Chapter 5

Techniques of Differentiation

In this chapter we focus on functions given by *formulas*. The derivatives of such functions are then also given by formulas. In chapter 4 we used information about the derivative of a function to recover the function itself; now we go *from* the function *to* its derivative. We develop the rules for *differentiating* a function: computing the formula for its derivative from the formula for the function. Then we use differentiation to investigate the properties of functions, especially their *extreme values*. Finally we examine a powerful method for solving equations that depends on being able to find a formula for a derivative.

5.1 The Differentiation Rules

There are three kinds of differentiation rules. First, any basic function has a specific rule giving its derivative. Second, the *chain rule* will find the derivative of a *chain* of functions. Third, there are general rules that allow us to calculate the derivatives of algebraic combinations—e.g., sums, products, and quotients—of any functions provided we know the derivatives of each of the component functions. To obtain all three kinds of rules we will typically start with the analytic definition of the derivative as the limit of a quotient of differences:

Definition. The **derivative** of the function f at x is the value of the limit

$$\lim_{\Delta x \to 0} \frac{f(x + \Delta x) - f(x)}{\Delta x} = f'(x).$$

In this chapter we will look at the cases where this limit can be evaluated exactly. Although using this definition of derivative usually leads to many algebraic manipulations, the other interpretations of derivatives as slopes, rates, and multipliers will still be helpful in visualizing what's going on. The process of calculating the derivative of a function is called **differentiation**. For this reason, functions which are locally linear and not locally vertical (so they do have slopes, and hence derivatives at every point) are called **differentiable** functions. Our goal in this chapter is to differentiate functions given by formulas.

Derivatives of Basic Functions

Functions given by formulas have derivatives given by formulas

When a function is given by a *formula*, there is in fact a formula for its derivative. We have already seen several examples in chapters 3 and 4. These examples include all of what we may consider the **basic functions**. We collect these formulas in the following table.

Rules for Derivatives of Basic Functions

function	derivative
$mx + b$	m
x^r	rx^{r-1}
$\sin x$	$\cos x$
$\cos x$	$-\sin x$
e^x	e^x
$\ln x$	$1/x$

In the case of the linear function $mx + b$, we obtained the derivative by using its geometric description as the *slope* of the graph of the function. The derivatives of the exponential and logarithm functions came from the definition of the exponential function as the solution of an initial value problem. To find the derivatives of the other functions we will need to start from the definition.

An example: $f(x) = x^3$

We begin by examining the calculation of the derivative of $f(x) = x^3$ using the definition. The change Δy in $y = f(x)$ corresponding to a change Δx in x is given by

$$\begin{aligned}\Delta y &= f(x+\Delta x) - f(x)\\ &= (x+\Delta x)^3 - x^3\\ &= 3x^2\cdot\Delta x + 3x(\Delta x)^2 + (\Delta x)^3.\end{aligned}$$

From this we get

$$\begin{aligned}f'(x) &= \lim_{\Delta x\to 0}\frac{\Delta y}{\Delta x}\\ &= \lim_{\Delta x\to 0} 3x^2 + 3x\cdot\Delta x + (\Delta x)^2.\end{aligned}$$

To see what's happening with this expression, let's consider the specific value $x = 2$ and evaluate the corresponding values of $\Delta y/\Delta x$ for successively smaller Δx.

Δx	$2^2 + 6\Delta x + (\Delta x)^2$	$\Delta y/\Delta x$
.1	12 + .6 + .01	12.61
.01	12 + .06 + .0001	12.0601
.001	12 + .006 + .000001	12.006001
.0001	12 + .0006 + .00000001	12.00060001
.00001	12 + .00006 + .0000000001	12.0000600001

The value of $\Delta y/\Delta x$ gets closer and closer to 12 as Δx gets smaller and smaller

It is clear from this table that we can make $\Delta y/\Delta x$ as close to 12 as we like by making Δx small enough. Therefore $f'(2) = 12$.

Note that in the table above we have used positive values of Δx. You should check to convince yourself that if we had used negative values of Δx we would have come up with a different set of approximations $\Delta y/\Delta x$, but that the limit would still be the same, namely 12—it doesn't matter whether we use positive or negative values for Δx, or a mixture of the two, so long as $\Delta x \to 0$.

In general, for any given x, the second and third terms in the expansion for $\Delta y/\Delta x$ become vanishingly small as $\Delta x \to 0$, so that $\Delta y/\Delta x$ can be made as close to $3x^2$ as we like by making Δx small enough. For this reason, we say that the derivative $f'(x)$ is *exactly* $3x^2$:

$$f'(x) = \lim_{\Delta x\to 0} 3x^2 + 3x\cdot\Delta x + (\Delta x)^2 = 3x^2.$$

In other words, given the function f specified by the formula $f(x) = x^3$ we have found the formula for its derivative function f': $f'(x) = 3x^2$. Note that this general formula agrees with the specific value $f'(2) = 12$ we have already obtained.

Notice the difference between the statements

$$f'(x) \approx \Delta y/\Delta x \qquad \text{and} \qquad f'(x) = 3x^2.$$

For a particular value of Δx, the corresponding value of $\Delta y/\Delta x$ is an approximation of $f'(x)$. We can obtain another, better approximation by computing $\Delta y/\Delta x$ for a smaller Δx. The successively better approximations differ from one another by less and less. In particular, they differ less and less from the *limit value* $3x^2$. The value of the derivative $f'(x)$ is *exactly* $3x^2$.

More generally, for any function $y = f(x)$, a particular difference quotient $\Delta y/\Delta x$ is an approximation of $f'(x)$. Successively smaller values of Δx give successively better approximations of $f'(x)$. Again $f'(x)$ *exactly* equals the limiting value of these successive approximations. In some cases, however, we are only able to approximate that limiting value, as we often did in chapter 3, and for many purposes the approximation is entirely satisfactory. In this chapter we will concentrate on the exact statements that are possible for functions given by formulas.

The other basic functions

Our formula for the derivative of the function $f(x) = x^3$ is one instance of the general rule for the derivative of $f(x) = x^r$.

The rule for the derivative of a power function

For every real number r, the derivative of $f(x) = x^r$ is $f'(x) = r\,x^{r-1}$.

We can prove this rule for the case when r is a positive integer using algebraic manipulations very like the ones carried out for x^3; see the exercises for verifications of this and the other differentiation rules in this section. Using a rule for quotients of functions (coming later in this section), we can show that this rule also holds for *negative* integer exponents. Further arguments using the chain rule show that the pattern still holds for *rational* exponents. We can eliminate this case-by-case approach, though, by recalling the approach developed in chapter 4. We saw that we can give meaning to b^r for any positive base b and any real number r by defining

$$b^r = e^{r\ln(b)}.$$

Using the formulas for the derivatives of e^x and $\ln x$ together with the chain rule, we can prove the rule for $x > 0$ and for arbitrary real exponent r directly, without first proving the special cases for integer or rational exponents. See the exercises for details. Arguments justifying the formulas for the derivatives of the trigonometric functions are also in the exercises.

Combining Functions

We can form new functions by combining functions. We have already studied one of the most useful ways of doing this in chapter 3 when we looked at forming "chains" of functions and developed the **chain rule** for taking the derivative of such a chain. Suppose $u = f(x)$ and $y = g(u)$. Chaining these two functions together we have y as a function of x:

Functions combined by chains...

$$y = h(x) = g(f(x)).$$

The chain rule tells us how to find the derivative of y with respect to x. In function notation it takes the form

$$h'(x) = g'(f(x)) \cdot f'(x).$$

In Leibniz notation, using $f(x) = u$ we can write the chain rule as

$$\frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx}.$$

The chain rule

We also saw in chapter 3 that the polynomial $5x^3-7x^2+3$ can be thought of as an *algebraic* combination of simple functions. We can build an even more complicated function by forming a quotient with this polynomial in the numerator and the difference of the functions $\sin x$ and e^x in the denominator. The result is

...and algebraically

$$\frac{5x^3 - 7x^2 + 3}{\sin x - e^x}.$$

The derivative of this function, as well as of other functions formed by adding, subtracting, multiplying and dividing simpler functions, is obtained by use of the following rules for the derivatives of algebraic combinations of differentiable functions.

Rules for Algebraic Combinations of Functions

Combining functions by adding, subtracting, multiplying and dividing

function	derivative
$f(x)+g(x)$	$f'(x)+g'(x)$
$f(x)-g(x)$	$f'(x)-g'(x)$
$cf(x)$	$cf'(x)$
$f(x)\cdot g(x)$	$f'(x)\cdot g(x)+f(x)\cdot g'(x)$
$\dfrac{f(x)}{g(x)}$	$\dfrac{g(x)\cdot f'(x)-f(x)\cdot g'(x)}{[g(x)]^2}$

Notice the signs in the rules

Notice carefully that the product rule has a plus sign but the quotient rule has a *minus* sign. You can remember these formulas better if you think about where these signs come from. Increasing *either* factor increases a (positive) product, so the derivative of *each* factor appears with a plus sign in the formula for the derivative of a product. Similarly, increasing the numerator increases a positive quotient, so the derivative of the numerator appears with a plus sign in the formula for the derivative of a quotient. However, increasing the denominator *decreases* a positive quotient, so the derivative of the denominator appears with a minus sign.

Let's now use the rules to differentiate the quotient

$$\frac{5x^3-7x^2+3}{\sin x-e^x}.$$

First, the derivative of the numerator $5x^3-7x^2+3$ is

$$5(3x^2)-7(2x)+0=15x^2-14x.$$

Similarly the derivative of $\sin x-e^x$ is $\cos x-e^x$. Finally, the derivative of the quotient function is obtained by using the rule for quotients:

$$\frac{(\sin x-e^x)(15x^2-14x)-(5x^3-7x^2+3)(\cos x-e^x)}{(\sin x-e^x)^2}.$$

The following examples further illustrate the use of the rules for algebraic combinations of functions.

function	derivative
$-3e^t + \sqrt[3]{t}$	$-3e^t + (1/3)t^{-2/3}$
$\dfrac{5}{x^3} - 7x^4 + \ln x$	$5(-3)x^{-4} - 7(4x^3) + 1/x$
$7\sqrt{x}\cos x$	$7(\dfrac{1}{2\sqrt{x}})\cos x + 7\sqrt{x}(-\sin x)$
$\left(\dfrac{4}{3}\right)\pi r^3$	$\left(\dfrac{4}{3}\right)\pi 3r^2$
$\dfrac{3s^6}{s^2 - s}$	$\dfrac{(s^2 - s)3(6s^5) - 3s^6(2s - 1)}{(s^2 - s)^2}$

For another kind of example, suppose the per capita daily energy consumption in a country is currently 800,000 BTU, and, due to energy conservation efforts, it is falling at the rate of 1,000 BTU per year. Suppose too that the population of the country is currently 200,000,000 people and is rising at the rate of 1,000,000 people per year. Is the total daily energy consumption of this country rising or falling? By how much?

Three different quantities vary with time in this example: daily per capita energy consumption, population and total daily energy consumption. We can model this situation with three functions $C(t)$, $P(t)$ and $E(t)$.

$C(t)$: per capita consumption at time t

$P(t)$: population at time t

$E(t)$: total energy consumption at time t

Since the per capita consumption times the number of people in the population gives the total energy consumption, these three functions are related algebraically:

$$E(t) = C(t) \cdot P(t).$$

If $t = 0$ represents today, then we are given the two rates of change

$$C'(0) = -1,000 = -10^3 \text{ BTU per person per year, and}$$
$$P'(0) = 1,000,000 = 10^6 \text{ persons per year.}$$

Using the product rule we can compute the current rate of change of the total daily energy consumption:

$$\begin{aligned} E'(0) &= C(0) \cdot P'(0) + C'(0) \cdot P(0) \\ &= (8 \times 10^5) \cdot (10^6) + (-10^3) \cdot (2 \times 10^8) \\ &= (8 \times 10^{11}) - (2 \times 10^{11}) \\ &= 6 \times 10^{11} \text{ BTU per year.} \end{aligned}$$

So the total daily energy consumption is currently rising at the rate of 6×10^{11} BTU per year. Thus the growth in the population more than offsets the efforts to conserve energy.

Checking units

Finally, it is a useful exercise to check that the units make sense in this computation. Recall that $C(t)$ represents *per capita* daily energy consumption, so the units for $C(0) \cdot P'(0)$ are

$$\frac{\text{BTU}}{\text{person}} \cdot \frac{\text{persons}}{\text{year}} = \frac{\text{BTU}}{\text{year}},$$

and, similarly, the units for $C'(0) \cdot P(0)$ are

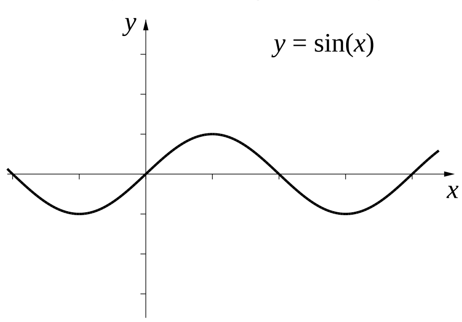
$$\frac{\text{BTU}}{\text{person}} \cdot \frac{1}{\text{year}} \cdot \text{persons} = \frac{\text{BTU}}{\text{year}}.$$

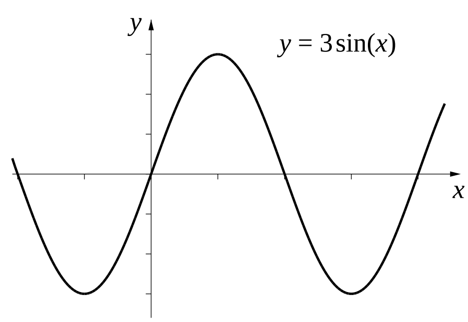
Informal Arguments

All of the rules for differentiating algebraic combinations of functions can be proved by using the algebraic definition of the derivative as a limit of a difference quotient. In fact, we will examine such a formal proof below. However, informal arguments based on geometric ideas or other intuitive insights are also valuable aids to understanding. Here are three examples of such arguments.

Stretching y-coordinates

- If a new function g is obtained from f by multiplying by a positive constant c, so $g(x) = cf(x)$, what is the relationship between the graphs of $y = f(x)$ and of $y = g(x)$? Stretching (or compressing, if c is less than 1) the y-coordinates of the points of the graph of f by a factor of c yields the graph of g.

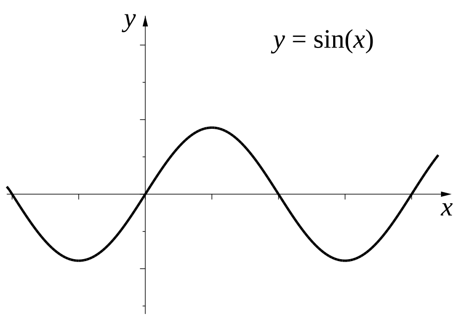


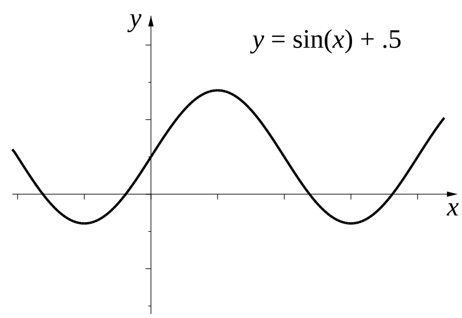


What then is the relationship between the slopes $f'(x)$ and $g'(x)$? If the y-coordinates are tripled, the slope will be three times as great. If they are halved, the slope will also be half as much. More generally, the elongated (or compressed) graph of g has a slope equal to c times the slope of the original graph of f. In other words, $g(x) = cf(x)$ implies $g'(x) = cf'(x)$.

- Now suppose instead that g is obtained from f by adding a constant b, so $g(x) = f(x) + b$. This time the graph of $y = g(x)$ is obtained from the graph of $y = f(x)$ by shifting up or down (according to the sign of c) by $|c|$ units. What is the relationship between the slopes $f'(x)$ and $g'(x)$? The shifted graph has exactly the same slope as the original graph, so in this case, $g(x) = f(x) + b$ implies $g'(x) = f'(x)$.

Shifting y-coordinates





There is a similar pattern when the coordinates of the input variable are stretched or shifted—that is when $y = f(u)$ and u is *rescaled* by the linear relation $u = mx + b$. These results depend on the chain rule and appear in the exercises.

The fact that the derivative of $f(x) + b$ is the same as the derivative of $f(x)$ is a special case of the general *addition rule*, which says *the derivative of a sum is the sum of the derivatives*. In the special case, the derivative of the constant function b is zero, so adding a constant leaves the derivative unchanged. To see that how natural it is to add rates in the general case, consider the following example:

Adding flows

Suppose we are diluting concentrated orange juice by mixing it with water in a big tub. We may let $f(t)$ be the amount (in gallons) of concentrate in the tub and $g(t)$ be the amount of water in the tub at time t. Then $f'(t)$ is the rate at which concentrate is being added at time t (measured in gallons per minute), and $g'(t)$ is the rate at which water is flowing into the tub. The formula $F(t) = f(t) + g(t)$ then gives the total amount of liquid in the tub at time t, and $F'(t)$ is the rate by which that total amount of liquid is changing at time t. Clearly that rate is the sum of the rates of flow of concentrate and water into the tank. If at some particular moment we are adding concentrate at the rate of 3.2 gal/min and water at the rate of 1.1 gal/min, the liquid in the tub is increasing by 4.3 gal/min at that moment.

A Formal Proof: the Product Rule

We include here the algebraic calculations yielding the rule for the derivative of the product of two arbitrary functions—just to give the flavor of these arguments. Algebraic arguments for the rest of these rules may be found in the exercises.

The Product Rule:

$$F(x) = f(x) \cdot g(x) \textit{ implies } F'(x) = f'(x) \cdot g(x) + f(x) \cdot g'(x)$$

To save some writing, let

$$\begin{aligned} \Delta F &= F(x + \Delta x) - F(x), \\ \Delta f &= f(x + \Delta x) - f(x), \\ \text{and} \qquad \Delta g &= g(x + \Delta x) - g(x). \end{aligned}$$

Rewrite the last two equations as

$$\begin{aligned} f(x + \Delta x) &= f(x) + \Delta f \\ g(x + \Delta x) &= g(x) + \Delta g. \end{aligned}$$

Now we can write

$$\begin{aligned} F(x + \Delta x) &= f(x + \Delta x) \cdot g(x + \Delta x) \\ &= (f(x) + \Delta f) \cdot (g(x) + \Delta g) \\ &= f(x) \cdot g(x) + f(x) \cdot \Delta g + \Delta f \cdot g(x) + \Delta f \cdot \Delta g \end{aligned}$$

Simplifying ΔF

This gives us a simple expression for

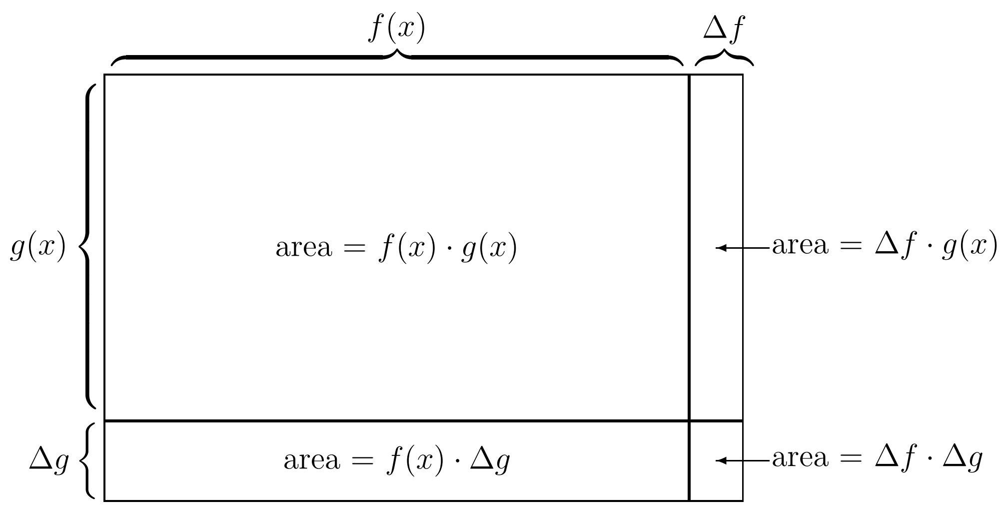
$$\Delta F = F(x + \Delta x) - F(x)$$

namely,

$$\Delta F = f(x) \cdot \Delta g + \Delta f \cdot g(x) + \Delta f \cdot \Delta g$$

These quantities all have nice geometric interpretations. First, think of the numbers $f(x)$ and $g(x)$ as lengths that depend on x; then $F(x)$ naturally stands for the area of the rectangle whose sides are $f(x)$ and $g(x)$. If the sides of the rectangle grow by the amounts Δf and Δg, then the area F grows by ΔF. As the following diagram shows, ΔF has three parts, corresponding to the three terms in the expression we derived algebraically for ΔF.

Interpret Δf and Δg as lengths



Now we divide ΔF by Δx and finish the argument:

$$\begin{aligned}\frac{\Delta F}{\Delta x} &= \frac{f(x)\cdot\Delta g + \Delta f\cdot g(x) + \Delta f\cdot\Delta g}{\Delta x}\\ &= f(x)\cdot\frac{\Delta g}{\Delta x} + \frac{\Delta f}{\Delta x}\cdot g(x) + \frac{\Delta f\cdot\Delta g}{\Delta x}\end{aligned}$$

Consider what happens to each of the three terms as Δx gets smaller and smaller. In the first term, the second factor $\Delta g/\Delta x$ approaches $g'(x)$—by the *definition* of the derivative. The first factor, $f(x)$ doesn't change at all as Δx shrinks. So the first term approaches $f(x)\cdot g'(x)$. Similarly, in the second term, the quotient $\Delta f/\Delta x$ approaches $f'(x)$, and the second term approaches $f'(x)\cdot g(x)$.

Finally, look at the third term. We would know what to expect if we had another factor of Δx in the denominator. We can put ourselves in familiar territory by the "trick" of multiplying the third term by $\Delta x/\Delta x$:

$$\frac{\Delta f\cdot\Delta g}{\Delta x} = \frac{\Delta f}{\Delta x}\cdot\frac{\Delta g}{\Delta x}\cdot\Delta x$$

Thus we can see that as Δx approaches zero, the third term itself approaches $f'(x) \cdot g'(x) \cdot 0 = 0$.

We may summarize our calculation by writing

$$\begin{aligned}\lim_{\Delta x \to 0} \frac{\Delta F}{\Delta x} &= f(x) \cdot \left(\lim_{\Delta x \to 0} \frac{\Delta g}{\Delta x}\right) + \left(\lim_{\Delta x \to 0} \frac{\Delta f}{\Delta x}\right) \cdot g(x) \\ &\quad + \left(\lim_{\Delta x \to 0} \frac{\Delta f}{\Delta x}\right) \cdot \left(\lim_{\Delta x \to 0} \frac{\Delta g}{\Delta x}\right) \cdot \left(\lim_{\Delta x \to 0} \Delta x\right)\end{aligned}$$

from which we have

$$\begin{aligned}\lim_{\Delta x \to 0} \frac{\Delta F}{\Delta x} &= f(x) \cdot g'(x) + f'(x) \cdot g(x) + f'(x) \cdot g'(x) \cdot 0 \\ &= f(x) \cdot g'(x) + f'(x) \cdot g(x).\end{aligned}$$

This completes the proof of the product rule. Other formal arguments are left to the exercises.

Exercises

Finding Derivatives

1. Find the derivative of each of the following functions.

a) $3x^5 - 10x^2 + 8$
b) $(5x^{12} + 2)(\pi - \pi^2 x^4)$
c) $\sqrt{u} - 3/u^3 + 2u^7$
d) $mx + b$ (m, b constant)
e) $.5 \sin x + \sqrt[3]{x} + \pi^2$
f) $\dfrac{\pi - \pi^2 x^4}{5x^{12} + 2}$
g) $2\sqrt{x} - \dfrac{1}{\sqrt{x}}$
h) $\tan z\,(\sin z - 5)$
i) $\dfrac{\sin x}{x^2}$
j) $x^2 e^x$
k) $\cos x + e^x$
l) $\sin x / \cos x$
m) $e^x \ln x$
n) $\dfrac{2^x}{10 + \sin x}$
o) $\sin(e^x \cos x)$
p) $6e^{\cos t} / 5\sqrt[3]{t}$
q) $\ln(x^2 + xe^x)$
r) $\dfrac{5x^2 + \ln x}{7\sqrt{x} + 5}$

2. Suppose f and g are functions and that we are given

$$\begin{array}{lll} f(2) = 3, & g(2) = 4, & g(3) = 2, \\ f'(2) = 2, & g'(2) = -1, & g'(3) = 17. \end{array}$$

Evaluate the derivative of each of the following functions at $t = 2$:

a) $f(t) + g(t)$
b) $5f(t) - 2g(t)$
c) $f(t)g(t)$
d) $\dfrac{f(t)}{g(t)}$
e) $g(f(t)$
f) $\sqrt{g(t)}$
g) $t^2 f(t)$
h) $(f(t))^2 + (g(t))^2$
i) $\dfrac{1}{f(t)}$
j) $f(3t - (g(1+t))^2)$

k) What additional piece of information would you need to calculate the derivative of $f(g(t))$ at $t = 2$?

l) Estimate the value of $f(t)/g(t)$ at $t = 1.95$

3. a) Extend the product rule to express $(f(t)g(t)h(t))'$ in terms of f, g, and h.

b) If the length, width, and height of a rectangular box are changing at the rates of 3, 6, and -5 inches/minute at the moment when all three dimensions happen to be 10 inches, at what rate is the volume of the box changing then?

c) If the length, width, and height of a box are 10 inches, 12 inches, and 8 inches, respectively, and if the length and height of the box are changing at the rates of 3 inches/minute and -2 inches/minute, respectively, at what rate must the width be changing to keep the volume of the box constant?

4. In this problem we examine the effect of stretching or shifting the coordinates of the input variable of a function. Your answers should address both the *algebra* and the *geometry* of the problem to show how the algebraic relations between the functions are manifested in their graphs.

a) Suppose $f(x) = \sin(x)$ and $g(x) = \sin(mx)$, where m is a constant stretching factor. What is the relation between $f'(x)$ and $g'(x)$?

b) As in (a), suppose $f(x) = \sin(x)$, but this time $g(x) = \sin(x + b)$ where b is the size of a (constant) shift. What is the relation between $f'(x)$ and $g'(x)$ this time?

c) Now consider the general case: $f(x)$ in an unspecified differentiable function and $g(x) = f(mx + b)$, where the input variable is stretched by the constant factor m and shifted by the constant amount b. What is the relation between $f'(x)$ and $g'(x)$ in this general case?

5. Which of the following functions has a derivative which is always positive (except at $x = 0$, where neither the function nor its derivative is defined)?

$$1/x \qquad -1/x \qquad 1/x^2 \qquad -1/x^2$$

6. a) As a function of its radius r, the volume of a sphere is given by the formula $V(r) = \frac{4}{3}\pi r^3$. If r is measured in centimeters, what are the units for $V'(r)$?

b) Explain why square cm are *not* the appropriate units for $V'(r)$, even though dimensionally correct.

7. Do the following.

a) Show that $\dfrac{1}{1-x^2}$ and $\dfrac{x^2}{1-x^2}$ have the same derivative.

b) If $f'(x) = g'(x)$ for every x, what can be concluded about the relationship between f and g? (Hint: What is $(f(x) - g(x))'$?)

c) Show that $\dfrac{1}{1-x^2} = \dfrac{x^2}{1-x^2} + C$ by finding C.

8. Suppose that the current total daily energy consumption in a particular country is 16×10^{13} BTU and is rising at the rate of 6×10^{11} BTU per year. Suppose that the current population is 2×10^8 people and is rising at the rate of 10^6 people per year. What is the current daily per capita energy consumption? Is it rising or falling? By how much?

9. The population of a particular country is 15,000,000 people and is growing at the rate of 10,000 people per year. In the same country the per capita yearly expenditure for energy is $1,000 per person and is growing at the rate of $8 per year. What is the country's current total yearly energy expenditure? How fast is the country's total yearly energy expenditure growing?

10. The population of a particular country is 30 million and is rising at the rate of 4,000 people per year. The total yearly personal income in the country is 20 billion dollars, and it is rising at the rate of 500 million dollars per year. What is the current per capita personal income? Is it rising or falling? By how much?

11. An explorer is marooned on an iceberg. The top of the iceberg is shaped like a square with sides of length 100 feet. The length of the sides is shrinking

at the rate of two feet per day. How fast is the area of the top of the iceberg shrinking? Assuming the sides continue to shrink at the rate of two feet per day, what will be the dimensions of the top of the iceberg in five days? How fast will the area of the top of the iceberg be shrinking then?

12. Suppose the iceberg of problem 9 is shaped like a cube. How fast is the volume of the cube shrinking when the sides have length 100 feet? How fast after five days?

Deriving Differentiation Rules

13. In this problem we calculate the derivative of $f(x) = x^4$.

a) Expand $f(x + \Delta x) = (x + \Delta x)^4 = (x + \Delta x)(x + \Delta x)(x + \Delta x)(x + \Delta x)$ as a sum of 16 terms. (Don't collect "like" terms yet.)

b) How many terms in part a involve *no* Δx's? What form do such terms have?

c) How many terms in part a involve exactly *one* Δx? What form do such terms have?

d) Group the terms in part a so that $f(x + \Delta x)$ has the form

$$Ax^4 + B\Delta x + R(\Delta x)^2 ,$$

where there are no Δx's among the terms in A or B, but R has several terms, some involving Δx. Use part b to check your value of A; use part c to check your value of B.

e) Compute the quotient $\dfrac{f(x + \Delta x) - f(x)}{\Delta x}$, taking advantage of part d.

f) Now find

$$\lim_{\Delta x \to 0} \frac{f(x + \Delta x) - f(x)}{\Delta x};$$

this is the derivative of x^4. Is your result here compatible with the rule for the derivative of x^n ?

14. In this problem we calculate the derivative of $f(x) = x^n$, where n is any positive integer.

a) First show that you can write

$$f(x + \Delta x) = x^n + nx^{n-1}\Delta x + R(\Delta x)^2$$

by developing the following line of argument. Write $(x + \Delta x)^n$ as a product of n identical factors:

$$(x + \Delta x)^n = \underbrace{(x + \Delta x)}_{1\text{-st}}\underbrace{(x + \Delta x)}_{2\text{-nd}}\underbrace{(x + \Delta x)}_{3\text{-rd}}\cdots\underbrace{(x + \Delta x)}_{n\text{-th}}$$

But now, before tackling this general case, look at the following examples. In the examples we use notation to help us keep track of which factors are contributing to the final result.

i) Consider the product $(a + b)(\underline{a} + \underline{b}) = a\underline{a} + a\underline{b} + b\underline{a} + b\underline{b}$. There are four individual terms. Each term contains one of the entries in the first factor (namely a or b) and one of the entries in the second factor (namely $\underline{a}$ or $\underline{b}$). The four terms represent thereby all possible ways of choosing one entry in the first factor and one entry in the second factor.

ii) Multiply out the product $(a+b)(\underline{a}+\underline{b})(A+B)$. (Don't combine like terms yet.) Does each term contain one entry from the first factor, one from the second, and one from the third? How many terms did you get? In fact there are two ways to choose an entry from the first factor, two ways to choose an entry from the second factor, and two ways to choose an entry from the third factor. Therefore, how many ways can you make a choice consisting of one entry from the first, one from the second, and one from the third?

Now return to the general case:

$$(x + \Delta x)^n = \underbrace{(x + \Delta x)}_{1\text{-st}}\underbrace{(x + \Delta x)}_{2\text{-nd}}\underbrace{(x + \Delta x)}_{3\text{-rd}}\cdots\underbrace{(x + \Delta x)}_{n\text{-th}}$$

How many ways can you choose an entry from each factor and *not* get any Δx's? Multiply these chosen entries together; what does the product look like (apart from having no Δx's in it)?

How many ways can you choose an entry from each factor in such a way that the resulting product has *precisely one* Δx? Describe all the various choices which give that result. What does a product that contains precisely one Δx factor look like? What do you obtain for the sum of *all* such terms with precisely one Δx factor?

What is the minimum number of Δx factors in any of the remaining terms in the full expansion of $(x + \Delta x)^n$?

Do your calculations agree with this summary:

$$(x + \Delta x)^n = x^n + nx^{n-1}\Delta x + R(\Delta x)^2 \ ?$$

b) Now find the value of $\dfrac{f(x+\Delta x)-f(x)}{\Delta x}$.

c) Finally, find

$$\lim_{\Delta x \to 0} \frac{f(x+\Delta x)-f(x)}{\Delta x}.$$

Do you get nx^{n-1}?

15. In this problem we give another derivation of the power rule based on writing

$$x^r = e^{r\ln(x)}.$$

Use the chain rule to differentiate $e^{r\ln(x)}$. Explain why your answer equals rx^{r-1}.

16. Does the rule for the derivative of x^r hold for $r = 0$? Why or why not?

17. In this exercise we prove the Addition Rule: $F(x) = f(x)+g(x)$ implies $F'(x) = f'(x) + g'(x)$.

a) Show $F(x+\Delta x) - F(x) = f(x+\Delta x) - f(x) + g(x+\Delta x) - g(x)$

b) Divide by Δx and finish the argument.

18. In this exercise we prove the Quotient Rule: $F(x) = f(x)/g(x)$ implies

$$F'(x) = \frac{g(x)f'(x) - f(x)g'(x)}{(g(x))^2}$$

a) Rewrite $F(x) = f(x)/g(x)$ as $f(x) = g(x)F(x)$. Pretend for the moment that you know what $F'(x)$ is and apply the Product Rule to find $f'(x)$ in terms of $F(x)$, $g(x)$, $F'(x)$, $g'(x)$.

b) Replace $F(x)$ by $f(x)/g(x)$ in your expression for $f'(x)$ in part a.

c) Solve the equation in part b for $F'(x)$ in terms of $f(x)$, $g(x)$, $f'(x)$ and $g'(x)$.

19. In this problem we calculate the derivative of $f(x) = x^n$ when n is a negative integer. First write $n = -m$, so m is a positive integer. Then $f(x) = x^{-m} = 1/x^m$.

a) Use the Quotient Rule and this new expression for f to find $f'(x)$.

b) Do the algebra to re-express $f'(x)$ as nx^{n-1}.

20. In this problem we calculate the derivatives of $\sin x$ and $\cos x$. We will need the **addition formulas**:

$$\sin(A+B) = \sin A \cos B + \cos A \sin B$$
$$\cos(A+B) = \cos A \cos B - \sin A \sin B$$

First tackle $f(x) = \sin x$:

a) Use the addition formula for $\sin(A+B)$ to rewrite $f(x+\Delta x)$ in terms of $\sin(x)$, $\cos(x)$, $\sin(\Delta x)$, and $\cos(\Delta x)$.

b) The quotient $\dfrac{f(x+\Delta x) - f(x)}{\Delta x}$ can now be written in the form

$$P(\Delta x) \cdot \sin x + Q(\Delta x) \cdot \cos x \,,$$

where P and Q are specific functions of Δx. What are the formulas for those functions?

c) Use a calculator or computer to estimate the limits

$$\lim_{\Delta x \to 0} P(\Delta x) \qquad \text{and} \qquad \lim_{\Delta x \to 0} Q(\Delta x) \ .$$

(Try $\Delta x = .1$, .01, .001, .0001. Be sure your calculator is set on radians, not degrees.) Using part b you should now be able to determine the limit

$$\lim_{\Delta x \to 0} \frac{f(x+\Delta x) - f(x)}{\Delta x}$$

by writing it in the form

$$\left(\lim_{\Delta x \to 0} P(\Delta x)\right) \cdot \sin x + \left(\lim_{\Delta x \to 0} Q(\Delta x)\right) \cdot \cos x \ .$$

d) What is $f'(x)$?

e) Proceed similarly to find the derivative of $g(x) = \cos x$.

21. In this problem we calculate the derivatives of the other circular functions. Use the quotient rule together with the derivatives of $\sin x$ and $\cos x$ to verify that the derivatives of the other four circular functions are as given

in the table below:

function	derivative
$\tan x = \dfrac{\sin x}{\cos x}$	$\sec^2 x$
$\csc x = \dfrac{1}{\sin x}$	$-\cot x \csc x$
$\sec x = \dfrac{1}{\cos x}$	$\sec x \tan x$
$\cot x = \dfrac{1}{\tan x}$	$-\csc^2 x$

Differential Equations

22. If $y = f(x)$ then the **second derivative** of f is just the derivative of the derivative of f; it is denoted $f''(x)$ or d^2y/dx^2. Find the second derivative of each of the following functions.

a) $f(x) = e^{3x-2}$

b) $f(x) = \sin \omega x$, where ω is a constant

c) $f(x) = x^2 e^x$

23. Show that e^{3x} and e^{-3x} both satisfy the (*second order*) differential equation

$$f''(x) = 9f(x).$$

Furthermore, show that *any* function of the form $g(x) = \alpha e^{3x} + \beta e^{-3x}$ satisfies this differential equation. Here α and β are arbitrary constants. Finally, choose α and β so that $g(x)$ also satisfies the two conditions $g(0) = 12$ and $g'(0) = 15$.

24. Show that $y = \sin x$ satisfies the differential equation $y'' + y = 0$. Show that $y = \cos x$ also satisfies the differential equation. Show that, in fact, $y = a \sin x + b \cos x$ satisfies the differential equation for any choice of constants a and b. Can you find a function $g(x)$ that satisfies these three conditions:

$$\begin{aligned} g''(x) + g(x) &= 0 \\ g(0) &= 1 \\ g'(0) &= 4? \end{aligned}$$

25. Show that $\sin \omega x$ satisfies the differential equation $y'' + \omega^2 y = 0$. What other solutions can you find to this differential equation? Can you find a function $L(x)$ that satisfies these three conditions:

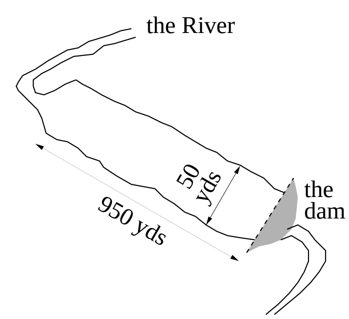
$$\begin{aligned} L''(x) + 4L(x) &= 0 \\ L(0) &= 36 \\ L'(0) &= 64? \end{aligned}$$

The Colorado River Problem

· Make your answer to this sequence of questions an essay. Identify all the variables you consider (e.g., "A stands for the area of the lake"), and indicate the functional relationships between them ("A depends on time t, measured in weeks from the present"). Identify the derivatives of those functions, as necessary.

The Colorado River—which excavated the Grand Canyon, among others—used to empty into the Gulf of California. It no longer does. Instead, it runs into a marshy area some miles from the Gulf and stops. One of the major reasons for this change is the construction of dams—notably the Hoover Dam. Every dam creates a lake behind it, and every lake increases the total surface area of the river. Since the rate at which water evaporates is proportional to the area of the water surface exposed to air, the lakes along the Colorado have increased the loss of river water through evaporation. Over the years, these losses (in conjunction with other factors, like increased usage by a rapidly growing population) have been significant enough to dry up the river at its mouth.

26. Let us analyze the evaporation rate along a river that was recently dammed. Suppose the lake is currently 50 yards wide, and getting wider at a rate of 3 yards per week. As the lake fills, it gets longer, too. Suppose it is currently 950 yards long, and it is extending upstream at a rate of 15 yards per week. Assuming the lake remains approximately rectangular as it grows, find



a) the current area of the lake, in square yards;

b) the rate at which the surface of the lake is currently growing, in square yards per week.

27. Suppose the lake continues to spread sideways at the rate of 3 yards per week, and it continues to extend upstream at the rate of 15 yards per week.

a) Express the area of the lake as a (quadratic!) function of time, where time is measured from the present, in weeks, and where the lake's area is as given in problem 25.

b) How many weeks will it take for the lake to cover 30 acres (= 145,200 square yards)?

c) At what rate is the lake surface growing when it covers 30 acres?

28. Compare the rates at which the surface of the lake is growing in problem 25 (which is the "current" rate) and in problem 26 (which is the rate when the lake covers 30 acres). Are these rates the same? If they are not, how do you account for the difference? In particular, the width and length grow at fixed rates, so why doesn't the area? Use what you know about derivatives to answer the question.

29. Suppose the local climate causes water to evaporate from the surface of the lake at the rate of 0.22 cubic yards per week, for each square yard of surface. Write a formula that expresses total evaporation per week in terms of area. Use E to denote total evaporation.

30. The lake is fed by the river, and that in turn is fed by rainwater and groundwater from its watershed. (The **watershed**, or basin, of a river is that part of the countryside containing the ponds and streams which drain into the river.) Suppose the watershed provides the lake, on average, with 25,000 cubic yards of new water each week.

Assuming, as we did in problem 25, that the lake widens at the constant rate of 3 yards per week, and lengthens at the rate of 15 yards per week, will the time ever come that the water being added to the lake from its watershed balances the water being removed by evaporation? In other words, will the lake ever stop filling?

5.2 Finding Partial Derivatives

We know from Chapter 3 that no additional formulas are needed to calculate *partial* derivatives. We simply use the usual differentiation formulas, treating all the variables except one—the one with respect to which the partial derivative is formed—as if they were constants. If we do this we get new techniques for analyzing rates of change in problems that involve functions of several variables.

Some Examples

Here are two examples to illustrate the technique for calculating partial derivatives:

Finding formulas for partial derivatives

1. Suppose $f(x, y) = x^2y + 5x^3 - \sqrt{x+y}$. Then

$$f_x(x, y) = 2xy + 15x^2 - \frac{1}{2\sqrt{x+y}}, \quad \text{and}$$
$$f_y(x, y) = x^2 - \frac{1}{2\sqrt{x+y}}.$$

2. Suppose $g(u, v) = e^{uv} + \dfrac{u}{v}$. Then

$$g_u(u, v) = ve^{uv} + \frac{1}{v}, \quad \text{and}$$
$$g_v(u, v) = ue^{uv} - \frac{u}{v^2}.$$

Eradication of Disease

Controlling—or, better still, eradicating—a communicable disease depends first on the development of a vaccine. But even after this step has been accomplished, public health officials must still answer important questions, including:

- What proportion of the population must be vaccinated in order to eliminate the disease?
- At what age should people be vaccinated?

In their 1982 article, "Directly Transmitted Infectious Diseases: Control by Vaccination," (*Science*, Vol. 215, 1053–1060), Roy Anderson and Robert May formulate a model for the spread of disease that permits them to answer these and other questions. For a particular disease in a particular environment, the important variables in their model are

1. The average human life expectancy L, in years;
2. The average age A at which individuals catch the disease, in years;
3. The average age V at which individuals are vaccinated against the disease, in years.

Anderson and May deduce from their model that in order to eradicate the disease, the proportion of the population that is vaccinated must exceed p, where p is given by

$$p = \frac{L+V}{L+A}.$$

For a disease like measles, public health officials can directly affect the variable V, for example by the recommendations they make to physicians about immunization schedules for children. They may also indirectly affect the variables A and L, because public health policy influences factors which can modify the age at which children catch the disease or the overall life expectancy of the population. (Many other factors affect these variables as well.) Which of these three variables has the greatest effect on the proportion of the population that must be vaccinated?

Partial derivatives can tell us which variables are most significant

In other words, which is largest: $\partial p/\partial L$, $\partial p/\partial A$, or $\partial p/\partial V$?

Using the rules, we compute:

$$\frac{\partial p}{\partial L} = \frac{1\cdot(L+A) - 1\cdot(L+V)}{(L+A)^2} = \frac{A-V}{(L+A)^2},$$
$$\frac{\partial p}{\partial A} = \frac{-(L+V)}{(L+A)^2}, \quad \text{and}$$
$$\frac{\partial p}{\partial V} = \frac{1}{L+A}.$$

For measles in the United States, reasonable values of the variables are L = 70 years, A = 5 years and V = 1 year. Using these values, the crucial proportion of the population needing to be vaccinated is $p = 71/75 = .947$, and the partial derivatives are

$$\frac{\partial p}{\partial L} = \frac{4}{(75)^2} = .0007,$$
$$\frac{\partial p}{\partial A} = \frac{-71}{(75)^2} = -.0126,$$
$$\frac{\partial p}{\partial V} = \frac{1}{75} = .0133.$$

Determining units

A comment is in order here on units. While the input variables L, A and V are all measured in years—so the rates are *per year*, the output variable p is dimensionless: it is the ratio of persons vaccinated to persons not vaccinated. It would be reasonable to write p as a percentage. Then we can attach the units *percent per year* to each of the three partial derivatives. Thus we have:

$$\frac{\partial p}{\partial L} = .07\% \qquad \text{per year}$$
$$\frac{\partial p}{\partial A} = -1.26\% \qquad \text{per year}$$
$$\frac{\partial p}{\partial V} = 1.33\% \qquad \text{per year.}$$

It is not surprising that a change in average life expectancy has a negligible effect on the proportion p of the population that must be vaccinated in order to eradicate measles. Nor is it surprising that changing the age of vaccination has the greatest effect on p. But it is not obvious ahead of time that changing the age at which children catch the disease has nearly as large an effect on p:

- Decreasing the age of vaccination decreases the proportion p by 1.33% per year of decrease.

- Increasing the age at which children catch measles decreases the proportion p by 1.26% per year of increase.

Changes can also go the "wrong" way. For example, in an area where use of communal child care facilities is growing, contact among very young children increases, and the age at which children are exposed to—and can catch—communicable diseases like measles falls. The Anderson–May model tells us that immunization practices must change to compensate: either the age of vaccination must drop a like amount, or the fraction of the population that is vaccinated must grow by 1.26% per year of decrease in the average age at infection.

Exercises

Finding Partial Derivatives

1. Find the partial derivatives of the following functions.

a) x^2y.

b) $\sqrt{x+y}$

c) e^{xy}

d) $\dfrac{y}{x}$

e) $\dfrac{x+y}{y+z}$

f) $\sin\dfrac{y}{x}$

2. a) Suppose $f(x,y) = e^{-(x+2y)}(2x-5y)$. Find $f_x(x,y)$ and $f_y(x,y)$.

b) Find a point (a,b) at which $f_x(a,b) = 0$. At such a point a small change in x leaves the value of f virtually unchanged.

c) Find a point (a,b) at which a small *increase* in the x-value would produce the same change in $f(a,b)$ as would the same-sized *decrease* in the y-value.

3. Suppose $g(u,v) = \dfrac{\sin u + v^2 + 7uv}{1+u^2+v^4}$. Find $g_u(u,v)$ and $g_v(u,v)$.

4. The **second partial derivatives** of $z = f(x,y)$ are the partial derivatives of $\partial f/\partial x$ and $\partial f/\partial y$, namely:

$$\frac{\partial^2 f}{\partial x^2} = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial x}\right)$$

$$\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right)$$

$$\frac{\partial^2 f}{\partial y^2} = \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial y}\right)$$

Find the three second partial derivatives of the following functions.

a) x^2y.

b) $\sqrt{x+y}$

c) e^{xy}

d) $\dfrac{y}{x}$

e) $\sin\dfrac{y}{x}$

Eradication of Disease

5. Suppose you were dealing with measles in a developing country where $L = 50$ years, $A = 4$ years, and $L = 2$ years. Discuss the impact on measles control if increased public health efforts increase L to 55 years, A to 5 years, and decrease V to 1.5 years.

Partial differential equations

6. Show that the function $z = \dfrac{1}{\sqrt{t}} \exp \dfrac{-x^2}{4t}$ satisfies the **partial differential equation**

$$\frac{\partial^2 z}{\partial x^2} = \frac{\partial z}{\partial t}.$$

7. Show that *every* linear function of the form $z = px + qy + c$ satisfies the partial differential equation

$$\frac{\partial^2 z}{\partial x^2} + \frac{\partial^2 z}{\partial y^2} = 0.$$

Here p, q, and c are arbitrary constants.

8. Show that the function $z = e^x \sin y$ also satisfies the partial differential equation

$$\frac{\partial^2 z}{\partial x^2} + \frac{\partial^2 z}{\partial y^2} = 0.$$

5.3 The Shape of the Graph of a Function

We know from chapter 3 that the derivative gives us qualitative information about the shape of the graph of a differentiable function.

function	derivative
increasing	positive
decreasing	negative
level	zero
steep (rising or falling)	large (positive or negative)
gradual (rising or falling)	small (positive or negative)
straight	constant

Having a formula for the derivative of a function will thus give us a great deal of information about the behavior of the function itself. In particular we will be interested in using the derivative to solve **optimization problems**—finding maximum or minimum values of a function. Such problems occur frequently in many fields.

Contexts for optimization problems

- Economists actually define human rationality in terms of optimization. Each person is assumed to have a *utility function*, a function that assigns to each of many possible outcomes its *utility*, a numerical measure of its value to her. (Different people may have different utility functions, depending on their personal value systems.) A rational person is one who acts to maximize her utility. Some utilities are expressed in terms of money. For example, a rational manufacturer will seek to maximize her profit (in dollars). Her profit will depend on—that is, be a function of—such variables as the cost of her raw materials and the unit price she charges for her product.

- Many physical laws are expressed as minimum principles. Ordinary soap bubbles exhibit one of these principles. A soap film has a *surface energy* which is proportional to its surface area. For almost any physical system, its stable state is one which minimizes its energy. Stable soap films are thus examples of *minimal surfaces.* Interfaces involving crystals also have surface energies, leading to the study of crystalline minimal surfaces.

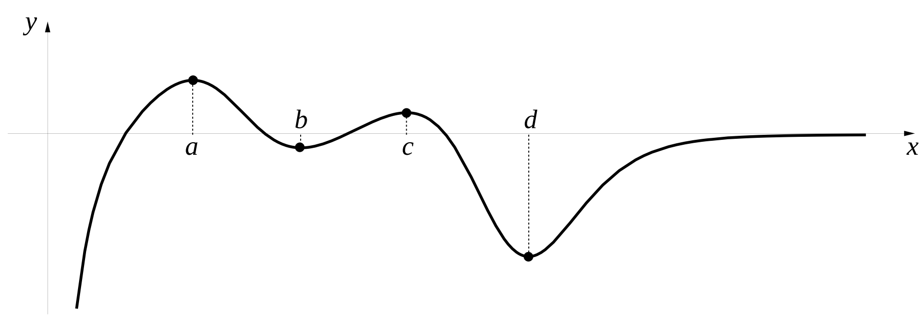
- Statisticians develop mathematical summaries for data—in other words, mathematical models. For example, a relationship between two numerical variables may be summarized by a linear function, say $y = mx + b$, where x and y are the variables of interest. It would be very rare to find data that were exactly linear. In a particular case, the statistician chooses the linear model that minimizes the *discrepancy* between the actual values of y and the theoretical values obtained from the linear function. Statisticians frequently measure this discrepancy by summing the squares of the differences between the actual and the theoretical values of y for each data point. The *best-fitting line* or *regression line* is the graph of the linear function which is optimal in this sense.

- Psychologists who study decision-making have found that some people are "risk-averse"; they make their decisions primarily to avoid risks. If we regard risk as a function of the various outcomes under consideration (a bit like a utility function), such a person acts to minimize this function.

The derivative is the key tool here. We will develop a general procedure for using the derivative of a function to locate its extremes.

Language

Here is a graph of what we might consider a "generic" differentiable function.



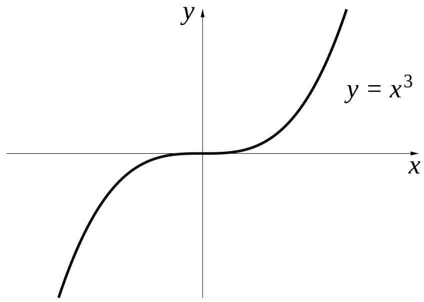
Local extremes and global extremes

The most distinctive features are the hill tops and valley bottoms, points where the graph levels and the derivative is zero. We distinguish between **local** extremes, like those occurring at the points $x = b$, $x = c$, and $x = d$ and a **global** extreme, like the global maximum at the point $x = a$. The function has a **local minimum** at $x = b$ because $f(x) \geq f(b)$ for all x *sufficiently near* b. The function has a **global maximum** at $x = a$ because $f(x) \leq f(a)$ for *all* x. Notice that this particular function does not have a global minimum. What kinds of local extremes does the function have at $x = c$ and $x = d$?

The convention is to say that *all* extremes are local extremes, and a local extreme may or may not also be a global extreme.

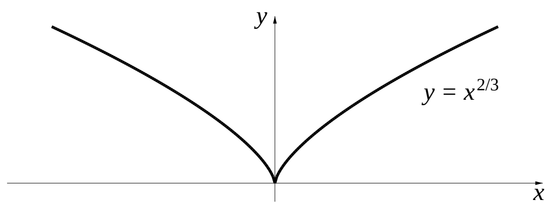
A function may have no extreme at a point where its derivative equals zero

Examining the graph of as simple a function as $f(x) = x^3$ shows us that a function need not have any extremes at all. Moreover, since $f'(0) = 0$ for this function, a zero derivative doesn't necessarily identify a point where an extreme occurs.



An extreme can occur at a cusp

Can a function have an extreme at a point *other* than where the derivative is zero? Consider the graph of $f(x) = x^{2/3}$ below.



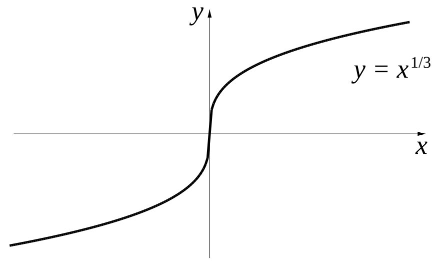
This function is differentiable everywhere except at the point $x = 0$. And it is at this very point, where

$$f'(x) = \frac{2}{3}x^{-1/3} = \frac{2}{3\sqrt[3]{x}}$$

is undefined, that the function has its global minimum. For this reason, points where the derivative fails to exist (or is infinite) are as important as points where the derivative equals zero. All of these kinds of points are called **critical points** for the function.

> A **critical point** for a function f is a point on the graph of f where f' equals zero or infinity or fails to exist.

How can we recognize, from looking at its graph, that a function fails to be differentiable at a point? We know that when the graph has a sharp corner or **cusp** it isn't locally linear at that point, and so has no derivative there. Thus critical points occur at these places. When the graph is locally linear but vertical, the slope cannot be given a finite numerical value. Critical points also occur at these vertical places where the slope is infinite. The graph below is an example of such a curve.



Continuous functions

A weaker condition than differentiability, but one that is useful, especially in this context, is **continuity**. We say that a function f is **continuous at a point** $x = a$ if

- it is defined at the point, and
- we can achieve changes in the output that are arbitrarily small by restricting changes in the input to be sufficiently small.

This second condition can also be expressed in the following form:

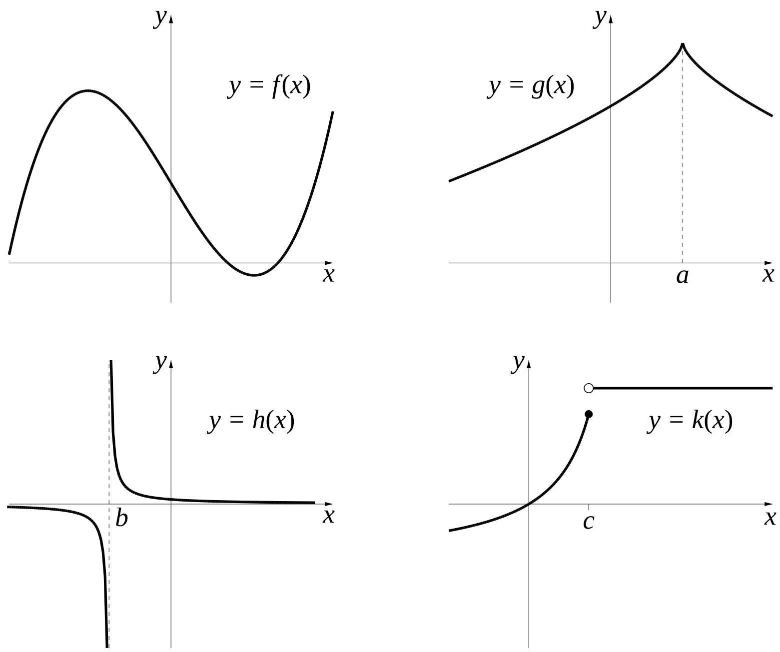
> Given any positive number ϵ (the proposed limit on the change in the output is traditionally designated by the Greek letter ϵ, pronounced 'epsilon'), there is always a positive number δ (the Greek letter 'delta'), such that whenever the change in the input is less than δ, then the corresponding change in the output will be less than ϵ.

Graphs of continuous functions have no gaps or jumps

A function is continuous on a set of real numbers if it is continuous at each point of the set. A natural way to think of (and recognize) a function which is continuous on an interval is by its graph on that interval, which is continuous in the usual sense: it has no gaps or jumps in it. You can draw it without picking up your pencil. Of the four functions whose graphs appear on the next page, f is continuous (and differentiable); g is continuous, but not differentiable (because of the cusp at $x = a$); h is not continuous, because h is undefined at $x = b$; and k is not continuous, because of the "jump" at $x = c$.

Luckily, the functions that we are likely to encounter are continuous on their natural domains. Among functions given by formulas, the only exceptions we have to worry about are quotients, like $f(x) = 1/x$, which have gaps in their natural domains at points where the denominator vanishes (at $x = 0$ in the case of $f(x) = 1/x$), so their domains are not single intervals. For convenience, we usually confine our attention to functions continuous on an interval.

A quotient isn't continuous where its denominator is zero



The Existence of Extremes

Not every function has extremes, as the example of $f(x) = x^3$ shows. We are, however, guaranteed extremes for certain functions.

Principle I. We are guaranteed that a function has a global maximum and a global minimum if we know:

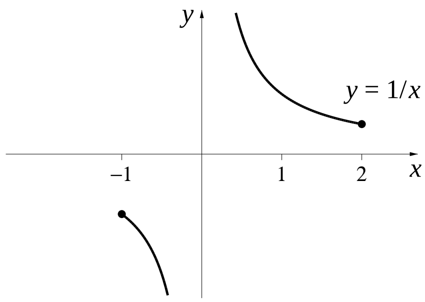
A continuous function on a finite closed interval has extremes

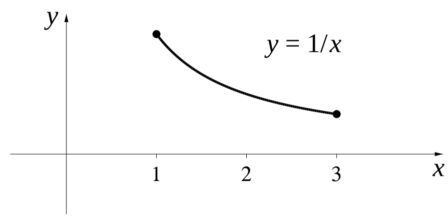
- the domain of the function is a *finite closed interval*, and
- the function is *continuous* on this interval.

The domain restriction in most optimization problems is likely to come from the physical constraints of the problem, not the mathematics. A **finite**

closed interval, written $[a, b]$, is the set of all real numbers between a and b, including the endpoints a and b. Other kinds of finite intervals are $(a, b]$, (this excludes the endpoint a), $[a, b)$, and (a, b). Infinite intervals include open and closed "rays" like $(a, +\infty)$, $[a, +\infty)$, $(-\infty, b)$ and $(-\infty, b]$, and the entire real line.

To illustrate Principle I, we note first that it does not apply to $f(x) = 1/x$ on the finite closed interval $[-1, 2]$, because this function isn't continuous at every point on this interval—in fact, the function isn't even defined for $x = 0$.





By contrast, Principle I *does* apply to the same function if we change its domain to a finite closed interval that does not include $x = 0$. In the figure above we use $[1, 3]$.

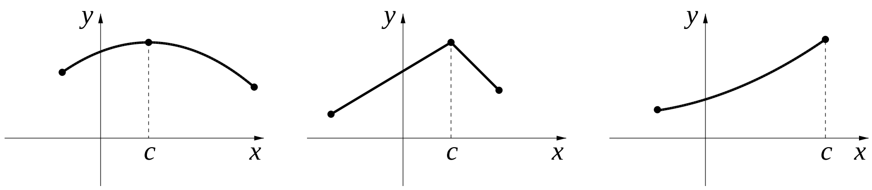
A function hat fails to satisfy the first condition of Principle I can still have global extremes. For any function continuous on an interval we can apply the following principle.

Continuous functions have extremes only at endpoints or critical points

Principle II. If a function f is continuous on an interval and has a local extreme at a point $x = c$ of the interval, then either $x = c$ is a critical point for the function, or $x = c$ is an endpoint of the interval. In other words, we are guaranteed that one of the following three conditions holds:

- $f'(c) = 0$,
- $f'(c)$ is undefined,
- $x = c$ is an endpoint of the interval.

The following graphs illustrate three local maxima, each satisfying one of these three conditions.



When we apply Principle II to optimization problems, an important part of the task will be ascertaining which, if any, of the critical points or endpoints we find actually gives the extreme we're looking for. We'll examine a variety of techniques, graphical and analytical, for locating critical points and determining what kind of extreme point (if any) they are.

Finding Extremes

Using a graphical approach

If we can use a computer to examine the graph of the function of interest, we can determine the existence and location of extremes by inspection. However, every graphing utility requires the user to specify the interval on which the function will be graphed, and careful analysis may be required in order to choose an interval that contains all the extremes of interest.

Computer graphing can be easy if the general location of the extremes is known

For functions given by data, whose graphs have only finitely many points, we can zoom in to find the exact coordinates of the extreme datapoints. For a function given by a formula, we can estimate the coordinates of an extreme to arbitrary accuracy by zooming in on the point as closely as desired. This is the method we used in some of the exercises of Chapter 1, and it is quite satisfactory in many situations.

Using the formula for the derivative.

In this chapter we are concentrating on functions given by formulas. For these functions we may want a method other than the approximation using a graphing utility described above.

Formulas can give exact answers and can handle parameters

- For some functions, the determination of extremes using a formula for the derivative is at least as easy as using the computer.
- Some functions are described in terms of a *parameter*, a constant whose value may vary from one problem to another. For example, the rate

equation from the S-I-R model for change in the number of infected, $I' = aSI - bI$, involves two parameters a and b, the transmission and recovery coefficients. For such a function, we cannot use the computer unless we specify a numerical value for each parameter, thus limiting the generality of our results.

- The computer gives only an approximation, while a precise answer may convey important additional information.

We will assume that the function we are studying is continuous on the domain of interest and that its domain is an interval. Our procedure for finding extremes of a function given by a formula $y = f(x)$ is thus a direct application of Principle II.

The search for local extremes

1. Determine the domain of the function and identify its endpoints, if any. Keep in mind that in an applied context, the domain may be determined by physical or other restrictions.

2. Find a formula for $f'(x)$.

3. Find any roots of $f'(x) = 0$ in the domain. An estimation procedure, for example Newton's Method (see the end of this chapter), may be used for complicated equations.

4. Find any points in the domain where $f'(x)$ is undefined.

5. Determine the shape of the graph to locate any local extremes. Find the shape either by looking directly at the graph of $y = f(x)$ or by analyzing the sign of $f'(x)$ on either side of each critical point $x = c$. (Sometimes the second derivative $f''(c)$ aids the analysis; see the exercises for details.)

The search for global extremes

1. Find the local extremes, as above.

2. If the domain is a finite closed interval, it is only necessary to compare the values of f at each of the critical points and at the endpoints to determine which is the global maximum and which is the global minimum.

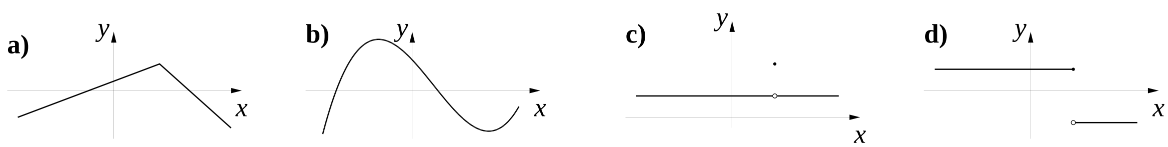
3. More generally, use the shape of the graph to ascertain whether the desired global extreme exists and to identify it.

In the succeeding sections and in the exercises we will carry out these search procedures in a variety of situations. Since there are often substantial algebraic difficulties in analyzing the sign of the derivative, or even determining when it is equal to zero, in realistic problems, in section 5 we will develop some numerical methods for handling such complications.

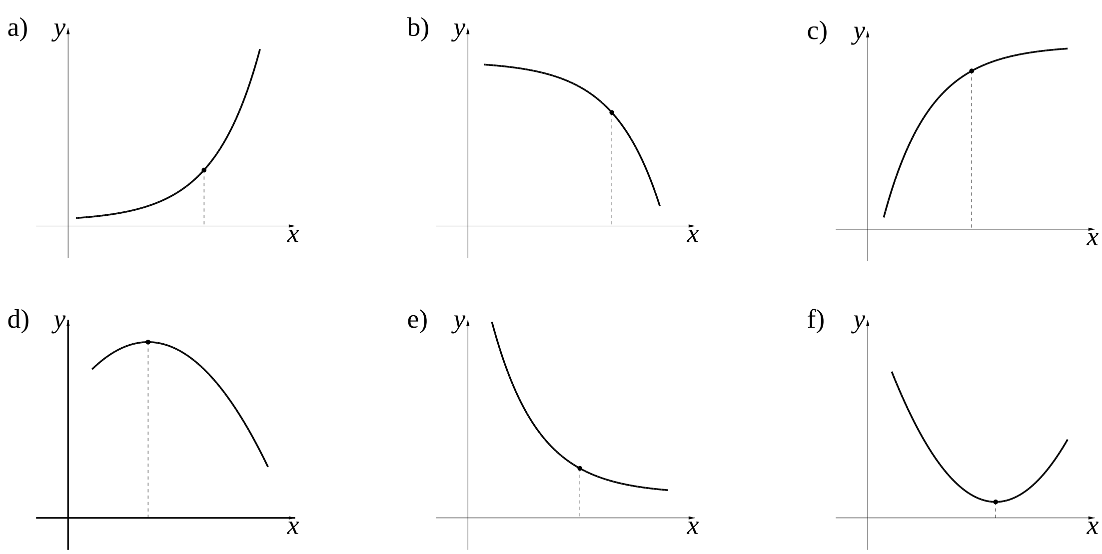
Exercises

Describing functions

1. For each of the following graphs, is the function continuous? Is the function differentiable?



2. For each of the following graphs of a function $y = f(x)$, is f' increasing or decreasing? At the indicated point, what is the sign of f'? Is f' (*not* f) increasing or decreasing at the point? What does this then say about the sign of f'' (the derivative of f') at the point?



3. For each of the following, sketch a graph of $y = f(x)$ that is consistent with the given information. On each graph, mark any critical points or extremes.

a) $f'(x) > 0$ for $x < 1$; $f'(1) = 0$; $f'(x) < 0$ for $1 < x < 2$; $f'(2) = 0$; $f'(x) > 0$ for $x > 2$.

b) $f'(x) > 0$ for $x < 2$; $f'(2) = 0$; $f'(x) > 0$ for $x > 2$.

c) $f'(x) > 0$ for $x < 2$; $f'(2) = 0$; $f'(x) < 0$ for $x > 2$.

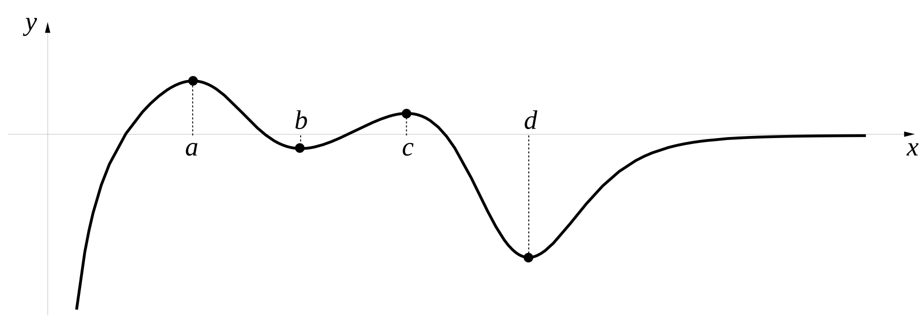
d) $f'(3) = 0$; $f''(3) > 0$.

4. **The geometric meaning of the second derivative** If f is any function and if $f''(x)$ is positive over some interval, then f' is increasing over that interval, and we say the curve is **concave upward** over that interval. If $f''(x)$ is negative over some interval, then f' is decreasing over that interval, and we say the curve is **concave downward** over that interval. You should study the graphs in problem 2 until you are clear what this means geometrically.

a) Suppose there were four functions f, g, h, and k, and that $f(0) = g(0) = h(0) = k(0) = 0$, and $f'(0) = g'(0) = h'(0) = k'(0) = 1$. Suppose, moreover, that $f''(0) = 1$, $g''(0) = 5$, $h''(0) = -1$, and $k''(0) = -5$. Sketch possible graphs of these functions near the origin.

b) We know that the magnitude of the first derivative tells us how steep the curve is—the greater the value of f', positive or negative, the steeper the curve. What does the magnitude of the second derivative tell us geometrically about the shape of the curve? Complete the sentence: "The greater the value of f'', the _______________."

5. a) Here is the graph you saw back on page 302:



At which point is the second derivative greater, b or d, and why?

b) Reproduce a sketch of this curve and indicate where the curve is concave up and where it is concave down.

c) What must be true about the second derivative at the points where the curve changes concavity from up to down, or vice versa? Give a clear justification for your answer.

d) What must be true about the second derivative near the right-hand end of the graph, and why?

e) Put all this together to sketch the graph of the second derivative of this function. Label the values $a - d$ on your sketch.

6. **Second derivative test for maxima and minima** Explain why the following test works.

- If $f'(c) = 0$ and $f''(c) > 0$, then f has a *local minimum* at $x = c$.
- If $f'(c) = 0$ and $f''(c) < 0$, then f has a *local maximum* at $x = c$.

(Hint: If $f'(c) = 0$ and $f''(c) > 0$, what can you say about the value of $f'(x)$—and hence the geometry of the graph of f—on either side of c?)

Finding critical points

7. For each of the following functions, find the critical points, if any, without using a computer or calculator. Can you sketch the graph of the function near the critical point? Use the second derivative test if you can't figure the behavior out from a simpler inspection.

a) $f(x) = x^{1/3}$

b) $f(x) = x^3 + \dfrac{3}{2}x^2 - 6x + 5$

c) $f(x) = \dfrac{2x+1}{x-1}$

d) $f(x) = \sin x$

e) $f(x) = \sqrt{1-x^2}$

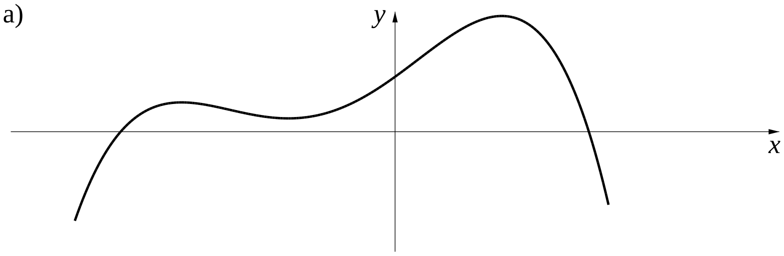
f) $f(x) = \dfrac{e^x}{x}$

g) $f(x) = x \ln x$

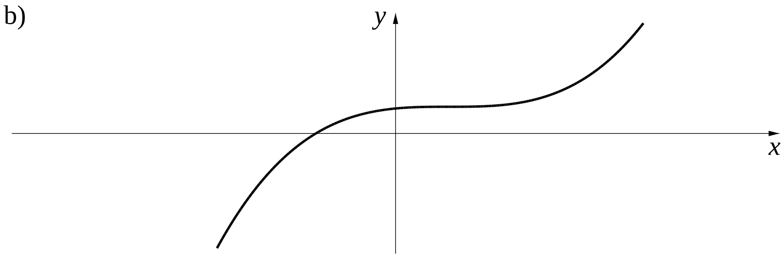
h) $f(x) = x^c + \dfrac{1}{x^c}$ where c is some constant

8. For each of the following graphs, mark any critical points or extremes. Indicate which extremes are local and which are global. (Assume that at their ends the curves continue in the direction they are headed.)

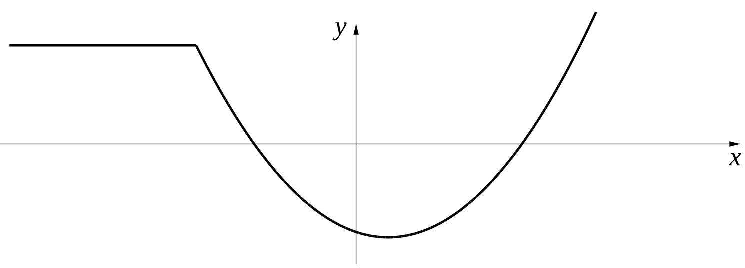
a)



b)



c)



Finding extremes

Except where indicated, you should not use a computer or calculator to solve the following problems.

9. For what positive value of x does $f(x) = x + \dfrac{7}{x}$ attain its minimum value? Explain how you found this value.

10. For what value of x in the interval $[1, 2]$ does $f(x) = x + \dfrac{7}{x}$ attain its minimum value? Explain how you found this value.

11. Use a graphing program to make a sketch of the function $y = f(x) =$

$x^2 2^{-x}$ on the interval $0 \leq x \leq 10$. From the graph, estimate the value of x which makes y largest, accurately to four decimal places. Then find where y takes on its maximum by setting the derivative $f'(x)$ equal to 0. You should find that the maximum occurs at $x = 2/\ln 2$. Finally, what is the numerical value of this estimate of $2/\ln 2$, accurate to seven decimal places?

12. Use a graphing program to sketch the graph of $y = \dfrac{1}{x^2} + x$ on the interval $0 < x < 4$ and estimate the value of x that makes y smallest on this interval. Then use the derivative of y to find the exact value of x that makes y smallest on the same interval. Compare the estimated and exact values.

13. What is the smallest value $y = \dfrac{4}{x^2} + x$ takes on when x is a positive number? Explain how you found this value.

14. The function $y = x^4 - 42x^2 - 80x$ has two local minima. Where are they (that is, what are their x coordinates), and what are their values? Which is the lower of the two minima? In this problem you will have to solve a cubic equation. You can use an estimation procedure, but it is also possible to solve by factoring the cubic. (The roots are integers.)

15. The function $y = x^4 - 6x^2 + 7$ has two local minima. Where are they, and which of the two is lower? In this problem you will have to solve a cubic equation. Do this by factoring and then approximating the x values to 4 decimal places.

16. Sketch the graph of $y = x \ln x$ on the interval $0 < x < 1$. Where does this function have its minimum, and what is the minimum value?

17. Let x be a positive number; if $x > 1$ then the cube of x is larger than its square. However, if $0 < x < 1$ then the square is larger than the cube. How *much* larger can it be; that is, what is the greatest amount by which x^2 can exceed x^3? For which x does this happen?

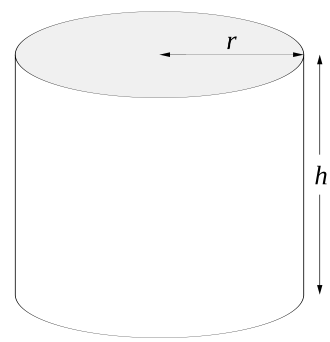
18. By how much can x^p exceed x^q, when $0 < p < q$ and $0 < x$? For which x does this happen?

5.4 Optimal Shapes

The Problem of the Optimal Tin Can

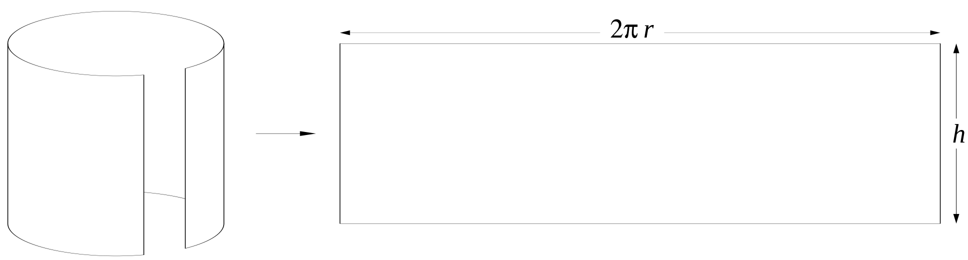
What is the minimum surface area?

Suppose you are a tin can manufacturer. You must make a can to hold a certain volume V of canned tomatoes. Naturally, the can will be a cylinder, but the proportions, the height h and the radius r, can vary. Your task is to choose the proportions so that you use the least amount of tin to make the can. In other words, you want the surface area of that can to be as small as possible.



The Solution

The surface area is the sum of the areas of the two circles at the top and bottom of the can, plus the area of the rectangle that would be obtained if the top and bottom were removed and the side cut vertically.



Thus A depends on r and h:

$$A = 2\pi r^2 + 2\pi rh.$$

Finding A as a function of r

However, r and h cannot vary independently. Because the volume V is fixed, r and h are related by

$$V = \pi r^2 h.$$

Solving the equation above for h in terms of r,

$$h = \frac{V}{\pi r^2},$$

we may express A as a function of r alone:

$$A = f(r) = 2\pi r^2 + 2\pi r \frac{V}{\pi r^2} = 2\pi r^2 + \frac{2V}{r}$$

Notice that the formula for this function involves the parameter V. The mathematical description of our task is to find the value of r that makes $A = f(r)$ a minimum.

V is a parameter

Following the procedure of the previous section, we first determine the domain of the function. Clearly this problem makes physical sense only for $r > 0$. Looking at the equation

Finding the domain of $f(r)$

$$h = \frac{V}{\pi r^2},$$

we see that although V is fixed, r can be arbitrarily large provided h is sufficiently small (resulting in a can that looks like an elephant stepped on it). Thus the domain of our function is $r > 0$, which is not a closed interval, so we have no guarantee that a minimum exists.

Next we compute $f'(r)$, keeping in mind that the symbols V and π represent constants and that we are differentiating with respect to the variable r.

$$f'(r) = 4\pi r - \frac{2V}{r^2} = \frac{4\pi r^3 - 2V}{r^2}$$

The derivative is undefined at $r = 0$, which is outside the domain under consideration. So now we set the derivative equal to zero and solve for any possible critical points.

Looking for critical points

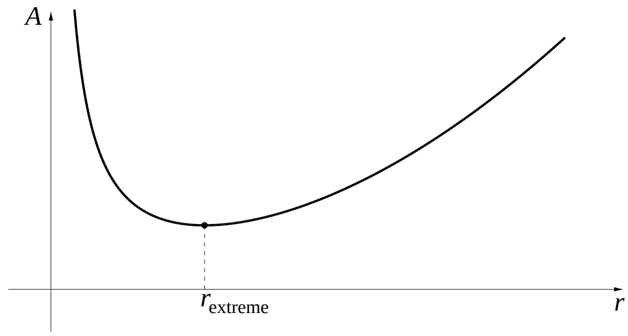
$$\begin{aligned} f'(r) &= \frac{4\pi r^3 - 2V}{r^2} \\ 0 &= \frac{4\pi r^3 - 2V}{r^2} \\ 0 &= 4\pi r^3 - 2V \\ r &= \sqrt[3]{V/2\pi} \end{aligned}$$

Thus $r = \sqrt[3]{V/2\pi}$ is the only critical point.

We can actually sketch the shape of the graph of A versus r based on this analysis of $f'(r)$. The sign of $f'(r)$ is determined by its numerator, since the denominator r^2 is always positive.

Finding the shape of the graph of A

- When $r < \sqrt[3]{V/2\pi}$, the numerator is negative, so the graph of A is falling.
- When $r > \sqrt[3]{V/2\pi}$, the numerator is positive, so the graph is rising.



Obviously, A has a *minimum* at

$$r_{\text{extreme}} = \sqrt[3]{V/2\pi}.$$

A has a minimum but no maximum

It is also obvious that there is no *maximum* area, since

$$\lim_{r\to 0} A = \infty \quad \text{and} \quad \lim_{r\to\infty} A = \infty.$$

Thus we see again that not *every* optimization problem has a solution.

The Mathematical Context: Optimal Shapes

It is interesting to find the value of the height $h = h_{\text{extreme}}$ when the area is a minimum. Since

$$h_{\text{extreme}} = \frac{V}{\pi r_{\text{extreme}}^2},$$

we can use $r_{\text{extreme}} = \sqrt[3]{V/2\pi}$ to write $V = 2\pi r_{\text{extreme}}^3$ and thus find

$$h_{\text{extreme}} = \frac{2\pi r_{\text{extreme}}^3}{\pi r_{\text{extreme}}^2} = 2r_{\text{extreme}}.$$

In other words, the height of the optimal tin can exactly equals its diameter. Campbell soup cans are far from optimal, but a can of Progresso plum tomatoes has diameter 4 inches and height 4.5 inches. Does someone at Progresso know calculus?

Problems like the one above—as well as others that you will find in the exercises—have been among the mainstays of calculus courses for generations. However, if you really had to face the problem of optimizing the shape of a tin can as a practical matter, you would probably have a numerical value for V specified in advance. In that case, your first impulse might be to use a graphing utility to approximate $r_{extreme}$, as accurately as your needs warrant. This is entirely appropriate. Using a graphing utility gives the answer as quickly as taking derivatives, and you are less likely to make mistakes in algebra. Moreover, the functions encountered in many other real problems are often complicated and messy, and their derivatives are hard to analyze. A graphing utility may provide the only practical course open to you.

So why do things the hard way?

So why insist that you use derivatives on these problems? It is because the actual context of these examples is not really saving money for food canners. The real context is geometry. If we had used a graphing facility to solve the can problem, we might have noticed that the optimal radius was about half the height, but we wouldn't have known the relationship is exact. Nor would we have recognized that the relationship holds for cylinders of arbitrary volume.

Using a graphing utility is essentially mindless, which is part of its virtue if we just need a specific answer to a particular problem. Part of the purpose of a calculus course, though, is to develop the concepts, tools, and the geometric vocabulary for thinking about problems in general, to see the connecting threads between apparently disparate settings. The more clearly we can think within a general framework, not just that of the specific problem being addressed, the better intuitions we will develop that will help us see unsuspected relationships in the problem at hand or think more creatively about other problems in other settings. The ability to give precise expression to our intuitions can lead to deeper insights. You should thus try to see the exercises that follow as not just about fences and storage bins—they are opportunities to observe that, often, the geometric regularity that pleases the eye is also optimal.

Symmetric shapes are often optimal

Exercises

For each of the following problems, find a function relating the variables and use differentiation to find the optimal value specified. Some of the problems (1, 3, 4, and 5) are expressed in terms of a general parameter—P, L, A, and L again—rather than specific numerical values. If this gives you trouble, try doing the problem using the specific parameter value given at the end of the problem for computer verification. Use other specific values as needed until you can do the problem in terms of the general parameter, which should behave exactly like any of the specific values you tried.

1. Show that the rectangle of perimeter P whose area is a maximum is a square. Use a graphing utility to check your answer for the special case when P=100 feet.

2. An open rectangular box is to be made from a piece of cardboard 8 inches wide and 15 inches long by cutting a square from each corner and bending up the sides. Find the dimensions of the box of largest volume. Use a graphing utility to check your answer.

3. One side of an open field is bounded by a straight river. A farmer has L feet of fencing. How should the farmer proportion a rectangular plot along the river in order to enclose as great an area as possible? Use a graphing utility to check your answer for the special case when $L = 100$ feet.

4. An open storage bin with a square base and vertical sides is to be constructed from A square feet of wood. Determine the dimensions of the bin if its volume is to be a maximum. (Neglect the thickness of the wood and any waste in construction.) Use a graphing utility to check your answer for the special case when A=100 square feet.

5. A roman window is shaped like a rectangle surmounted by a semicircle. If the perimeter of the window is L feet, what are the dimensions of the window of maximum area? Use a graphing utility to check your answer for the special case when $L = 100$ feet.

6. Suppose the roman window of problem 5 has clear glass in its rectangular part and colored glass in its semicircular part. If the colored glass transmits only half as much light per square foot as the clear glass does, what are the

dimensions of the window that transmits the most light? Use a graphing utility to check your answer for the special case when $L = 100$ feet.

7. A cylindrical oil can with radius r inches and height h inches is made with a steel top and bottom and cardboard sides. The steel costs 3 cents per square inch, the cardboard costs 1 cent per square inch, and rolling the crimp around the top and bottom edges costs 1/2 cent per linear inch. (Both crimps are done at the same time, so only count the contribution of one circumference.)

a) Express the cost C of the can as a function of r and h.

b) Find the dimensions of the cheapest can holding 100 cubic inches of oil. (You'll need to solve a cubic equation to find the critical point. Use a graphing utility or an estimation procedure to approximate the critical point to 3 decimal places.)

5.5 Newton's Method

Finding Critical Points

When we solve optimization problems for functions given by formulas, we begin by calculating the derivative and using the derivative formula to find critical points. Almost always the derivative is defined for all elements of the domain, and we find the critical points by determining the roots of the equation obtained by setting the derivative equal to zero.

Finding the roots of this equation often requires an estimation procedure. For example, consider the function

$$f(x) = x^4 + x^3 + x^2 + x + 1.$$

The derivative of f is

$$f'(x) = 4x^3 + 3x^2 + 2x + 1,$$

which is certainly defined for all x.

In order to use a graphing utility to find the roots of $f'(x) = 0$, we need to choose an interval that will contain the roots we seek. Since $f'(0) = 1 > 0$ and $f'(-1) = -2 < 0$, we know f' has at least one root on $[-1, 0]$. (Why is this?) But might there be other roots outside this interval?

Solving $f'(x) = 0$

It is easy to see that $f'(x)$ is positive for all $x > 0$, so there are no roots to the right of $[-1, 0]$. What about $x < -1$? Rewriting the derivative as

$$f'(x) = (2x + 1)(2x^2 + 1) + x^2$$

lets us see that $f'(x)$ is negative for all $x < -1$ (check this for yourself), so there are no roots to the left of $[-1, 0]$ either.

Examining the graph of $y = 4x^3 + 3x^2 + 2x + 1$ on $[-1, 0]$, we see that it crosses the x-axis exactly once, so there is a unique critical point. Progressively shrinking the interval, we find that, to eight decimal places, this critical point is

$$c = -.60582958\ldots.$$

Notice that it in this method it requires roughly the same amount of work to get each additional digit in the answer.

A good algorithm is a convenient tool

There is, however, another approximation procedure we can use, called Newton's method. This has wide applicability and usually converges very rapidly to solutions—additional digits are much easier to get than in the method we just looked at. It also has the virtue of being algorithmic, so that we can write a single computer program which can then be used for any root-finding problem we may encounter without a great deal of thought and decision-making. Newton's method is also an interesting application of both local linearity and successive approximation, two important themes of this course.

Local Linearity and the Tangent Line

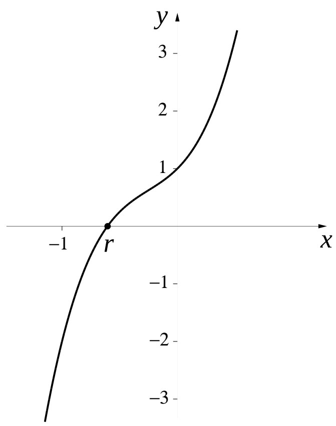
The root is an x-intercept of the graph

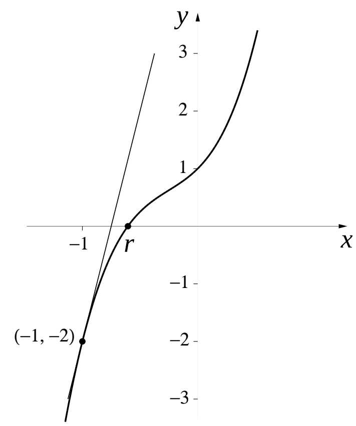
Let's give the derivative function the new name g to emphasize that we now want to consider it as a function in its own right, out of the context of the function f from which it was derived. Look at the following graph of the function

$$y = g(x) = 4x^3 + 3x^2 + 2x + 1\,.$$

We are seeking the number r such that $g(r) = 0$. The root r is the x-coordinate of the point where the graph crosses the x-axis.

The basic plan of attack in Newton's method is to replace the graph of $y = g(x)$ by a straight line that looks reasonably like the graph near the root r.Then, the x-intercept of that line will be a reasonably good estimate for r. The graph below includes such a line.





The function g is locally linear at $x = -1$, and we have drawn an extension of the local linear approximation of g at this point. This line is called the **tangent line** to the graph of $y = g(x)$ at $x = -1$, by analogy to the tangent line to a circle. To find the x-intercept of the tangent line, we must know its equation. Clearly the line passes through the **point of tangency** $(-1, g(-1)) = (-1, -2)$. What is its slope? It is the same as the slope of the local linear approximation at $(-1, g(-1))$, namely

The tangent line extends the local linear approximation

$$g'(-1) = 12(-1)^2 + 6(-1) + 2 = 8.$$

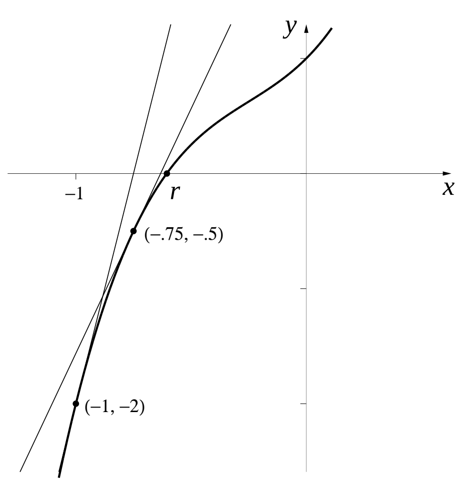
Thus the equation of the tangent line is

The equation of the tangent line

$$y + 2 = 8(x + 1).$$

To find the x-intercept of this line, we must set y equal to zero and solve for x: $0 + 2 = 8(x + 1)$ gives us $x = -0.75$. Of course, this x-intercept is not equal to r, but it's a better approximation than, say, –1. To get an even better approximation, we repeat this process, starting with the line tangent to the graph of g at $x = -0.75$ instead of at $x = -1$.

Finding the x-intercept of the tangent line



The slope of this new tangent line is $g'(-0.75) = 4.25$, and it passes through $(-0.75, -0.5)$, so its equation is

$$y + 0.5 = 4.25(x + 0.75).$$

Setting $y = 0$ and solving for x gives a new x-intercept whose value is equal to -0.6323529, closer still to r.

It seems reasonable to repeat the process yet again, using the tangent line at $x = -0.6323529$. Let's first introduce some notation to keep track of our computations. Call the original point of tangency x_0, so $x_0 = -1$. Let x_1 be the x-intercept of the tangent line at $x = x_0$. Draw the tangent line at $x = x_1$, and call its x-intercept x_2. Continuing in this way, we get a sequence of points $x_0, x_1, x_2, x_3, \ldots$ which appear to approach nearer and nearer to r. That is, they appear to approach r as their limit.

Successive approximations get closer to the root r

The Algorithm

This process of using one number to determine the next number in the sequence is the heart of Newton's method—it is an iterative method. Moreover, it turns out to be quite simple to calculate each new estimate in terms of the previous one. To see how this works, let's compute x_1 in terms of x_0. We know x_1 is the x-intercept of the line tangent to the graph of g at $(x_0, g(x_0))$. The slope of this line is $g'(x_0)$, so

The general equation of the tangent line

$$y - g(x_0) = g'(x_0)(x - x_0)$$

is the **equation of the tangent line**. Since this line crosses the x-axis at the point $(x_1, 0)$, we set $x = x_1$ and $y = 0$ in the equation to obtain

$$0 - g(x_0) = g'(x_0)(x_1 - x_0).$$

Now it is easy to solve for x_1:

$$\begin{aligned} g'(x_0)(x_1 - x_0) &= -g(x_0) \\ x_1 - x_0 &= \frac{-g(x_0)}{g'(x_0)} \\ x_1 &= x_0 - \frac{g(x_0)}{g'(x_0)}. \end{aligned}$$

In the same way we get

$$x_2 = x_1 - \frac{g(x_1)}{g'(x_1)}, \qquad x_3 = x_2 - \frac{g(x_2)}{g'(x_2)},$$

and so on.

To summarize, suppose that x_0 is given some value START. Then Newton's method is the computation of the sequence of numbers determined by

$$\begin{aligned} x_0 &= \text{START} \\ x_{n+1} &= x_n - \frac{g(x_n)}{g'(x_n)}, \qquad n = 0, 1, 2, 3, \ldots \end{aligned}$$

As we have seen many times, the sequence

$$x_1, x_2, x_3, \ldots, x_n, \ldots$$

is a list of numbers to which we can always add a new term—by iterating our method yet again. For most functions, if we begin with an appropriate starting value of x_0, there is another number r that is the *limit* of this list of numbers, in the sense that the difference between x_n and r becomes as small as we wish as n increases without bound,

The limit of the successive approximations is the root

$$r = \lim_{n \to \infty} x_n.$$

The numbers $x_1, x_2, x_3, \ldots, x_n, \ldots$ constitute a sequence of *successive approximations* for the root r of the equation $g(x) = 0$. We can write a computer program to carry out this algorithm for as many steps as we choose. The program NEWTON does just that for $g(x) = 4x^3 + 3x^2 + 2x + 1$.

Program: NEWTON
Newton's method for solving $g(x) = 4x^3 + 3x^2 + 2x + 1 = 0$

```
start = -1
numberofsteps = 8
x = start
FOR n = 0 to numberofsteps
   print n, x
   g = 4 * x^3 + 3 * x^2 + 2 * x + 1
   gprime = 12 * x^2 + 6 * x + 2
    x = x - g/gprime
 NEXT n
```

If we program a computer using this algorithm with START = −1, then we get

$$\begin{aligned}
x_0 &= -1.00000000000 \\
x_1 &= -0.75000000000 \\
x_2 &= -0.63235294118 \\
x_3 &= -0.60687911790 \\
x_4 &= -0.60583128240 \\
x_5 &= -0.60582958619 \\
x_6 &= -0.60582958619 \\
x_7 &= -0.60582958619 \\
x_8 &= -0.60582958619.
\end{aligned}$$

Thus we have found the root of $g(x) = 0$—the critical point we were looking for. In fact, after only 6 steps we could see that the value of the critical point was specified to at least ten decimal places. Also at the sixth step, we had the eight decimal places obtained with the use of the graphing utility. In fact, it turns out that the number of decimal places fixed roughly doubles with each round. In the above list, for instance, x_2 fixed one decimal, x_3 fixed two decimals, x_4 fixed four, x_5 fixed ten (the eleventh digit of the root is really an 8, which gets rounded to a 9 in $x_6 - x_8$.), and x_6 would have fixed at least twenty if we had printed them all out! Moreover, by changing only three lines of the program—the first, sixth, and seventh—we can use NEWTON for any other function. With the use of the program NEWTON, we will see that in most cases we can obtain results more quickly and to a higher degree of accuracy with Newton's method than by using a graphing utility, although it is still sometimes helpful to use a graphing utility to get a reasonable starting value.

Examples

Example 1. Start with $\cos x = x$. The solution(s) to this equation (if any) will be the x-coordinates of any points of intersection of the graphs of $y = \cos x$ and $y = x$. Sketch these two graphs and convince yourself that there is one solution, between 0 and $\pi/2$. The equation $\cos x = x$ is not in the form $g(x) = 0$, so rewrite it as $\cos x - x = 0$. Now we can apply Newton's method with $g(x) = \cos x - x$. Try starting with $x_0 = 1$. This gives the iteration scheme

$$x_0 = 1$$
$$x_{n+1} = x_n - \frac{\cos x_n - x_n}{-\sin x_n - 1}, \qquad n = 0, 1, 2, \ldots$$

The numbers we get are

$$\begin{aligned} x_0 &= 1.000000000 \\ x_1 &= \ .750363868\ldots \\ x_2 &= \ .739112891\ldots \\ x_3 &= \ .739085133\ldots \\ x_4 &= \ .739085133\ldots \end{aligned}$$

We have the solution to 9 decimal places in only four steps. Not only does Newton's method work, it works fast!

Example 2. Suppose we continue with the equation $\cos x = x$, but this time choose $x_0 = 0$. What will we find? The numbers we get are

$$\begin{aligned} x_0 &= 0.000000000 \\ x_1 &= 1.000000000 \\ x_2 &= 0.750363868 \end{aligned}$$

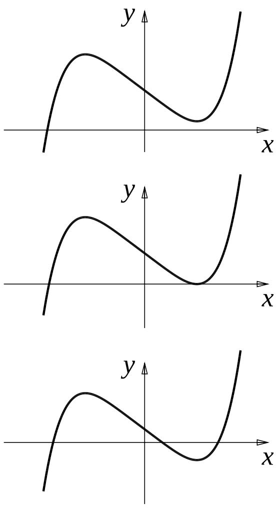
There is no need to continue; we can see that we will again obtain $r = 0.739085133\ldots$ as in Example 1. Look again at your sketch and see why you might have predicted this result.

Example 3. Next, let's find the roots of the polynomial $x^5 - 3x + 1$. This means solving the equation $x^5 - 3x + 1 = 0$. We know the necessary derivative, so we're ready to apply Newton's method, except for one thing: which starting value x_0 do we pick? This is the part of Newton's method that leaves us on our own.

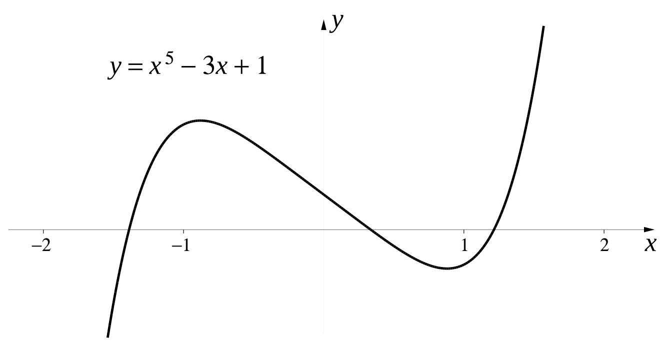
Finding the starting value x_0 can be hard

Assuming that some graphing software is available, the best thing to do is graph the function. But for most graphing utilities, we need to choose an interval. How do we choose one which is sure to include all the roots of the polynomial? The derivative $5x^4 - 3$ of this polynomial is simple enough that we can use it to get an idea of the shape of the graph of $y = x^5 - 3x + 1$ before we turn to the computer. Clearly the derivative is zero only for $x = \pm\sqrt[4]{3/5}$, and the derivative is positive except between these two values of x. In other words, we know the shape of the graph of $y = g(x)$—it is increasing, then decreases for a bit, then increases from there on out. This still is not enough information to tell us how many roots g has, though; its graph might lie in any one of the following configurations and so have 1, 2, or 3 roots (there are two other possibilities not shown—one has 2 roots and one has 1 root). We can thus say that $g(x) = 0$ has at least one and at most three real roots. However, if we further observe that $g(-2) = -25, g(-1) = 3, g(0) = 1, g(1) = -1$, and $g(2) = 27$, we see that the graph of g must cross the x–axis at some value of x between -2 and -1, between 0 and 1, and again between 1 and 2. Therefore the right–hand sketch above must be the correct one.

Using the derivative to find the shape of the graph



Or we can almost as easily turn to a graphing utility. If we try the interval $[-5, 5]$, we see again that the graph crosses the x-axis in exactly three points.



Finding the root between -2 and -1

One of the roots is between -2 and -1, one is between 0 and 1, and the third is between 1 and 2. To find the first, we apply Newton's method with $x_0 = -2$. Then we get

$$
\begin{aligned}
x_0 &= -2.000000000 \\
x_1 &= -1.675324675\ldots \\
x_2 &= -1.478238029\ldots \\
x_3 &= -1.400445373\ldots \\
x_4 &= -1.389019863\ldots \\
x_5 &= -1.388792073\ldots \\
x_6 &= -1.388791984\ldots \\
x_7 &= -1.388791984\ldots
\end{aligned}
$$

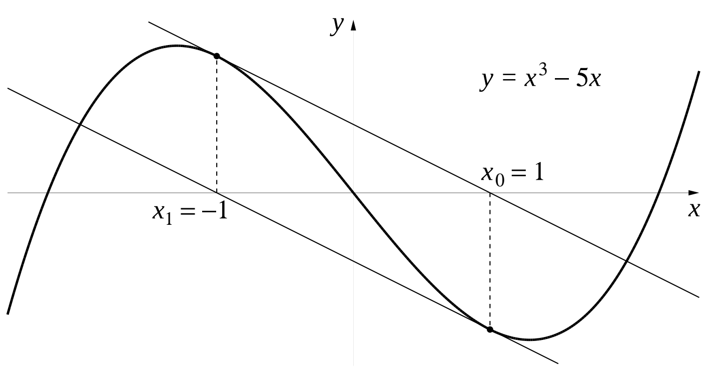
This took a few more steps than the other examples, but not a lot. Notice again that once there are any decimals fixed at all, the number of decimals fixed roughly doubles in the next approximation. In the exercises you will be asked to compute the other two roots.

Example 4. Let's use Newton's method to find the obvious solution $r = 0$ of $x^3 - 5x = 0$. If we choose x_0 sufficiently close to 0, Newton's method should work just fine. But what does "sufficiently close" mean? Suppose we try $x_0 = 1$. Then we get

$$\begin{aligned} x_0 &= 1 \\ x_1 &= -1 \\ x_2 &= 1 \\ x_3 &= -1 \\ x_4 &= 1 \\ &\vdots \end{aligned}$$

The x_n's oscillate endlessly, never getting close to 0. Going back to the geometric interpretation of Newton's method, this oscillation can be explained by the graph below.

Newton's method can fail



Using more advanced methods, it is possible to get precise estimates for how close x_0 needs to be to r in order for Newton's method to succeed. For now, we'll just have to rely on common sense and trial and error.

One important thing to note is the relation between algebra and Newton's method. Although we can now solve many more equations than we could earlier, this doesn't mean that we can abandon algebra. In fact, given a new equation, you should first try to solve it algebraically, for exact solutions are often better. Only when this fails should you look for approximate solutions using Newton's method. So don't forget algebra—you'll still need it!

Exercises

1. **The Babylonian algorithm.** Show that the Babylonian algorithm of chapter 2 is the same as Newton's method applied to the equation $x^2 - a = 0$.

2. When Newton introduced his method, he did so with the example $x^3 - 2x - 5 = 0$. Show that this equation has only one root, and find it.

This example appeared in 1669 in an unpublished manuscript of Newton's (a published version came later, in 1711). The interesting fact is that Newton's method *differs* from the one presented here: his scheme was more complicated, requiring a different formula to get each approximation. In 1690, Joseph Raphson transformed Newton's scheme into the one used above. Thus, "Newton's method" is more properly called the "Newton–Raphson method", and many modern texts use this more accurate name.

3. Use Newton's method to find a solution of $x^3 + 2x^2 + 10x = 20$ near the point $x = 1$.

The approximate solution 1;22,7,42,33,4,40 of this equation appears in a book written in 1228 by Leonardo of Pisa (also known as Fibonacci). This number looks odd because it's written in sexagesimal notation: it translates into

$$1 + \frac{22}{60} + \frac{7}{60^2} + \frac{42}{60^3} + \frac{33}{60^4} + \frac{4}{60^5} + \frac{40}{60^6}.$$

This solution is accurate to 10 decimal places, which is not bad for 750 years ago. In the Middle Ages, there was a lot of interest in solving equations. There were even contests, with a prize going to the person who could solve the most. The quadratic formula, which expresses algebraically the roots of any second degree equation, had been known for thousands of years, but there were no general methods for finding roots of higher degree equations in the 13th century. We don't know how Leonardo found his solution—why give away your secrets to your competitors!

4. Use Newton's method to find a solution of $x^3 + 3x^2 = 5$.

In 1530, Nicolo Tartaglia was challenged to solve this equation algebraically. Five years later, in 1535, he found the solution

$$x = \sqrt[3]{\frac{3+\sqrt{5}}{2}} + \sqrt[3]{\frac{3-\sqrt{5}}{2}} - 1\,.$$

Initially, Tartaglia could solve only certain types of cubic equations, but this was enough to let him win some famous contests with other mathematicians of the time. By 1541, he knew the general solution, but he made the mistake of telling Geronimo Cardano. Cardano published the solution in 1545 and the resulting formulas are called "Cardan's Formulas".

The above solution of $x^3 + 3x^2 = 5$ is called a **solution by radicals** because it is obtained by extracting various roots or radicals. Similarly, some time before 1545, Luigi Ferrari showed that any fourth degree equation can be solved by radicals. This led to an intense interest in the fifth degree equation. To see what happens in this case, read the next problem.

5. In Example 3, we saw that one root of $x^5 - 3x + 1$ was –1.39887919.... Use Newton's method to find the other two roots.

> In 1826, Niels Henrik Abel proved that the general polynomial of degree 5 or greater cannot be solved by radicals. Using the work of Evariste Galois (done around 1830, but not understood until many years later), it can be shown that the roots of the equation $x^5 - 3x + 1 = 0$ cannot be expressed by any combination of radicals. Thus algebra can't solve this equation—some kind of successive approximation technique is unavoidable!

6. One of the more surprising applications of Newton's method is to compute reciprocals. To make things more concrete, we will compute 1/3.4567 . Note that this number is the root of the equation $1/x = 3.4567$.

a) Show that the formula of Newton's method gives us

$$x_{n+1} = 2x_n - 3.4567x_n^2.$$

b) Using $x_0 = .5$ and the formula from (a), compute 1/3.4567 to a high degree of accuracy.

c) Try starting with $x_0 = 1$. What happens? Explain graphically what goes wrong.

This method for computing reciprocals is important because it involves only *multiplication* and *subtraction*. Since $a/b = a \cdot (1/b)$, this implies that division can likewise be built from multiplication and subtraction. Thus, when designing a computer, the division routine doesn't need to be built from scratch—the designer can use the method illustrated here. There are some computers that do division this way.

7. In this problem we will determine the maximum value of the function

$$f(x) = \frac{x+1}{x^4+1}.$$

a) Graph $f(x)$ and convince yourself that the maximum value occurs somewhere around $x = .5$. Of course, the exact location is where the slope of the graph is zero, i.e., where $f'(x) = 0$. So we need to solve this equation.

b) Compute $f'(x)$.

c) Since the answer to (b) is a fraction, it vanishes when its numerator does. Setting the numerator equal to 0 gives a fourth degree equation. Use Newton's method to find a solution near $x = .5$.

d) Compute the maximum value of $f(x)$.

8. Consider the hyperbola $y = 1/x$ and the circle $x^2 - 4x + y^2 + 3 = 0$.

a) By graphing the circle and the hyperbola, convince yourself that there are two points of intersection.

b) By substituting $y = 1/x$ into the equation of the circle, obtain a fourth degree equation satisfied by the x–coordinate of the points of intersection.

c) Solve the equation from (b) by Newton's method, and then determine the points of intersection.

9. Sometimes Newton's method doesn't work so nicely. For example, consider the equation $\sin x = 0$.

a) Compute x_1 using Newton's method for each of the four starting values $x_0 = 1.55, 1.56, 1.57$ and 1.58.

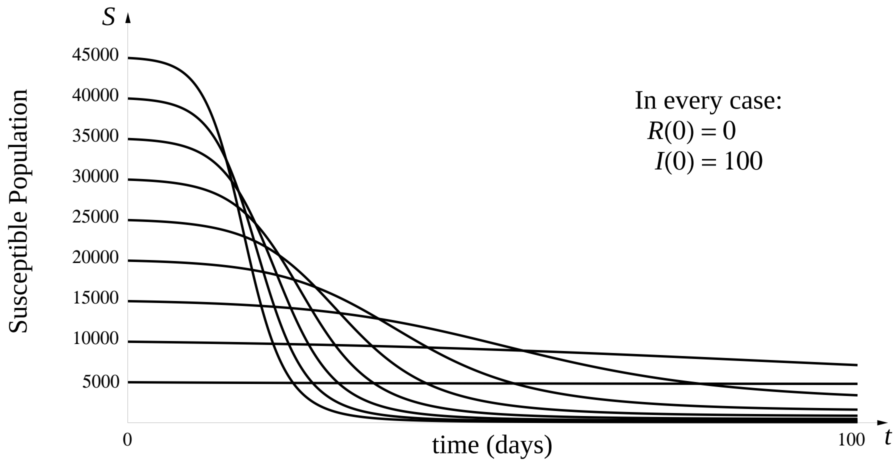
b) The answers you get are wildly different. Using the basic formula

$$x_{n+1} = x_n - \frac{g(x_n)}{g'(x_n)}$$

explain why.

The epidemic runs its course

We return to the epidemiology example we have studied since chapter 1. Recall that our S-I-R model keeps track of three subgroups of the population: the susceptible, the infected, and the recovered. One of the interesting features of the model is that the larger the initial susceptible population, the more rapidly the epidemic runs its course. We observe this by choosing fixed values of $R_0 = R(0)$ and $I_0 = I(0)$ and looking at graphs of $S(t)$ versus t for various values of $S_0 = S(0)$.



We see in each case that for sufficiently large t the graph of S levels off, approaching a value we'll call S_∞:

$$S_\infty = \lim_{t\to\infty} S(t).$$

What we mean by the epidemic "running its course" is that $S(t)$ reaches this limit value. We can see from the graphs that the value of S_0 affects the number S_∞ of individuals who escape the disease entirely. It turns out that we can actually find the value of S_∞ if we know the values of S_0, I_0, and the parameters a and b.

Recall that a is the *transmission coefficient*, and b is the *recovery coefficient* for the disease. The differential equations of the S-I-R model are

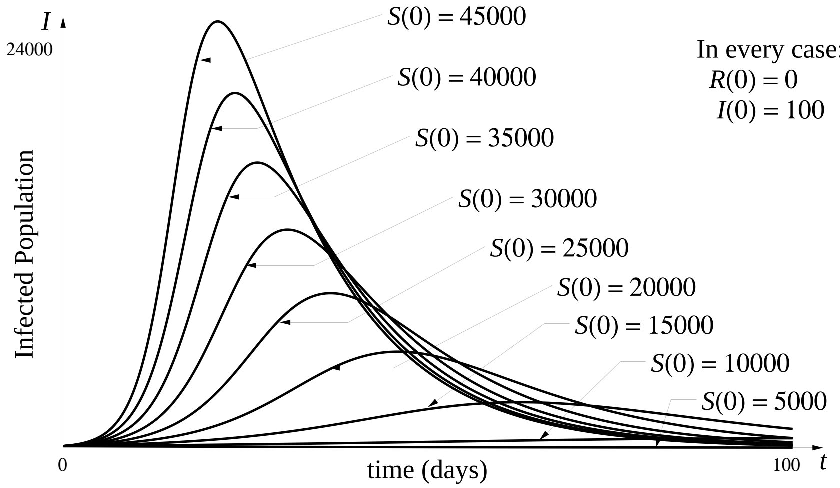
$$\begin{aligned} S' &= -aSI, \\ I' &= aSI - bI, \\ R' &= bI. \end{aligned}$$

10. Use the differentiation rules together with these differential equations to show that

$$(I + S - (b/a) \cdot \ln S)' = 0.$$

11. Explain why the result of problem 10 means that $I + S - (b/a) \cdot \ln S$ has the same value—call it C—for every value of t.

12. Look at the graphs of the solutions $I(t)$ for various values of $S(0)$ below.



Write $\lim_{t\to\infty} I(t) = I_\infty$. What is the value of I_∞ for all values of $S(0)$?

13. Use the results of problems 11 and 12 to explain why

$$S_\infty - \frac{b}{a}\ln(S_\infty) = I_0 + S_0 - \frac{b}{a}\ln(S_0).$$

This equation determines S_∞ implicitly as a function of I_0 and S_0. For particular values of I_0 and S_0 (and of the parameters), you can use Newton's method to find S_∞.

14. Use the values

$$\begin{aligned} a &= .00001 \text{ (person-days)}^{-1} \\ b &= .08 \text{ day}^{-1} \\ I_0 &= 100 \text{ persons} \\ S_0 &= 35{,}000 \text{ persons} \end{aligned}$$

Writing x instead of S_∞ gives

$$x - 8000\ln(x) = -48{,}605.$$

Apply Newton's method to find $x = S_\infty$. Judging from the graph for $S_0 = 35000$, it looks like a reasonable first estimate for S_∞ might be 100.

15. Using the same values of a, b, and I_0 as in problem 14, determine the value of S_∞ for each of the following initial population sizes:

a) $S_0 = 45,000$.

b) $S_0 = 25,000$.

c) $S_0 = 5,000$.

5.6 Chapter Summary

The Main Ideas

- **Formulas** for **derivatives** can be calculated for functions given by formulas using the definition of the derivative as the limit of difference quotients
$$\lim_{\Delta x \to 0} \frac{f(x + \Delta x) - f(x)}{\Delta x}.$$

- Each particular difference quotient is an **approximation** to the derivative. Successive approximations, for smaller and smaller Δx, approach the derivative as a limit. The formula for the derivative gives its *exact* value.

- Formulas for **partial derivatives** are obtained using the formulas for derivatives of functions of a single variable by simply treating all variables other than the one of interest as if they were constants.

- **Optimization** problems occur in many different contexts. For instance, we seek to maximize benefits, minimize energy, and minimize error.

- The **sign** of the derivative indicates where a graph rises and where it falls.

- Functions which are **continuous** on a **finite closed interval** have **global extremes**.

- For a function continuous on an interval, its **local extremes** occur at **critical points**—points where the derivative equals zero or fails to exist—or at endpoints.

- Local linearity permits us to replace the graph of a function $y = g(x)$ by its **tangent line** at a point near a root of $g(x) = 0$, and then the x-intercept of the tangent line is a better approximation to the root. Successive approximations, obtained by iterating this procedure, yield **Newton's method** for solving the equation $g(x) = 0$.

Self-Testing

- You should be able to **differentiate** a function given by a formula.
- You should be able to use differentiation formulas to **calculate** partial derivatives.
- In most cases, you should be able to determine from a graph of a function on an interval whether that function is **continuous** and/or **differentiable** on that interval.
- You should be able to find **critical points** for a function of one or several variables given by a formula.
- You should be able to use the formula for the derivative of a function of a single variable to find **local and global extremes**.
- You should be able to use **Newton's method** to solve an equation of the form $g(x) = 0$.

Chapter Exercises

Prices, demand and profit

Suppose the demand D (in units sold) for a particular product is determined by its price p (in dollars), $D = f(p)$. It is reasonable to assume that when the price is low, the demand will be high, but as the price rises, the demand will fall. In other words, we assume that the slope of the demand function is negative. If the manufacturing cost for each unit of the product is c dollars, then the profit per unit at price p is $p - c$. Finally the total profit T gained at the unit price p will be the number of units sold at the price p (that is, the demand $D(p)$) multiplied by the profit per unit $p - c$.

$$T = g(p) = D(p)\,\text{units} \times (p - c)\,\frac{\text{dollars}}{\text{unit}}$$

In this series of problems we will determine the effect of the demand function and of the unit manufacturing cost on the maximum total profit.

1. Suppose the demand function is linear

$$D = f(p) = 1000 - 500p \text{ units},$$

and the unit manufacturing cost is .20, so the total profit is

$$T = g(p) = (1000 - 500p)(p - .20) \text{ dollars}.$$

Find the "best" price – that is, find the price that yields the maximum total profit.

2. Suppose the demand function is the same as in problem 1, but the unit manufacturing cost rises to .30. What is the "best" price now? How much of the rise in the unit manufacturing cost is passed on to the consumer if the manufacturer charges this best price?

3. Suppose the demand function for a particular product is $D(p) = 2000 - 500p$, and that the unit manufacturing cost is .30. What price should the manufacturer charge to maximize her profit? Suppose the unit manufacturing cost rises to .50. What price should she charge to maximize her profit now? How much of the rise in the unit manufacturing cost should she pass on to the consumer?

4. If the demand function for a product is $D(p) = 1500 - 100p$, compare the "best" price for unit manufacturing costs of .30 and .50. How much of the rise in cost should the manufacturer pass on to the consumer?

5. As you may have noticed, problems 2–4 illustrate an interesting phenomenon. In each case, exactly half of the rise in the unit manufacturing cost should be passed on to the consumer. Is this a coincidence?

a) Consider the most general case of a linear demand function

$$D = f(p) = a - mp.$$

and unit cost c. What is the "best" price? Is exactly half the unit manufacturing cost passed on to the consumer? Explain your answer.

b) Now consider a non-linear demand function

$$D = f(p) = \frac{1000}{1 + p^2}.$$

Find the “best” price for unit costs

i) .50 dollar per unit;

ii) 1.00 dollar per unit;

iii) 1.50 dollars per unit.

How much of the price increase is passed on to the consumer in cases (ii) and (iii)?