ is a straight line parallel to the horizontal axis, at distance c from it (Figure 5).

The functions $f(x) = cx$ also have particularly simple graphs—straight lines through $(0, 0)$, as in Figure 6. A proof of this fact is indicated in Figure 7:

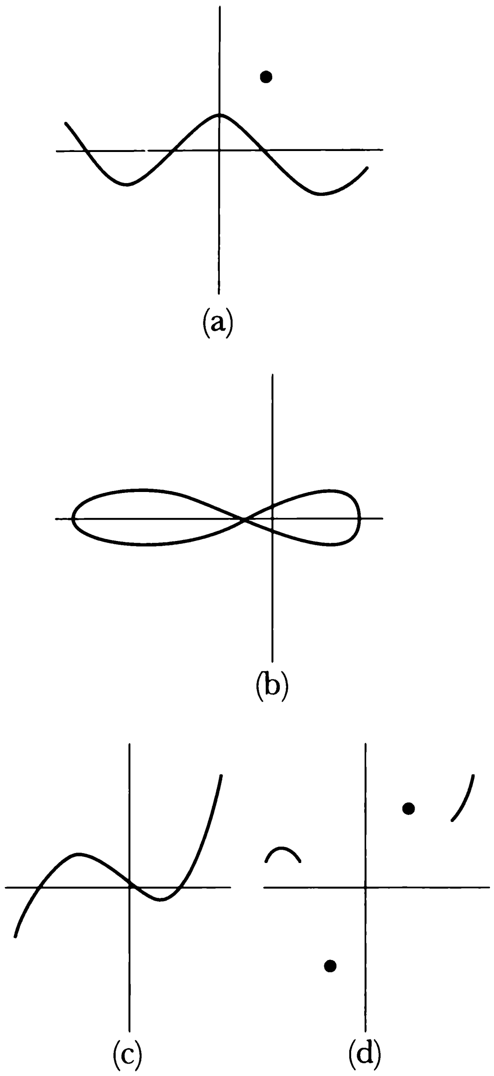


FIGURE 10

f is to be of the form $f(x) = \alpha x + \beta$, then we must have

$$\alpha a + \beta = b,$$
$$\alpha c + \beta = d;$$

therefore $\alpha = (d-b)/(c-a)$ and $\beta = b - [(d-b)/(c-a)]a$, so

$$f(x) = \frac{d-b}{c-a}x + b - \frac{d-b}{c-a}a = \frac{d-b}{c-a}(x-a) + b,$$

a formula most easily remembered by using the "point-slope form" (see Problem 6).

Of course, this solution is possible only if $a \neq c$; the graphs of linear functions account only for the straight lines which are not parallel to the vertical axis. The vertical straight lines are not the graph of *any* function at all; in fact, the graph of a function can never contain even two distinct points on the same vertical line. This conclusion is immediate from the definition of a function—two points on the same vertical line correspond to pairs of the form (a, b) and (a, c) and, by definition, a function cannot contain (a, b) and (a, c) if $b \neq c$. Conversely, if a set of points in the plane has the property that no two points lie on the same vertical line, then it is surely the graph of a function. Thus, the first two sets in Figure 10 are not graphs of functions and the last two are; notice that the fourth is the graph of a function whose domain is not all of **R**, since some vertical lines have no points on them at all.

After the linear functions the simplest is perhaps the function $f(x) = x^2$. If we draw some of the pairs in f, i.e., some of the pairs of the form (x, x^2), we obtain a picture like Figure 11.

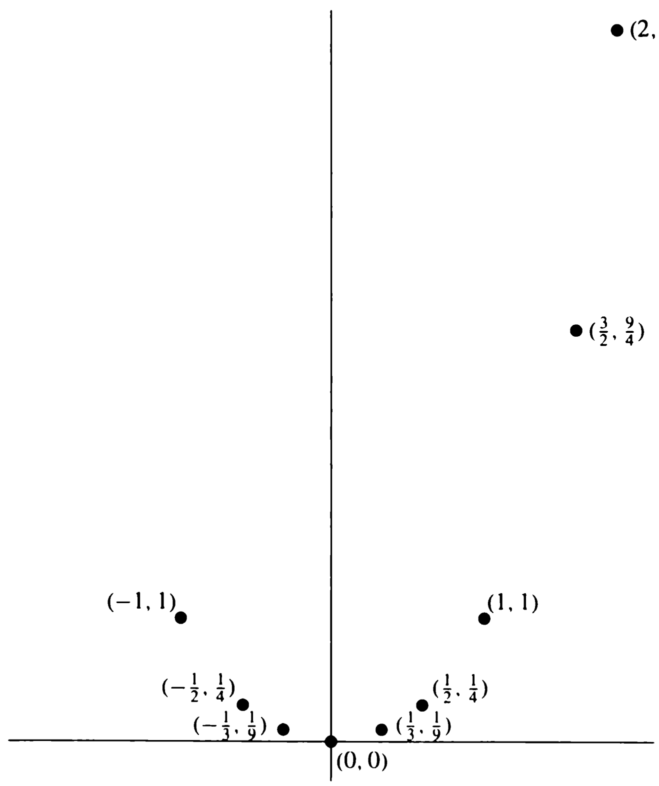


FIGURE 11

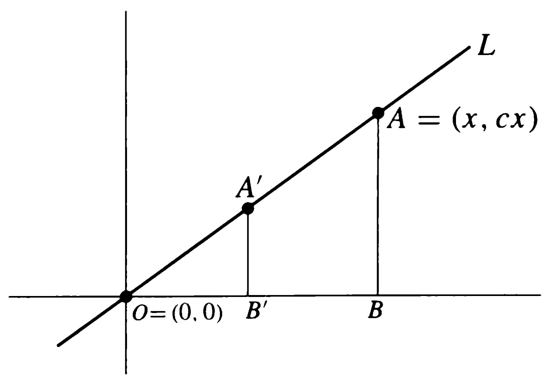


FIGURE 7

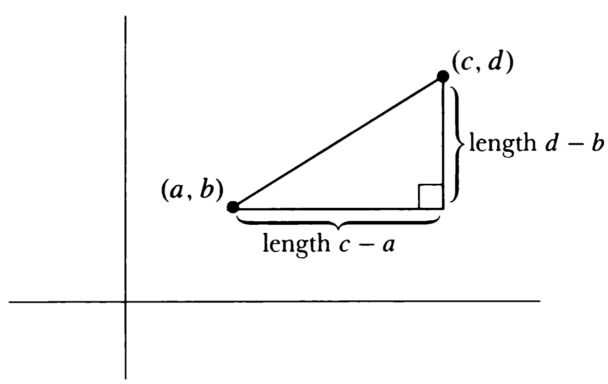


FIGURE 8

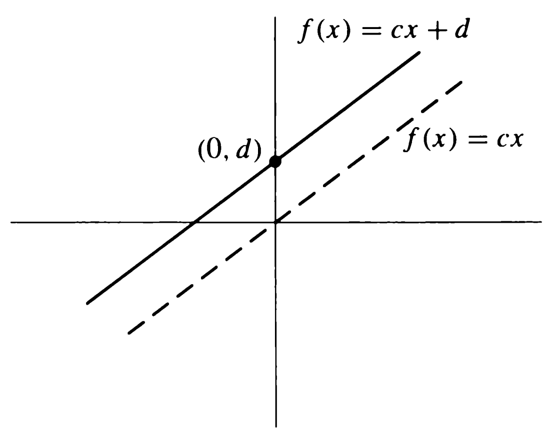


FIGURE 9

Let x be some number not equal to 0, and let L be the straight line which passes through the origin O, corresponding to $(0, 0)$, and through the point A, corresponding to (x, cx). A point A', with first coordinate y, will lie on L when the triangle $A'B'O$ is similar to the triangle ABO, thus when

$$\frac{A'B'}{OB'} = \frac{AB}{OB} = c;$$

this is precisely the condition that A' corresponds to the pair (y, cy), i.e., that A' lies on the graph of f. The argument has implicitly assumed that $c > 0$, but the other cases are treated easily enough. The number c, which measures the ratio of the sides of the triangles appearing in the proof, is called the *slope* of the straight line, and a line parallel to this line is also said to have slope c.

This demonstration has neither been labeled nor treated as a formal proof. Indeed, a rigorous demonstration would necessitate a digression which we are not at all prepared to follow. The rigorous proof of *any* statement connecting geometric and algebraic concepts would first require a real proof (or a precisely stated assumption) that the points on a straight line correspond in an exact way to the real numbers. Aside from this, it would be necessary to develop plane geometry as precisely as we intend to develop the properties of real numbers. Now the detailed development of plane geometry is a beautiful subject, but it is by no means a prerequisite for the study of calculus. We shall use geometric pictures only as an aid to intuition; for our purposes (and for most of mathematics) it is perfectly satisfactory to *define* the plane to be the set of all pairs of real numbers, and to *define* straight lines as certain collections of pairs, including, among others, the collections $\{(x, cx) : x \text{ a real number}\}$. To provide this artificially constructed geometry with all the structure of geometry studied in high school, one more definition is required. If (a, b) and (c, d) are two points in the plane, i.e., pairs of real numbers, we *define* the **distance** between (a, b) and (c, d) to be

$$\sqrt{(a-c)^2 + (b-d)^2}.$$

If the motivation for this definition is not clear, Figure 8 should serve as adequate explanation—with this definition the Pythagorean theorem has been built into our geometry.*

Reverting once more to our informal geometric picture, it is not hard to see (Figure 9) that the graph of the function $f(x) = cx + d$ is a straight line with slope c, passing through the point $(0, d)$. For this reason, the functions $f(x) = cx + d$ are called **linear functions**. Simple as they are, linear functions occur frequently, and you should feel comfortable working with them. The following is a typical problem whose solution should not cause any trouble. Given two distinct points (a, b) and (c, d), find the linear function f whose graph goes through (a, b) and (c, d). This amounts to saying that $f(a) = b$ and $f(c) = d$. If

*The fastidious reader might object to this definition on the grounds that nonnegative numbers are not yet known to have square roots. This objection is really unanswerable at the moment—the definition will just have to be accepted with reservations, until this little point is settled.

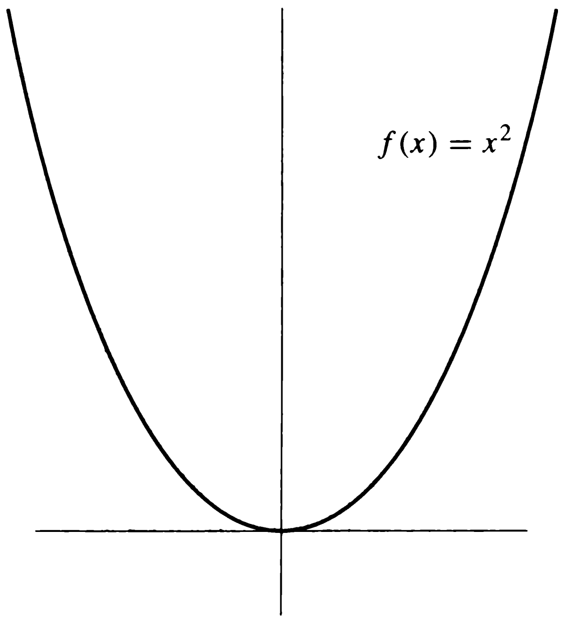


FIGURE 12

It is not hard to convince yourself that all the pairs (x, x^2) lie along a curve like the one shown in Figure 12; this curve is known as a **parabola**.

Since a graph is just a drawing on paper, made (in this case) with printer's ink, the question "Is this what the graph really looks like?" is hard to phrase in any sensible manner. No drawing is ever *really* correct since the line has thickness. Nevertheless, there are some questions which one *can* ask: for example, how can you be sure that the graph does not look like one of the drawings in Figure 13? It is easy to see, and even to prove, that the graph cannot look like (a); for if $0 < x < y$, then $x^2 < y^2$, so the graph should be higher at y than at x, which is not the case in (a) . It is also easy to see, simply by drawing a very accurate graph, first plotting many pairs (x, x^2), that the graph cannot have a large "jump" as in (b) or a "corner" as in (c). In order to prove these assertions, however, we first need to say, in a mathematical way, what it means for a function not to have a "jump" or "corner"; these ideas already involve some of the fundamental concepts of calculus. Eventually we will be able to define them rigorously, but meanwhile you may amuse yourself by attempting to define these concepts, and then examining your definitions critically. Later these definitions may be compared with the ones mathematicians have agreed upon. If they compare favorably, you are certainly to be congratulated!

The functions $f(x) = x^n$, for various natural numbers n, are sometimes called **power functions**. Their graphs are most easily compared as in Figure 14, by drawing several at once.

The power functions are only special cases of polynomial functions, introduced in the previous chapter. Two particular polynomial functions are graphed in

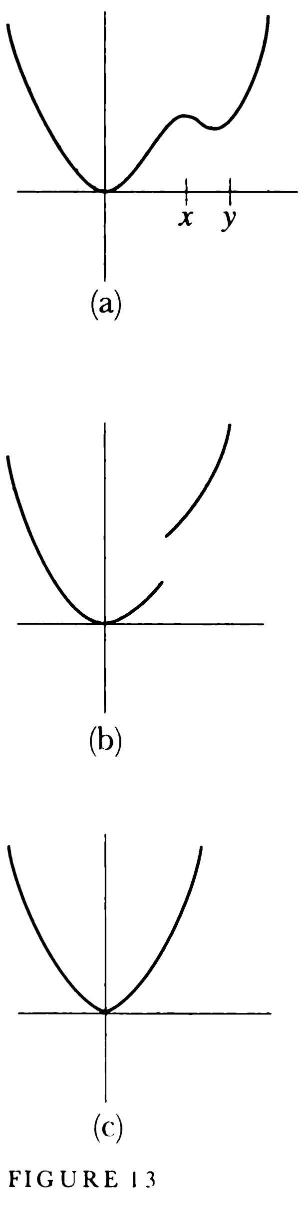


FIGURE 13

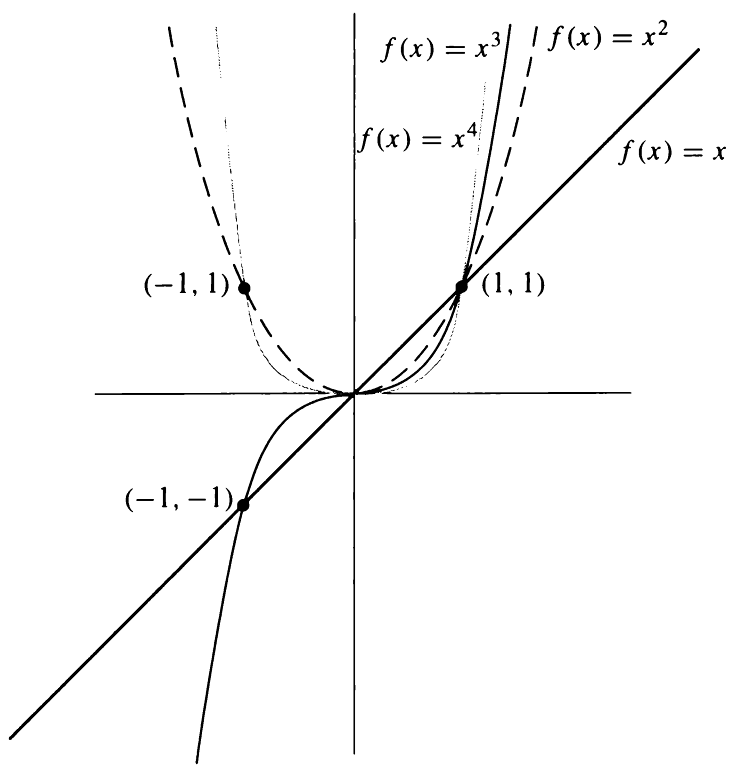
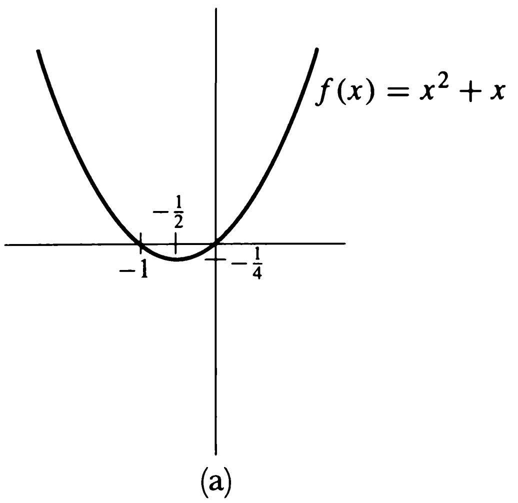


FIGURE 14



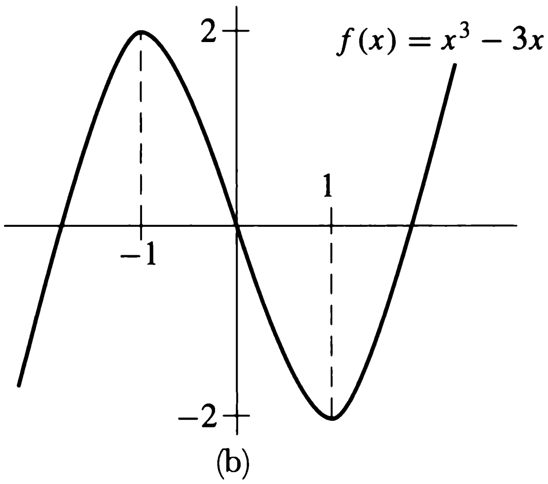


FIGURE 15

Figure 15, while Figure 16 is meant to give a general idea of the graph of the polynomial function

$$f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_0,$$

in the case $a_n > 0$.

In general, the graph of f will have at most $n - 1$ "peaks" or "valleys" (a "peak" is a point like $(x, f(x))$ in Figure 16, while a "valley" is a point like $(y, f(y))$. The number of peaks and valleys may actually be much smaller (the power functions, for example, have at most one valley). Although these assertions are easy to make, we will not even contemplate giving proofs until Part III (once the powerful methods of Part III are available, the proofs will be very easy).

Figure 17 illustrates the graphs of several rational functions. The rational functions exhibit even greater variety than the polynomial functions, but their behavior will also be easy to analyze once we can use the derivative, the basic tool of Part III.

Many interesting graphs can be constructed by "piecing together" the graphs of functions already studied. The graph in Figure 18 is made up entirely of straight lines. The function f with this graph satisfies

$$f\left(\frac{1}{n}\right) = (-1)^{n+1},$$

$$f\left(\frac{-1}{n}\right) = (-1)^{n+1},$$

$$f(x) = 1, \qquad |x| \geq 1,$$

and is a linear function on each interval $[1/(n+1), 1/n]$ and $[-1/n, -1/(n+1)]$. (The number 0 is not in the domain of f.) Of course, one can write out an explicit formula for $f(x)$, when x is in $[1/(n+1), 1/n]$; this is a good exercise in the use of linear functions, and will also convince you that a picture is worth a thousand words.

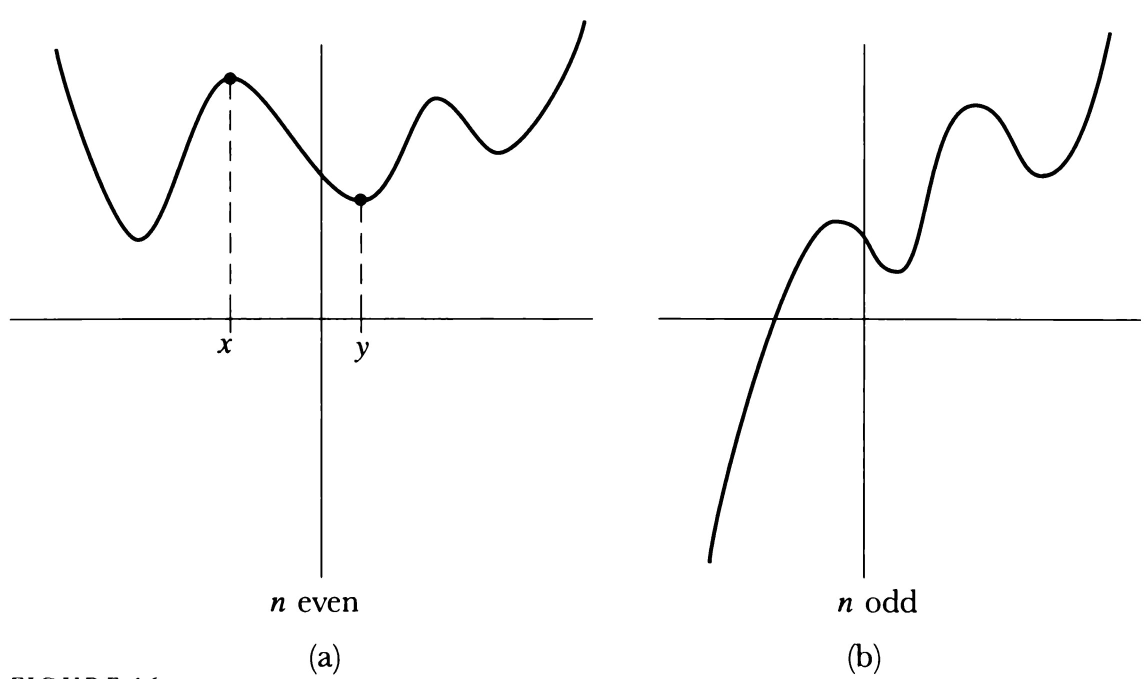


FIGURE 16

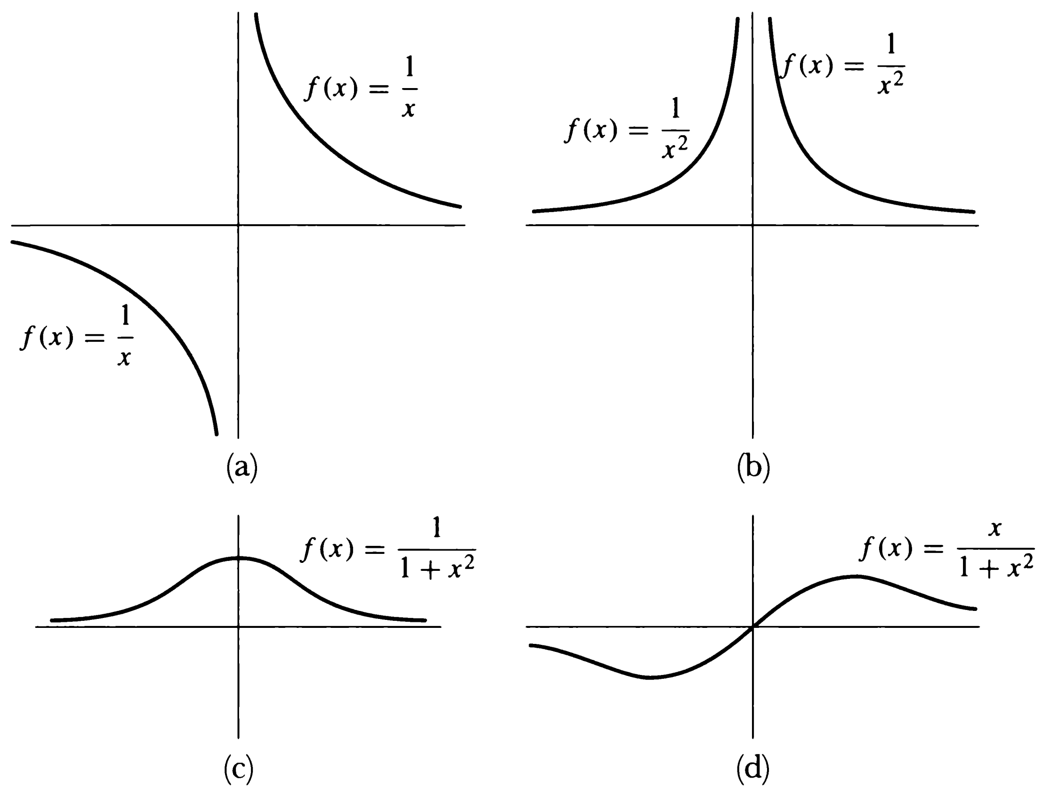


FIGURE 17

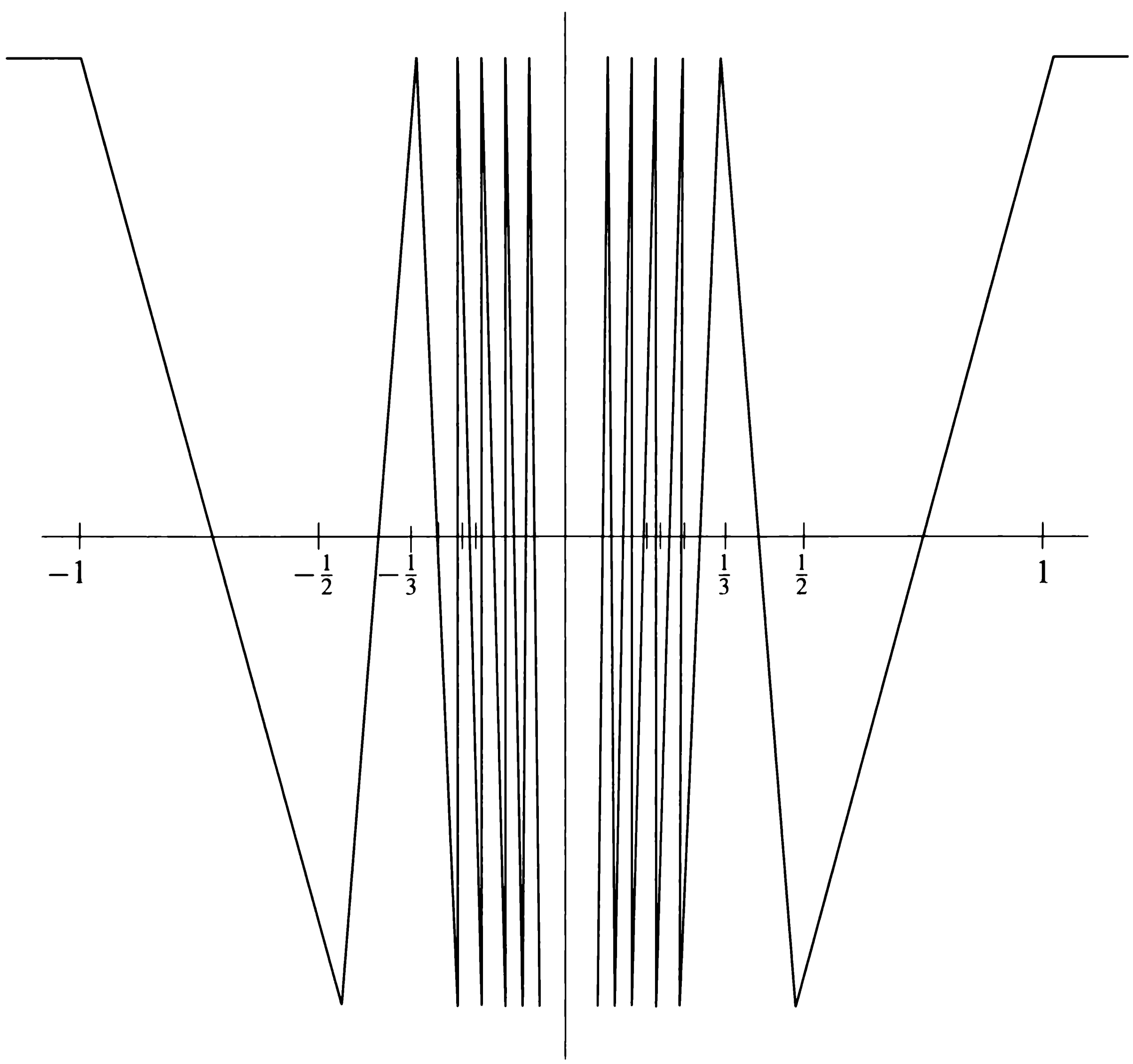


FIGURE 18

It is actually possible to define, in a much simpler way, a function which exhibits this same property of oscillating infinitely often near 0, by using the sine function, which we will discuss in detail in Chapter 15. As usual, we are using radian measure, so an angle of 2π means an angle "all the way around" a circle, an angle of π an angle half way around (or 180° in layman's terms), an angle of $\pi/2$ a right angle, etc.

The graph of the sine function is shown in Figure 19.

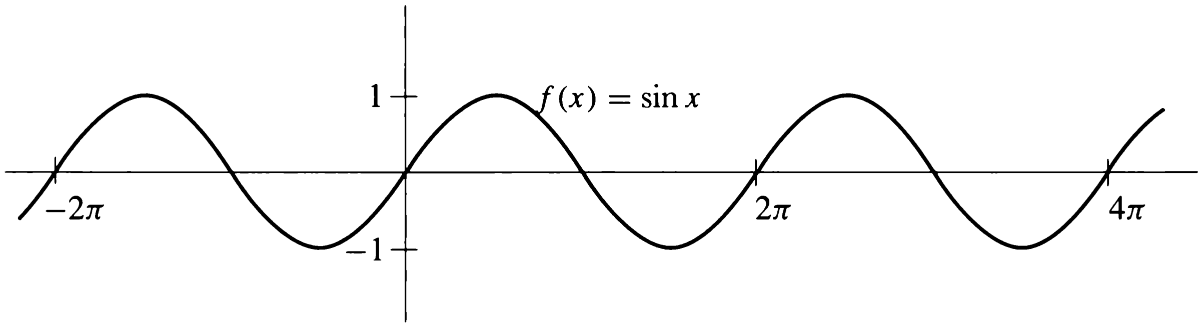


FIGURE 19

Now consider the function $f(x) = \sin 1/x$. The graph of f is shown in Figure 20. To draw this graph it helps to first observe that

$$f(x) = 0 \qquad \text{for } x = \frac{1}{\pi}, \frac{1}{2\pi}, \frac{1}{3\pi}, \dots,$$

$$f(x) = 1 \qquad \text{for } x = \frac{1}{\frac{1}{2}\pi}, \frac{1}{\frac{1}{2}\pi + 2\pi}, \frac{1}{\frac{1}{2}\pi + 4\pi}, \dots,$$

$$f(x) = -1 \qquad \text{for } x = \frac{1}{\frac{3}{2}\pi}, \frac{1}{\frac{3}{2}\pi + 2\pi}, \frac{1}{\frac{3}{2}\pi + 4\pi}, \dots.$$

Notice that when x is large, so that $1/x$ is small, $f(x)$ is also small; when x is

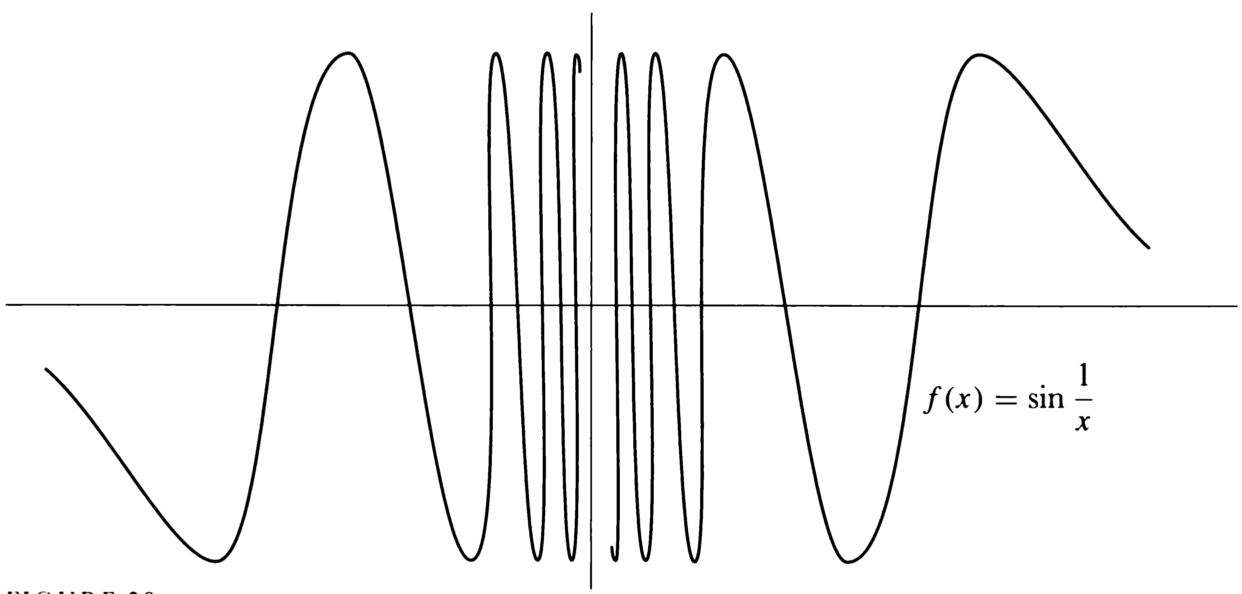


FIGURE 20

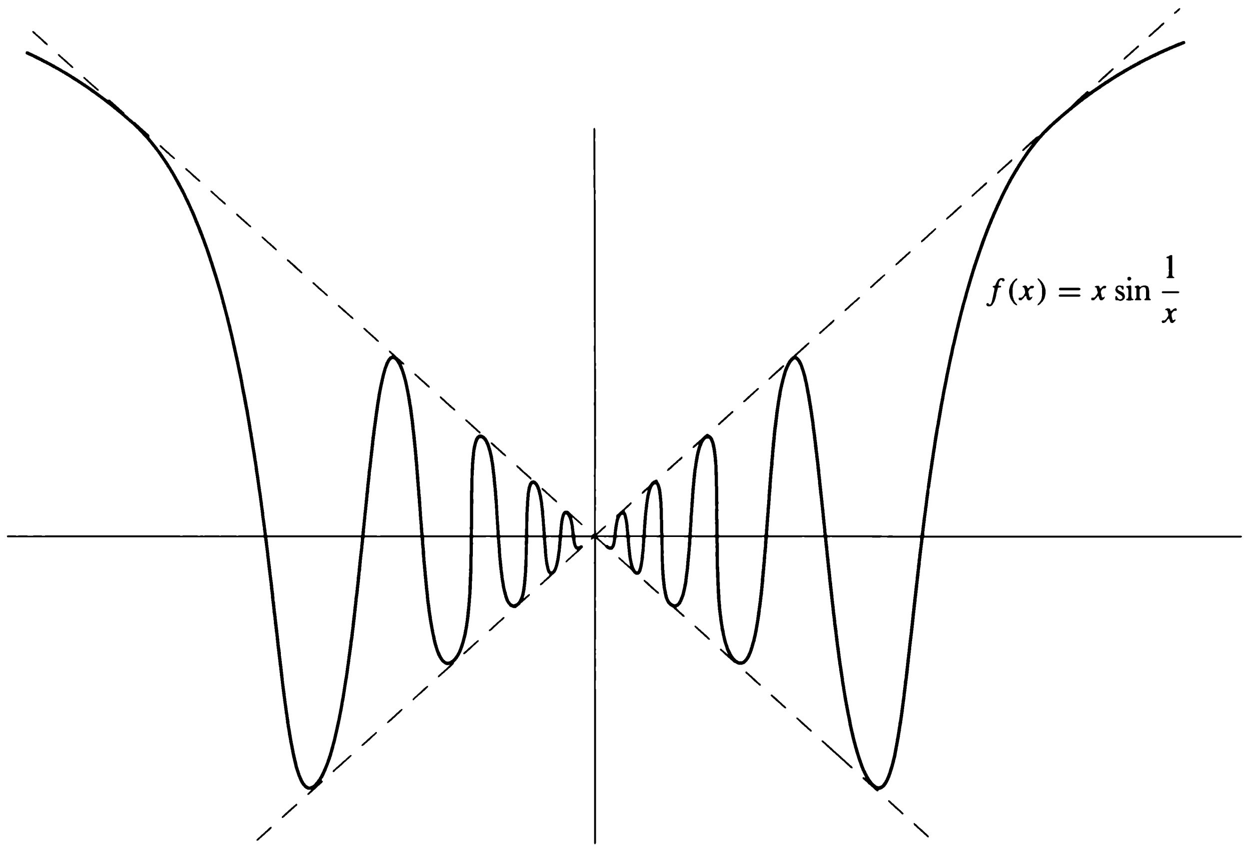
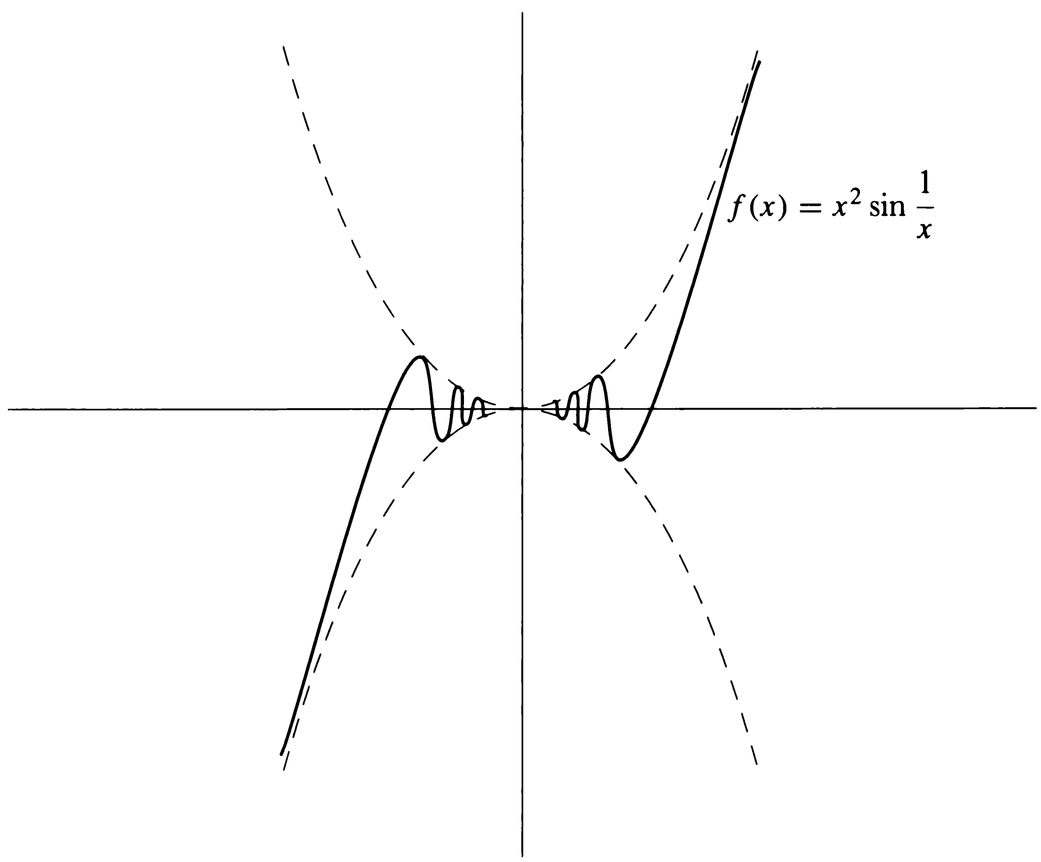


FIGURE 21

"large negative," that is, when $|x|$ is large for negative x, again $f(x)$ is close to 0, although $f(x) < 0$.

An interesting modification of this function is $f(x) = x \sin 1/x$. The graph of this function is sketched in Figure 21. Since $\sin 1/x$ oscillates infinitely often near 0 between 1 and -1, the function $f(x) = x \sin 1/x$ oscillates infinitely often between x and $-x$. The behavior of the graph for x large or large negative is harder to



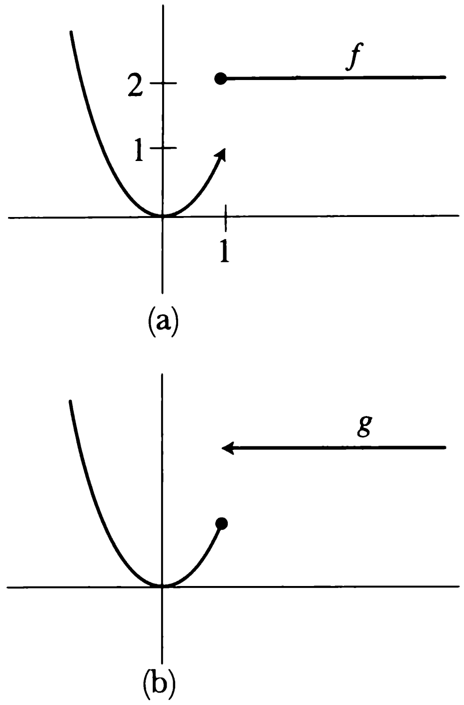


FIGURE 23

analyze. Since $\sin 1/x$ is getting close to 0, while x is getting larger and larger, there seems to be no telling what the product will do. It *is* possible to decide, but this is another question that is best deferred to Part III. The graph of $f(x) = x^2 \sin 1/x$ has also been illustrated (Figure 22).

For these infinitely oscillating functions, it is clear that the graph cannot hope to be really "accurate." The best we can do is to show part of it, and leave out the part near 0 (which is the interesting part). Actually, it is easy to find much simpler functions whose graphs cannot be "accurately" drawn. The graphs of

$$f(x) = \begin{cases} x^2, & x < 1 \\ 2, & x \geq 1 \end{cases} \qquad \text{and} \qquad g(x) = \begin{cases} x^2, & x \leq 1 \\ 2, & x > 1 \end{cases}$$

can only be distinguished by some convention similar to that used for open and closed intervals (Figure 23).

Out last example is a function whose graph is spectacularly nondrawable:

$$f(x) = \begin{cases} 0, & x \text{ irrational} \\ 1, & x \text{ rational.} \end{cases}$$

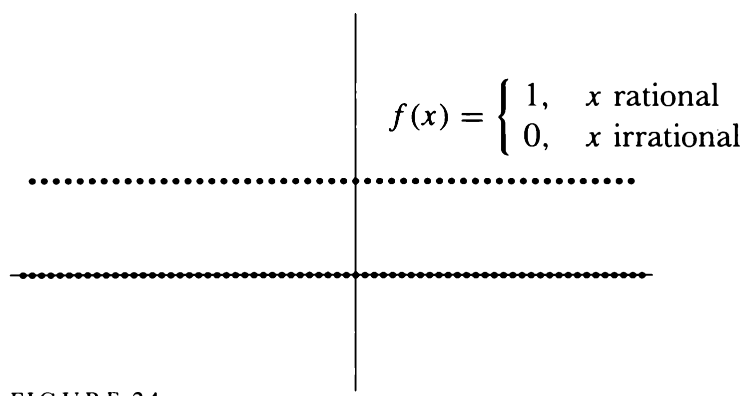


FIGURE 24

The graph of f must contain infinitely many points on the horizontal axis and also infinitely many points on a line parallel to the horizontal axis, but it must not contain either of these lines entirely. Figure 24 shows the usual textbook picture of the graph. To distinguish the two parts of the graph, the dots are placed closer together on the line corresponding to irrational x. (There is actually a mathematical reason behind this convention, but it depends on some sophisticated ideas, introduced in Problems 21-5 and 21-6.)

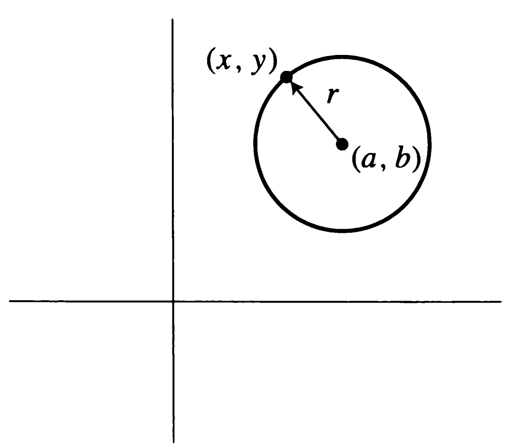


FIGURE 25

The peculiarities exhibited by some functions are so engrossing that it is easy to forget some of the simplest, and most important, subsets of the plane, which are not the graphs of functions. The most important example of all is the **circle**. A circle with center (a, b) and radius $r > 0$ contains, by definition, all the points (x, y) whose distance from (a, b) is equal to r. The circle thus consists (Figure 25) of all points (x, y) with

$$\sqrt{(x-a)^2 + (y-b)^2} = r$$

or

$$(x-a)^2 + (y-b)^2 = r^2.$$

The circle with center (0,0) and radius 1, often regarded as a sort of standard copy, is called the *unit circle.*

A close relative of the circle is the **ellipse**. This is defined as the set of points, the *sum* of whose distances from two "focus" points is a constant. (When the two foci are the same, we obtain a circle.) If, for convenience, the focus points are taken to be $(-c, 0)$ and $(c, 0)$, and the sum of the distances is taken to be $2a$ (the factor 2 simplifies some algebra), then (x, y) is on the ellipse if and only if

$$\sqrt{(x-(-c))^2+y^2}+\sqrt{(x-c)^2+y^2}=2a$$

or

$$\sqrt{(x+c)^2+y^2}=2a-\sqrt{(x-c)^2+y^2}$$

or

$$x^2+2cx+c^2+y^2=4a^2-4a\sqrt{(x-c)^2+y^2}+x^2-2cx+c^2+y^2$$

or

$$4(cx-a^2)=-4a\sqrt{(x-c)^2+y^2}$$

or

$$c^2x^2-2cxa^2+a^4=a^2(x^2-2cx+c^2+y^2)$$

or

$$(c^2-a^2)x^2-a^2y^2=a^2(c^2-a^2)$$

or

$$\frac{x^2}{a^2}+\frac{y^2}{a^2-c^2}=1.$$

This is usually written simply

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}=1,$$

where $b=\sqrt{a^2-c^2}$ (since we must clearly choose $a>c$, it follows that $a^2-c^2>0$). A picture of an ellipse is shown in Figure 26. The ellipse intersects the horizontal axis when $y=0$, so that

$$\frac{x^2}{a^2}=1, \qquad x=\pm a,$$

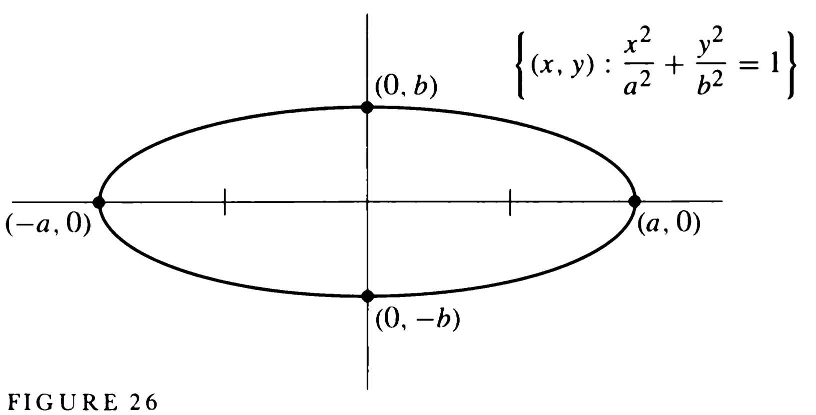


FIGURE 26

and it intersects the vertical axis when $x = 0$, so that

$$\frac{y^2}{b^2} = 1, \qquad y = \pm b.$$

The **hyperbola** is defined analogously, except that we require the *difference* of the two distances to be constant. Choosing the points $(-c, 0)$ and $(c, 0)$ once again, and the constant difference as $2a$, we obtain, as the condition that (x, y) be on the hyperbola,

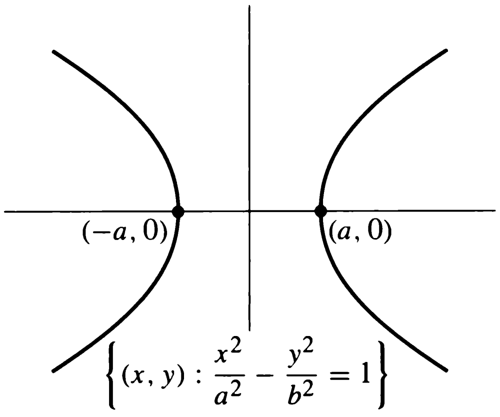


FIGURE 27

$$\sqrt{(x+c)^2 + y^2} - \sqrt{(x-c)^2 + y^2} = \pm 2a,$$

which may be simplified to

$$\frac{x^2}{a^2} + \frac{y^2}{a^2 - c^2} = 1.$$

In this case, however, we must clearly choose $c > a$, so that $a^2 - c^2 < 0$. If $b = \sqrt{c^2 - a^2}$, then (x, y) is on the hyperbola if and only if

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1.$$

The picture is shown in Figure 27. It contains two pieces, because the difference between the distances of (x, y) from $(-c, 0)$ and $(c, 0)$ may be taken in two different orders. The hyperbola intersects the horizontal axis when $y = 0$, so that $x = \pm a$, but it never intersects the vertical axis.

It is interesting to compare (Figure 28) the hyperbola with $a = b = \sqrt{2}$ and the graph of the function $f(x) = 1/x$. The drawings look quite similar, and the two sets are actually identical, except for a rotation through an angle of $\pi/4$ (Problem 23).

Clearly no rotation of the plane will change circles or ellipses into the graphs of functions. Nevertheless, the study of these important geometric figures can often be reduced to the study of functions. Ellipses, for example, are made up of the

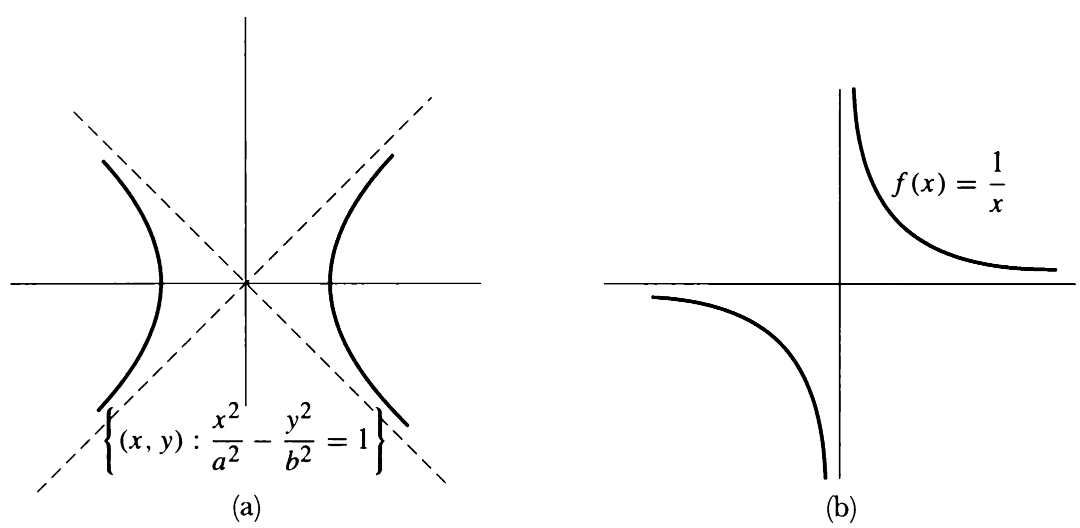


FIGURE 28

graphs of two functions,

$$f(x) = b\sqrt{1-(x^2/a^2)}, \qquad -a \le x \le a$$

and

$$g(x) = -b\sqrt{1-(x^2/a^2)}, \qquad -a \le x \le a.$$

Of course, there are many other pairs of functions with this same property. For example, we can take

$$f(x) = \begin{cases} b\sqrt{1-(x^2/a^2)}, & 0 < x \le a \\ -b\sqrt{1-(x^2/a^2)}, & -a \le x \le 0 \end{cases}$$

and

$$g(x) = \begin{cases} -b\sqrt{1-(x^2/a^2)}, & 0 < x \le a \\ b\sqrt{1-(x^2/a^2}, & -a \le x \le 0. \end{cases}$$

We could also choose

$$f(x) = \begin{cases} b\sqrt{1-(x^2/a^2)}, & x \text{ rational}, \ -a \le x \le a \\ -b\sqrt{1-(x^2/a^2)}, & x \text{ irrational}, -a \le x \le a \end{cases}$$

and

$$g(x) = \begin{cases} -b\sqrt{1-(x^2/a^2)}, & x \text{ rational}, \ -a \le x \le a \\ b\sqrt{1-(x^2/a^2)}, & x \text{ irrational}, -a \le x \le a. \end{cases}$$

But all these other pairs necessarily involve unreasonable functions which jump around. A proof, or even a precise statement of this fact, is too difficult at present. Although you have probably already begun to make a distinction between those functions with reasonable graphs, and those with unreasonable graphs, you may find it very difficult to state a reasonable definition of reasonable functions. A mathematical definition of this concept is by no means easy, and a great deal of this book may be viewed as successive attempts to impose more and more conditions that a "reasonable" function must satisfy. As we define some of these conditions, we will take time out to ask if we have really succeeded in isolating the functions which deserve to be called reasonable. The answer, unfortunately, will always be "no," or at best, a qualified "yes."

PROBLEMS

1. Indicate on a straight line the set of all x satisfying the following conditions. Also name each set, using the notation for intervals (in some cases you will also need the $\cup$ sign).

(i) $|x-3| < 1$.
(ii) $|x-3| \le 1$.
(iii) $|x-a| < \varepsilon$.
(iv) $|x^2-1| < \frac{1}{2}$.

(v) $\dfrac{1}{1+x^2} \geq \frac{1}{5}$.

(vi) $\dfrac{1}{1+x^2} \leq a$ (give an answer in terms of a, distinguishing various cases).

(vii) $x^2 + 1 \geq 2$.

(viii) $(x+1)(x-1)(x-2) > 0$.

2. There is a very useful way of describing the points of the closed interval $[a, b]$ (where we assume, as usual, that $a < b$).

(a) First consider the interval $[0, b]$, for $b > 0$. Prove that if x is in $[0, b]$, then $x = tb$ for some t with $0 \leq t \leq 1$. What is the significance of the number t? What is the mid-point of the interval $[0, b]$?

(b) Now prove that if x is in $[a, b]$, then $x = (1-t)a + tb$ for some t with $0 \leq t \leq 1$. Hint: This expression can also be written as $a + t(b-a)$. What is the midpoint of the interval $[a, b]$? What is the point $1/3$ of the way from a to b?

(c) Prove, conversely, that if $0 \leq t \leq 1$, then $(1-t)a + tb$ is in $[a, b]$.

(d) The points of the *open* interval (a, b) are those of the form $(1-t)a + tb$ for $0 < t < 1$.

3. Draw the set of all points (x, y) satisfying the following conditions. (In most cases your picture will be a sizable portion of a plane, not just a line or curve.)

(i) $x > y$.

(ii) $x + a > y + b$.

(iii) $y < x^2$.

(iv) $y \leq x^2$.

(v) $|x - y| < 1$.

(vi) $|x + y| < 1$.

(vii) $x + y$ is an integer.

(viii) $\dfrac{1}{x+y}$ is an integer.

(ix) $(x-1)^2 + (y-2)^2 < 1$.

(x) $x^2 < y < x^4$.

4. Draw the set of all points (x, y) satisfying the following conditions:

(i) $|x| + |y| = 1$.

(ii) $|x| - |y| = 1$.

(iii) $|x - 1| = |y - 1|$.

(iv) $|1 - x| = |y - 1|$.

(v) $x^2 + y^2 = 0$.

(vi) $xy = 0$.

(vii) $x^2 - 2x + y^2 = 4$.

(viii) $x^2 = y^2$.

5. Draw the set of all points (x, y) satisfying the following conditions:

   (i) $x = y^2$.

   (ii) $\dfrac{y^2}{a^2} - \dfrac{x^2}{b^2} = 1$.

   (iii) $x = |y|$.

   (iv) $x = \sin y$.

   Hint: You already know the answers when x and y are interchanged.

6. (a) Show that the straight line through (a, b) with slope m is the graph of the function $f(x) = m(x - a) + b$. This formula, known as the "point-slope form" is far more convenient than the equivalent expression $f(x) = mx + (b - ma)$; it is immediately clear from the point-slope form that the slope is m, and that the value of f at a is b.

   (b) For $a \neq c$, show that the straight line through (a, b) and (c, d) is the graph of the function

   $$f(x) = \frac{d-b}{c-a}(x - a) + b.$$

   (c) When are the graphs of $f(x) = mx + b$ and $g(x) = m'x + b'$ parallel straight lines?

7. (a) For any numbers A, B, and C, with A and B not both 0, show that the set of all (x, y) satisfying $Ax + By + C = 0$ is a straight line (possibly a vertical one). Hint: First decide when a vertical straight line is described.

   (b) Show conversely that every straight line, including vertical ones, can be described as the set of all (x, y) satisfying $Ax + By + C = 0$.

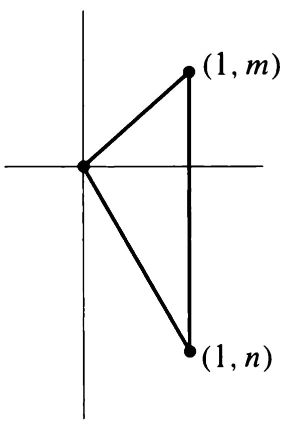


FIGURE 29

8. (a) Prove that the graphs of the functions

   $$f(x) = mx + b,$$
   $$g(x) = nx + c,$$

   are perpendicular if $mn = -1$, by computing the squares of the lengths of the sides of the triangle in Figure 29. (Why is this special case, where the lines intersect at the origin, as good as the general case?)

   (b) Prove that the two straight lines consisting of all (x, y) satisfying the conditions

   $$Ax + By + C = 0,$$
   $$A'x + B'y + C' = 0,$$

   are perpendicular if and only if $AA' + BB' = 0$.

9. (a) Prove, using Problem 1-19, that

   $$\sqrt{(x_1 + y_1)^2 + (x_2 + y_2)^2} \le \sqrt{x_1{}^2 + x_2{}^2} + \sqrt{y_1{}^2 + y_2{}^2}.$$

(b) Prove that

$$\sqrt{(x_3 - x_1)^2 + (y_3 - y_1)^2} \leq \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} + \sqrt{(x_3 - x_2)^2 + (y_3 - y_2)^2}.$$

Interpret this inequality geometrically (it is called the "triangle inequality"). When does strict inequality hold?

10. Sketch the graphs of the following functions, plotting enough points to get a good idea of the general appearance. (Part of the problem is to make a reasonable decision how many is "enough"; the queries posed below are meant to show that a little thought will often be more valuable than hundreds of individual points.)

(i) $f(x) = x + \dfrac{1}{x}$. (What happens for x near 0, and for large x? Where does the graph lie in relation to the graph of the identify function? Why does it suffice to consider only positive x at first?)

(ii) $f(x) = x - \dfrac{1}{x}$.

(iii) $f(x) = x^2 + \dfrac{1}{x^2}$.

(iv) $f(x) = x^2 - \dfrac{1}{x^2}$.

11. Describe the general features of the graph of f if

(i) f is even.
(ii) f is odd.
(iii) f is nonnegative.
(iv) $f(x) = f(x + a)$ for all x (a function with this property is called **periodic**, with **period** a.

12. Graph the functions $f(x) = \sqrt[m]{x}$ for $m = 1, 2, 3, 4$. (There is an easy way to do this, using Figure 14. Be sure to remember, however, that $\sqrt[m]{x}$ means the *positive* mth root of x when m is even; you should also note that there will be an important difference between the graphs when m is even and when m is odd.)

13. (a) Graph $f(x) = |x|$ and $f(x) = x^2$.
(b) Graph $f(x) = |\sin x|$ and $f(x) = \sin^2 x$. (There is an important difference between the graphs, which we cannot yet even describe rigorously. See if you can discover what it is; part (a) is meant to be a clue.)

14. Describe the graph of g in terms of the graph of f if

(i) $g(x) = f(x) + c$.
(ii) $g(x) = f(x + c)$. (It is easy to make a mistake here.)
(iii) $g(x) = cf(x)$.
(iv) $g(x) = f(cx)$. (Distinguish the cases $c = 0$, $c > 0$, $c < 0$.)
(v) $g(x) = f(1/x)$.
(vi) $g(x) = f(|x|)$.
(vii) $g(x) = |f(x)|$.
(viii) $g(x) = \max(f, 0)$.
(ix) $g(x) = \min(f, 0)$.
(x) $g(x) = \max(f, 1)$.

15. Draw the graph of $f(x) = ax^2 + bx + c$. Hint: Use the methods of Problem 1-18.

16. Suppose that A and C are not both 0. Show that the set of all (x, y) satisfying

$$Ax^2 + Bx + Cy^2 + Dy + E = 0$$

is either a parabola, an ellipse, or an hyperbola (or a "degenerate case": two lines [either intersecting or parallel], one line, a point, or $\emptyset$). Hint: The case $C = 0$ is essentially Problem 15, and the case $A = 0$ is just a minor variant. Now consider separately the cases where A and B are both positive or negative, and where one is positive while the other is negative. When do we have a circle?

17. The symbol $[x]$ denotes the largest integer which is $\leq x$. Thus, $[2.1] = [2] = 2$ and $[-0.9] = [-1] = -1$. Draw the graph of the following functions (they are all quite interesting, and several will reappear frequently in other problems).

(i) $f(x) = [x]$.
(ii) $f(x) = x - [x]$.
(iii) $f(x) = \sqrt{x - [x]}$.
(iv) $f(x) = [x] + \sqrt{x - [x]}$.
(v) $f(x) = \left[\dfrac{1}{x}\right]$.
(vi) $f(x) = \dfrac{1}{\left[\dfrac{1}{x}\right]}$.

18. Graph the following functions.

(i) $f(x) = \{x\}$, where $\{x\}$ is defined to be the distance from x to the nearest integer.
(ii) $f(x) = \{2x\}$.
(iii) $f(x) = \{x\} + \frac{1}{2}\{2x\}$.
(iv) $f(x) = \{4x\}$.
(v) $f(x) = \{x\} + \frac{1}{2}\{2x\} + \frac{1}{4}\{4x\}$.

Many functions may be described in terms of the decimal expansion of a number. Although we will not be in a position to describe infinite decimals rigorously until Chapter 23, your intuitive notion of infinite decimals should suffice to carry you through the following problem, and others which occur before Chapter 23. There is one ambiguity about infinite decimals which must be eliminated: Every decimal ending in a string of 9's is equal to another ending in a string of 0's (e.g., $1.23999\ldots = 1.24000\ldots$). We will always use the one ending in 9's.

***19.** Describe as best you can the graphs of the following functions (a complete picture is usually out of the question).

(i) $f(x) =$ the 1st number in the decimal expansion of x.
(ii) $f(x) =$ the 2nd number in the decimal expansion of x.
(iii) $f(x) =$ the number of 7's in the decimal expansion of x if this number is finite, and 0 otherwise.
(iv) $f(x) = 0$ if the number of 7's in the decimal expansion of x is finite, and 1 otherwise.
(v) $f(x) =$ the number obtained by replacing all digits in the decimal expansion of x which come after the first 7 (if any) by 0.
(vi) $f(x) = 0$ if 1 never appears in the decimal expansion of x, and n if 1 first appears in the nth place.

***20.** Let

$$f(x) = \begin{cases} 0, & x \text{ irrational} \\ \dfrac{1}{q}, & x = \dfrac{p}{q} \text{ rational in lowest terms.} \end{cases}$$

(A number p/q is in **lowest terms** if p and q are integers with no common factor, and $q > 0$). Draw the graph of f as well as you can (don't sprinkle points randomly on the paper; consider first the rational numbers with $q = 2$, then those with $q = 3$, etc.).

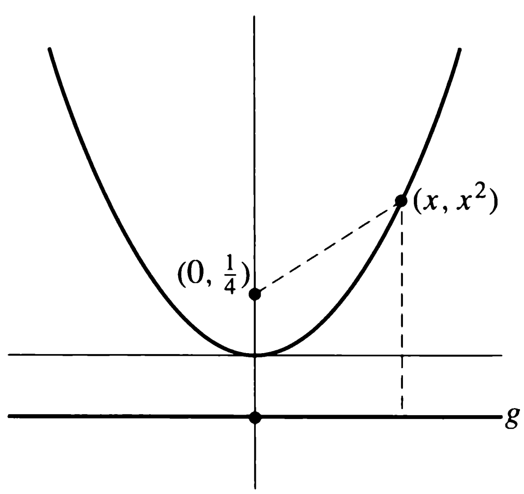


FIGURE 30

21. (a) The points on the graph of $f(x) = x^2$ are the ones of the form (x, x^2). Prove that each such point is equidistant from the point $(0, \frac{1}{4})$ and the graph of $g(x) = -\frac{1}{4}$. (See Figure 30.)

(b) Given a horizontal line L, the graph of $g(x) = \gamma$, and a point $P = (\alpha, \beta)$ not on L, so that $\gamma \neq \beta$, show that the set of all points (x, y) equidistant from P and L is the graph of a function of the form $f(x) = ax^2 + bx + c$. What is this set if $\gamma = \beta$?

***22.** (a) Show that the square of the distance from (c, d) to (x, mx) is

$$x^2(m^2 + 1) + x(-2md - 2c) + d^2 + c^2.$$

Using Problem 1-18 to find the minimum of these numbers, show that the distance from (c, d) to the graph of $f(x) = mx$ is

$$|cm - d|/\sqrt{m^2 + 1}.$$

(b) Find the distance from (c, d) to the graph of $f(x) = mx + b$. (Reduce this case to part (a).)

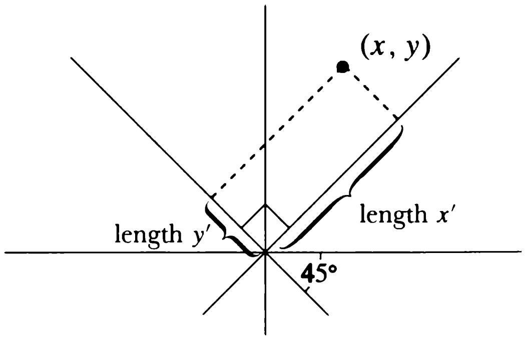


FIGURE 31

***23.** (a) Using Problem 22, show that the numbers x' and y' indicated in Figure 31 are given by

$$x' = \frac{1}{\sqrt{2}}x + \frac{1}{\sqrt{2}}y,$$
$$y' = -\frac{1}{\sqrt{2}}x + \frac{1}{\sqrt{2}}y.$$

(b) Show that the set of all (x, y) with $(x'/\sqrt{2})^2 - (y'/\sqrt{2})^2 = 1$ is the same as the set of all (x, y) with $xy = 1$.

APPENDIX 1. VECTORS

Suppose that v is a point in the plane; in other words, v is a pair of numbers

$$v = (v_1, v_2).$$

For convenience, we will use this convention that subscripts indicate the first and second pairs of a point that has been described by a single letter. Thus, if we mention the points w and z, it will be understood that w is the pair (w_1, w_2), while z is the pair (z_1, z_2).

Instead of the actual pair of numbers (v_1, v_2), we often picture v as an arrow from the origin O to this point (Figure 1), and we refer to these arrows as *vectors* in the plane. Of course, we've haven't really said anything new yet, we've simply introduced an alternate term for a point of the plane, and another mental picture. The real point of the new terminology is to emphasize that we are going to do some new things with points in the plane.

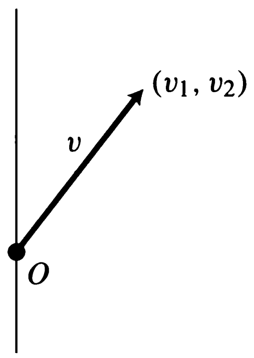


FIGURE 1

For example, suppose that we have two vectors (i.e., points) in the plane,

$$v = (v_1, v_2), \qquad w = (w_1, w_2).$$

Then we can define a new vector (a new point of the plane) $v + w$ by the equation

$$v + w = (v_1 + w_1, v_2 + w_2). \tag{1}$$

Notice that all the letters on the right side of this equation are numbers, and the $+$ sign is just our usual addition of numbers. On the other hand, the $+$ sign on the left side is new: previously, the sum of two points in the plane wasn't defined, and we've simply used equation (1) as a *definition*.

A very fussy mathematician might want to use some new symbol for this newly defined operation, like

$$v \boldsymbol{+} w, \qquad \text{or perhaps} \qquad v \oplus w,$$

but there's really no need to insist on this; since $v + w$ hasn't been defined before, there's no possibility of confusion, so we might as well keep the notation simple.

Of course, any one can make new notation; for example, since it's our definition, we could just as well have defined $v + w$ as $(v_1 + w_1 \cdot w_2,\ v_2 + {w_1}^2)$, or by some other equally weird formula. The real question is, does our new construction have any particular significance?

Figure 2 shows two vectors v and w, as well as the point

$$(v_1 + w_1, v_2 + w_2),$$

which, for the moment, we have simply indicated in the usual way, without drawing an arrow. Note that it is easy to compute the slope of the line L between v and our new point: as indicated in Figure 2, this slope is just

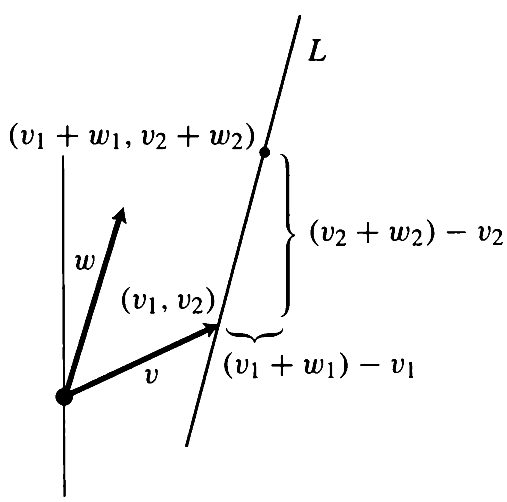


FIGURE 2

$$\frac{(v_2 + w_2) - v_2}{(v_1 + w_1) - v_1} = \frac{w_2}{w_1},$$

and this, of course, is the slope of our vector w, from the origin O to (w_1, w_2). In other words, the line L is parallel to w.

Similarly, the slope of the line M between (w_1, w_2) and our new point is

$$\frac{(v_2 + w_2) - w_2}{(v_1 + w_1) - v_2} = \frac{v_2}{v_1},$$

which is the slope of the vector v; so M is parallel to v. In short, the new point $v + w$ lies on the parallelogram having v and w as sides. When we draw $v + w$ as an arrow (Figure 3), it points along the diagonal of this parallelogram. In physics, vectors are used to symbolize forces, and the sum of two vectors represents the resultant force when two different forces are applied simultaneously to the same object.

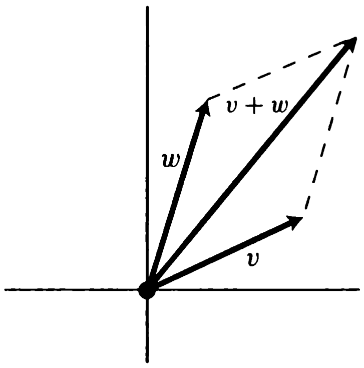


FIGURE 3

Figure 4 shows another way of visualizing the sum $v+w$. If we use "w" to denote an arrow parallel to w, and having the same length, but starting at v instead of at the origin, then $v + w$ is the vector from O to the final endpoint; thus we get to $v + w$ by first following v, and then following w.

Many of the properties of $+$ for ordinary numbers also hold for this new $+$ for vectors. For example, the "commutative law"

$$v + w = w + v,$$

is obvious from the geometric picture, since the parallelogram spanned by v and w is the same as the parallelogram spanned by w and v. It is also easily checked analytically, since it states that

$$(v_1 + w_1, v_2 + w_2) = (w_1 + v_1, w_2 + v_2),$$

and thus simply depends on the commutative law for numbers:

$$v_1 + w_1 = w_1 + v_1,$$
$$v_2 + w_2 = w_2 + v_2.$$

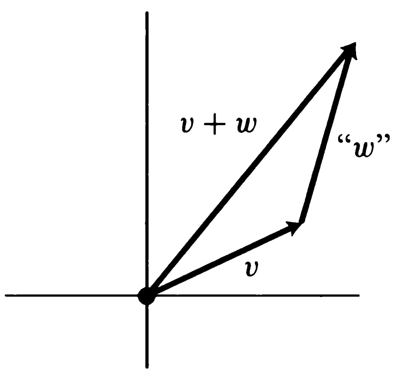


FIGURE 4

Similarly, unraveling definitions, we find the "associative law"

$$[v + w] + z = v + [w + z].$$

Figure 5 indicates a method of finding $v + w + z$.

The origin $O = (0, 0)$ is an "additive identity,"

$$O + v = v + O = v,$$

and if we define

$$-v = (-v_1, -v_2),$$

then we also have

$$v + (-v) = -v + v = O.$$

Naturally we can also define

$$w - v = w + (-v),$$

exactly as with numbers; equivalently,

$$w - v = (w_1 - v_1, w_2 - v_2).$$

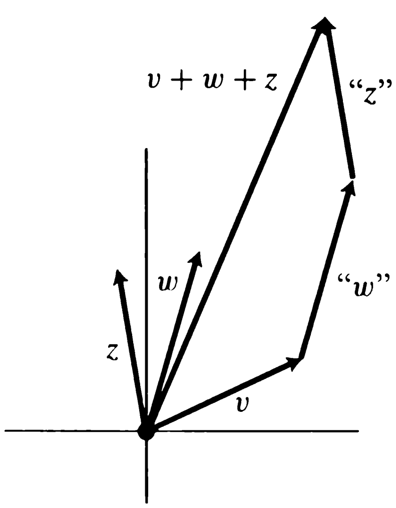


FIGURE 5

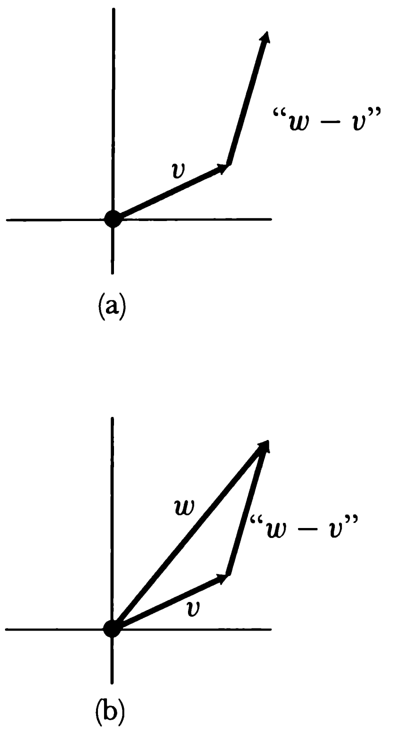


FIGURE 6

Just as with numbers, our definition of $w - v$ simply means that it satisfies

$$v + (w - v) = w.$$

Figure 6(a) shows v and an arrow "$w - v$" that is parallel to $w - v$ but that starts at the endpoint of v. As we established with Figure 4, the vector from the origin to the endpoint of this arrow is just $v + (w - v) = w$ (Figure 6(b)). In other words, we can picture $w - v$ geometrically as the arrow that goes from v to w (except that it must then be moved back to the origin).

There is also a way of multiplying a number by a vector: For a number a and a vector $v = (v_1, v_2)$, we define

$$a \cdot v = (av_1, av_2)$$

(We sometimes simply write av instead of $a \cdot v$; of course, it is then especially important to remember that v denotes a vector, rather than a number.) The vector $a \cdot v$ points in the same direction as v when $a > 0$ and in the opposite direction when $a < 0$ (Figure 7).

You can easily check the following formulas:

$$\begin{aligned} a \cdot (b \cdot v) &= (ab) \cdot v, \\ 1 \cdot v &= v, \\ 0 \cdot v &= O, \\ -1 \cdot v &= -v. \end{aligned}$$

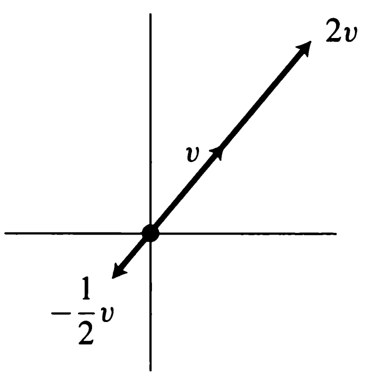


FIGURE 7

Notice that we have only defined a product of a number and a vector, we have not defined a way of 'multiplying' two vectors to get another vector.* However, there are various ways of 'multiplying' vectors to get numbers, which are explored in the following problems.

PROBLEMS

1. Given a point v of the plane, let $R_\theta(v)$ be the result of rotating v around the origin through an angle of θ (Figure 8). The aim of this problem is to obtain a formula for R_θ, with minimal calculation.

(a) Show that

$R_\theta(1, 0) = (\cos\theta, \sin\theta)$, [we should really write $R_\theta((1, 0))$, etc.]
$R_\theta(0, 1) = (-\sin\theta, \cos\theta)$.

(b) Explain why we have

$$\begin{aligned} R_\theta(v + w) &= R_\theta(v) + R_\theta(w), \\ R_\theta(a \cdot w) &= a \cdot R_\theta(w). \end{aligned}$$

(c) Now show that for any point (x, y) we have

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

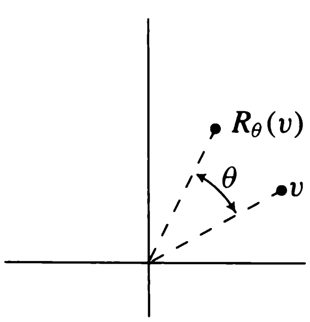


FIGURE 8

*If you jump to Chapter 25, you'll find that there is an important way of defining a product, but this is something very special for the plane—it doesn't work for vectors in 3-space, for example, even though the other constructions do.

(d) Use this result to give another solution to Problem 4-23.

2. Given v and w, we define the *number*

$$v \cdot w = v_1 w_1 + v_2 w_2;$$

this is often called the 'dot product' or 'scalar product' of v and w ('scalar' being a rather old-fashioned word for a number, as opposed to a vector).

(a) Given v, find a vector w such that $v \cdot w = 0$. Now describe the set of all such vectors w.

(b) Show that

$$v \cdot w = w \cdot v$$
$$v \cdot (w + z) = v \cdot w + v \cdot z$$

and that

$$a \cdot (v \cdot w) = (a \cdot v) \cdot w = v \cdot (a \cdot w).$$

Notice that the last of these equations involves *three* products: the dot product $\bullet$ of two vectors; the product $\cdot$ of a number and a vector; and the ordinary product $\cdot$ of two numbers.

(c) Show that $v \cdot v \geq 0$, and that $v \cdot v = 0$ only when $v = O$. Hence we can define the *norm* $\|v\|$ as

$$\|v\| = \sqrt{v \cdot v},$$

which will be 0 only for $v = O$. What is the geometric interpretation of the norm?

(d) Prove that

$$\|v + w\| \leq \|v\| + \|w\|,$$

and that equality holds if and only if $v = 0$ or $w = 0$ or $w = a \cdot v$ for some number $a > 0$.

(e) Show that

$$v \cdot w = \frac{\|v + w\|^2 - \|v - w\|^2}{4}.$$

3. (a) Let R_θ be rotation by an angle of θ (Problem 1). Show that

$$R_\theta(v) \cdot R_\theta(w) = v \cdot w.$$

(b) Let $e = (1, 0)$ be the vector of length 1 pointing along the first axis, and let $w = (\cos\theta, \sin\theta)$; this is a vector of length 1 that makes an angle of θ with the first axis (compare Problem 1). Calculate that

$$e \cdot w = \cos\theta.$$

Conclude that in general

$$v \cdot w = \|v\| \cdot \|w\| \cdot \cos\theta,$$

where θ is the angle between v and w.

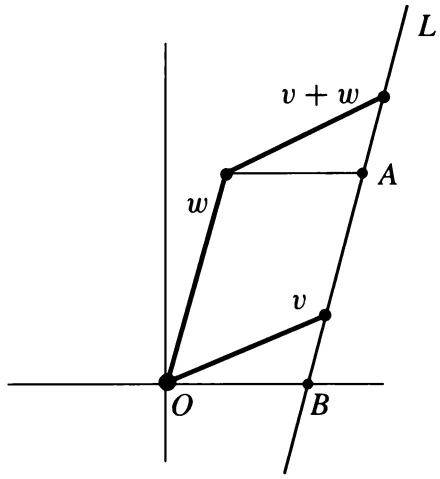


FIGURE 9

4. Given two vectors v and w, we'd expect to have a simple formula, involving the coordinates v_1, v_2, w_1, w_2, for the area of the parallelogram they span. Figure 9 indicates a strategy for finding such a formula: since the triangle with vertices $w, A, v+w$ is congruent to the triangle OBv, we can reduce the problem to an easier one where one side of the parallelogram lies along the horizontal axis:

(a) The line L passes through v and is parallel to w, so has slope w_2/w_1. Conclude that the point B has coordinate

$$\frac{v_1w_2 - w_1v_2}{w_2},$$

and that the parallelogram therefore has area

$$\det(v, w) = v_1w_2 - w_1v_2.$$

This formula, which defines the *determinant* det, certainly seems to be simple enough, but it can't really be true that $\det(v, w)$ always gives the area. After all, we clearly have

$$\det(w, v) = -\det(v, w),$$

so sometimes det will be negative! Indeed, it is easy to see that our "derivation" made all sorts of assumptions (that w_2 was positive, that B had a positive coordinate, etc.) Nevertheless, it seems likely that $\det(v, w)$ is $\pm$ the area; the next problem gives an independent proof.

5. (a) If v points along the positive horizontal axis, show that $\det(v, w)$ is the area of the parallelogram spanned by v and w for w above the horizontal axis ($w_2 > 0$), and the negative of the area for w below this axis.

(b) If R_θ is rotation by an angle of θ (Problem 1), show that

$$\det(R_\theta v, R_\theta w) = \det(v, w).$$

Conclude that $\det(v, w)$ is the area of the parallelogram spanned by v and w when the rotation from v to w is counterclockwise, and the negative of the area when it is clockwise.

6. Show that

$$\det(v, w+z) = \det(v, w) + \det(v, z)$$
$$\det(v+w, z) = \det(v, z) + \det(w, z)$$

and that

$$a\det(v, w) = \det(a \cdot v, w) = \det(v, a \cdot w).$$

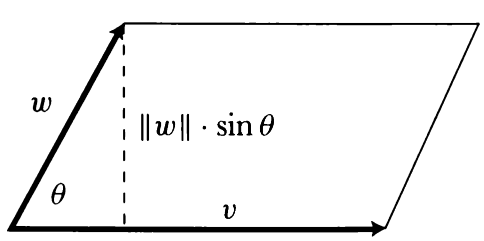


FIGURE 10

7. Using the method of Problem 3, show that

$$\det(v, w) = \|v\| \cdot \|w\| \cdot \sin\theta,$$

which is also obvious from the geometric interpretation (Figure 10).

APPENDIX 2. THE CONIC SECTIONS

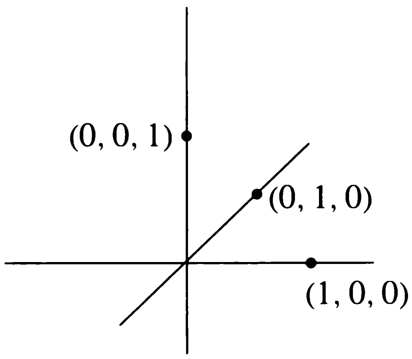


FIGURE 1

Although we will be concerned almost exclusively with figures in the plane, defined formally as the set of all pairs of real numbers, in this Appendix we want to consider three-dimensional space, which we can describe in terms of triples of real numbers, using a "three-dimensional coordinate system," consisting of three straight lines intersecting at right angles (Figure 1). Our *horizontal* and *vertical* axes now mutate to two axes in a horizontal plane, with the third axis perpendicular to both.

One of the simplest subsets of this three-dimensional space is the (infinite) *cone* illustrated in Figure 2; this cone may be produced by rotating a "generating line," of slope C say, around the third axis.

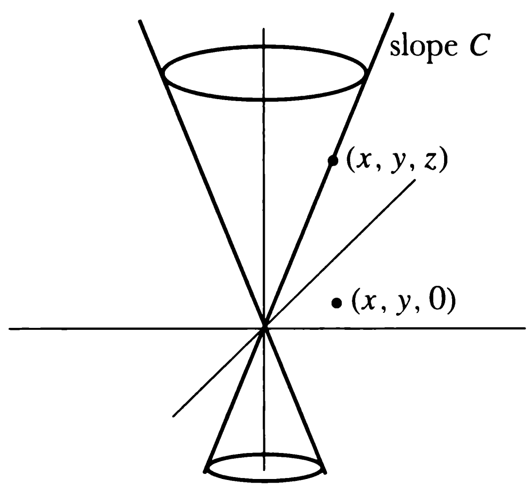


FIGURE 2

For any given first two coordinates x and y, the point $(x, y, 0)$ in the horizontal plane has distance $\sqrt{x^2 + y^2}$ from the origin, and thus

(1) $\qquad (x, y, z)$ is on the cone if and only if $z = \pm C\sqrt{x^2 + y^2}$.

We can descend from these three-dimensional vistas to the more familiar two-dimensional one by asking what happens when we intersect this cone with some plane P (Figure 3).

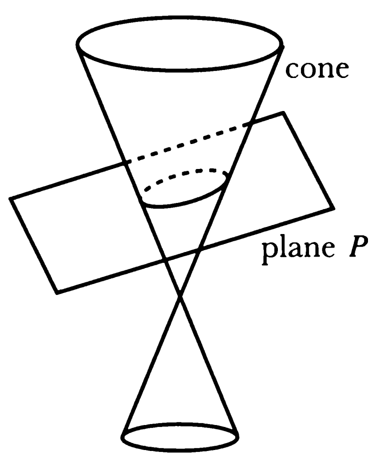


FIGURE 3

If the plane is parallel to the horizontal plane, there's certainly no mystery—the intersection is just a circle. Otherwise, the plane P intersects the horizontal plane in a straight line. We can make things a lot simpler for ourselves if we rotate everything around the vertical axis so that this intersection line points straight out from the plane of the paper, while the first axis is in the usual position that we are familiar with. The plane P is thus viewed "straight on," so that all we see (Figure 4) is its intersection L with the plane of the first and third axes; from this view-point the cone itself simply appears as two straight lines.

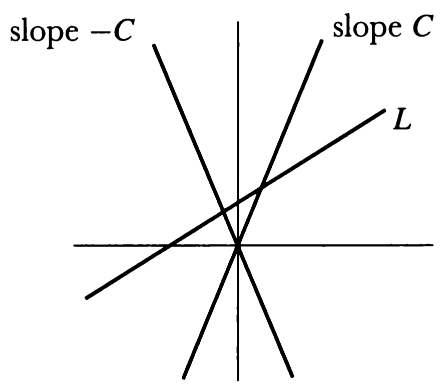


FIGURE 4

If this line L happens to be vertical, consisting of all points (a, z) for some a, then equation (1) says that the intersection of the cone and the plane consists of all points (a, y, z) with

$$z^2 - C^2y^2 = C^2a^2,$$

which is an hyperbola.

Otherwise, in the plane of the first and third axes, the line L can be described as the collection of all points of the form

$$(x, Mx + B),$$

where M is the slope of L. For an arbitrary point (x, y, z) it follows that

(2) $\qquad (x, y, z)$ is in the plane P if and only if $z = Mx + B$.

Combining (1) and (2), we see that (x, y, z) is in the intersection of the cone and the plane if and only if

$$(*) \qquad Mx + B = \pm C\sqrt{x^2 + y^2}.$$

Now we have to choose coordinate axes in the plane P. We can choose L as the first axis, measuring distances from the intersection Q with the horizontal plane (Figure 5); for the second axis we just choose the line through Q parallel to our original second axis. If the first coordinate of a point in P with respect to these axes is x, then the first coordinate of this point with respect to the original axes can be written in the form

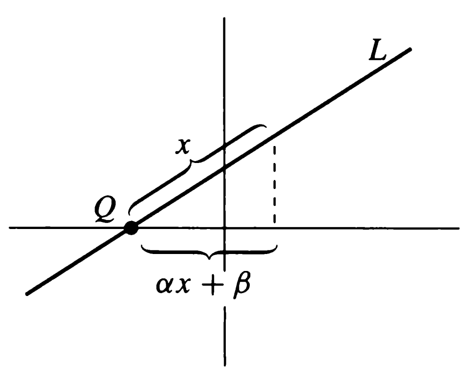


FIGURE 5

$$\alpha x + \beta$$

for some α and β. On the other hand, if the second coordinate of the point with respect to these axes is y, then y is also the second coordinate with respect to the original axes.

Consequently, $(*)$ says that the point lies on the intersection of the plane and the cone if and only if

$$M(\alpha x + \beta) + B = \pm C\sqrt{(\alpha x + \beta)^2 + y^2}.$$

Although this looks fairly complicated, after squaring we can write this as

$$C^2y^2 - \alpha^2(M^2 - C^2)x^2 + Ex + F = 0$$

for some E and F that we won't bother writing out.

Now Problem 4-16 indicates that this is either a parabola, an ellipse, or an hyperbola. In fact, looking a little more closely at the solution, we see that the values of E and F are irrelevant:

(1) If $M = \pm C$ we obtain a parabola;
(2) If $C^2 > M^2$ we obtain an ellipse;
(3) If $C^2 < M^2$ we obtain an hyperbola.

These analytic conditions are easy to interpret geometrically (Figure 6):

(1) If our plane is parallel to one of the generating lines of the cone we obtain a parabola;
(2) If our plane slopes less than the generating line of the cone (so that our intersection omits one half of the cone) we obtain an ellipse;
(3) If our plane slopes more than the generating line of the cone we obtain an hyperbola.

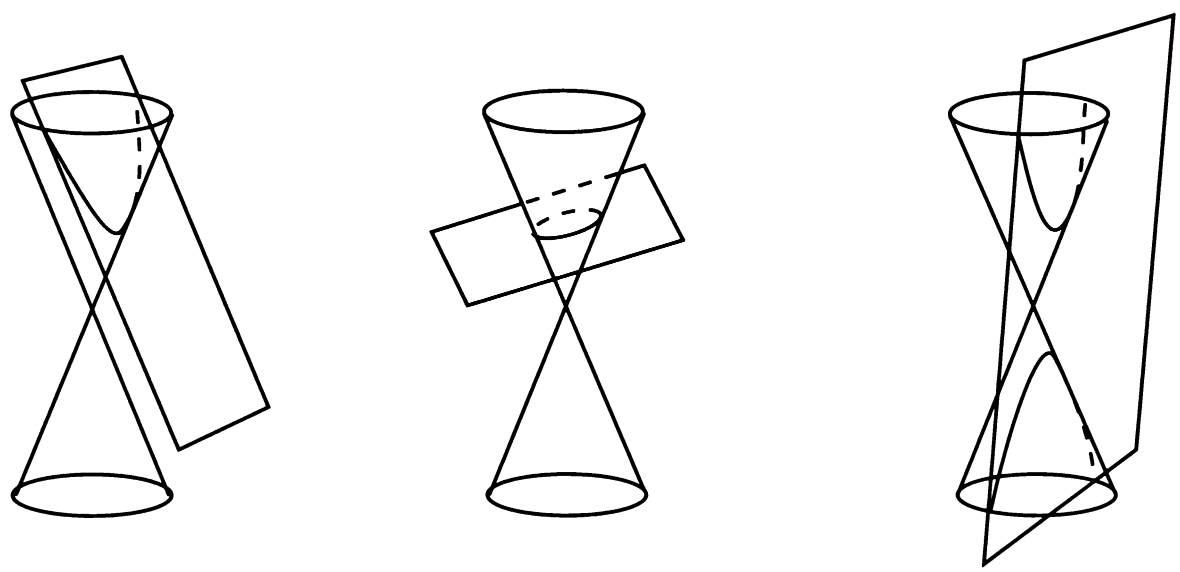


FIGURE 6

In fact, the very names of these "conic sections" are related to this description. The word *parabola* comes from a Greek root meaning 'alongside,' the same root that appears in parable, not to mention paradigm, paradox, paragon, paragraph, paralegal, parallax, parallel, and even parachute. *Ellipse* comes from a Greek root meaning 'defect,' or omission, as in ellipsis (an omission, ... or the dots that indicate it). And *hyperbola* comes from a Greek root meaning 'throwing beyond,' or excess. With the currency of words like hyperactive, hypersensitive, and hyperventilate, not to mention hype, one can probably say, without risk of hyperbole, that this root is familiar to almost everyone.*

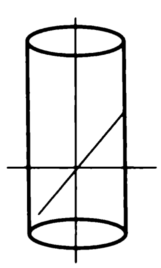


FIGURE 7

PROBLEMS

1. Consider a cylinder with a generator perpendicular to the horizontal plane (Figure 7); the only requirement for a point (x, y, z) to lie on this cylinder is

*Although the correspondence between these roots and the geometric picture correspond so beautifully, for the sake of dull accuracy it has to be reported that the Greeks originally applied the words to describe features of certain equations involving the conic sections.

that (x, y) lies on a circle:

$$x^2 + y^2 = C^2.$$

Show that the intersection of a plane with this cylinder can be described by an equation of the form

$$(\alpha x + \beta)^2 + y^2 = C^2.$$

What possibilities are there?

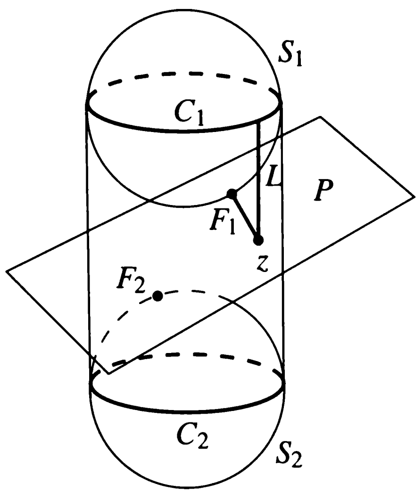


FIGURE 8

2. In Figure 8, the sphere S_1 has the same diameter as the cylinder, so that its equator C_1 lies along the cylinder; it is also tangent to the plane P at F_1. Similarly, the equator C_2 of S_2 lies along the cylinder, and S_2 is tangent to P at F_2.

(a) Let z be any point on the intersection of P and the cylinder. Explain why the length of the line from z to F_1 is equal to the length of the vertical line L from z to C_1.

(b) By proving a similar fact for the length of the line from z to F_2, show that the distance from z to F_1 plus the distance from z to F_2 is a constant, so that the intersection is an ellipse, with foci F_1 and F_2.

3. Similarly, use Figure 9(a) to prove geometrically that the intersection of a plane and a cone is an ellipse when the plane intersects just one half of the cone. Similarly, use (b) to prove that the intersection is an hyperbola when the plane intersects both halves of the cone.

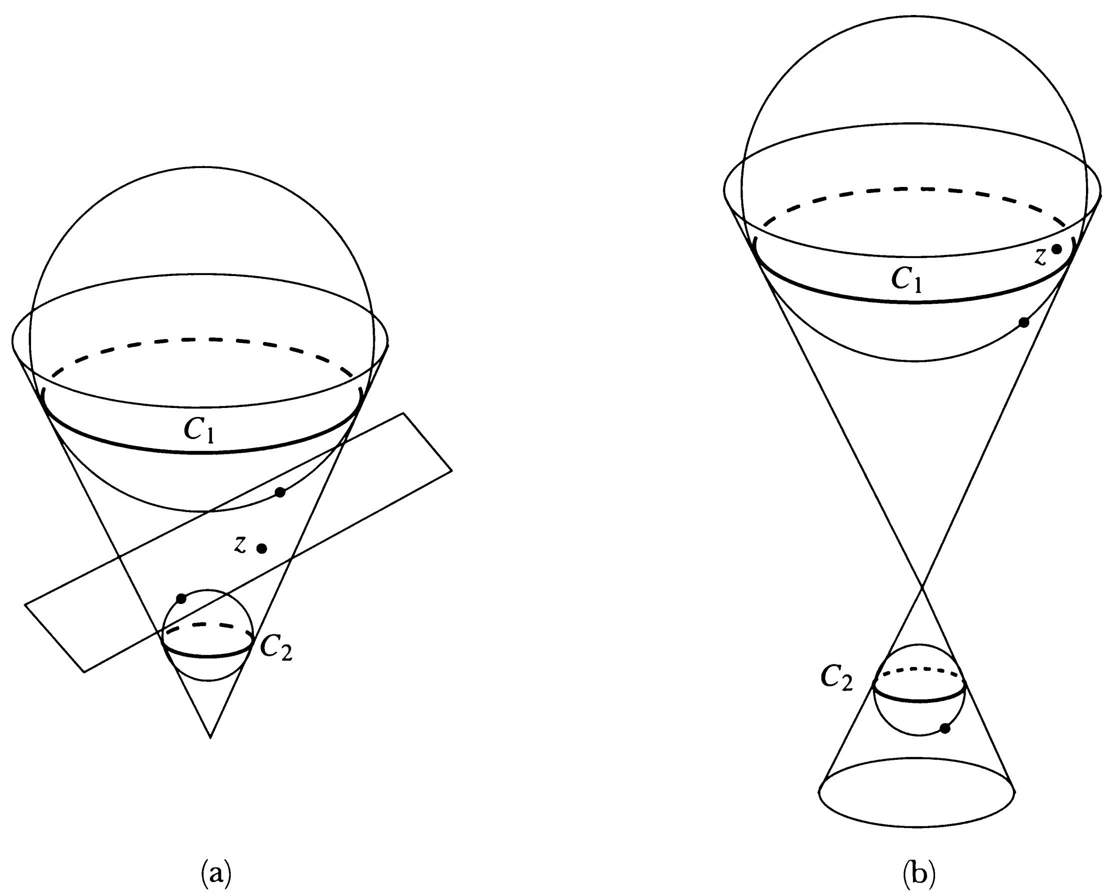


FIGURE 9

APPENDIX 3. POLAR COORDINATES

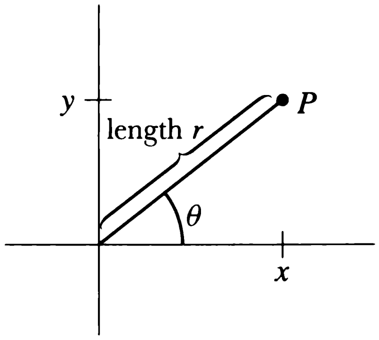
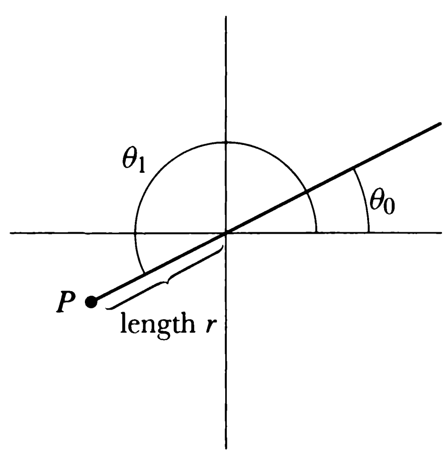


FIGURE 1

In this chapter we've been acting all along as if there's only one way to label points in the plane with pairs of numbers. Actually, there are many different ways, each giving rise to a different "coordinate system." The usual coordinates of a point are called its cartesian coordinates, after the French mathematician and philosopher René Descartes (1596–1650), who first introduced the idea of coordinate systems. In many situations it is more convenient to introduce polar coordinates, which are illustrated in Figure 1. To the point P we assign the polar coordinates (r, θ), where r is the distance from the origin O to P, and θ is the measure, in radians, of the angle between the horizontal axis and the line from O to P. This θ is not determined unambiguously. For example, points on the right side of the horizontal axis could have either $\theta = 0$ or $\theta = 2\pi$; moreover, θ is completely ambiguous at the origin O. So it is necessary to exclude some ray through the origin if we want to assign a unique pair (r, θ) to each point under consideration.

On the other hand, there is no problem associating a unique point to any pair (r, θ). In fact, it is possible (though not approved of by all) to associate a point to (r, θ) when $r < 0$, according to the scheme indicated in Figure 2. Thus, it always makes sense to talk about "the point with polar coordinates (r, θ)," (with or without the possibility of $r < 0$), even though there is some ambiguity when we talk about "the polar coordinates" of a given point.



P is the point with polar coordinates (r, θ_1) and also the point with polar coordinates $(-r, \theta_2)$.

FIGURE 2

It is clear from Figure 1 (and Figure 2) that the point with polar coordinates (r, θ) has cartesian coordinates (x, y) given by

$$x = r\cos\theta, \qquad y = r\sin\theta.$$

Conversely, if a point has cartesian coordinates (x, y), then (any of) its polar coordinates (r, θ) satisfy

$$r = \pm\sqrt{x^2 + y^2}$$
$$\tan\theta = \frac{y}{x} \qquad \text{if } x \neq 0.$$

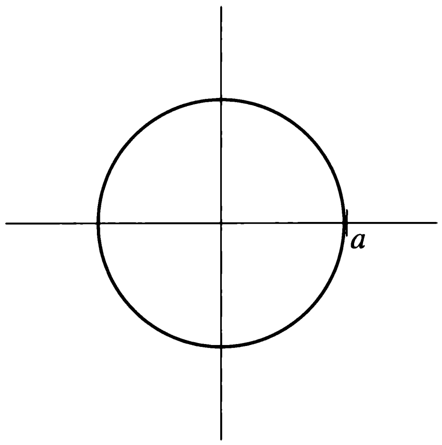


FIGURE 3

Now suppose that f is a function. Then by the **graph of f in polar coordinates** we mean the collection of all points P with polar coordinates (r, θ) satisfying $r = f(\theta)$. In other words, the graph of f in polar coordinates is the collection of all points with polar coordinates $(f(\theta), \theta)$. No special significance should be attached to the fact that we are considering pairs $(f(\theta), \theta)$, with $f(\theta)$ first, as opposed to pairs $(x, f(x))$ in the usual graph of f; it is purely a matter of convention that r is considered the first polar coordinate and θ is considered the second.

The graph of f in polar coordinates is often described as "the graph of the equation $r = f(\theta)$." For example, suppose that f is a constant function, $f(\theta) = a$ for all θ. The graph of the equation $r = a$ is simply a circle with center O and radius a (Figure 3). This example illustrates, in a rather blatant way, that polar coordinates are likely to make things simpler in situations that involve symmetry with respect to the origin O.

The graph of the equation $r = \theta$ is shown in Figure 4. The solid line corresponds to all values of $\theta \geq 0$, while the dashed line corresponds to values of $\theta \leq 0$.

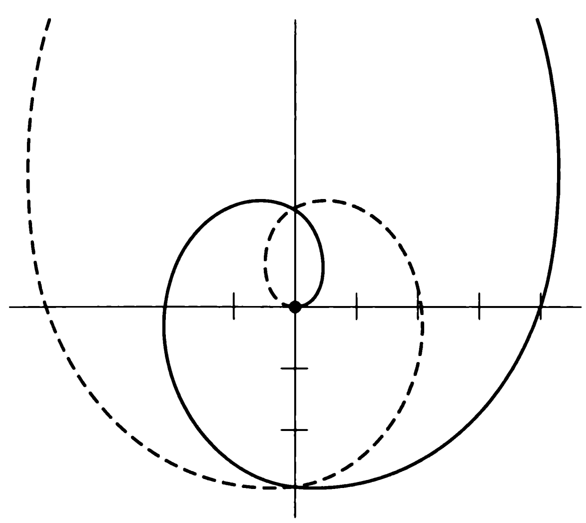
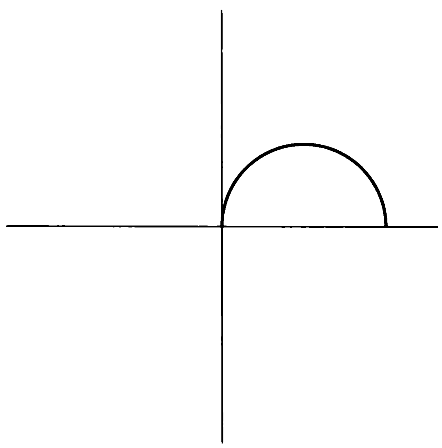
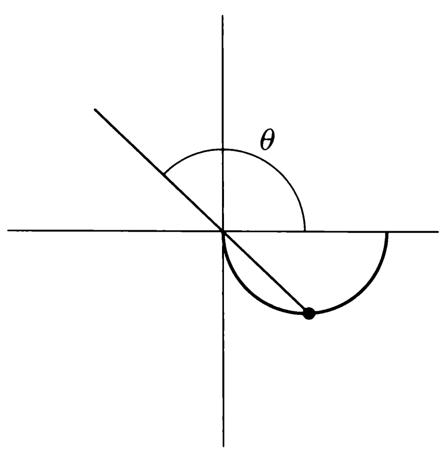


FIGURE 4 Spiral of Archimedes



(a)



(b)

FIGURE 5

As another example involving both positive and negative r, consider the graph of the equation $r = \cos\theta$. Figure 5(a) shows the part that corresponds to $0 \leq \theta \leq \pi/2$ Figure 5(b) shows the part corresponding to $\pi/2 \leq \theta \leq \pi$; here $r < 0$. You can check that no new points are added for $\theta > \pi$ or $\theta < 0$. It is easy to describe this same graph in terms of the cartesian coordinates of its points. Since the polar coordinates of any point on the graph satisfy

$$r = \cos\theta,$$

and hence

$$r^2 = r\cos\theta,$$

its cartesian coordinates satisfy the equation

$$x^2 + y^2 = x$$

which describes a circle (Problem 4-16). [Conversely, it is clear that if the cartesian coordinates of a point satisfy $x^2 + y^2 = x$, then it lies on the graph of the equation $r = \cos\theta$.]

Although we've now gotten a circle in two different ways, we might well be hesitant about trying to find the equation of an ellipse in polar coordinates. But it turns out that we can get a very nice equation if we choose one of the *foci* as the origin. Figure 6 shows an ellipse with one focus at O, with the sum of the distances of all points from O and the other focus $\mathbf{f}$ being $2a$. We've chosen $\mathbf{f}$ to the left of O, with coordinates written as

$$(-2\varepsilon a, 0).$$

(We have $0 \le \varepsilon < 1$, since we must have $2a >$ distance from $\mathbf{f}$ to O).

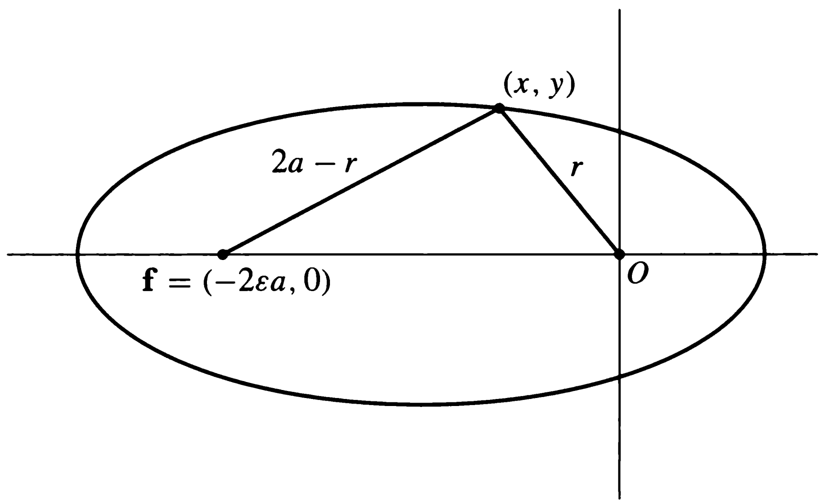


FIGURE 6

The distance r from (x, y) to O is given by

(1) $$r^2 = x^2 + y^2.$$

By assumption, the distance from (x, y) to $\mathbf{f}$ is $2a - r$, hence

$$(2a - r)^2 = (x - [-2\varepsilon a])^2 + y^2,$$

or

(2) $$4a^2 - 4ar + r^2 = x^2 + 4\varepsilon a x + 4\varepsilon^2 a^2 + y^2.$$

Subtracting (1) from (2), and dividing by $4a$, we get

$$a - r = \varepsilon x + \varepsilon^2 a,$$

or

$$\begin{aligned} r &= a - \varepsilon x - \varepsilon^2 a \\ &= (1 - \varepsilon^2)a - \varepsilon x, \end{aligned}$$

which we can write as

$$r = \Lambda - \varepsilon x, \qquad \text{for } \Lambda = (1 - \varepsilon^2)a. \tag{3}$$

Substituting $r \cos\theta$ for x, we have

$$\begin{aligned} r &= \Lambda - \varepsilon r \cos\theta, \\ r(1 + \varepsilon \cos\theta) &= \Lambda, \end{aligned}$$

and thus

$$r = \frac{\Lambda}{1 + \varepsilon \cos\theta}. \tag{4}$$

In Chapter 4 we found that

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1 \tag{5}$$

is the equation in cartesian coordinates for an ellipse with $2a$ as the sum of the distances to the foci, but with the foci at $(-c, 0)$ and $(c, 0)$, where

$$b = \sqrt{a^2 - c^2}.$$

Since the distance between the foci is $2c$, when this ellipse is moved left by c units, so that the focus $(c, 0)$ is now at the origin, we get the ellipse (4) when we take $c = \varepsilon a$ or $\varepsilon = c/a$ (with equation (3) determining Λ). Conversely, given the ellipse described by (4), for the corresponding equation (5) the value of a is determined by (3),

$$a = \frac{\Lambda}{1 - \varepsilon^2},$$

and again using $c = \varepsilon a$, we get

$$b = \sqrt{a^2 - c^2} = \sqrt{a^2 - \varepsilon^2 a^2} = a\sqrt{1 - \varepsilon^2} = \frac{\Lambda}{\sqrt{1 - \varepsilon^2}}.$$

Thus, we can obtain a and b, the lengths of the major and minor axes, immediately from ε and Λ.

The number

$$\varepsilon = \frac{c}{a} = \frac{\sqrt{a^2 - b^2}}{a} = \sqrt{1 - \left(\frac{b}{a}\right)^2},$$

the *eccentricity* of the ellipse, determines the "shape" of the ellipse (the ratio of the major and minor axes), while the number Λ determines its "size," as shown by (4).

PROBLEMS

1. If two points have polar coordinates (r_1, θ_1) and (r_2, θ_2), show that the distance d between them is given by

$$d^2 = {r_1}^2 + {r_2}^2 - 2r_1r_2\cos(\theta_1 - \theta_2).$$

What does this say geometrically?

2. Describe the general features of the graph of f in polar coordinates if

(i) f is even.
(ii) f is odd.
(iii) $f(\theta) = f(\theta + \pi)$.

3. Sketch the graphs of the following equations.

(i) $r = a\sin\theta$.
(ii) $r = a\sec\theta$. Hint: It is a very simple graph!
(iii) $r = \cos 2\theta$. Good luck on this one!
(iv) $r = \cos 3\theta$.
(v) $r = |\cos 2\theta|$.
(vi) $r = |\cos 3\theta|$.

4. Find equations for the cartesian coordinates of points on the graphs (i), (ii) and (iii) in Problem 3.

5. Consider a hyperbola, where the difference of the distance between the two foci is the constant $2a$, and choose one focus at O and the other at $(-2\varepsilon a, 0)$. (In this case, we must have $\varepsilon > 1$). Show that we obtain the exact same equation in polar coordinates

$$r = \frac{\Lambda}{1 + \varepsilon\cos\theta}$$

as we obtained for an ellipse.

6. Consider the set of points (x, y) such that the distance (x, y) to O is equal to the distance from (x, y) to the line $y = a$ (Figure 7). Show that the distance to the line is $a - r\cos\theta$, and conclude that the equation can be written

$$a = r(1 + \cos\theta).$$

Notice that this equation for a parabola is again of the same form as (4).

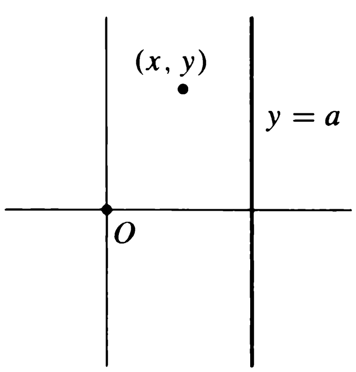


FIGURE 7

7. Now, for any Λ and ε, consider the graph in polar coordinates of the equation (4), which implies (3). Show that the points satisfying this equation satisfy

$$(1 - \varepsilon^2)x^2 + y^2 = \Lambda^2 - 2\Lambda\varepsilon x.$$

Using Problem 4-16, show that this is an ellipse for $\varepsilon < 1$, a parabola for $\varepsilon = 1$, and a hyperbola for $\varepsilon > 1$.

8. (a) Sketch the graph of the *cardioid* $r = 1 - \sin\theta$.
(b) Show that it is also the graph of $r = -1 - \sin\theta$.
(c) Show that it can be described by the equation

$$x^2 + y^2 = \sqrt{x^2 + y^2} - y,$$

and conclude that it can be described by the equation

$$(x^2 + y^2 + y)^2 = x^2 + y^2$$

9. Sketch the graphs of the following equations.

(i) $r = 1 - \frac{1}{2}\sin\theta$.
(ii) $r = 1 - 2\sin\theta$.
(iii) $r = 2 + \cos\theta$.

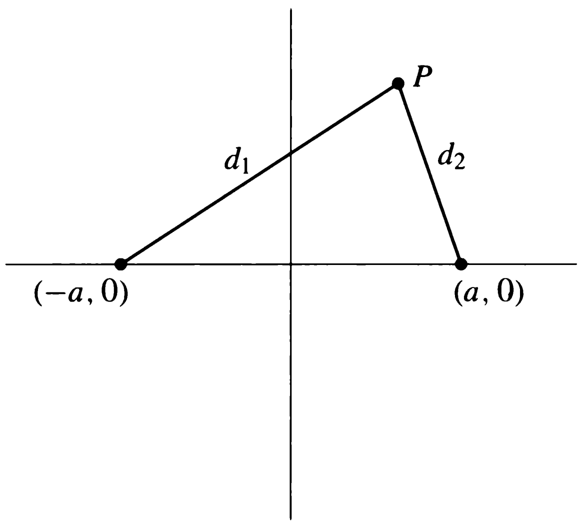


FIGURE 8

10. (a) Sketch the graph of the *lemniscate*

$$r^2 = 2a^2 \cos 2\theta.$$

(b) Find an equation for its cartesian coordinates.
(c) Show that it is the collection of all points P in Figure 8 satisfying $d_1 d_2 = a^2$.
(d) Make a guess about the shape of the curves formed by the set of all P satisfying $d_1 d_2 = b$, when $b > a^2$ and when $b < a^2$.