often the area function of the curve is also transcendental. Such functions cannot be fully investigated by algebra, and eventually one turns to calculus for further enlightenment. But there is no need to rush into advanced methods. The most common transcendental functions arise from conic sections and inherit many of their properties from elementary geometry and algebra. For example, the addition formulas for cos and sin are inherited from the geometry of the circle, as we saw in Section 5.3.

In the next two sections we shall see that the most important transcendental function, the exponential, inherits its basic properties from the geometry of the hyperbola. The exponential function can be defined as a function of area, and the area under the hyperbola has properties that are clear from geometry. Yet the exponential function also reveals properties of the hyperbola that are not otherwise obvious. In particular, we shall see that it highlights the integer points in a remarkable way, so once again there is an unexpected rapprochement between arithmetic and geometry.

Exercises

Arc length and area happen to be essentially the same function for the unit circle, because the area of a sector with angle θ is $\theta/2$. This can be seen by approximating the area by triangles with unit sides and angle θ/n.

9.1.1. Show that the area of an isosceles triangle with two unit sides and angle θ/n between them is $\frac{1}{2}\sin\frac{\theta}{n}$.

We have not yet given a general definition of area for curved figures, but the area of a sector is naturally interpreted as a limit of polygon areas.

9.1.2. Show, using Exercise 5.2.7, that $n \times \frac{1}{2}\sin\frac{\theta}{n} \to \frac{\theta}{2}$ as $n \to \infty$, and explain why this limit can be interpreted as the area of the unit sector with angle θ.

This result shows that the parameter θ in the equations $x = \cos\theta$, $y = \sin\theta$ can also be interpreted as (twice) an area. For other curves, the area and arc length functions are not so closely related, and area usually turns out to be more manageable. This is certainly true for the hyperbola, and it is the reason we use area to define the transcendental functions used in the remainder of this chapter.

9.2 The Area Bounded by a Curve

The idea of approximating a curve by polygons can be used to define the area of a curved region, just as easily as we define the length of a curve. The case that interests us is where the curve is the graph $y = f(x)$ of an algebraic function f, and the region is bounded by the curve, the x-axis, and the vertical lines $x = a$ and $x = b$ (Figure 9.1).

The area of the region is the least upper bound of the areas of all polygons Π contained in it. It exists if there is an upper bound to these polygon areas, as there is, for example, if there is an upper bound to $f(x)$ itself over the interval from a to b.

Students of calculus will recognize this definition as essentially the definite integral of $f(x)$ from a to b, but we shall not need calculus to derive the basic properties of area. For example, the following property of curved areas is inherited from the corresponding property of polygon areas: *if a region is magnified by a factor M in the x-direction and a factor N in the y-direction, then its area is magnified by a factor MN.* Magnification by MN is true for the areas of polygons Π (for example, by cutting them into triangles with bases in

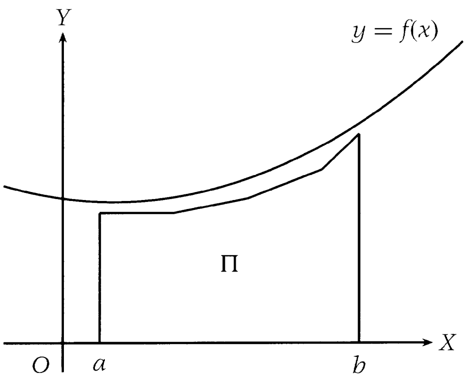

FIGURE 9.1 Area under a curve.

the x-direction and heights in the y-direction), and hence it is true for their least upper bound.

An important special case of this result is that magnification by M in all directions magnifies area by M^2. This is the underlying reason why the area of a circle of radius R is proportional to R^2.

Exercises

The first known determination of a curved area was made by Hippocrates of Chios around 430 B.C. He found the area of the region between the two circular arcs shown in Figure 9.2: one a quarter circle with radius OB, the other a semicircle with diameter AB. The region is called a *lune* because of its resemblance to a crescent moon, and Hippocrates showed that it has the same area as triangle AOB. Approximation by polygons is not required, except to show that the area of a circle is proportional to the square of its radius. (Hippocrates probably just assumed this; the idea of proving it rigorously using approximation by polygons is credited to Eudoxus, around 350 B.C.).

9.2.1. Show that the quarter circle with radius OB has the same area as the semicircle with diameter AB, and deduce Hippocrates' result.

Hippocrates made history by showing that curved areas are not beyond the grasp of mathematics, but his actual result is not very informative. It says nothing about the area of the circle, because the areas of the semicircle and quarter circle cancel out. Indeed, because we now

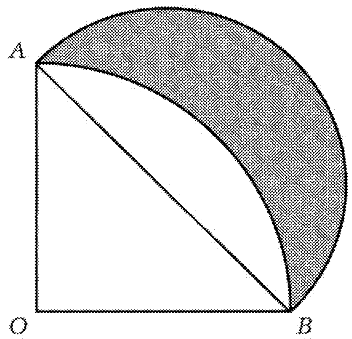

FIGURE 9.2 The lune of Hippocrates.

know that the area function of the circle is transcendental, an algebraic area (like that of triangle AOB) can only be obtained by subtracting one transcendental area from another.

The simplest curve whose area function happens to be algebraic is the parabola, and this area function was also the first to be discovered, by Archimedes around 250 B.C. A modern proof of his result may be based on approximation by polygons like those shown in Figure 9.3.

For simplicity, we take the parabola to be $y = x^2$ and find the area it bounds with the x-axis and the line $x = 1$. The proof is easily generalized to find the area up to an arbitrary line $x = a$.

9.2.2. Show that the area of a polygon with steps of width $1/n$ as in Figure 9.3 is $(1^2 + 2^2 + 3^2 + \cdots + (n-1)^2)/n^3$. We shall call this the nth *lower step polygon*.

This raises the problem of summing the series $1^2+2^2+3^2+\cdots+(n-1)^2$, which in fact was solved by Archimedes for another purpose. (Strangely enough, he used this series to find the area of a spiral.)

9.2.3. Show by induction on n that

$$1^2 + 2^2 + 3^2 + \cdots + (n-1)^2 = \frac{2n^3 - 3n^2 + n}{6}.$$

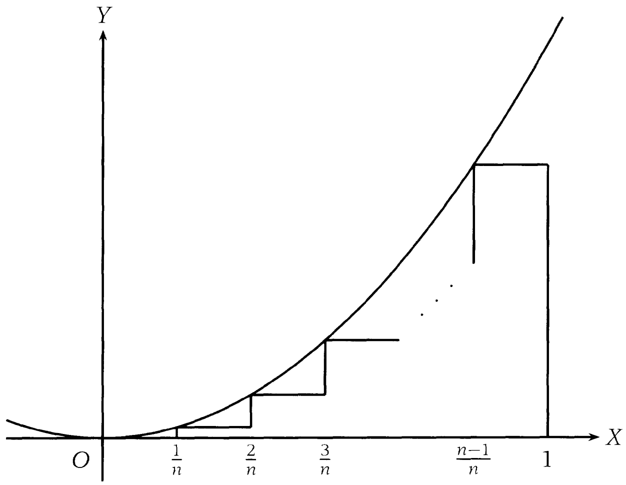

FIGURE 9.3 Area under a parabola.

9.2.4. Deduce from Exercise 9.2.3 that the area of the nth lower step polygon $\to 1/3$ as $n \to \infty$.

It looks like we have found the area under the curve between 0 and 1, but to be sure we should check that $\frac{1}{3}$ is really the least upper bound of the areas of all polygons in this region.

9.2.5. Find an nth *upper* step polygon that contains the region and differs from the nth lower step polygon by area $\frac{1}{n}$. Conclude that $\frac{1}{3}$ is the only number between the areas of lower and upper step polygons, and hence that it is the least upper bound of the areas of all polygons in the region below the curve between 0 and 1.

9.3 The Natural Logarithm and the Exponential

The algebraic area function for the parabola is another instance where the parabola is "just right." We already know that the circle has a transcendental area function, and the same is true of other ellipses. If we attempt to find the area under the hyperbola $xy = 1$, between $x = 1$ and $x = t$, say, we are in for another disappointment. The resulting function of t is also transcendental, so we cannot expect to "see" the least upper bound of polygonal areas, as we could for the parabola. But while the area function itself is complicated, some of its *properties* are simple because they are inherited from the hyperbola.

The area under $xy = 1$ from $x = 1$ to $x = t$ is called the *natural logarithm* of t and is written $\log t$. It follows that $\log 1 = 0$, and it is natural to suppose $\log t$ is negative for $t < 1$. We ensure this by taking the area with a negative sign when $0 < t < 1$. Figure 9.4 shows the graph of $\log t$ for $t > 0$. For the moment we do not attempt to define log for other values of t.

The most important property of the logarithm is the following, which follows very directly from the fact that $y = 1/x$.

Additive property of log. *If a and b are positive real numbers, then*

$$\log ab = \log a + \log b.$$

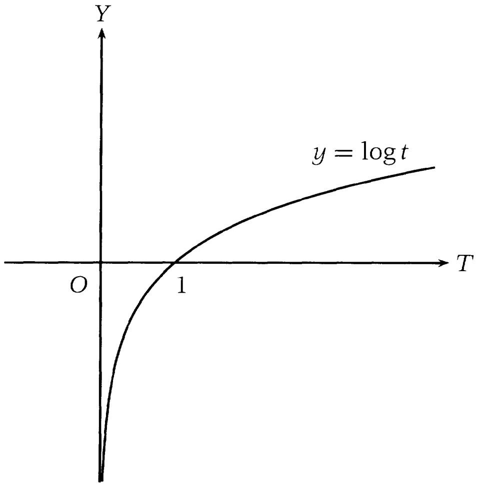

FIGURE 9.4 Graph of the logarithm function.

Proof We consider the area under the curve $xy = 1$ between $x = 1$ and $x = ab$, which is $\log ab$ by definition, and we split it by the line $x = a$ (Figure 9.5). This cuts off a region of area $\log a$, again by definition, so it remains to show that the region between $x = a$ and $x = ab$ has area $\log b$.

If we compare the region between a and ab with the region between 1 and b whose area is $\log b$ by definition (Figure 9.6), we see that the former region is a times as long as the latter. But it is also $1/a$ times as high. In fact, because $y = 1/x$, the height at the point $x = at$ between a and ab is $1/at$, which is $1/a$ times the height $1/t$ at the corresponding point $x = t$ between 1 and b.

Thus the region between a and ab is the result of magnifying a region of area $\log b$ by a factor a in the x-direction and a factor $1/a$ in the y-direction; hence its area is also $\log b$, as required.

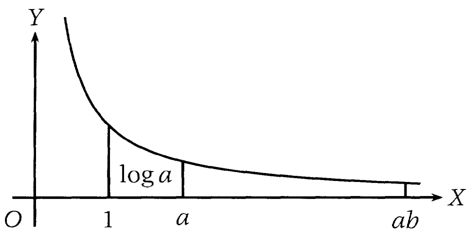

FIGURE 9.5 The area defining $\log a$.

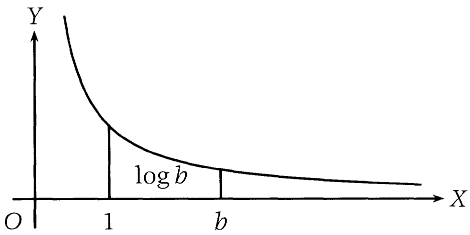

FIGURE 9.6 The area defining $\log b$.

This establishes the additive property when $a, b \geq 1$, but the argument is similar when either of them is between 0 and 1. One finds, for example, that the region from 1 to ab is smaller than the region from 1 to a when $a > 1$ and $b < 1$, and this is accounted for because the area from 1 to b has a negative sign. $\square$

It follows from the additive property of logarithms that $\log t$ behaves like an *exponent* of t, that is, as if $t = e^{\log t}$ for some number e. The number e is the value of t for which $\log t = 1$, and its basic properties (including its existence, which is not completely obvious) are included in the following theorem.

Consequences of the additive property. *The* log *function has the properties*

1. *For any real number a and integer n,* $\log a^n = n \log a$.
2. *The* log *function takes each real value exactly once.*
3. *There is a number e with* $\log e = 1$.
4. *Suppose a^r is defined, for any real number r, to be the number whose* log *is* $r \log a$. *Then $x = e^y$ if and only if $y = \log x$.*

Proof

1. When $n = 2, 3, \ldots$ the additive property gives

$$\log a^2 = \log a + \log a = 2 \log a,$$
$$\log a^3 = \log a^2 + \log a = 2 \log a + \log a = 3 \log a,$$

and so on (or more formally, by induction on n). When $n = 0$,

$$\log a^0 = \log 1 = 0,$$