Chapter 5
Applications of the Derivative

Abstract We present five applications of the calculus:

1. Atmospheric pressure in a gravitational field
2. Motion in a gravitational field
3. Newton's method for finding the zeros of a function
4. The reflection and refraction of light
5. Rates of change in economics

5.1 Atmospheric Pressure

If you have ever traveled in the mountains, you probably noticed that air pressure diminishes at higher altitudes. If you exert yourself, you get short of breath; if you cook, you notice that water boils at less than 100 °C. Our first application is to derive a differential equation to investigate the nature of air pressure as a function of altitude.

Let $P(y)$ be the air pressure [force/area] at altitude y above sea level. The force of air pressure at altitude y supports the weight of air above y. Imagine a column of air of unit cross section. The volume of air in the column between the altitudes y and $y+h$ (see Fig. 5.1) is h unit volumes. The weight of this air is $h\overline{\rho}g$, where h is the volume of the column, $\overline{\rho}$ is the average density [mass/volume] of air between altitudes y and $y+h$, and g is the acceleration due to gravity. This weight is supported by the force of air pressure at y minus the force of air pressure at $y+h$. Therefore,

$$h\overline{\rho}g = P(y) - P(y+h).$$

Dividing by h, we get

$$\overline{\rho}g = \frac{P(y) - P(y+h)}{h}.$$

P.D. Lax and M.S. Terrell, *Calculus With Applications*, Undergraduate Texts in Mathematics, 217
DOI 10.1007/978-1-4614-7946-8_5, © Springer Science+Business Media New York 2014

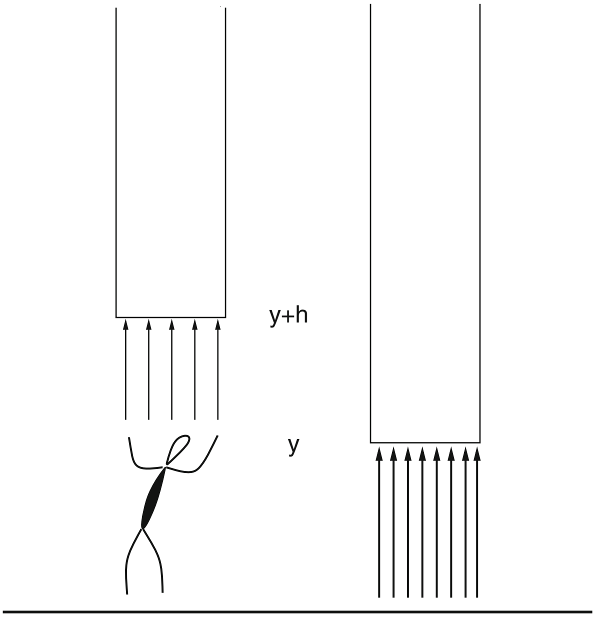

Fig. 5.1 The column of atmosphere above altitude y is heavier than that above $y+h$

As h tends to zero, the average density of the air, $\overline{\rho}$, tends to the density $\rho(y)$ at y, and the difference quotient on the right-hand side tends to $-P'(y)$, giving us the differential equation

$$\rho(y)g = -P'(y). \tag{5.1}$$

When gas is compressed, both its density and its pressure are increased. If we assume that air pressure and density are linearly related,

$$\rho = kP, \qquad k \text{ some constant,}$$

and set this into the differential equation (5.1), we get

$$kgP(y) = -P'(y).$$

According to Theorem 3.11, the solutions of this equation are the exponential functions

$$P(y) = P(0)e^{-kgy},$$

where $P(0)$ is atmospheric pressure at sea level.

So we have deduced that atmospheric pressure is an exponential function of altitude. Is there anything we can say about the constant k? We can determine the dimension of k from $k = \frac{\rho}{P}$. The dimension of k is then

$$\left[\frac{\text{density}}{\text{pressure}}\right] = \left[\frac{\frac{\text{mass}}{\text{volume}}}{\frac{\text{force}}{\text{area}}}\right].$$

Since force is equal to mass times acceleration, the dimension of k is

$$\left[\frac{\text{mass}/\text{length}^3}{\frac{\text{mass}\times\text{length}}{\text{time}^2}/\text{length}^2}\right] = \frac{\text{time}^2}{\text{length}^2} = \frac{1}{\text{velocity}^2}.$$

What is the velocity that appears in k? What velocity can one associate with the atmosphere? It turns out that $1/k$ is the square of the speed of *sound*. Let us check this value for k numerically using $P(y) = P(0)\mathrm{e}^{-kgy}$ to calculate air pressure at Denver, Colorado, at altitude 1 mile. The speed of sound at sea level is approximately 1000 ft/s. Therefore,

$$k = 10^{-6}\ \ (\text{s}/\text{ft})^2, \qquad g = 32\ \ \text{ft}/\text{s}^2, \qquad y = 5\,280\ \text{ft},$$

and $kgy = 10^{-6}(32)(5\,280) = 0.169$. Since $\mathrm{e}^{-0.169} = 0.844$, the air pressure formula gives

$$P(1\text{ mile}) = 0.844\,P(0).$$

Atmospheric pressure at sea level is 14.7 psi (pounds per square inch). Using our formula, we get $(0.844)(14.7) = 12.4$ psi for air pressure at Denver. The measured value of atmospheric pressure in Denver is 12.1 psi, so our formula gives a good approximation.

Problems

5.1. Compare the atmospheric pressure at your city to the approximate value determined by the equation in this section.

5.2. While investigating atmospheric pressure, we assumed that air density is proportional to air pressure. In this problem, consider pressure in the ocean, where water density is nearly constant. You may ignore atmospheric pressure at the surface.

(a) Derive a differential equation for ocean pressure, similar to Eq. (5.1), in two different ways: In one equation, assume that y is measured from the surface down, and in the other, assume that y is measured from the bottom up. How do the resulting equations compare? What are the values of $P(0)$ in each case? Are there advantages of one equation over the other?
(b) Solve the "surface down" differential equation.
(c) Take the density of ocean water to be 1025 [kg/m^3], and atmospheric pressure 10^5 [N/m^2]. Divers use a rule of thumb that the pressure increases by one atmosphere for each 10 m of descent. Does this agree with the "surface down" pressure function you found in part (b)?

5.2 Laws of Motion

In this section, we see how calculus is used to derive differential equations of motion of idealized particles along straight lines. The beauty of these equations is their universality; we can use them to deduce how high we can jump on the surface of the Earth, as well as on the Moon. A *particle* is an idealization in physics, an indivisible point with no extent in space, so that a single coordinate gives the position. It has a *mass*, usually denoted by the letter m. In this simple case, the position of a particle is completely described by its distance y from an arbitrarily chosen point (the origin) on the line; y is taken to be positive if y lies to one side (chosen arbitrarily) of the origin and negative if the particle is located on the other side of the origin.

Since the particle moves, y is a function of time t. The derivative $y'(t)$ of this function is the *velocity* of the particle, a quantity usually denoted by v:

$$v = y' = \frac{\mathrm{d}y}{\mathrm{d}t}. \tag{5.2}$$

Note that v is positive if the y-coordinate of the particle increases during the motion.

The velocity of a particle may, of course, change as time changes. The rate at which it changes is called the *acceleration*, and is usually denoted by the letter a:

$$a = v' = \frac{\mathrm{d}v}{\mathrm{d}t} = \frac{\mathrm{d}^2y}{\mathrm{d}t^2}. \tag{5.3}$$

Newton's laws of motion relate the acceleration of a particle to its mass and the force acting on it as follows: A force f acting along the y-axis causes a particle of mass m to accelerate at the rate a, and

$$f = ma. \tag{5.4}$$

According to Eq. (5.4), a force acting along the y-axis is negative if it imparts a negative acceleration to a particle traveling along the y-axis. There is nothing mysterious about this negative sign. It merely means that the force is acting in the negative direction along the y-axis.

If a number of different forces act on a particle, as they do in most realistic situations, the effective force acting on the particle is the *sum* of the separate forces. For example, a body might be subject to the force of gravity f_g, the force of air resistance f_a, an electric force f_e, and a magnetic force f_m; the effective force f is then

$$f = f_g + f_a + f_e + f_m, \tag{5.5}$$

and the equations governing motions under this combination of forces are

$$y' = v, \qquad mv' = f. \tag{5.6}$$

There is a tremendous advantage in being able to synthesize the force acting on a particle from various simpler forces, each of which can be analyzed separately.

Although we can write down the equations of motion (5.6) governing particles subject to any combination of forces (5.5), we cannot, except in simple situations, write down formulas for the particle's position as a function of time. In this chapter, we show how to solve the simplest version of the problem when f is a constant due to gravity. In Chap. 10, we describe numerical methods to calculate the position of a particle as a function of time to a high degree of accuracy when the force is not constant. We also investigate equations of motion for a particle subject to a combination of forces.

Gravity. We illustrate how Newton's second law (5.4) can be used to describe motion in the specific situation in which the force is that of gravity at the surface of the Earth. According to a law that is again Newton's, the magnitude f_g of the force of gravity exerted on a particle of mass m is proportional to its mass:

$$f_g = gm. \tag{5.7}$$

The constant of proportionality g and the direction of the force depend on the gravitational pull of other masses acting on the particle. At a point near the surface of the Earth, the direction of the force is toward the center of the Earth, and the value of g is approximately

$$g = 9.81\,\mathrm{m/s^2}, \tag{5.8}$$

where m is meters and s is seconds. Near the surface of the Moon, the value of g is approximately

$$g = 1.6\,\mathrm{m/s^2}. \tag{5.9}$$

For the moment, let us stay near the Earth. Denote by y the distance from the surface of the Earth of a falling body, and denote by v the vertical velocity of this falling body. Since y was chosen to increase upward, and the force of gravity is downward, the force of gravity is $-gm$. Substituting this into Newton's law (5.4), we see that

$$-gm = ma,$$

where a is the acceleration of the falling body. Divide by m:

$$-g = a.$$

Recalling the definitions of velocity and acceleration, we can write $a = y''$, so Newton's law for a falling body is

$$y'' = -g. \tag{5.10}$$

We claim that all solutions of this equation are of the form

$$y = -\frac{1}{2}gt^2 + v_0 t + b, \tag{5.11}$$

where b and v_0 are constants. To see this, rewrite Eq. (5.10) as $0 = y'' + g = (y' + gt)'$, from which we conclude that $y' + gt$ is a constant; call it v_0. Then $y' + gt - v_0 = 0$. We rewrite this equation as $\left(y + \frac{1}{2}gt^2 - v_0 t\right)' = 0$. From this, we conclude that $y + \frac{1}{2}gt^2 - v_0 t$ is a constant; call it b. This proves that all solutions of Eq. (5.10) are of the form (5.11).

The significance of the constant v_0 is this: it is the particle's initial velocity; that is, $y' = v_0$ when $t = 0$. Similarly, b is the initial position of the particle, i.e., $y = b$ when $t = 0$. So we see that initial position and initial velocity can be prescribed arbitrarily, but thereafter, the motion is uniquely determined.

How High Can You Jump? Suppose that a student, let us call her Anya, idealized as a particle, can jump k meters vertically starting from a crouching position. How much force is exerted?

Denote by h Anya's height and by m her mass. Denote by f the force that her feet in crouching position exert on the ground. As long as Anya's feet are on the ground, the total upward force acting on her body is the force exerted by her feet minus the force of gravity:

$$f - gm.$$

According to Newton's law, Anya's motion, considered as the motion of a particle, is governed by Eq. (5.4),

$$my'' = f - gm,$$

where $y(t)$ is the distance of Anya's head from the ground at time t. We divide by m and rewrite the result as

$$y'' = p - g,$$

where $p = \dfrac{f}{m}$ is force per unit mass. As we have seen, all solutions of this equation are of the form (5.11):

$$y(t) = \frac{1}{2}(p-g)t^2 + v_0 t + b.$$

Setting $t = 0$, we get $y(0) = b$, the distance of Anya's head from the ground in crouching position. Differentiating y and setting $t = 0$, we get $y'(0) = v_0$. Since at the outset, the body is at rest, we have $v_0 = 0$. So

$$y(t) = \frac{1}{2}(p-g)t^2 + b. \tag{5.12}$$

The description (5.12) is valid until the time t_1 when Anya's feet leave the ground. That occurs when $y(t_1)$, the position of her head, equals her height h, that is, when $y(t_1) = h$. Setting this into Eq. (5.12), we get $\frac{1}{2}(p-g)t_1^2 = h - b$. Denote by c the difference between the position of Anya's head in the standing and crouching positions: $c = h - b$. Solving for t_1, the time at which Anya's feet leave the ground, we get

$$t_1 = \sqrt{\frac{2c}{p-g}}. \tag{5.13}$$

What is Anya's upward velocity v_1 at time t_1? Since velocity is the time derivative of position, we have $y'(t) = (p-g)t$. Using Eq. (5.13), we get

$$v_1 = y'(t_1) = \sqrt{2c(p-g)}. \tag{5.14}$$

After her feet leave the ground, the only force acting on Anya is gravity. So for $t > t_1$, the equation governing the position of her head is

$$y'' = -g.$$

The solutions of this equation are of the form (5.11). We rewrite it with t replaced by $t - t_1$ and b replaced by h:

$$y(t) = -\frac{1}{2}g(t-t_1)^2 + v_1(t-t_1) + h. \tag{5.15}$$

Here h is the position at time t_1. The greatest height reached by the trajectory (5.15) is time t_2, when the velocity is zero. Differentiating Eq. (5.15), we get

$$y'(t_2) = -g(t_2 - t_1) + v_1 = 0,$$

which gives $t_2 - t_1 = \frac{v_1}{g}$. Setting this into formula (5.15) for $y(t)$ gives

$$y(t_2) = -\frac{1}{2}g(t_2-t_1)^2 + v_1(t_2-t_1) + h = -\frac{v_1^2}{2g} + \frac{v_1^2}{g} + h = \frac{v_1^2}{2g} + h.$$

Using formula (5.14) for v_1, we get

$$y(t_2) = c\frac{p-g}{g} + h.$$

So the height of the jump $k = y(t_2) - h$ is given by

$$k = c\left(\frac{p}{g} - 1\right). \tag{5.16}$$

Using this relation, we can express the jumping force p per unit mass as

$$p = g\left(1 + \frac{k}{c}\right). \tag{5.17}$$

Notice that in order to be able to jump at all, the force per mass exerted has to be greater than the acceleration of gravity g.

Anya is rather tall, so we shall take her height h to be 2 m and b to be 1.5 m, making $c = h - b = 0.5$ m. We take the height of the jump k to be 0.25 m. Then by Eq. (5.17), $p = 1.5g$ (Fig. 5.2).

Let us see how far such a force would carry us on the Moon. The distance k_m jumped on the moon is given by formula (5.16), where g is replaced by g_m, the acceleration due to gravity on the Moon:

$$k_m = c\left(\frac{p}{g_m} - 1\right).$$

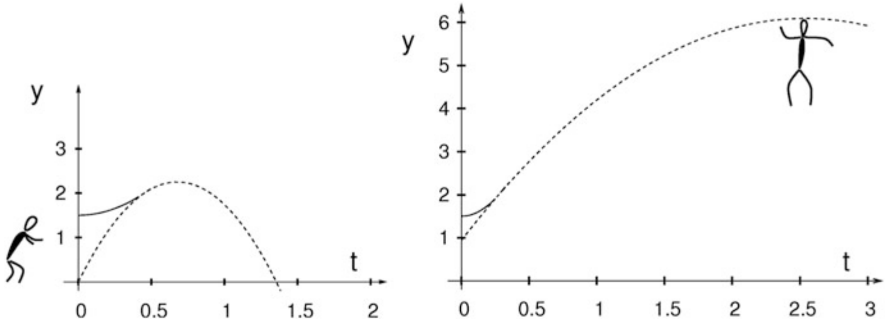

Fig. 5.2 Jumping on the surface of the Earth (*left*) and Moon (*right*) with the same force, plotting height as a function of time. The convex parabola indicates head position while the feet are pushing

Using $p = 1.5g$ and $c = 0.5$, we get

$$k_m = \frac{1}{2}\left(1.5\frac{g}{g_m} - 1\right). \tag{5.18}$$

Since $g = 9.8\ \text{m/s}^2$, and $g_m = 1.6\ \text{m/s}^2$, their ratio is 6.125. Setting these values into formula (5.18), we deduce that on the Moon, Anya can jump

$$k_m = \frac{1}{2}\big(1.5(6.125) - 1\big) = 4.1\,\text{m}.$$

Problems

5.3. Suppose that the initial position and velocity of a particle subject to Earth's gravity are $y(0) = 0$ and $y'(0) = 10$ (m/s). Calculate position and velocity at time $t = 1$ and $t = 2$.

5.4. What is the largest value of $y(t)$ during the motions described in Problem 5.3?

5.5. Write Newton's law (5.4) as a differential equation for the position $y(t)$ of a particle of mass m in the following situation: (1) $y > 0$ is the distance down to a horizontal surface at $y = 0$. (2) There are two forces on the particle; one is downward due to constant gravitational acceleration g as we have discussed, and the other is a constant upward force f_{up}.

5.6. We have said that forces may be added, and that positions can be more difficult to find than forces. In this problem, a particle of mass m at position $y(t)$ moves in six different cases according to

$$\begin{aligned} my'' &= f, \\ y(0) &= 1, \\ y'(0) &= v, \end{aligned}$$

where v is 0 or 3, and f is 5 or 7 or $5+7$. Solve for the position functions $y(t)$ in the six cases. Is it ever true that positions can be added when the forces are added?

5.7. What theorem did we use to deduce that all solutions of Eq. (5.10) are of the form (5.11)?

5.3 Newton's Method for Finding the Zeros of a Function

In the preceding two sections, we applied calculus to problems in the physical sciences. In this section, we apply calculus to solving mathematical problems.

Many mathematical problems are of the following form: we are seeking a number, called "unknown" and denoted by, say, the letter z, which has some desirable property expressed in an equation. Such an equation can be written in the form

$$f(z) = 0,$$

where f is some function. Very often, additional restrictions are placed on the number z. In many cases, these restrictions require z to lie in a certain interval. So the task of "solving an equation" is really nothing but finding a number z for which a given function f vanishes, that is, where the value of f is zero. Such a number z is called a *zero* of the function f. In some problems, we are content to find *one* zero of f in a specified interval, while in other problems, we are interested in finding *all* zeros of f in an interval.

What does it mean to "find" a zero of a function? It means to devise a procedure that gives as good an approximation as desired, of a number z where the given function f vanishes. There are two ways of measuring the goodness of an approximation z_{approx}: one is to demand that z_{approx} differ from an exact zero z by, say, less than $\frac{1}{100}$, or $\frac{1}{1000}$, or 10^{-m}. Another way of measuring the goodness of an approximation is to insist that the value of f at z_{approx} be very small, say less than $\frac{1}{100}$, $\frac{1}{1000}$, or in general, less than 10^{-m}. Of course, these notions go hand in hand: if z_{approx} is close to the true zero z, then $f(z_{\text{approx}})$ will be close to $f(z) = 0$, provided that the function f is continuous.

In this section, we describe a method for finding approximations to zeros of functions f that are not only continuous but differentiable, preferably twice differentiable. The basic step of the method is this: starting with some fairly good approximation to a zero of f, we use the derivative to produce a much better one. If the approximation is not yet good enough, we repeat the basic step as often as necessary to produce an approximation that is good enough according to either of the two criteria mentioned earlier. There are two ways of describing the basic step, geometrically and analytically. We start with the geometric description.

Denote by z_{old} the starting approximation. We assume—and this is *crucial* for the applicability of this method—that $f'(z_{\text{old}}) \neq 0$. This guarantees that the line

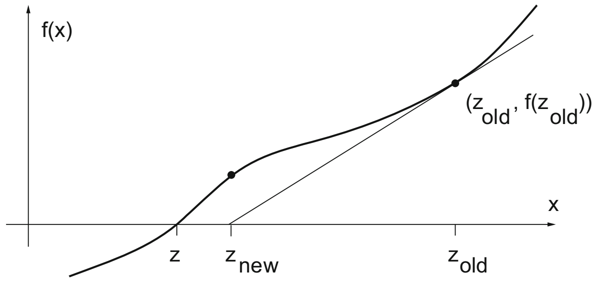

Fig. 5.3 Newton's method to approximate a root z

tangent to the graph of f at $(z_{\text{old}}, f(z_{\text{old}}))$ (see Fig. 5.3) is not parallel to the x-axis, and so intersects the x-axis at some point. This point of intersection is our new approximation z_{new}. We now calculate z_{new}. Since the slope of the tangent is $f'(z_{\text{old}})$, we have

$$f'(z_{\text{old}}) = \frac{f(z_{\text{old}}) - f(z_{\text{new}})}{z_{\text{old}} - z_{\text{new}}} = \frac{f(z_{\text{old}})}{z_{\text{old}} - z_{\text{new}}}.$$

From this relation we can determine z_{new}:

$$z_{\text{new}} = z_{\text{old}} - \frac{f(z_{\text{old}})}{f'(z_{\text{old}})}. \tag{5.19}$$

The rationale behind this procedure is this: if the graph of f were a straight line, z_{new} would be an exact zero of f. In reality, the graph of f is not a straight line, but if f is differentiable, its graph over a short enough interval is *nearly* straight, and so z_{new} can reasonably be expected to be a good approximation to the exact zero z, provided that the interval (z, z_{old}) is short enough.

We shall show at the end of this section that if z_{old} is a good enough approximation to a zero of the function f (in a sense made precise there), then z_{new} is a much better one, and that if we repeat the procedure, the resulting sequence of approximations will converge very rapidly to a zero of the function f. But first we give some examples.

The method described above has been devised, like so much else in calculus, by Newton and is therefore called *Newton's method*.

5.3a Approximation of Square Roots

Let $f(x) = x^2 - 2$. We seek a positive solution of

$$f(z) = z^2 - 2 = 0.$$

Let us see how closely we can approximate the exact solution, which is $z = \sqrt{2}$. For $f(x) = x^2 - 2$, we have $f'(x) = 2x$, so if z_{old} is an approximation to $\sqrt{2}$, Newton's recipe (5.19) yields

$$z_{\text{new}} = z_{\text{old}} - \frac{z_{\text{old}}^2 - 2}{2z_{\text{old}}} = \frac{z_{\text{old}}}{2} + \frac{1}{z_{\text{old}}}. \tag{5.20}$$

Notice that this relation is just Eq. (1.7) revisited. Let us take $z_{\text{old}} = 2$ as a first approximation to $\sqrt{2}$. Using formula (5.20), we get

$$z_{\text{new}} = 1.5.$$

We then repeat the process, with $z_{\text{new}} = 1.5$ now becoming z_{old}. Thus, we construct a sequence $z_1, z_2, \ldots$ of (hopefully) better approximations to $\sqrt{2}$ by choosing $z_1 = 2$ and setting

$$z_{n+1} = \frac{z_n}{2} + \frac{1}{z_n}.$$

The first six approximations are

$$\begin{aligned} z_1 &= 2.0 \\ z_2 &= 1.5 \\ z_3 &= 1.4166\ldots \\ z_4 &= 1.414215686\ldots \\ z_5 &= 1.414213562\ldots \\ z_6 &= 1.414213562\ldots \end{aligned}$$

Since z_5 and z_6 agree up to the first eight decimal places after the decimal point, we surmise that z_5 gives the first eight decimal places of $\sqrt{2}$ correctly. Indeed,

$$(1.41421356)^2 = 1.999999993\ldots$$

is very near and slightly less than 2, while

$$(1.41421357)^2 = 2.000000021\ldots$$

is very near, but slightly more than, 2. It follows from the intermediate value theorem that $z^2 = 2$ at some point between these numbers, i.e. that

$$1.41421356 < \sqrt{2} < 1.41421357.$$

When we previously encountered the sequence $\{z_n\}$, where it was constructed in a somewhat ad hoc fashion, the sequence was shown to converge.

5.3b Approximation of Roots of Polynomials

The beauty of Newton's method is its universality. It can be used to find the zeros of not only quadratic functions, but functions of all sorts.

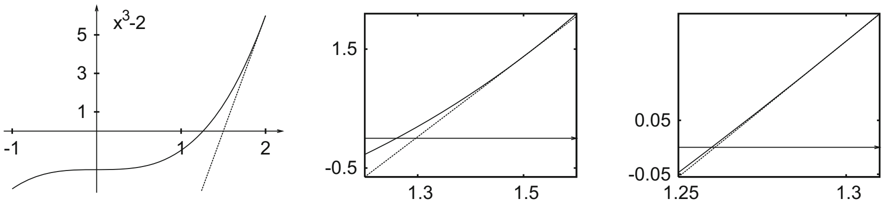

Fig. 5.4 Three steps of Newton's method to approximate a root of $f(x) = x^3 - 2$, drawn at different magnifications. The three tangent lines shown are at z_1, z_2, and z_3 of Example 5.1

Example 5.1. Let $f(x) = x^3 - 2$. We seek a sequence of approximations to a solution of

$$z^3 - 2 = 0, \tag{5.21}$$

i.e., to the number $\sqrt[3]{2}$. Since $f'(x) = 3x^2$, Newton's recipe gives the following sequence of approximations (Fig. 5.4):

$$z_{n+1} = z_n - \frac{z_n^3 - 2}{3z_n^2} = \frac{2z_n}{3} + \frac{2}{3z_n^2}. \tag{5.22}$$

Starting with $z_1 = 2$ as a first approximation, we have

$$\begin{aligned} z_1 &= 2.0 \\ z_2 &= 1.5 \\ z_3 &= 1.2962962\ldots \\ z_4 &= 1.2609322\ldots \\ z_5 &= 1.2599218\ldots \\ z_6 &= 1.2599210\ldots \end{aligned}$$

Since z_5 and z_6 agree up to the sixth digit after the decimal, we surmise that

$$\sqrt[3]{2} = 1.259921\ldots$$

Indeed, $(1.259921)^3 = 1.9999997\ldots$, while $(1.259922)^3 = 2.000004\ldots$, so that

$$1.259921 < \sqrt[3]{2} < 1.259922.$$

Next we find all zeros of

$$f(x) = x^3 - 6x^2 - 2x + 12. \tag{5.23}$$

Since f is a polynomial of degree 3, an odd number, $f(x)$ is very large and positive when x is very large and positive, and very large and negative when x is very large and negative. So by the intermediate value theorem, $f(x)$ is zero somewhere. To get a better idea where the zero, or zeros, might be located, we calculate the value of f at integers ranging from $x = -2$ to $x = 6$:

x	-2	-1	0	1	2	3	4	5	6
$f(x)$	-16	7	12	5	-8	-21	-28	-23	0

This table shows that f has a zero at $z = 6$, and since the value of f at $x = -2$ is negative, at $x = -1$ positive, f has a zero in the interval $(-2,-1)$. Similarly, f has a zero in the interval $(1,2)$.

According to a theorem of algebra, if z is a zero of a polynomial, $x - z$ is a factor. Indeed, we can write our f in the factored form

$$f(x) = (x-6)(x^2-2).$$

This form for f shows that its other zeros are $z = \pm\sqrt{2}$, and that there are no others.

Let us ignore this exact knowledge of the zeros of f (which, after all, was due to a lucky accident). Let us see how well Newton's general method works in this case. The formula, for this particular function, reads

$$z_{n+1} = z_n - \frac{z_n^3 - 6z_n^2 - 2z_n + 12}{3z_n^2 - 12z_n - 2}.$$

Starting with $z_1 = 5$ as a first approximation to the exact root 6, we get the following sequence of approximations:

$$\begin{aligned} z_1 &= 5 \\ z_2 &= 6.76\ldots \\ z_3 &= 6.147\ldots \\ z_4 &= 6.007\ldots \\ z_5 &= 6.00001\ldots \end{aligned}$$

Similar calculations show that if we start with a guess z *close enough* to one of the other two zeros $\sqrt{2}$ and $-\sqrt{2}$, we get a sequence of approximations that converges *rapidly* to the exact zeros.

5.3c The Convergence of Newton's Method

How rapid is rapid, and how close is close? In the last example, starting with an initial guess that was off by 1, we obtained, after four steps of the method, an approximation that differs from the exact zero $z = 6$ by 0.007.

Furthermore, perusal of the examples presented so far indicates that Newton's method works faster, the closer z_n gets to the zero! We shall analyze Newton's method to explain its extraordinary efficiency and also to determine its limitations.

Newton's method is based on a linear approximation. If there were no error in this approximation—i.e., if f had been a linear function—then Newton's method would have furnished in one step the exact zero of f. Therefore, in analyzing the error inherent in Newton's method, we must start with the deviation of the function f from its linear approximation. The deviation is described by the linear approximation theorem, Theorem 4.5:

$$f(x) = f(z_{\text{old}}) + f'(z_{\text{old}})(x - z_{\text{old}}) + \frac{1}{2} f''(c)(x - z_{\text{old}})^2, \tag{5.24}$$

where c is some number between z_{old} and x. Let us introduce for simplicity the abbreviations

$$f''(c) = s \qquad \text{and} \quad f'(z_{\text{old}}) = m,$$

and let z be the exact zero of f. Then Eq. (5.24) yields for $x = z$,

$$f(z) = 0 = f(z_{\text{old}}) + m(z - z_{\text{old}}) + \frac{1}{2} s(z - z_{\text{old}})^2.$$

Divide by m and use Newton's recipe (5.19) to get

$$0 = \frac{f(z_{\text{old}})}{m} + (z - z_{\text{old}}) + \frac{1}{2}\frac{s}{m}(z - z_{\text{old}})^2 = -z_{\text{new}} + z_{\text{old}} + (z - z_{\text{old}}) + \frac{1}{2}\frac{s}{m}(z - z_{\text{old}})^2.$$

We can rewrite this as

$$z_{\text{new}} - z = \frac{1}{2}\frac{s}{m}(z_{\text{old}} - z)^2. \tag{5.25}$$

We are interested in finding out under what conditions z_{new} is a better approximation to z than z_{old}. Formula (5.25) is ideal for deciding this, since it asserts that $(z_{\text{new}} - z)$ is the product of $(z - z_{\text{old}})$ and $\frac{1}{2}\frac{s}{m}(z - z_{\text{old}})$. There is an improvement if and only if that second factor is less than 1 in absolute value, i.e., if

$$\frac{1}{2}\left|\frac{s}{m}\right| |z - z_{\text{old}}| < 1. \tag{5.26}$$

Suppose now that $f'(z) \neq 0$. Since f' is continuous, f' is bounded away from zero at all points close to z, and since f'' is continuous, s does not vary too much. Also, Eq. (5.26) holds if z_{old} is close enough to z. In fact, for z_{old} close enough, we have

$$\frac{1}{2}\left|\frac{s}{m}\right| |z - z_{\text{old}}| < \frac{1}{2}. \tag{5.27}$$

If Eq. (5.27) holds, we deduce from Eq. (5.25) that

$$|z_{\text{new}} - z| \leq \frac{1}{2} |z_{\text{old}} - z|. \tag{5.28}$$

Now let $z_1, z_2, \ldots$ be a sequence of approximations generated by repeated applications of Newton's recipe. Suppose z_1 is so close to z that Eq. (5.27) holds for $z_{\text{old}} = z_1$ and for all z_{old} that are as close or closer to z than z_1. Then it follows from Eq. (5.28) that z_2 is closer to z than z_1 and, in general, that each z_{n+1} is closer to z than the previous z_n, and so Eq. (5.27) holds for all subsequent z_n. Repeated application of Eq. (5.28) shows that

$$|z_{n+1} - z| \le \frac{1}{2}|z_n - z| \le \left(\frac{1}{2}\right)^2 |z_{n-1} - z| \le \cdots \le \left(\frac{1}{2}\right)^n |z_1 - z|. \tag{5.29}$$

This proves the following theorem.

Theorem 5.1. Convergence theorem for Newton's method. *Let f be twice continuously differentiable on an open interval and z a zero of f such that*

$$f'(z) \neq 0. \tag{5.30}$$

Then repeated applications of Newton's recipe

$$z_{n+1} = z_n - \frac{f(z_n)}{f'(z_n)} \tag{5.31}$$

yields a sequence of approximations $z_1,\ z_2,\ \ldots$ that converges to z, provided that the first approximation z_1 is close enough to z.

A few comments are in order:

1. The proof that z_n tends to z is based on inequality (5.29), according to which $|z_{n+1} - z|$ is less than a constant times $\left(\frac{1}{2}\right)^n$. This is a gross overestimate; to understand the true rate at which z_n converges to z, we have to go back to relation (5.25). For z_{old} close to z, the numbers m and s differ little from $f'(z)$ and $f''(z)$ respectively, so that Eq. (5.25) asserts that $|z_{\text{new}} - z|$ is practically a constant multiple of $(z_{\text{old}} - z)^2$. Now if $|z_{\text{old}} - z|$ is small, its square is enormously small! To give an example, suppose that $\left|\frac{f''(z)}{2f'(z)}\right| \le 1$ and that $|z_{\text{old}} - z| \le 10^{-3}$. Then by Eq. (5.25), we conclude that

$$|z_{\text{new}} - z| \approx (z_{\text{old}} - z)^2 = 10^{-6}.$$

In words: If the first approximation lies within one-thousandth of an exact zero, and if $\left|\frac{f''(z)}{2f'(z)}\right| < 1$, then Newton's method takes us *in one step* to a new approximation that lies within one millionth of that exact zero.

Example 5.2. In Example 5.1, we have $\frac{1}{2}\frac{s}{m} < 1$ and

$$z_5 = \underline{1.25992}186056593,$$
$$z_6 = \underline{1.25992104989}539,$$

where the underlined digits are correct, an increase of about twice as many in one step.

2. It is necessary to start close enough to z, not only to achieve rapid convergence, but to achieve convergence at all. Figure 5.5 shows an example in which Newton's method fails to get us any closer to a zero. The points z_{old} and z_{new} are chosen so

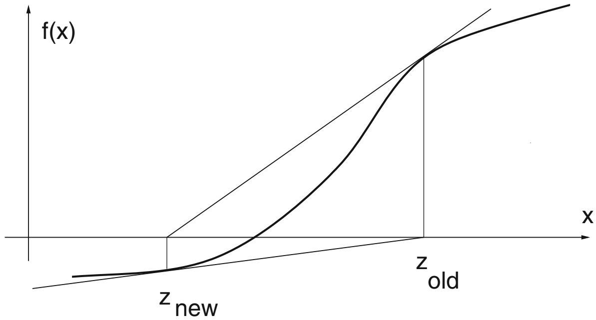

Fig. 5.5 Newton's method can fail by cycling

that the tangent to the graph of f at the point $(z_{\text{old}}, f(z_{\text{old}}))$ intersects the x-axis at z_{new}, and the tangent to the graph of f at $(z_{\text{new}}, f(z_{\text{new}}))$ intersects the x-axis at z_{old}. Newton's recipe brings us from z_{old} to z_{new}, then back to z_{old}, etc., without getting any closer to the zero between them.

3. Our analysis indicates difficulty with Newton's method when $f'(z) = 0$ at the zero z. Here is an example: the function $f(x) = (x-1)^2$ has a double zero at $z = 1$; therefore, $f'(z) = 0$. Newton's method yields the following sequence of iterates:

$$z_{n+1} = z_n - \frac{f(z_n)}{f'(z_n)} = z_n - \frac{(z_n - 1)^2}{2(z_n - 1)} = \frac{z_n + 1}{2}.$$

Subtracting 1 from both sides, we get $z_{n+1} - 1 = \dfrac{z_n - 1}{2}$. Using this relation repeatedly, we get

$$z_{n+1} - 1 = \frac{1}{2}(z_n - 1) = \frac{1}{4}(z_{n-1} - 1) = \cdots = \left(\frac{1}{2}\right)^n (z_1 - 1).$$

Thus z_n approaches the zero $z = 1$ at the rate of a constant times 2^{-n}, and not the super fast rate at which Newton's method converges when $f'(z) \neq 0$.

Problems

5.8. Use Newton's method to determine the first four digits after the decimal point of $3^{1/4}$ and of $\sqrt[3]{7}$.

Hint: Evaluating $c^{1/q}$ is equivalent to finding the zero of $z^q - c$.

5.9. Find all zeros of the following functions in the indicated domain:

(a) $f(x) = 1 + x^{1/3} - x^{1/2}$, $x \geq 0$. *Hint*: Try introducing $x = y^6$.
(b) $f(x) = x^3 - 3x^2 + 1$, $-\infty < x < \infty$.
(c) $f(x) = \dfrac{x}{x^2+1} + 1 - \sqrt{x}$, $x \geq 1$.

5.10. We claimed in the text that "if f had been a linear function, then Newton's method would have furnished in one step the exact zero of f." Show that this is true.

5.11. Show that you can find the largest value assumed by a function f in an interval $[a,b]$ by performing the following steps:

(a) Evaluate $f(x)$ at the endpoints a and b and at N equidistant points x_j between the endpoints, where N is to be specified. Set $x_0 = a,\ x_{N+1} = b$.
(b) If $f(x_j)$ is greater than both $f(x_{j-1})$ and $f(x_{j+1})$, there is a zero of $f'(x)$ in the interval $(x_{j-1},\ x_{j+1})$. Use Newton's method to find such a zero, and denote it by z_j.
(c) Determine the largest of the values $f(z_j)$ and compare to the values of f at the endpoints.

Why is it important to select N sufficiently large?

5.12. Let z be a zero of the function f, and suppose that neither f' nor f'' is zero. Show that all subsequent approximations $z_2,\ z_3,\ \ldots$ generated by Newton's method are

(a) Greater than z if $f'(z)$ and $f''(z)$ have like signs,
(b) Less than z if $f'(z)$ and $f''(z)$ have opposite signs.

Verify the truth of these assertions for examples presented in this section.

5.13. In this exercise, we ask you to investigate the following method designed to obtain a sequence of progressively better approximations to a zero of a function f:

$$z_{\text{new}} = z_{\text{old}} - af(z_{\text{old}}).$$

Here a is a number to be chosen in some suitable way. Clearly, if z_{old} happens to be the exact root z, then $z_{\text{new}} = z_{\text{old}}$. The question is this: if z_{old} is a good approximation to z, will z_{new} be a better approximation, and how much better?

(a) Use this method to construct a sequence $z_1,\ z_2,\ \ldots$ of approximations to the positive root of

$$f(z) = z^2 - 2 = 0,$$

starting with $z_1 = 2$. Observe that

- For $a = \frac{1}{2}$, $z_n \to \sqrt{2}$, but the z_n are alternately less than and greater than $\sqrt{2}$.
- For $a = \frac{1}{3}$, $z_n \to \sqrt{2}$ monotonically.
- For $a = 1$, the sequence $(z_n\}$ diverges.

(b) Prove, using the mean value theorem, that

$$z_{\text{new}} - z = (1 - am)(z_{\text{old}} - z), \tag{5.32}$$

where m is the value of f' somewhere between z and z_{old}. Prove that if a is chosen so that

$$|1-af'(z)|<1,$$

then $z_n \to z$, provided that z_1 is taken close enough to z.
Can you explain your findings under (a) in light of formula (5.32)?

(c) What would be the best choice for a, i.e., one that would yield the most rapidly converging sequence of approximations?
(d) Since the ideal value of a in part (c) is not practical, try using some reasonable estimate of $1/f'$, perhaps $1/(\text{secant slope})$ in any interval where f changes sign, for the functions in Eqs. (5.21) and (5.23).

5.4 Reflection and Refraction of Light

Mathematics can be used in science to derive laws from basic principles. In this section, we show how to use calculus to derive the laws of reflection and refraction from Fermat's principle.

Fermat's principle: *Among all possible paths PRQ connecting two points P and Q via a mirror, a ray of light travels along the one that takes the least time to traverse*.

Flat Mirrors. We wish to calculate the path of a ray of light that is reflected from a flat mirror. The path of a reflected ray going from point P to point Q is pictured in Fig. 5.6. The ray consists of two straight line segments, one, the incident ray, leading from P to the point of reflection R, the other, the reflected ray, leading from R to Q.

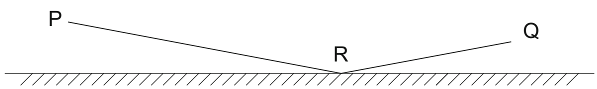

Fig. 5.6 A light ray reflects from a flat mirror

In a uniform medium like air, light travels with constant speed, so the time needed to traverse the path PRQ equals its length divided by the speed of light. So the path taken by the light ray will be the *shortest* path PRQ. We choose the mirror to be the axis of a Cartesian coordinate system. The coordinates of the point R are $(x,0)$. Denote the coordinates of the point P by (a,b), those of Q by (c,d). According to the Pythagorean theorem, the distances PR and RQ are

$$\ell_1(x)=PR=\sqrt{(x-a)^2+b^2}, \qquad \ell_2(x)=RQ=\sqrt{(c-x)^2+d^2}.$$

The total length ℓ of the path is

$$\ell(x)=\ell_1(x)+\ell_2(x)=\sqrt{(x-a)^2+b^2}+\sqrt{(c-x)^2+d^2}.$$

According to Fermat's principle, the coordinate x ought to minimize $\ell(x)$. The function ℓ is defined and differentiable for all real numbers x, not just those between a and c. For x large enough, positive or negative, $\ell(x)$ is also very large because $\ell(x) \geq |x-a| + |c-x|$. So by taking a large enough closed interval, ℓ assumes its maximum at an endpoint and its minimum at a point in the interior where $\ell'(x) = 0$.

Differentiating ℓ, we get

$$\ell'(x) = \frac{x-a}{\sqrt{(x-a)^2+b^2}} - \frac{c-x}{\sqrt{(c-x)^2+d^2}} = \frac{x-a}{\ell_1(x)} - \frac{c-x}{\ell_2(x)}.$$

Since $\ell'(x) = 0$, we have

$$\frac{x-a}{\ell_1(x)} = \frac{c-x}{\ell_2(x)}. \tag{5.33}$$

We ask you to check in Problem 5.14 that every x that satisfies Eq. (5.33) must be between a and c.

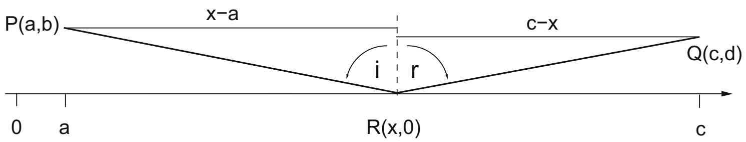

Fig. 5.7 Since the two right triangles are similar, $i = r$

The dashed line is perpendicular to the mirror at the point of reflection. The ratio $\frac{x-a}{\ell_1(x)}$ is the sine of the *angle of incidence*, defined as the angle i formed by the incident ray and the perpendicular to the mirror (Fig. 5.7). Similarly, $\frac{c-x}{\ell_2}$ is the sine of the *angle of reflection*, defined as the angle r formed by the reflected ray and the perpendicular to the mirror. Therefore,

$$\sin i = \sin r.$$

Since these angles are acute, this relation can be expressed by saying that *the angle of incidence equals the angle of reflection*. This is the celebrated *law of reflection*.

We now give a simple geometric derivation of the law of reflection from a plane mirror. See Fig. 5.8. We introduce the *mirror image* P' of the point P, so called because as will be apparent after this argument, it is the point you perceive if your eye is at Q. The mirror is the perpendicular bisector of the interval PP'. Then every point R of the mirror is equidistant from P and P',

$$PR = P'R,$$

so that

$$\ell(x) = PR + RQ = P'R + RQ.$$

The right side of this equation is the sum of two sides of the triangle $P'RQ$. According to the triangle inequality, that sum is at least as great as the third side:

$$\ell(x) \geq P'Q.$$

Equality holds in the triangle inequality only in the special case that the "triangle" is flat, that is, when P', R, and Q lie on a straight line. In that case, we see geometrically that the angle of incidence equals the angle of reflection.

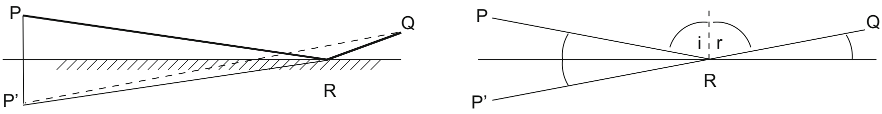

Fig. 5.8 A geometric argument to locate the path of shortest time

Curved Mirrors. We now turn to light reflection in a curved mirror. This case can no longer be handled by elementary geometry. Calculus, on the other hand, gives the answer, as we shall now demonstrate. Again according to Fermat's principle of least time, the point of reflection R can be characterized as that point on the mirror that minimizes the total length

$$\ell = PR + RQ.$$

We introduce Cartesian coordinates with R as the origin and the x-axis tangent to the mirror at R. See Fig. 5.9. In terms of these coordinates, the mirror is described by an equation of the form

$$y = f(x) \qquad \text{such that } f(0) = 0, \quad f'(0) = 0.$$

Denoting as before the coordinates of P by (a,b) and those of Q by (c,d), we see that the sum of the distances $\ell(x) = \ell_1(x) + \ell_2(x)$ from (c,d) to $(x, f(x))$ to (a,b) can be expressed as

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

We assumed that reflection occurs at $x = 0$. By Fermat's principle, the path from (a,b) to $(0,0)$ to (c,d) takes the least time and hence has the shortest total length ℓ. Therefore, $\ell'(0) = 0$. Differentiating, we get

$$\ell'(x) = \frac{(x-a) + f'(x)(f(x)-b)}{\ell_1(x)} + \frac{(x-c) + f'(x)(f(x)-d)}{\ell_2(x)}.$$

Since $f'(0) = 0$ at R, the value of $\ell'(x)$ at $x = 0$ simplifies to

$$\ell'(0) = \frac{-a}{\ell_1(0)} + \frac{-c}{\ell_2(0)} = \frac{-a}{\sqrt{a^2+b^2}} + \frac{-c}{\sqrt{c^2+d^2}} = 0.$$

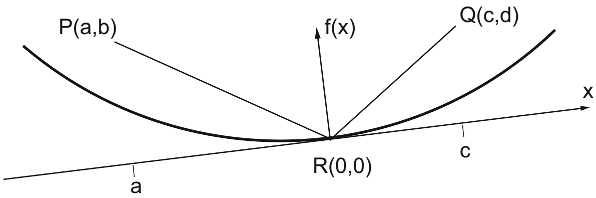

Fig. 5.9 The law of reflection also applies to a curved mirror

This equation agrees with the case of a straight mirror. We conclude as before that *the angle of incidence equals the angle of reflection*, except that in this case, these angles are defined as those formed by the rays with the line perpendicular to the tangent of the mirror at the point of reflection.

If we had not chosen the x-axis to be tangent to the mirror at the point of reflection, we would have had to use a fair amount of trigonometry to deduce the law of reflection from the relation $\ell'(x) = 0$. This shows that in calculus, as well as in analytic geometry, life can be made simpler by a wise choice of coordinate axes.

One difference between reflections from a straight mirror and from a curved mirror is that for a curved mirror, there may well be several points R that furnish a reflection. Only one of these is an absolute minimum, which points to an important modification of Fermat's principle: *Among all possible paths that connect two points P and Q via a mirror, light travels along those paths that take the least time to traverse compared to all nearby paths.* In other words, we do not care which one may be the absolute minimum; light seeks out those paths that are local, not absolute, minima. An observer located at P sees the object located at Q when he looks toward any of these points R; this can be observed in some funhouse mirrors as in Fig. 5.10.

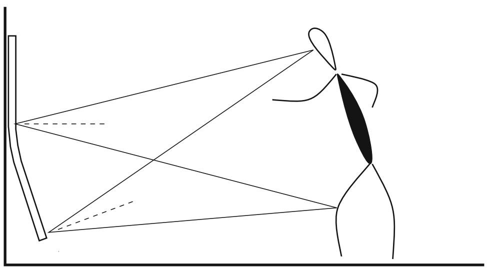

Fig. 5.10 The law of reflection. If the mirror is curved, you might see your knee in two places

Refraction of Light. We shall study the *refraction* of light rays, that is, their passage from one medium into another, in cases in which the propagation speed of light in the two media is different. A common example is the refraction at an air and water interface; see Fig. 5.11.

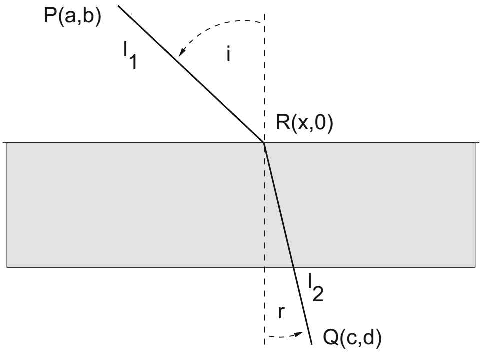

Fig. 5.11 Refraction of light traveling through air above and water below. On a straight path from P to Q, the light would spend more time in the water

We rely as before on Fermat's optical principle: among all possible paths PRQ, light travels along the one that takes the least time to traverse.

Denote by c_1 and c_2 the speed of light in air and water, respectively. The time it takes light to travel from P to R is $\frac{PR}{c_1}$, and from R to Q it takes $\frac{RQ}{c_2}$. The total time t is then

$$t = \frac{PR}{c_1} + \frac{RQ}{c_2}.$$

We introduce the line separating air and water as the x-axis. Denote as before the coordinates of P and of Q by (a,b) and (c,d) respectively, and the coordinate of R by $(x,0)$. Then

$$PR = \ell_1(x) = \sqrt{(x-a)^2+b^2}, \qquad RQ = \ell_2(x) = \sqrt{(c-x)^2+d^2},$$

and so

$$t(x) = \frac{\ell_1(x)}{c_1} + \frac{\ell_2(x)}{c_2}.$$

As before, we notice that $t(x) > \frac{|x-a|}{c_1} + \frac{|c-x|}{c_2}$. Therefore, for x large (positive or negative), $t(x)$ is large. It follows from the same argument we gave previously that $t(x)$ achieves its minimum at a point where $t'(x)$ is 0. The derivative is

$$t'(x) = \frac{\ell_1'(x)}{c_1} + \frac{\ell_2'(x)}{c_2} = \frac{x-a}{c_1\ell_1(x)} - \frac{c-x}{c_2\ell_2(x)}.$$

From the relation $t'(x) = 0$, we deduce that

$$\frac{c_2}{c_1}\frac{x-a}{\ell_1(x)} = \frac{c-x}{\ell_2(x)}.$$

As in Eq. (5.33), the ratios $\dfrac{x-a}{\ell_1(x)}$ and $\dfrac{c-x}{\ell_2(x)}$ can be interpreted geometrically (see Fig. 5.11) as the sine of the angle of incidence i and the sine of the angle of refraction r, respectively:

$$\frac{c_2}{c_1}\sin i = \sin r. \tag{5.34}$$

This is the *law of refraction*, named for the Dutch mathematician and astronomer Snell, and is often stated as follows: *when a light ray travels from a medium* 1 *into a medium* 2 *where the propagation speeds are c_1 and c_2, respectively, it is refracted so that the ratio of the sines of the angle of refraction to the angle of incidence is equal to the ratio $\frac{c_2}{c_1}$ of the propagation speeds*. The ratio $I = \frac{c_2}{c_1}$ is called the *index of refraction*. Since the sine function does not exceed 1, it follows from the law of refraction that $\sin r$ does not exceed the index of refraction I, i.e., from Eq. (5.34),

$$\sin r = I \sin i \le I. \tag{5.35}$$

The speed of light in water is less than that in air: $I = \frac{c_2}{c_1} < 1$. It follows from inequality (5.35) that r cannot exceed a critical angle r_{crit} defined by $\sin r_{\text{crit}} = I$. For water and air, the index of refraction is approximately $\frac{1}{1.33}$, so the critical angle is $\sin^{-1}\left(\frac{1}{1.33}\right) \approx 49^\circ$. This means that an underwater observer who looks in a direction that makes an angle greater than 49° with the perpendicular cannot see points above the water, since such a refracted ray would violate the law of refraction. He sees instead reflections of underwater objects. (See Fig. 5.12.) This phenomenon, well known to snorkelers, is called total internal reflection.

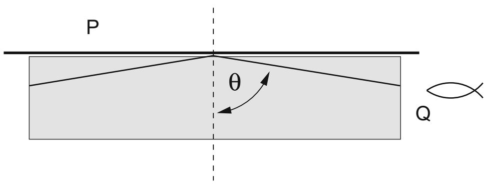

Fig. 5.12 A fish (or snorkeler) looking toward the surface with angle $\theta > 49^\circ$ can't see an object at point P in the air

Problems

5.14. We showed in the derivation of the law of reflection that when $\ell(x)$ is at a minimum, $\dfrac{x-a}{\ell_1(x)} = \dfrac{c-x}{\ell_2(x)}$. Show that if x satisfies this equation, then x must be strictly between a and c.

5.15. The derivation of the law of reflection from Fermat's principle used calculus. Find all the places in the derivation in which knowledge of calculus was used.

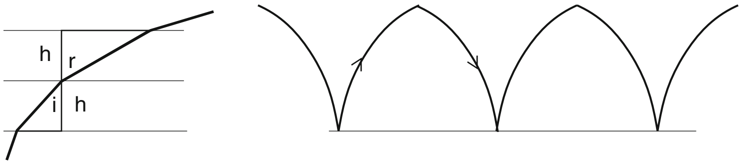

Fig. 5.13 *Left*: a light ray is refracted, in Problem 5.16. *Right*: with many thin layers, the ray may bend

5.16. On the left of Fig. 5.13, a ray has slope m_1 in the lower layer, and m_2 in the upper.

(a) If the light speeds are c_1 and c_2, use Snell's law to show that

$$\frac{c_2}{c_1}\frac{1}{\sqrt{m_1^2+1}} = \frac{1}{\sqrt{m_2^2+1}}.$$

(b) Suggest functions $c(y)$ and $y(x)$ for which $c(y)\sqrt{(y')^2+1}$ is constant, so that the graph of y produces one of the upward paths on the right of the figure. It is drawn to suggest repeated reflections between a mirror at the bottom and total internal reflection at the top.

5.5 Mathematics and Economics

Econometrics deals with measurable (and measured) quantities in economics. The basis of *econometric theory*, as of any theory, is the relations between such quantities. This section contains some brief remarks on the concepts of calculus applied to some of the functions that occur in economic theory.

Fixed and Variable Costs. Denote by $C(q)$ the total cost of producing q units of a certain commodity. Many ingredients make up the total cost; some, like raw materials needed, are *variable* and are dependent on the amount q produced. Others, like investment in a plant, are *fixed* and are independent of q. Now, $C(q)$ can be a rather complicated function of q, but it can be thought of as comprising two basic components, the *variable cost* $C_v(q)$ and the *fixed cost* $C_f(q) = F$, so that

$$C(q) = C_v(q) + F.$$

A manager who is faced with the decision whether to increase production has to know how much the additional production of h units will cost. The cost per additional unit is

$$\frac{C(q+h)-C(q)}{h}.$$

For reasonably small h, this is well approximated by $\frac{\mathrm{d}C}{\mathrm{d}q}$. This is called the *marginal cost of production*. Another function of interest is the *average cost function*

$$AC(q) = \frac{C(q)}{q} = \frac{C_v(q)+F}{q}.$$

Productivity. Let $G(L)$ be the amount of goods produced by a labor force of size L. A manager, in order to decide whether to hire more workers, wants to know how much additional goods will be produced by h additional laborers. The gain in production per laborer added is

$$\frac{G(L+h)-G(L)}{h}.$$

For reasonably small h, this is well approximated by $\frac{\mathrm{d}G}{\mathrm{d}L}$. This is called the *marginal productivity of labor*.

Demand. The consumer demand q for a certain product is a function of the price p of the product. The slope of the demand function, called the marginal demand, is the rate at which the demand changes given a change in price, $\frac{\mathrm{d}q}{\mathrm{d}p}$. The marginal demand is a measure of how responsive consumer demand is to a change in price. As you can see from the definition, the marginal demand depends on the units in which you measure quantity and price. For example, if you measure the quantity of oil in barrels rather than gallons, the marginal demand is $\frac{1}{42}$ as much, since there are 42 gallons to a barrel. Similarly with the price. Change the units from dollars to pesos, and the marginal demand will change depending on the exchange rate. Rather than specify units, economists define the *elasticity of demand*, ε, as

$$\varepsilon = \frac{p}{q}\frac{\mathrm{d}q}{\mathrm{d}p}.$$

First let us verify that ε is independent of a change in units. Suppose that the price is given as $P = kp$, and the quantity is given as $Q = cq$. Let the demand function be given by $q = f(p)$. Then $Q = cq = cf\left(\frac{P}{k}\right)$. By the chain rule,

$$\frac{\mathrm{d}Q}{\mathrm{d}P} = \frac{c}{k}f'\left(\frac{P}{k}\right) = \frac{c}{k}\frac{\mathrm{d}q}{\mathrm{d}p}.$$

It follows that ε is independent of units:

$$\varepsilon = \frac{P}{Q}\frac{\mathrm{d}Q}{\mathrm{d}P} = \frac{kp}{cq}\frac{\mathrm{d}Q}{\mathrm{d}P} = \frac{kp}{cq}\frac{c}{k}\frac{\mathrm{d}q}{\mathrm{d}p} = \frac{p}{q}\frac{\mathrm{d}q}{\mathrm{d}p}.$$

In Problem 5.18, we ask you to verify, using the chain rule, that an equivalent definition for the elasticity of demand is

$$\varepsilon = \frac{\mathrm{d}\ \log q}{\mathrm{d}\ \log p}.$$

Other Marginals. We give two further examples of derivatives in economics.

Example 5.3. Let $P(e)$ be the profit realized after the expense of e dollars. The added profit per dollar where h additional dollars are spent is

$$\frac{P(e+h)-P(e)}{h},$$

well approximated for small h by $\frac{\mathrm{d}P}{\mathrm{d}e}$, called the *marginal profit of expenditure*.

Example 5.4. Let $T(I)$ be the tax imposed on a taxable income I. The increase in tax per dollar on h additional dollars of taxable income is

$$\frac{T(I+h)-T(I)}{h}.$$

For moderate h and some values I, this is well approximated by $\frac{\mathrm{d}T}{\mathrm{d}I}$, called the *marginal rate of taxation*. However, T as prescribed by the tax code is only piecewise differentiable, in that the derivative $\frac{\mathrm{d}T}{\mathrm{d}I}$ fails to exist at certain points.

These examples illustrate two facts:

(a) The rate at which functions change is as interesting in economics, business, and finance as in every other kind of quantitative description.
(b) In economics, the rate of change of a function $y(x)$ is not called the derivative of y with respect to x but the marginal y of x.

Here are some examples of the uses to which the notion of derivative can be put in economic thinking. The managers of a firm would likely not hire additional workers when the going rate of pay exceeds the marginal productivity of labor, for the firm would lose money. Thus, declining productivity places a limitation on the size of a firm.

Actually, one can argue persuasively that efficiently run firms will stop hiring even before the situation indicated above is reached. The most efficient mode for a firm is one in which the cost of producing a unit of commodity is minimal. The cost of a unit commodity is

$$\frac{C(q)}{q}.$$

The derivative must vanish at the point where this is minimal. By the quotient rule, we get

$$\frac{q\frac{dC}{dq} - C(q)}{q^2} = 0,$$

which implies that at the point q_{max} of maximum efficiency,

$$\frac{dC}{dq}(q_{max}) = \frac{C(q_{max})}{q_{max}}. \tag{5.36}$$

In words: *At the peak of efficiency, the marginal cost of production equals the average cost of production*. The firm would still make more money by expanding production, but would not be as efficient as before, and so its relative position would be weakened.

Example 5.5. Let us see how this works in the simple case that the cost function is $C(q) = q^2 + 1$. Then the variable cost is $C_v(q) = q^2$, and the fixed cost is $C_f(q) = 1$. The average variable cost is $AC_v(q) = \frac{q^2}{q} = q$, the average fixed cost is $AC_f(q) = \frac{1}{q}$, the average cost is $AC(q) = \frac{q^2+1}{q} = q + \frac{1}{q}$, and the marginal cost is $C'(q) = 2q$. The average cost reaches its minimum when the average cost equals the marginal cost,

$$q + \frac{1}{q} = 2q,$$

which gives $q = 1$.

Equation (5.36) has the following geometric interpretation: The line connecting $(q_{max}, C(q_{max}))$ to the origin is tangent to the graph of C at $(q_{max}, C(q_{max}))$. Such a point does not exist for all functions, but does exist for functions $C(q)$ for which $\frac{C(q)}{q}$ tends to infinity as q tends to infinity. It has been remarked that the nonmonopolistic capitalistic system is possible precisely because the cost functions in capitalistic production have this property. We ask you for an example of this property in Problem 5.19.

We conclude this brief section by pointing out that to be realistic, economic theory has to take into account the enormous diversity and interdependence of economic activities. Any halfway useful models deal typically with functions of very many variables. These functions are not derived a priori from detailed theoretical considerations, but are empirically determined. For this reason, these functions are usually taken to be of very simple form, linear or quadratic; the coefficients are then determined by making a best fit to observed data. The fit that can be obtained in this way is as good as could be obtained by taking functions of more complicated forms. Therefore, there is no incentive or justification to consider more complicated functions. The mathematical theory of economics makes substantial use of statistical techniques for fitting linear and quadratic functions of many variables to recorded data and with maximizing or minimizing such functions when variables are subject to realistic restrictions.

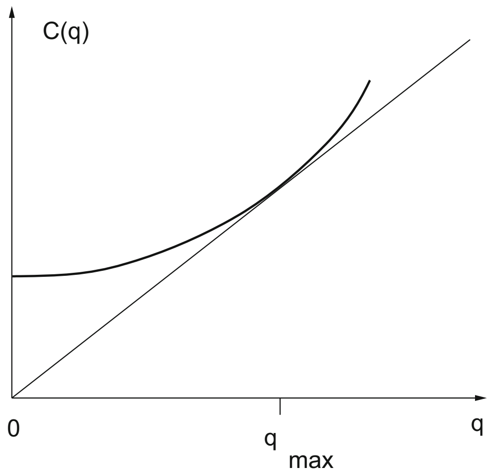

Fig. 5.14 Tangency at the peak of efficiency

Problems

5.17. Suppose you have two plants that have two different cost functions C_1 and C_2. You want to produce a total of Q units.

(a) Explain (including the meaning of q) why the total cost function C can be expressed as $C(q) = C_1(q) + C_2(Q - q)$.
(b) Show that the optimal division of production occurs when the marginal cost of production at plant 1 is *equal* to the marginal cost of production at plant 2.
(c) Suppose both costs are quadratic, $C_1(q) = aq^2$, $C_2(q) = bq^2$. Sketch a graph of C. If plant 2 is 20 % more expensive than plant 1, show that about 55 % of production should be done at plant 1.

If the costs were not equal, it would pay to switch some production from one plant to the other!

5.18. Elasticity of demand is defined as $\varepsilon = \frac{p}{q}\frac{\mathrm{d}q}{\mathrm{d}p}$. Use the chain rule to argue that an equivalent definition is $\varepsilon = \frac{\mathrm{d}\log q}{\mathrm{d}\log p}$.

5.19. Consider a cost function $C(q) = aq^k + b$, where a and b are positive. For what values of k is there a most efficient production level as illustrated in Fig. 5.14?