CHAPTER 12 INVERSE FUNCTIONS

We now have at our disposal quite powerful methods for investigating functions; what we lack is an adequate supply of functions to which these methods may be applied. We have studied various ways of forming new functions from old—addition, multiplication, division, and composition—but using these alone, we can produce only the rational functions (even the sine function, although frequently used for examples, has never been defined). In the next few chapters we will begin to construct new functions in quite sophisticated ways, but there is one important method which will practically double the usefulness of any other method we discover.

If we recall that a function is a collection of pairs of numbers, we might hit upon the bright idea of simply reversing all the pairs. Thus from the function

$$f = \{ (1, 2), (3, 4), (5, 9), (13, 8) \},$$

we obtain

$$g = \{ (2, 1), (4, 3), (9, 5), (8, 13) \}.$$

While $f(1) = 2$ and $f(3) = 4$, we have $g(2) = 1$ and $g(4) = 3$.

Unfortunately, this bright idea does not always work. If

$$f = \{ (1, 2), (3, 4), (5, 9), (13, 4) \},$$

then the collection

$$\{ (2, 1), (4, 3), (9, 5), (4, 13) \}$$

is not a function at all, since it contains both $(4, 3)$ and $(4, 13)$. It is clear where the trouble lies: $f(3) = f(13)$, even though $3 \neq 13$. This is the only sort of thing that can go wrong, and it is worthwhile giving a name to the functions for which this does not happen.

DEFINITION

A function f is **one-one** (read "one-to-one") if $f(a) \neq f(b)$ whenever $a \neq b$.

The identity function I is obviously one-one, and so is the following modification:

$$g(x) = \begin{cases} x, & x \neq 3, 5 \\ 3, & x = 5 \\ 5, & x = 3. \end{cases}$$

The function $f(x) = x^2$ is not one-one, since $f(-1) = f(1)$, but if we define

$$g(x) = x^2, \qquad x \geq 0$$

(and leave g undefined for $x < 0$), then g is one-one, because g is increasing (since $g'(x) = 2x > 0$, for $x > 0$). This observation is easily generalized: If n is a natural number and

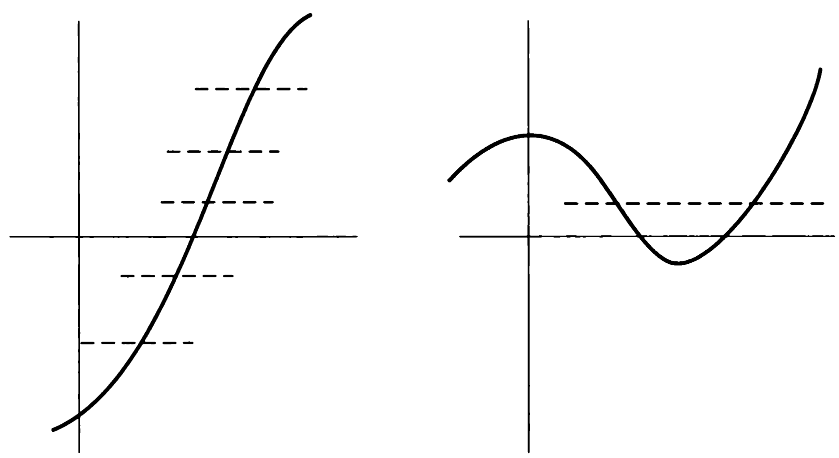
$$f(x) = x^n, \qquad x \geq 0,$$

then f is one-one. If n is odd, one can do better: the function

$$f(x) = x^n \qquad \text{for all } x$$

is one-one (since $f'(x) = nx^{n-1} > 0$, for all $x \neq 0$).

It is particularly easy to decide from the graph of f whether f is one-one: the condition $f(a) \neq f(b)$ for $a \neq b$ means that no *horizontal* line intersects the graph of f twice (Figure 1).



a one-one function (a) a function that is not one-one (b)

FIGURE 1

If we reverse all the pairs in (a not necessarily one-one function) f we obtain, in any case, some collection of pairs. It is popular to abstain from this procedure unless f is one-one, but there is no particular reason to do so—instead of a definition with restrictive conditions we obtain a definition and a theorem.

DEFINITION For any function f, the **inverse** of f, denoted by $\boldsymbol{f^{-1}}$, is the set of all pairs (a, b) for which the pair (b, a) is in f.

THEOREM 1 f^{-1} is a function if and only if f is one-one.

PROOF Suppose first that f is one-one. Let (a, b) and (a, c) be two pairs in f^{-1}. Then (b, a) and (c, a) are in f, so $a = f(b)$ and $a = f(c)$; since f is one-one this implies that $b = c$. Thus f^{-1} is a function.

Conversely, suppose that f^{-1} is a function. If $f(b) = f(c)$, then f contains the pairs $(b, f(b))$ and $(c, f(c)) = (c, f(b))$, so $(f(b), b)$ and $(f(b), c)$ are in f^{-1}. Since f^{-1} is a function this implies that $b = c$. Thus f is one-one. ▮

The graphs of f and f^{-1} are so closely related that it is possible to use the graph of f to visualize the graph of f^{-1}. Since the graph of f^{-1} consists of all pairs (a, b) with (b, a) in the graph of f, one obtains the graph of f^{-1} from the graph of f by interchanging the horizontal and vertical axes. If f has the graph shown in Figure 2(a),

FIGURE 2(b)

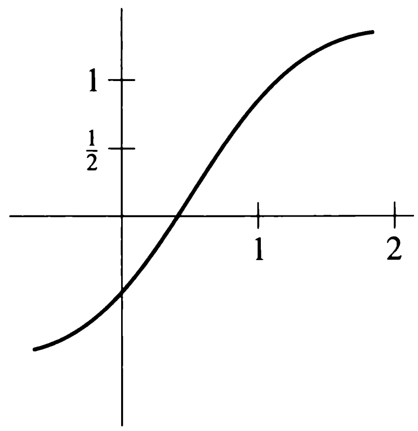


FIGURE 2(a)

and you rotate this page counter clockwise through a right angle, then the graph of f^{-1} appears on your left (Figure 2(b)). The only trouble is that the numbering on the horizontal axis goes in the wrong direction, so you must flip this picture over to get the usual picture of f^{-1}, which appears on your right (Figure 3).

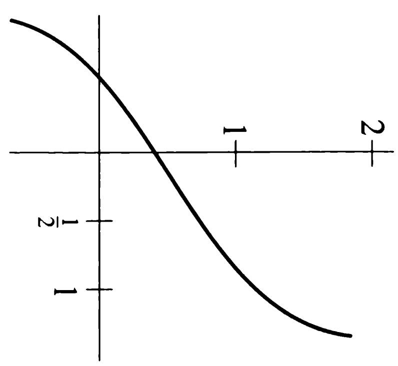


FIGURE 3

This procedure is awkward with books and impossible with blackboards, so it is fortunate that there is another way of constructing the graph of f^{-1}. The points

(a, b) and (b, a) are reflections of each other through the graph of $I(x) = x$, which is called the **diagonal** (Figure 4). To obtain the graph of f^{-1} we merely reflect the graph of f through this line (Figure 5).

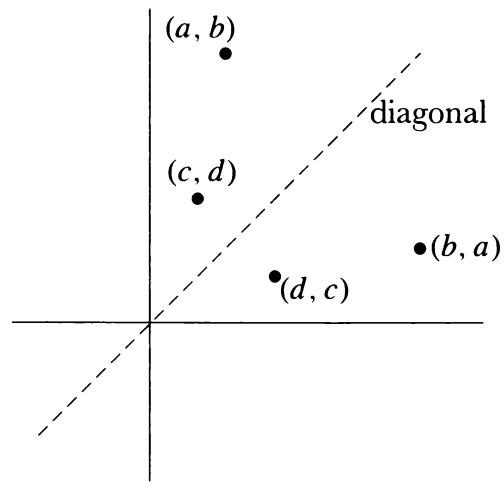


FIGURE 4

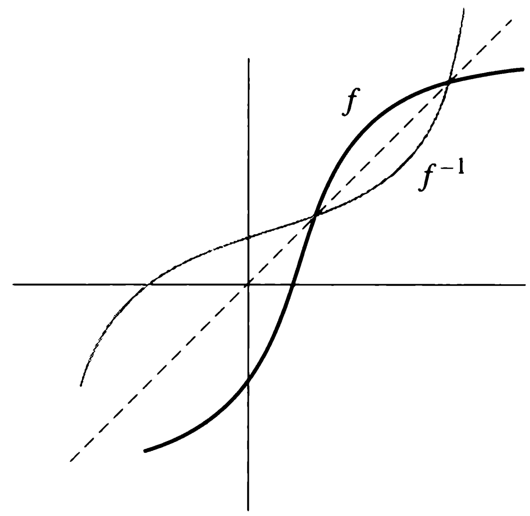


FIGURE 5

Reflecting through the diagonal twice will clearly leave us right back where we started; this means that $(f^{-1})^{-1} = f$, which is also clear from the definition. In conjunction with Theorem 1, this equation has a significant consequence: if f is a one-one function, then the function f^{-1} is also one-one (since $(f^{-1})^{-1}$ is a function).

There are a few other simple manipulations with inverse functions of which you should be aware. Since (a, b) is in f precisely when (b, a) is in f^{-1}, it follows that

$$b = f(a) \qquad \text{means the same as} \qquad a = f^{-1}(b).$$

Thus $f^{-1}(b)$ is the (unique) number a such that $f(a) = b$; for example, if $f(x) = x^3$, then $f^{-1}(b)$ is the unique number a such that $a^3 = b$, and this number is, by definition, $\sqrt[3]{b}$.

The fact that $f^{-1}(x)$ is the number y such that $f(y) = x$ can be restated in a much more compact form:

$$f(f^{-1}(x)) = x, \qquad \text{for all } x \text{ in the domain of } f^{-1}.$$

Moreover,

$$f^{-1}(f(x)) = x, \qquad \text{for all } x \text{ in the domain of } f;$$

this follows from the previous equation upon replacing f by f^{-1}. These two important equations can be written

$$\begin{aligned} f \circ f^{-1} &= I, \\ f^{-1} \circ f &= I \end{aligned}$$

(except that the right side will have a bigger domain if the domain of f or f^{-1} is not all of $\mathbf{R}$).

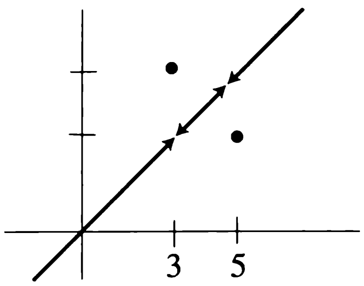


FIGURE 6

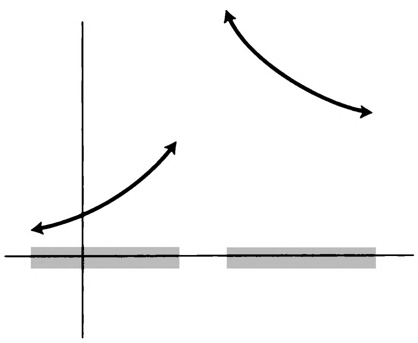


FIGURE 7

Since many standard functions will be defined as the inverses of other functions, it is quite important that we be able to tell which functions are one-one. We have already hinted which class of functions are most easily dealt with—increasing and decreasing functions are obviously one-one. Moreover, if f is increasing, then f^{-1} is also increasing, and if f is decreasing, then f^{-1} is decreasing (the proof is left to you). In addition, f is increasing if and only if $-f$ is decreasing, a very useful fact to remember.

It is certainly not true that every one-one function is either increasing or decreasing. One example has already been mentioned, and is now graphed in Figure 6:

$$g(x) = \begin{cases} x, & x \neq 3, 5 \\ 3, & x = 5 \\ 5, & x = 3. \end{cases}$$

Figure 7 shows that there are even continuous one-one functions which are neither increasing nor decreasing. But if you try drawing a few pictures you will soon suspect that every one-one continuous function defined on an interval is either increasing or decreasing.

THEOREM 2 If f is continuous and one-one on an interval, then f is either increasing or decreasing on that interval.

PROOF The proof proceeds in three easy steps:

(1) If $a < b < c$ are three points in the interval, then

$$\begin{array}{lll} \text{either} & \text{(i)} & f(a) < f(b) < f(c) \\ \text{or} & \text{(ii)} & f(a) > f(b) > f(c). \end{array}$$

Suppose, for example, that $f(a) < f(c)$. If we had $f(b) < f(a)$ (Figure 8), then the Intermediate Value Theorem applied to the interval $[b, c]$ would give an x with

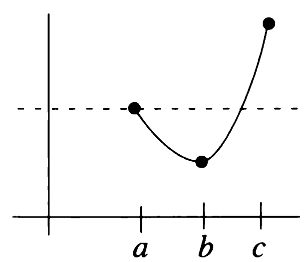


FIGURE 8

$b < x < c$ and $f(x) = f(a)$, contradicting the fact that f is one-one on $[a, c]$. Similarly, $f(b) > f(c)$ would lead to a contradiction, so $f(a) < f(b) < f(c)$.

Naturally, the same sort of argument works for the case $f(a) > f(c)$.

(2) If $a < b < c < d$ are four points in the interval, then

$$\begin{array}{lll} \text{either} & \text{(i)} & f(a) < f(b) < f(c) < f(d) \\ \text{or} & \text{(ii)} & f(a) > f(b) > f(c) > f(d). \end{array}$$

For we can apply (1) to $a < b < c$ and then to $b < c < d$.

(3) Take any $a < b$ in the interval, and suppose that $f(a) < f(b)$. Then f is increasing: For if c and d are any two points, we can apply (2) to the collection $\{a, b, c, d\}$ (after arranging them in increasing order). ∎

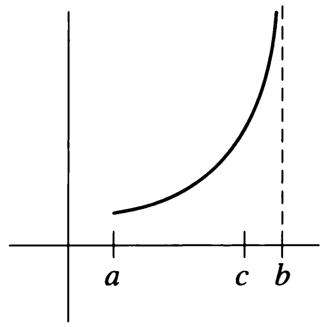


FIGURE 9

Henceforth we shall be concerned almost exclusively with continuous increasing or decreasing functions which are defined on an interval. If f is such a function, it is possible to say quite precisely what the domain of f^{-1} will be like.

Suppose first that f is a continuous increasing function on the closed interval $[a, b]$. Then, by the Intermediate Value Theorem, f takes on every value between $f(a)$ and $f(b)$. Therefore, the domain of f^{-1} is the closed interval $[f(a), f(b)]$. Similarly, if f is continuous and decreasing on $[a, b]$, then the domain of f^{-1} is $[f(b), f(a)]$.

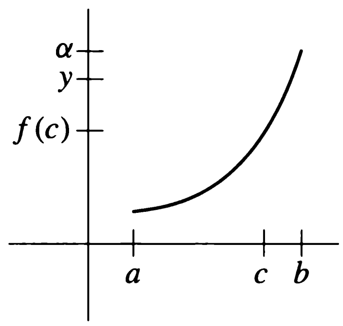


FIGURE 10

If f is a continuous increasing function on an *open* interval (a, b) the analysis becomes a bit more difficult. To begin with, let us choose some point c in (a, b). We will first decide which values $> f(c)$ are taken on by f. One possibility is that f takes on arbitrarily large values (Figure 9). In this case f takes on *all* values $> f(c)$, by the Intermediate Value Theorem. If, on the other hand, f does not take on arbitrarily large values, then $A = \{f(x) : c \le x < b\}$ is bounded above, so A has a least upper bound α (Figure 10). Now suppose y is any number with $f(c) < y < \alpha$. Then f takes on some value $f(x) > y$ (because α is the least upper bound of A). By the Intermediate Value Theorem, f actually takes on the value y. Notice that f cannot take on the value α itself; for if $\alpha = f(x)$ for $a < x < b$ and we choose t with $x < t < b$, then $f(t) > \alpha$, which is impossible.

Precisely the same arguments work for values less than $f(c)$: either f takes on all values less than $f(c)$ or there is a number $\beta < f(c)$ such that f takes on all values between β and $f(c)$, but not β itself.

This entire argument can be repeated if f is decreasing, and if the domain of f is $\mathbf{R}$ or (a, ∞) or $(-\infty, a)$. Summarizing: if f is a continuous increasing, or decreasing, function whose domain is an interval having one of the forms (a, b), $(-\infty, b)$, (a, ∞), or $\mathbf{R}$, then the domain of f^{-1} is also an interval which has one of these four forms, and we can easily fit the remaining types of intervals, $(a, b]$, $[a, b]$, $(-\infty, b]$, and $[a, \infty)$, into this discussion also.

Now that we have completed this preliminary analysis of continuous one-one functions, it is possible to begin asking which important properties of a one-one function are inherited by its inverse. For continuity there is no problem.

THEOREM 3 If f is continuous and one-one on an interval, then f^{-1} is also continuous.

PROOF We know by Theorem 2 that f is either increasing or decreasing. We might as well assume that f is increasing, since we can then take care of the other case by applying the usual trick of considering $-f$. We might as well assume our interval is open, since it is easy to see that a continuous increasing or decreasing function on any interval can be extended to one on a larger open interval.

We must show that

$$\lim_{x \to b} f^{-1}(x) = f^{-1}(b)$$

for each b in the domain of f^{-1}. Such a number b is of the form $f(a)$ for some a in the domain of f. For any $\varepsilon > 0$, we want to find a $\delta > 0$ such that, for all x,

$$\text{if } f(a) - \delta < x < f(a) + \delta, \text{ then } a - \varepsilon < f^{-1}(x) < a + \varepsilon.$$

Figure 11 suggests the way of finding δ (remember that by looking sideways you see the graph of f^{-1}): since

$$a - \varepsilon < a < a + \varepsilon,$$

it follows that

$$f(a - \varepsilon) < f(a) < f(a + \varepsilon);$$

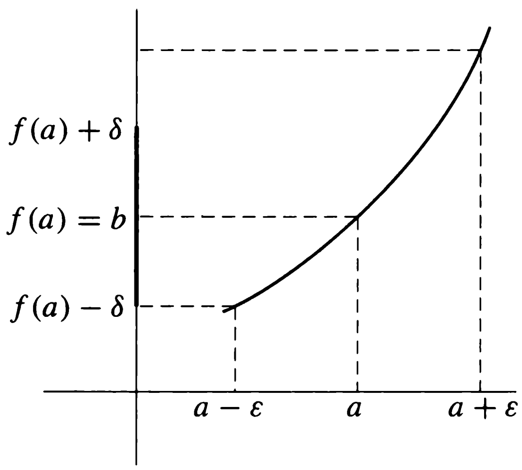


FIGURE 11

we let δ be the smaller of $f(a+\varepsilon) - f(a)$ and $f(a) - f(a-\varepsilon)$. Figure 11 contains the entire proof that this δ works, and what follows is simply a verbal account of the information contained in this picture.

Our choice of δ ensures that

$$f(a - \varepsilon) \le f(a) - \delta \text{ and } f(a) + \delta \le f(a + \varepsilon).$$

Consequently, if

$$f(a) - \delta < x < f(a) + \delta,$$

then

$$f(a - \varepsilon) < x < f(a + \varepsilon).$$

Since f is increasing, f^{-1} is also increasing, and we obtain

$$f^{-1}(f(a - \varepsilon)) < f^{-1}(x) < f^{-1}(f(a + \varepsilon)),$$

i.e.,

$$a - \varepsilon < f^{-1}(x) < a + \varepsilon,$$

which is precisely what we want. ▮

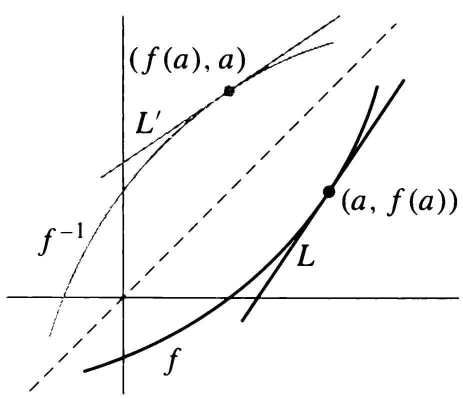


FIGURE 12

Having successfully investigated continuity of f^{-1}, it is only reasonable to tackle differentiability. Again, a picture indicates just what result ought to be true. Figure 12 shows the graph of a one-one function f with a tangent line L through $(a, f(a))$. If this entire picture is reflected through the diagonal, it shows the graph of f^{-1} and the tangent line L' through $(f(a), a)$. The slope of L' is the reciprocal of the slope of L. In other words, it appears that

$$(f^{-1})'(f(a)) = \frac{1}{f'(a)}.$$

This formula can equally well be written in a way which expresses $(f^{-1})'(b)$ directly, for each b in the domain of f^{-1}:

$$(f^{-1})'(b) = \frac{1}{f'(f^{-1}(b))}.$$

Unlike the argument for continuity, this pictorial "proof" becomes somewhat involved when formulated analytically. There is another approach which might be tried. Since we know that

$$f(f^{-1}(x)) = x,$$

it is tempting to prove the desired formula by applying the Chain Rule:

$$f'(f^{-1}(x)) \cdot (f^{-1})'(x) = 1,$$

so

$$(f^{-1})'(x) = \frac{1}{f'(f^{-1}(x))}.$$

Unfortunately, this is not a proof that f^{-1} is differentiable, since the Chain Rule cannot be applied unless f^{-1} is already known to be differentiable. But this argument does show what $(f^{-1})'(x)$ will have to be if f^{-1} *is* differentiable, and it can also be used to obtain some important preliminary information.

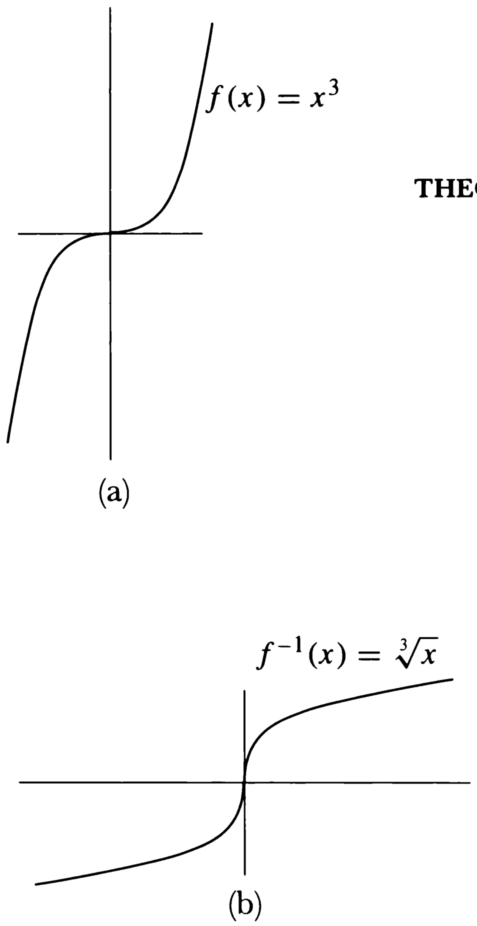


FIGURE 13

THEOREM 4 If f is a continuous one-one function defined on an interval and $f'(f^{-1}(a)) = 0$, then f^{-1} is *not* differentiable at a.

PROOF We have

$$f(f^{-1}(x)) = x.$$

If f^{-1} *were* differentiable at a, the Chain Rule would imply that

$$f'(f^{-1}(a)) \cdot (f^{-1})'(a) = 1,$$

hence

$$0 \cdot (f^{-1})'(a) = 1,$$

which is absurd. ▮

A simple example to which Theorem 4 applies is the function $f(x) = x^3$. Since $f'(0) = 0$ and $0 = f^{-1}(0)$, the function f^{-1} is not differentiable at 0 (Figure 13).

Having decided where an inverse function cannot be differentiable, we are now ready for the rigorous proof that in all other cases the derivative is given by the formula which we have already "derived" in two different ways. Notice that the following argument uses *continuity* of f^{-1}, which we have already proved.

THEOREM 5 Let f be a continuous one-one function defined on an interval, and suppose that f is differentiable at $f^{-1}(b)$, with derivative $f'(f^{-1}(b)) \neq 0$. Then f^{-1} is differentiable at b, and

$$(f^{-1})'(b) = \frac{1}{f'(f^{-1}(b))}.$$

PROOF Let $b = f(a)$. Then

$$\lim_{h\to 0} \frac{f^{-1}(b+h) - f^{-1}(b)}{h}$$
$$= \lim_{h\to 0} \frac{f^{-1}(b+h) - a}{h}$$

Now every number $b+h$ in the domain of f^{-1} can be written in the form

$$b+h = f(a+k)$$

for a unique k (we should really write $k(h)$, but we will stick with k for simplicity). Then

$$\begin{aligned}\lim_{h\to 0}\frac{f^{-1}(b+h)-a}{h} \\ &= \lim_{h\to 0}\frac{f^{-1}(f(a+k))-a}{f(a+k)-b} \\ &= \lim_{h\to 0}\frac{k}{f(a+k)-f(a)}.\end{aligned}$$

We are clearly on the right track! It is not hard to get an explicit expression for k; since

$$b+h = f(a+k)$$

we have

$$f^{-1}(b+h) = a+k$$

or

$$k = f^{-1}(b+h) - f^{-1}(b).$$

Now by Theorem 3, the function f^{-1} is continuous at b. This means that k approaches 0 as h approaches 0. Since

$$\lim_{k\to 0}\frac{f(a+k)-f(a)}{k} = f'(a) = f'(f^{-1}(b)) \neq 0,$$

this implies that

$$(f^{-1})'(b) = \frac{1}{f'(f^{-1}(b))}. \blacksquare$$

The work we have done on inverse functions will be amply repaid later, but here is an immediate dividend. For n odd, let

$$f_n(x) = x^n \qquad \text{for all } x;$$

for n even, let

$$f_n(x) = x^n, \qquad x \geq 0.$$

Then f_n is a continuous one-one function, whose inverse function is

$$g_n(x) = \sqrt[n]{x} = x^{1/n}.$$

By Theorem 5 we have, for $x \neq 0$,

$$\begin{aligned} g_n{}'(x) &= \frac{1}{f_n{}'(f_n{}^{-1}(x))} \\ &= \frac{1}{n(f_n{}^{-1}(x))^{n-1}} \\ &= \frac{1}{n(x^{1/n})^{n-1}} \\ &= \frac{1}{n} \cdot \frac{1}{x^{1-(1/n)}} \\ &= \frac{1}{n} \cdot x^{(1/n)-1}. \end{aligned}$$

Thus, if $f(x) = x^a$, and a is an integer or the reciprocal of a natural number, then $f'(x) = ax^{a-1}$. It is now easy to check that this formula is true if a is any rational number: Let $a = m/n$, where m is an integer, and n is a natural number; if

$$f(x) = x^{m/n} = \left(x^{1/n}\right)^m,$$

then, by the Chain Rule,

$$\begin{aligned} f'(x) &= m\left(x^{1/n}\right)^{m-1} \cdot \frac{1}{n} \cdot x^{(1/n)-1} \\ &= \frac{m}{n} \cdot x^{[(m/n)-(1/n)]+[(1/n)-1]} \\ &= \frac{m}{n} x^{(m/n)-1}. \end{aligned}$$

Although we now have a formula for $f'(x)$ when $f(x) = x^a$ and a is rational, the treatment of the function $f(x) = x^a$ for irrational a will have to be saved for later—at the moment we do not even know the *meaning* of a symbol like $x^{\sqrt{2}}$. Actually, inverse functions will be involved crucially in the definition of x^a for irrational a. Indeed, in the next few chapters several important functions will be defined in terms of their inverse functions.

PROBLEMS

1. Find f^{-1} for each of the following f.

(i) $f(x) = x^3 + 1$.

(ii) $f(x) = (x-1)^3$.

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

(iv) $f(x) = \begin{cases} -x^2 & x \geq 0 \\ 1 - x^3, & x < 0. \end{cases}$

(v) $f(x) = \begin{cases} x, & x \neq a_1, \dots, a_n \\ a_{i+1} & x = a_i, \quad i = 1, \dots, n-1 \\ a_1, & x = a_n. \end{cases}$

(vi) $f(x) = x + [x]$.

(vii) $f(0.a_1a_2a_3\ldots) = 0.a_2a_1a_3\ldots$. (Decimal representation is being used.)

(viii) $f(x) = \dfrac{x}{1-x^2}$, $-1 < x < 1$.

2. Describe the graph of f^{-1} when

(i) f is increasing and always positive.
(ii) f is increasing and always negative.
(iii) f is decreasing and always positive.
(iv) f is decreasing and always negative.

3. Prove that if f is increasing, then so is f^{-1}, and similarly for decreasing functions.

4. If f and g are increasing, is $f+g$? Or $f \cdot g$? Or $f \circ g$?

5. (a) Prove that if f and g are one-one, then $f \circ g$ is also one-one. Find $(f \circ g)^{-1}$ in terms of f^{-1} and g^{-1}. Hint: The answer is *not* $f^{-1} \circ g^{-1}$.
(b) Find g^{-1} in terms of f^{-1} if $g(x) = 1 + f(x)$.

6. Show that $f(x) = \dfrac{ax+b}{cx+d}$ is one-one if and only if $ad - bc \neq 0$, and find f^{-1} in this case.

7. On which intervals $[a, b]$ will the following functions be one-one?

(i) $f(x) = x^3 - 3x^2$.
(ii) $f(x) = x^5 + x$.
(iii) $f(x) = (1 + x^2)^{-1}$.
(iv) $f(x) = \dfrac{x+1}{x^2+1}$.

8. Suppose that f is differentiable with derivative $f'(x) = (1 + x^3)^{-1/2}$. Show that $g = f^{-1}$ satisfies $g''(x) = \frac{3}{2}g(x)^2$.

9. Suppose that f is a one-one function and that f^{-1} has a derivative which is nowhere 0. Prove that f is differentiable. Hint: There is a one-step proof.

10. As a follow up to Problem 10-17, what additional condition on g will insure that f is differentiable?

11. Find a formula for $(f^{-1})''(x)$.

***12.** Prove that if $f'(f^{-1}(x)) \neq 0$ and $f^{(k)}(f^{-1}(x))$ exists, then $(f^{-1})^{(k)}(x)$ exists.

13. The Schwarzian derivative $\mathscr{D}f$ was defined in Problem 10-19.

(a) Prove that if $\mathscr{D}f(x)$ exists for all x, then $\mathscr{D}f^{-1}(x)$ also exists for all x in the domain of f^{-1}.
(b) Find a formula for $\mathscr{D}f^{-1}(x)$.

***14.** (a) Prove that there is a differentiable function f such that $[f(x)]^5 + f(x) + x = 0$ for all x. Hint: Show that f can be expressed as an inverse function. The easiest way to do this is to find f^{-1}. And the easiest way to do *this* is to set $x = f^{-1}(y)$.
(b) Find f' in terms of f, using an appropriate theorem of this chapter.
(c) Find f' in another way, by simply differentiating the equation defining f.

The function in Problem 14 is often said to be **defined implicitly** by the equation $y^5+y+x = 0$. The situation for this equation is quite special, however. As the next problem shows, an equation does not usually define a function implicitly on the whole line, and in some regions more than one function may be defined implicitly.

15. (a) What are the two differentiable functions f which are defined implicitly on $(-1, 1)$ by the equation $x^2+y^2 = 1$, i.e., which satisfy $x^2+[f(x)]^2 = 1$ for all x in $(-1, 1)$? Notice that there are no solutions defined outside $[-1, 1]$.
(b) Which functions f satisfy $x^2 + [f(x)]^2 = -1$?
*(c) Which differentiable functions f satisfy $[f(x)]^3 - 3f(x) = x$? Hint: It will help to first draw the graph of the function $g(x) = x^3 - 3x$.

In general, determining on what intervals a differentiable function is defined implicitly by a particular equation may be a delicate affair, and is best discussed in the context of advanced calculus. If we *assume* that f is such a differentiable solution, however, then a formula for $f'(x)$ can be derived, exactly as in Problem 14(c), by differentiating both sides of the equation defining f (a process known as "implicit differentiation"):

16. (a) Apply this method to the equation $[f(x)]^2 + x^2 = 1$. Notice that your answer will involve $f(x)$; this is only to be expected, since there is more than one function defined implicitly by the equation $y^2 + x^2 = 1$.
(b) But check that your answer works for both of the functions f found in Problem 15(a).
(c) Apply this same method to $[f(x)]^3 - 3f(x) = x$.

17. (a) Use implicit differentiation to find $f'(x)$ and $f''(x)$ for the functions f defined implicitly by the equation $x^3 + y^3 = 7$.
(b) One of these functions f satisfies $f(-1) = 2$. Find $f'(-1)$ and $f''(-1)$ for this f.

18. The collection of all points (x, y) such that $3x^3 + 4x^2y - xy^2 + 2y^3 = 4$ forms a certain curve in the plane. Find the equation of the tangent line to this curve at the point $(-1, 1)$.

19. Leibnizian notation is particularly convenient for implicit differentiation. Because y is so consistently used as an abbreviation for $f(x)$, the equation in x and y which defines f implicitly will automatically stand for the equation

(c) Prove that if f is an increasing function such that $f = f^{-1}$, then $f(x) = x$ for all x. Hint: Although the geometric interpretation will be immediately convincing, the simplest proof (about 2 lines) is to rule out the possibilities $f(x) < x$ and $f(x) > x$.

***25.** Which functions have the property that the graph is still the graph of a function when reflected through the graph of $-I$ (the "antidiagonal")?

26. A function f is **nondecreasing** if $f(x) \leq f(y)$ whenever $x < y$. (To be more precise we should stipulate that the domain of f be an interval.) A **nonincreasing** function is defined similarly. Caution: Some writers use "increasing" instead of "nondecreasing," and "strictly increasing" for our "increasing."

(a) Prove that if f is nondecreasing, but not increasing, then f is constant on some interval. (Beware of unintentional puns: "not increasing" is not the same as "nonincreasing.")
(b) Prove that if f is differentiable and nondecreasing, then $f'(x) \geq 0$ for all x.
(c) Prove that if $f'(x) \geq 0$ for all x, then f is nondecreasing.

***27.** (a) Suppose that $f(x) > 0$ for all x, and that f is decreasing. Prove that there is a *continuous* decreasing function g such that $0 < g(x) \leq f(x)$ for all x.
(b) Show that we can even arrange that g will satisfy $\lim_{x \to \infty} g(x)/f(x) = 0$.

which f is supposed to satisfy. How would the following computation be written in our notation?

$$y^4 + y^3 + xy = 1,$$

$$4y^3\frac{dy}{dx} + 3y^2\frac{dy}{dx} + y + x\frac{dy}{dx} = 0,$$

$$\frac{dy}{dx} = \frac{-y}{4y^3 + 3y^2 + x}.$$

20. As long as Leibnizian notation has entered the picture, the Leibnizian notation for derivatives of inverse functions should be mentioned. If dy/dx denotes the derivative of f, then the derivative of f^{-1} is denoted by dx/dy. Write out Theorem 5 in this notation. The resulting equation will show you another reason why Leibnizian notation has such a large following. It will also explain at which point $(f^{-1})'$ is to be calculated when using the dx/dy notation. What is the significance of the following computation?

$$x = y^n,$$

$$y = x^{1/n},$$

$$\frac{dx^{1/n}}{dx} = \frac{dy}{dx} = \frac{1}{\dfrac{dx}{dy}} = \frac{1}{ny^{n-1}}.$$

21. Suppose that f is a differentiable one-one function with a nowhere zero derivative and that $f = F'$. Let $G(x) = xf^{-1}(x) - F(f^{-1}(x))$. Prove that $G'(x) = f^{-1}(x)$. (Disregarding details, this problem tells us a very interesting fact: if we know a function whose derivative is f, then we also know one whose derivative is f^{-1}. But how could anyone ever guess the function G? Two different ways are outlined in Problems 14-14 and 19-16.)

22. Suppose h is a function such that $h'(x) = \sin^2(\sin(x + 1))$ and $h(0) = 3$. Find

(i) $(h^{-1})'(3)$.
(ii) $(\beta^{-1})'(3)$, where $\beta(x) = h(x + 1)$.

23. (a) Prove that an increasing and a decreasing function intersect at most once.
(b) Find two continuous increasing functions f and g such that $f(x) = g(x)$ precisely when x is an integer.
(c) Find a continuous increasing function f and a continuous decreasing function g, defined on $\mathbf{R}$, which do not intersect at all.

***24.** (a) If f is a continuous function on $\mathbf{R}$ and $f = f^{-1}$, prove that there is at least one x such that $f(x) = x$. (What does the condition $f = f^{-1}$ mean geometrically?)
(b) Give several examples of continuous f such that $f = f^{-1}$ and $f(x) = x$ for exactly one x. Hint: Try decreasing f, and remember the geometric interpretation. One possibility is $f(x) = -x$.

APPENDIX. PARAMETRIC REPRESENTATION OF CURVES

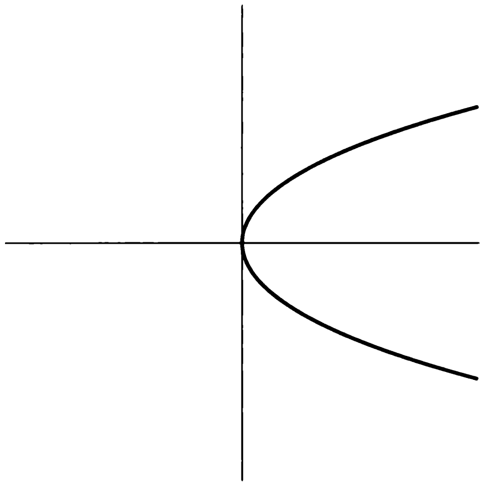


FIGURE 1

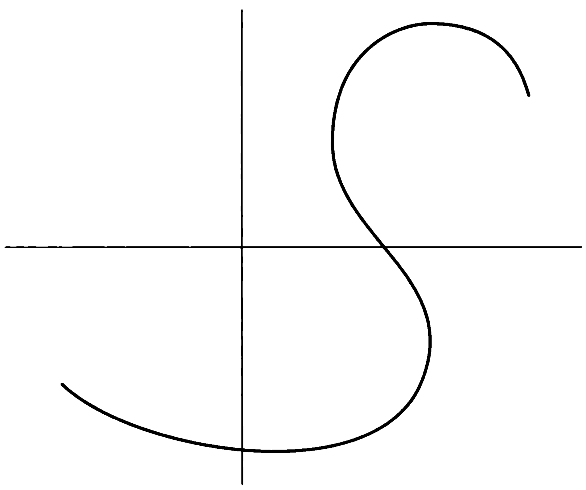


FIGURE 2

The material in this chapter serves to emphasize something that we noticed a long time ago—a perfectly nice looking curve need not be the graph of a function (Figure 1). In other words, we may not be able to describe it as the set of all points $(x, f(x))$. Of course, we might be able to describe the curve as the set of all points $(f(x), x)$; for example, the curve in Figure 1 is the set of all points (x^2, x). But even this trick doesn't work in most cases. It won't allow us to describe the circle, consisting of all points (x, y) with $x^2 + y^2 = 1$, or an ellipse, and it can't be used to describe a curve like the one in Figure 2.

The simplest way of describing curves in the plane in general harks back to the physical conception of a curve as the path of a particle moving in the plane. At each time t, the particle is at a certain point, which has two coordinates; to indicate the dependence of these coordinates on the time t, we can call them $u(t)$ and $v(t)$. Thus, we end up with *two* functions. Conversely, given two functions u and v, we can consider the curve consisting of all points $(u(t), v(t))$. This curve is said to be represented *parametrically* by u and v, and the pair of functions u, v is called a parametric representation of the curve. The curve represented parametrically by u and v thus consists of all pairs (x, y) with $x = u(t)$ and $y = v(t)$. It is often described briefly as "the curve $x = u(t)$, $y = v(t)$." Notice that the graph of a function f can always be described parametrically, as the curve $x = t$, $y = f(t)$.

Instead of considering a curve in the plane as defined by two functions, we can obtain a conceptually simpler picture if we broaden our original definition of function somewhat. Instead of considering a rule which associates a number with another number, we can consider a "function c from real numbers to the plane," i.e., a rule c that associates, to each number t, a *point in the plane*, which we can denote by $c(t)$. With this notion, a curve is just a function from some interval of real numbers to the plane.

Of course, these two different descriptions of a curve are essentially the same: A pair of (ordinary) functions u and v determines a single function c from the real numbers to the plane by the rule

$$c(t) = (u(t), v(t)),$$

and, conversely, given a function c from the real numbers to the plane, each $c(t)$ is a point in the plane, so it is a pair of numbers, which we can call $u(t)$ and $v(t)$, so that we have unique functions u and v satisfying this equation.

In Appendix 1 to Chapter 4, we used the term "vector" to describe a point in the plane. In conformity with this usage, a curve in the plane may also be called a "vector-valued function." The conventions of that Appendix would lead us to write $c(t) = (c_1(t), c_2(t))$, but in this Appendix we'll continue to use notation like $c(t) = (u(t), v(t))$ to minimize the use of subscripts.

A simple example of a vector-valued function that is quite useful is

$$\mathbf{e}(t) = (\cos t, \sin t),$$

which goes round and round the unit circle (Figure 3).

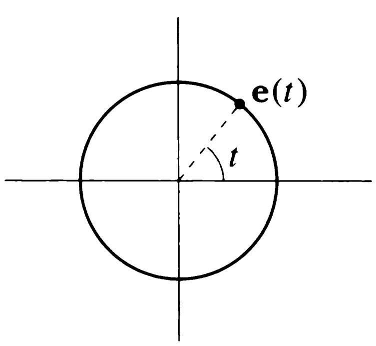


FIGURE 3

For two (ordinary) functions f and g, we defined new functions $f+g$ and $f \cdot g$ by the rules

$$(1) \qquad (f+g)(x) = f(x) + g(x),$$

$$(2) \qquad (f \cdot g)(x) = f(x) \cdot g(x).$$

Since we have defined a way of adding vectors, we can imitate the first of these definitions for vector-valued functions c and d: we define the vector-valued function $c+d$ by

$$(c+d)(t) = c(t) + d(t),$$

where the $+$ on the right-hand side is now *the sum of vectors.* This simply amounts to saying that if

$$c(t) = (u(t), v(t)),$$
$$d(t) = (w(t), z(t)),$$

then

$$(c+d)(t) = \big(u(t), v(t)\big) + \big(w(t), z(t)\big) = \big(u(t) + w(t), v(t) + z(t)\big).$$

Recall that we have also defined $a \cdot v$ for a number a and a vector v. To extend this to vector-valued functions, we want to consider an ordinary *function* α and a vector-valued function c, so that for each t we have a number $\alpha(t)$ and a vector $c(t)$. Then we can define a new vector-valued function $\alpha \cdot c$ by

$$(\alpha \cdot c)(t) = \alpha(t) \cdot c(t),$$

where the $\cdot$ on the right-hand side is the product of a number and a vector. This simply amounts to saying that

$$(\alpha \cdot c)(t) = \alpha(t) \cdot (u(t), v(t)) = \big(\alpha(t) \cdot u(t),\ \alpha(t) \cdot v(t)\big).$$

Notice that the curve $\alpha \cdot \mathbf{e}$,

$$(\alpha \cdot \mathbf{e})(t) = (\alpha(t)\cos t,\ \alpha(t)\sin t),$$

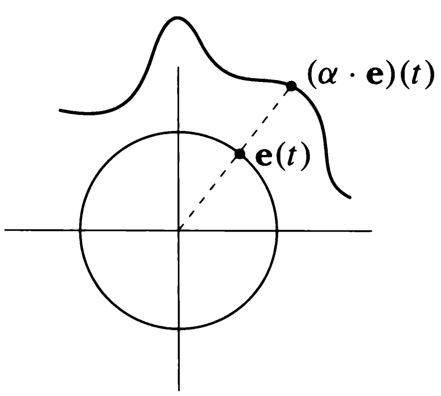


FIGURE 4

is already quite general (Figure 4). In the notation of Appendix 3 to Chapter 4, the point $(\alpha \cdot \mathbf{e})(t)$ has polar coordinates $\alpha(t)$ and t, so that $(\alpha \cdot \mathbf{e})(t)$ is the "graph of α in polar coordinates."

Even more generally, given any vector-valued function c, we can define new functions r and θ by

$$c(t) = r(t) \cdot \mathbf{e}(\theta(t)),$$

where $r(t)$ is just the distance from the origin to $c(t)$, and $\theta(t)$ is some choice of the angle of $c(t)$ (as usual, the function θ isn't defined unambiguously, so one has to be careful when using this way of writing an arbitrary curve c).

We aren't in a position to extend (2) to vector-valued functions in general, since we haven't defined the product of two vectors. However, Problems 2 and 4 of Appendix 1 to Chapter 4 define two *real-valued* products $v \bullet w$ and $\det(v, w)$. It should be clear, given vector-valued functions c and d, how we would define two ordinary (real-valued) functions

$$c \cdot d \qquad \text{and} \qquad \det(c, d).$$

Beyond imitating simple arithmetic operations on functions, we can consider more interesting problems, like limits. For $c(t) = (u(t), v(t))$, we can define

$$(*) \qquad \lim_{t \to a} c(t) = \lim_{t \to a} (u(t), v(t)) \quad \text{to be} \quad \left(\lim_{t \to a} u(t), \lim_{t \to a} v(t) \right).$$

Rules like

$$\lim_{t \to a} c + d = \lim_{t \to a} c + \lim_{t \to a} d,$$
$$\lim_{t \to a} \alpha \cdot c = \lim_{t \to a} \alpha(t) \cdot \lim_{t \to a} c$$

follow immediately. Problem 10 shows how to give an equivalent definition that imitates the basic definition of limits directly.

Limits lead us of course to derivatives. For

$$c(t) = (u(t), v(t))$$

we can define c' by the straightforward definition

$$c'(a) = \big(u'(a), v'(a)\big).$$

We could also try to imitate the basic definition:

$$c'(a) = \lim_{h \to 0} \frac{c(a+h) - c(a)}{h},$$

where the fraction on the right-hand side is understood to mean

$$\frac{1}{h} \cdot [c(a+h) - c(a)].$$

As a matter of fact, these two definitions are equivalent, because

$$\begin{aligned} \lim_{h \to 0} \frac{c(a+h) - c(a)}{h} &= \lim_{h \to 0} \left(\frac{u(a+h) - u(a)}{h}, \frac{v(a+h) - v(a)}{h} \right) \\ &= \left(\lim_{h \to 0} \frac{u(a+h) - u(a)}{h}, \lim_{h \to 0} \frac{v(a+h) - v(a)}{h} \right) \\ &\qquad \text{by our definition } (*) \text{ of limits} \\ &= \big(u'(a), v'(a)\big). \end{aligned}$$

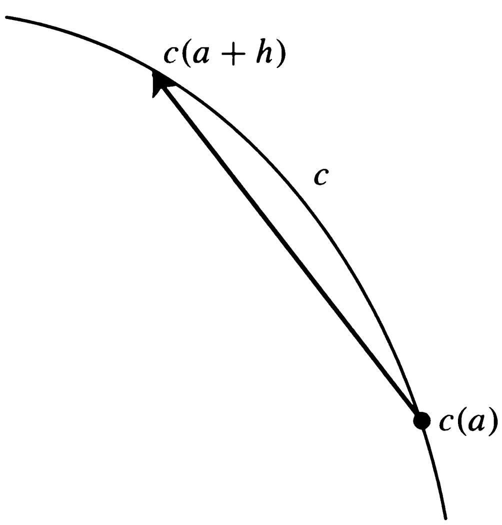


FIGURE 5

Figure 5 shows $c(a+h)$ and $c(a)$, as well as the arrow from $c(a)$ to $c(a+h)$; as we showed in Appendix 1 to Chapter 4, this arrow is $c(a+h) - c(a)$, except moved over so that it starts at $c(a)$. As $h \to 0$, this arrow would appear to move closer and closer to the tangent of our curve, so it seems reasonable to *define* the tangent line of c at $c(a)$ to be the straight line along $c'(a)$, when $c'(a)$ is moved over so that it starts at $c(a)$. In other words, we define the tangent line of c at $c(a)$ as the set of all points

$$c(a) + s \cdot c'(a);$$

for $s = 0$ we get the point $c(a)$ itself, for $s = 1$ we get $c(a) + c'(a)$, etc. (Note, however, that this definition does not make sense when $c'(a) = (0, 0)$.) Problem 1 shows that this definition agrees with the old one when our curve c is defined by

$$c(t) = (t, f(t)),$$

so that we simply have the graph of f.

Once again, various old formulas have analogues. For example,

$$\begin{aligned}(c + d)'(a) &= c'(a) + d'(a),\\ (\alpha \cdot c)'(a) &= \alpha'(a) \cdot c(a) + \alpha(a) \cdot c'(a),\end{aligned}$$

or, as equations involving functions,

$$\begin{aligned}(c + d)' &= c' + d',\\ (\alpha \cdot c)' &= \alpha' \cdot c + \alpha \cdot c'.\end{aligned}$$

These formulas can be derived immediately from the definition in terms of the component functions. They can also be derived from the definition as a limit, by imitating previous proofs; for the second, we would of course use the standard trick of writing

$$\begin{aligned}&\alpha(a + h)c(a + h) - \alpha(a)c(a) =\\ &\qquad \alpha(a + h) \cdot [c(a + h) - c(a)] + [\alpha(a + h) - \alpha(a)] \cdot c(a).\end{aligned}$$

We can also consider the function

$$d(t) = c(p(t)) = (c \circ p)(t),$$

where p is now an ordinary function, from numbers to numbers. The new curve d passes through the same points as c, except at different times; thus p corresponds to a "reparameterization" of c. For

$$\begin{aligned}c &= (u, v),\\ d &= (u \circ p, v \circ p),\end{aligned}$$

we obtain

$$\begin{aligned}d'(a) &= \big((u \circ p)'(a), (v \circ p)'(a)\big)\\ &= \big(p'(a)u'(p(a)), p'(a)v'(p(a))\big)\\ &= p'(a) \cdot \big(u'(p(a)), v'(p(a))\big)\\ &= p'(a) \cdot c'(p(a)),\end{aligned}$$

or simply

$$d' = p' \cdot (c' \circ p).$$

Notice that if $p(a) = a$, so that d and c actually pass through the same point at time a, then $d'(a) = p'(a) \cdot c'(a)$, so that the tangent vector $d'(a)$ is just a multiple of $c'(a)$. This means that the tangent *line* to c at $c(a)$ is the same as the tangent line to the reparameterized curve d at $d(a) = c(a)$. The one exception occurs when $p'(a) = 0$, since the tangent line for d is then undefined, even though the tangent line for c may be defined. For example, $d(t) = c(t^3)$ won't have a tangent line defined at $t = 0$, even though it's merely a reparameterization of c.

Finally, since we can define real-valued functions

$$(c \cdot d)(t) = c(t) \cdot d(t),$$
$$\det(c, d)(t) = \det(c(t), d(t)),$$

we ought to have formulas for the derivatives of these new functions. As you might guess, the proper formulas are

$$(c \cdot d)'(a) = c(a) \cdot d'(a) + c'(a) \cdot d(a),$$
$$[\det(c, d)]'(a) = \det(c', d)(a) + \det(c, d')(a).$$

These can be derived by straightforward calculations from the definitions in terms of the component functions. But it is more elegant to imitate the proof of the ordinary product rule, using the simple formulas in Problems 2 and 4 of Appendix 1 to Chapter 4, and, of course, the "standard trick" referred to above.

PROBLEMS

1. (a) For a function f, the "point-slope form" (Problem 4-6) of the tangent line at $(a, f(a))$ can be written as $y - f(a) = (x - a)f'(a)$, so that the tangent line consists of all points of the form

$$\big(x, f(a) + (x - a)f'(a)\big).$$

Conclude that the tangent line consists of all points of the form

$$\big(a + s, f(a) + sf'(a)\big).$$

(b) If c is the curve $c(t) = (t, f(t))$, conclude that the tangent line of c at $(a, f(a))$ [using our new definition] is the same as the tangent line of f at $(a, f(a))$.

2. Let $c(t) = (f(t), t^2)$, where f is the function shown in Figure 21 of Chapter 9. Show that c lies along the graph of the non-differentiable function $h(x) = |x|$, but that $c'(0) = (0, 0)$. In other words, a reparameterization can "hide" a corner. For this reason, we are usually only interested in curves c with c' never equal to $(0, 0)$.

3. Suppose that $x = u(t)$, $y = v(t)$ is a parametric representation of a curve, and that u is one-one on some interval.

(a) Show that on this interval the curve lies along the graph of $f = v \circ u^{-1}$.

(b) If u is differentiable on this interval and $u'(t) \neq 0$, show that at the point $x = u(t)$ we have

$$f'(x) = \frac{v'(t)}{u'(t)}.$$

In Leibnizian notation this is often written suggestively as

$$\frac{dy}{dx} = \frac{\dfrac{dy}{dt}}{\dfrac{dx}{dt}}.$$

(c) We also have

$$f''(x) = \frac{u'(t)v''(t) - v'(t)u''(t)}{(u'(t))^3}.$$

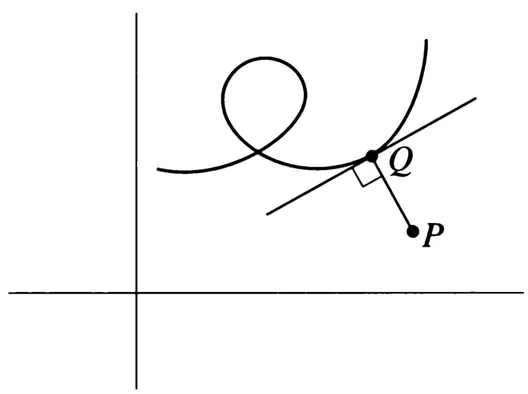


FIGURE 6

4. Consider a function f defined implicitly by the equation $x^{2/3} + y^{2/3} = 1$. Compute $f'(x)$ in two ways:

(i) By implicit differentiation.

(ii) By considering the parametric representation $x = \cos^3 t$, $y = \sin^3 t$.

5. Let $x = u(t)$, $y = v(t)$ be the parametric representation of a curve, with u and v differentiable, and let $P = (x_0, y_0)$ be a point in the plane. Prove that if the point $Q = (u(\bar{t}), v(\bar{t}))$ on the curve is closest to (x_0, y_0), and $u'(\bar{t})$ and $v'(\bar{t})$ are not both 0, then the line from P to Q is perpendicular to the tangent line of the curve at Q (Figure 6). The same result holds if Q is furthest from (x_0, y_0).

We've seen that the "graph of f in polar coordinates" is the curve

$$(f \cdot \mathbf{e})(t) = (f(t)\cos t, f(t)\sin t);$$

in other words, the graph of f in polar coordinates is the curve with the parametric representation

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

6. (a) Show that for the graph of f in polar coordinates the slope of the tangent line at the point with polar coordinates $(f(\theta), \theta)$ is

$$\frac{f(\theta)\cos\theta + f'(\theta)\sin\theta}{-f(\theta)\sin\theta + f'(\theta)\cos\theta}.$$

(b) Show that if $f(\theta) = 0$ and f is differentiable at θ, then the line through the origin making an angle of θ with the positive horizontal axis is a tangent line of the graph of f in polar coordinates. Use this result to add some details to the graph of the Archimedean spiral in Appendix 3 of Chapter 4, and to the graphs in Problems 3 and 10 of that Appendix as well.

(c) Suppose that the point with polar coordinates $(f(\theta), \theta)$ is further from the origin O than any other point on the graph of f. What can you say about the tangent line to the graph at this point? Compare with Problem 5.

(d) Suppose that the tangent line to the graph of f at the point with polar coordinates $(f(\theta), \theta)$ makes an angle of α with the horizontal axis (Figure 7), so that $\alpha - \theta$ is the angle between the tangent line and the ray from O to the point. Show that

$$\tan(\alpha - \theta) = \frac{f(\theta)}{f'(\theta)}.$$

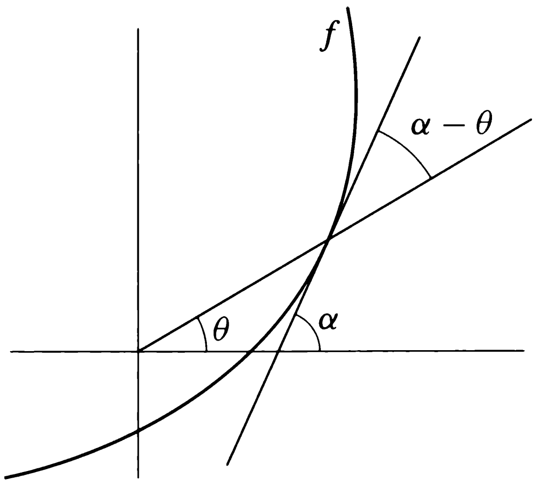


FIGURE 7

7. (a) In Problem 8 of Appendix 3 to Chapter 4 we found that the cardioid $r = 1 - \sin\theta$ is also described by the equation $(x^2 + y^2 + y)^2 = x^2 + y^2$. Find the slope of the tangent line at a point on the cardioid in two ways:

(i) By implicit differentiation.
(ii) By using the previous problem.

(b) Check that at the origin the tangent lines are vertical, as they appear to be in Figure 8.

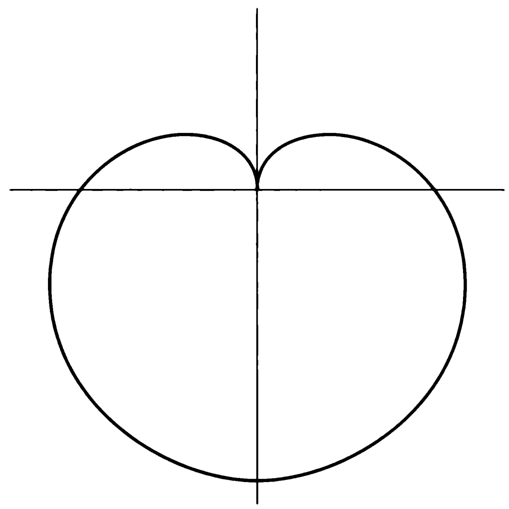


FIGURE 8

The next problem uses the material from Chapter 15, in particular, radian measure, and the inverse trigonometric functions and their properties.

8. A *cycloid* is defined as the path traced out by a point on the rim of a rolling wheel of radius a. You can see a beautiful cycloid by pasting a reflector on the edge of a bicycle wheel and having a friend ride slowly in front of the headlights of your car at night. Lacking a car, bicycle, or trusting friend, you can settle instead for Figure 9.

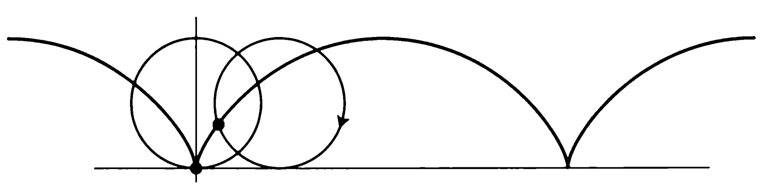


FIGURE 9

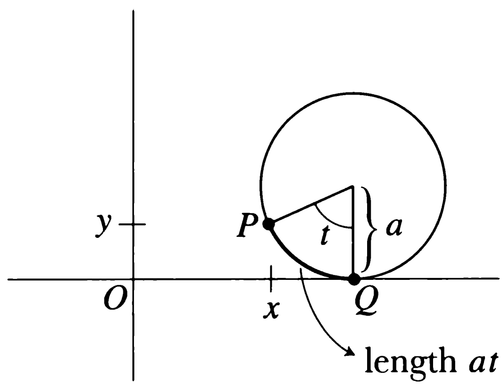
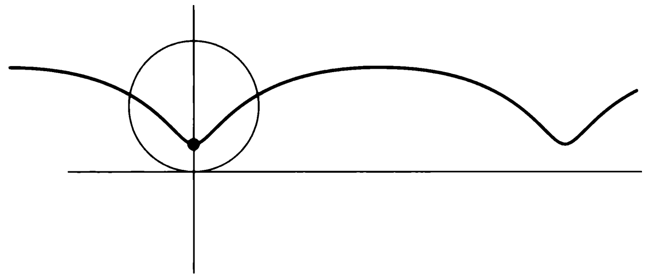


FIGURE 10

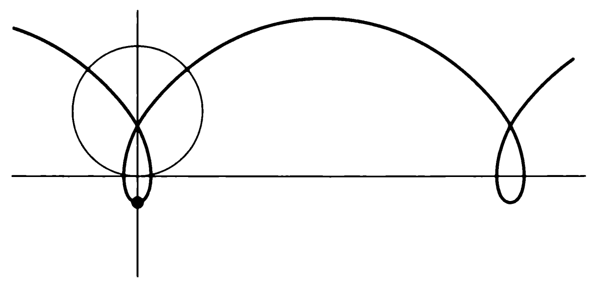
(a) Let $u(t)$ and $v(t)$ be the coordinates of the point on the rim after the wheel has rotated through an angle of t (radians). This means that the arc of the wheel rim from P to Q in Figure 10 has length at. Since the wheel is rolling, at is also the distance from O to Q. Show that we have the parametric representation of the cycloid

$$u(t) = a(t - \sin t)$$
$$v(t) = a(1 - \cos t).$$

Figure 11 shows the curves we obtain if the distance from the point to the center of the wheel is (a) less than the radius or (b) greater than the radius. In the latter case, the curve is not the graph of a function; at certain times the point is moving backwards, even though the wheel is moving forwards!



(a)



(b)

FIGURE 11

In Figure 9 we drew the cycloid as the graph of a function, but we really need to check that this is the case:

(b) Compute $u'(t)$ and conclude that u is increasing. Problem 3 then shows that the cycloid is the graph of $f = v \circ u^{-1}$, and allows us to compute $f'(t)$.

(c) Show that the tangent lines of the cycloid at the "vertices" are vertical.

It isn't possible to get an explicit formula for f, but we can come close.

(d) Show that

$$u(t) = a \arccos \frac{a - v(t)}{a} \pm \sqrt{[2a - v(t)]v(t)}.$$

Hint: first solve for t in terms of $v(t)$.

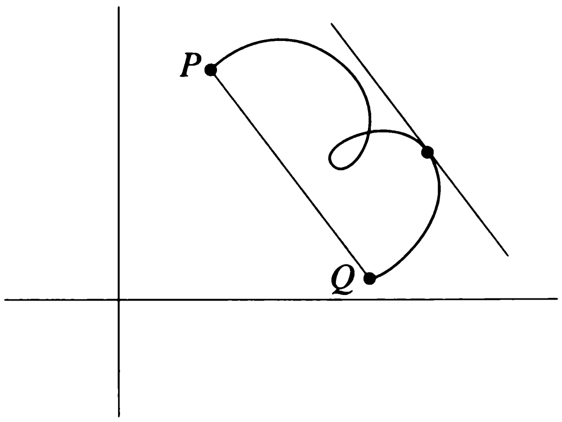


FIGURE 12

(e) The first half of the first arch of the cycloid is the graph of g^{-1}, where

$$g(y) = a \arccos \frac{a-y}{a} - \sqrt{(2a-y)y}.$$

9. Let u and v be continuous on $[a,b]$ and differentiable on (a,b); then u and v give a parametric representation of a curve from $P = (u(a), v(a))$ to $Q = (u(b), v(b))$. Geometrically, it seems clear (Figure 12) that at some point on the curve the tangent line is parallel to the line segment from P to Q. Prove this analytically. Hint: This problem will give a geometric interpretation for one of the theorems in Chapter 11. You will also need to assume that we don't have $u'(x) = v'(x) = 0$ for any x in (a,b) (compare Problem 2).

10. The following definition of a limit for a vector-valued function is the direct analogue of the definition for ordinary functions:

$\lim_{t\to a} c(t) = l$ means that for every $\varepsilon > 0$ there is some $\delta > 0$ such that, for all t, if $0 < |t-a| < \delta$, then $\|c(t) - l\| < \varepsilon$.

Here $\| \ \|$ is the *norm*, defined in Problem 2 of Appendix 1 to Chapter 4. If $l = (l_1, l_2)$, then

$$\|c(t) - l\|^2 = |u(t) - l_1|^2 + |v(t) - l_2|^2.$$

(a) Conclude that

$$|u(t) - l_1| \le \|c(t) - l\| \qquad \text{and} \qquad |v(t) - l_2| \le \|c(t) - l\|,$$

and show that if $\lim_{t\to a} c(t) = l$ according to the above definition, then we also have

$$\lim_{t\to a} u(t) = l_1 \qquad \text{and} \qquad \lim_{t\to a} v(t) = l_2,$$

so that $\lim_{t\to a} c(t) = l$ according to our definition ($*$) in terms of component functions, on page 246.

(b) Conversely, show that if $\lim_{t\to a} c(t) = l$ according to the definition in terms of component functions, then also $\lim_{t\to a} c(t) = l$ according to the above definition.