Contents

CHAPTER 3

Applications of the Derivative

Chapter 2 concentrated on computing derivatives. This chapter concentrates on *using* them. Our computations produced dy/dx for functions built from x^n and $\sin x$ and $\cos x$. Knowing the slope, and if necessary also the second derivative, we can answer the questions about $y = f(x)$ that this subject was created for:

1. How does y change when x changes?
2. What is the maximum value of y? Or the minimum?
3. How can you tell a maximum from a minimum, using derivatives?

The information in dy/dx is entirely *local*. It tells what is happening close to the point and nowhere else. In Chapter 2, Δx and Δy went to zero. Now we want to get them back. The local information explains the larger picture, *because Δy is approximately dy/dx times Δx.*

The problem is to connect the finite to the infinitesimal—the average slope to the instantaneous slope. Those slopes are close, and occasionally they are equal. Points of equality are assured by the Mean Value Theorem—which is the local-global connection at the center of differential calculus. But we cannot predict *where dy/dx* equals $\Delta y/\Delta x$. Therefore we now find other ways to recover a function from its derivatives—or to estimate distance from velocity and acceleration.

It may seem surprising that we learn about y from dy/dx. All our work has been going the other way! We struggled with y to squeeze out dy/dx. Now we use dy/dx to study y. That's life. Perhaps it really is life, to understand one generation from later generations.

3.1 Linear Approximation

The book started with a straight line $f = vt$. The distance is linear when the velocity is constant. As soon as v begins to change, $f = vt$ falls apart. Which velocity do we choose, when $v(t)$ is not constant? The solution is to take very short time intervals,

in which v is nearly constant:

$$f = vt \quad \text{is completely false}$$

$$\Delta f = v\Delta t \quad \text{is nearly true}$$

$$df = vdt \quad \text{is exactly true.}$$

For a brief moment the function $f(t)$ is linear—and stays near its tangent line.

In Section 2.3 we found the tangent line to $y = f(x)$. At $x = a$, the slope of the curve and the slope of the line are $f'(a)$. For points on the line, start at $y = f(a)$. Add the slope times the "increment" $x - a$:

$$Y = f(a) + f'(a)(x - a). \tag{1}$$

We write a capital Y for the line and a small y for the curve. The whole point of tangents is that they are close (*provided we don't move too far from* a):

$$y \approx Y \quad \text{or} \quad f(x) \approx f(a) + f'(a)(x - a). \tag{2}$$

That is the all-purpose **linear approximation**. Figure 3.1 shows the square root function $y = \sqrt{x}$ and its tangent line at $x = a = 100$. At the point $y = \sqrt{100} = 10$, the slope is $1/2\sqrt{x} = 1/20$. The table beside the figure compares $y(x)$ with $Y(x)$.
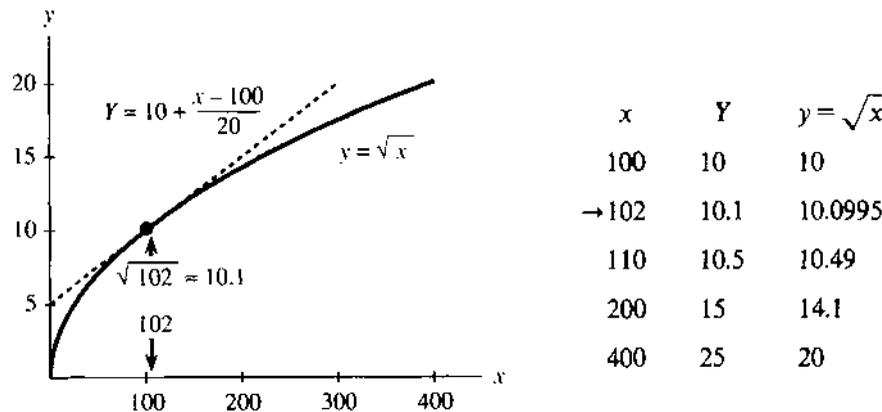

x	Y	$y = \sqrt{x}$
100	10	10
→102	10.1	10.0995
110	10.5	10.49
200	15	14.1
400	25	20

Fig. 3.1 $Y(x)$ is the linear approximation to $\sqrt{x}$ near $x = a = 100$.

The accuracy gets worse as x departs from 100. The tangent line leaves the curve. The arrow points to a good approximation at 102, and at 101 it would be even better. In this example Y is larger than y—the straight line is above the curve. The slope of the line stays constant, and the slope of the curve is decreasing. Such a curve will soon be called "concave downward," and its tangent lines are above it.

Look again at $x = 102$, where the approximation is good. In Chapter 2, when we were approaching dy/dx, we started with $\Delta y/\Delta x$:

$$\text{slope} \approx \frac{\sqrt{102} - \sqrt{100}}{102 - 100}. \tag{3}$$

Now that is turned around! The slope is $1/20$. *What we don't know is* $\sqrt{102}$:

$$\sqrt{102} \approx \sqrt{100} + (\text{slope})(102 - 100). \tag{4}$$

You work with what you have. Earlier we didn't know dy/dx, so we used (3). Now we are experts at dy/dx, and we use (4). After computing $y' = 1/20$ once and for

all, the tangent line stays near $\sqrt{x}$ for every number near 100. When that nearby number is $100 + \Delta x$, notice the error as the approximation is squared:

$$\left(\sqrt{100} + \frac{1}{20}\Delta x\right)^2 = 100 + \Delta x + \frac{1}{400}(\Delta x)^2.$$

The desired answer is $100 + \Delta x$, and we are off by the last term involving $(\Delta x)^2$. The whole point of linear approximation is to ignore every term after Δx.

There is nothing magic about $x = 100$, except that it has a nice square root. Other points and other functions allow $y \approx Y$. I would like to express this same idea in different symbols. *Instead of starting from a and going to x, we start from x and go a distance Δx to $x + \Delta x$.* The letters are different but the mathematics is identical.

3A At any point x, and for any smooth function $y = f(x)$,

$$\text{slope at } x \approx \frac{f(x + \Delta x) - f(x)}{\Delta x}. \tag{5}$$

For the approximation to $f(x + \Delta x)$, multiply both sides by Δx and add $f(x)$:

$$f(x + \Delta x) \approx f(x) + (\text{slope at } x)(\Delta x). \tag{6}$$

EXAMPLE 1 *An important linear approximation:* $(1 + x)^n \approx 1 + nx$ for x near zero.

EXAMPLE 2 *A second important approximation:* $1/(1 + x)^n \approx 1 - nx$ for x near zero.

Discussion Those are really the same. By changing n to $-n$ in Example 1, it becomes Example 2. These are linear approximations using the slopes n and $-n$ at $x = 0$:

$$(1 + x)^n \approx 1 + (\text{slope at zero}) \text{ times } (x - 0) = 1 + nx.$$

Here is the same thing with $f(x) = x^n$. The basepoint in equation (6) is now 1 or x:

$$(1 + \Delta x)^n \approx 1 + n\Delta x \qquad (x + \Delta x)^n \approx x^n + nx^{n-1}\Delta x.$$

Better than that, here are numbers. For $n = 3$ and -1 and 100, take $\Delta x = .01$:

$$(1.01)^3 \approx 1.03 \qquad \frac{1}{1.01} \approx .99 \qquad \left(1 + \frac{1}{100}\right)^{100} \approx 2$$

Actually that last number is no good. The 100th power is too much. Linear approximation gives $1 + 100\Delta x = 2$, but a calculator gives $(1.01)^{100} = 2.7\ldots$. This is close to e, the all-important number in Chapter 6. The binomial formula shows why the approximation failed:

$$(1 + \Delta x)^{100} = 1 + 100\Delta x + \frac{(100)(99)}{(2)(1)}(\Delta x)^2 + \cdots.$$

Linear approximation forgets the $(\Delta x)^2$ term. For $\Delta x = 1/100$ that error is nearly $\frac{1}{2}$. It is too big to overlook. The exact error is $\frac{1}{2}(\Delta x)^2 f''(c)$, where the Mean Value Theorem in Section 3.8 places c between x and $x + \Delta x$. You already see the point:

$$y - Y \text{ is of order } (\Delta x)^2. \text{ Linear approximation, quadratic error.}$$

DIFFERENTIALS

There is one more notation for this linear approximation. It has to be presented, because it is often used. The notation is suggestive and confusing at the same time—

it keeps the same symbols dx and dy that appear in the derivative. Earlier we took great pains to emphasize that dy/dx is not an ordinary fraction.† Until this paragraph, dx and dy have had no independent meaning. Now they become separate variables, like x and y but with their own names. These quantities dx and dy are called *differentials*.

The symbols dx and dy measure changes *along the tangent line*. They do for the approximation $Y(x)$ exactly what Δx and Δy did for $y(x)$. Thus dx and Δx both measure distance across.

Figure 3.2 has $\Delta x = dx$. But the change in y does not equal the change in Y. One is Δy (exact for the function). The other is dy (exact for the tangent line). *The differential dy is equal to ΔY, the change along the tangent line*. Where Δy is the true change, dy is its linear approximation $(dy/dx)dx$.

You often see dy written as $f'(x)dx$.
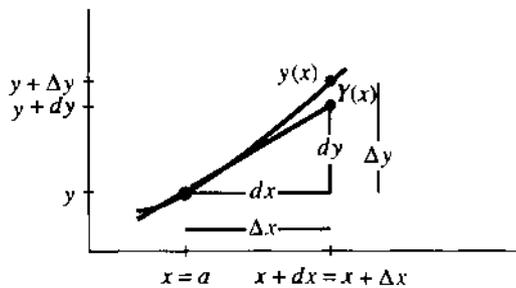
$\Delta y =$ change in y (along curve)

$dy =$ change in Y (along tangent)

Fig. 3.2   The linear approximation to Δy is

$$dy = f'(x)dx.$$

EXAMPLE 3 $y = x^2$ has $dy/dx = 2x$ so $dy = 2x\,dx$. The table has basepoint $x = 2$. The prediction dy differs from the true Δy by exactly $(\Delta x)^2 = .01$ and $.04$ and $.09$.

	dx	dy	Δx	Δy	
$y = x^2$	.1	0.4	.1	0.41	$\Delta y = (2 + \Delta x)^2 - 2^2$
$dy = 4dx$	.2	0.8	.2	0.84	$\Delta y = 4\Delta x + (\Delta x)^2$
	.3	1.2	.3	1.29	

The differential $dy = f'(x)dx$ is consistent with the derivative $dy/dx = f'(x)$. We finally have $dy = (dy/dx)dx$, but this is not as obvious as it seems! It looks like cancellation—it is really a definition. Entirely new symbols could be used, but dx and dy have two advantages: They suggest small steps and they satisfy $dy = f'(x)dx$. Here are three examples and three rules:

$$d(x^n) = nx^{n-1}dx \qquad d(f + g) = df + dg$$

$$d(\sin x) = \cos x\,dx \qquad d(cf) = c\,df$$

$$d(\tan x) = \sec^2 x\,dx \qquad d(fg) = f\,dg + g\,df$$

Science and engineering and virtually all applications of mathematics depend on linear approximation. The true function is "*linearized*," using its slope v:

Increasing the time by Δt increases the distance by $\approx v\Delta t$

Increasing the force by Δf increases the deflection by $\approx v\Delta f$

Increasing the production by Δp increases its value by $\approx v\Delta p$.

†Fraction or not, it is absolutely forbidden to cancel the d's.

The goal of dynamics or statics or economics is to predict this multiplier v—the derivative that equals the slope of the tangent line. The multiplier gives a *local prediction* of the change in the function. The exact law is nonlinear—but Ohm's law and Hooke's law and Newton's law are linear approximations.

ABSOLUTE CHANGE, RELATIVE CHANGE, PERCENTAGE CHANGE

The change Δy or Δf can be measured in three ways. So can Δx:

Absolute change	Δf	Δx
Relative change	$\dfrac{\Delta f}{f(x)}$	$\dfrac{\Delta x}{x}$
Percentage change	$\dfrac{\Delta f}{f(x)} \times 100$	$\dfrac{\Delta x}{x} \times 100$

Relative change is often more realistic than absolute change. If we know the distance to the moon within three miles, that is more impressive than knowing our own height within one inch. Absolutely, one inch is closer than three miles. Relatively, three miles is much closer:

$$\frac{3 \text{ miles}}{300,000 \text{ miles}} < \frac{1 \text{ inch}}{70 \text{ inches}} \quad \text{or} \quad .001\% < 1.4\%.$$

EXAMPLE 4 The radius of the Earth is within 80 miles of $r = 4000$ miles.
(a) Find the variation dV in the volume $V = \frac{4}{3}\pi r^3$, using linear approximation.
(b) Compute the relative variations dr/r and dV/V and $\Delta V/V$.

Solution The job of calculus is to produce the derivative. After $dV/dr = 4\pi r^2$, its work is done. The variation in volume is $dV = 4\pi(4000)^2(80)$ cubic miles. A 2% relative variation in r gives a 6% relative variation in V:

$$\frac{dr}{r} = \frac{80}{4000} = 2\% \qquad \frac{dV}{V} = \frac{4\pi(4000)^2(80)}{4\pi(4000)^3/3} = 6\%.$$

Without calculus we need the exact volume at $r = 4000 + 80$ (also at $r = 3920$):

$$\frac{\Delta V}{V} = \frac{4\pi(4080)^3/3 - 4\pi(4000)^3/3}{4\pi(4000)^3/3} \approx 6.1\%$$

One comment on $dV = 4\pi r^2 dr$. This is (area of sphere) times (change in radius). It is the volume of a thin shell around the sphere. The shell is added when the radius grows by dr. The exact $\Delta V/V$ is 3917312/640000%, but calculus just calls it 6%.

3.1 EXERCISES

Read-through questions

On the graph, a linear approximation is given by the __a__ line. At $x = a$, the equation for that line is $Y = f(a) + $ __b__ . Near $x = a = 10$, the linear approximation to $y = x^3$ is $Y = 1000 + $ __c__ . At $x = 11$ the exact value is $(11)^3 = $ __d__ . The approximation is $Y = $ __e__ . In this case $\Delta y = $ __f__ and $dy = $ __g__ . If we know $\sin x$, then to estimate $\sin(x + \Delta x)$ we add __h__ .

In terms of x and Δx, linear approximation is $f(x + \Delta x) \approx f(x) + $ __i__ . The error is of order $(\Delta x)^p$ or $(x - a)^p$ with $p = $ __j__ . The differential dy equals __k__ times the differential __l__ . Those movements are along the __m__ line, where Δy is along the __n__ .

Find the linear approximation Y to $y = f(x)$ near $x = a$:

1 $f(x) = x + x^4, a = 0$ 2 $f(x) = 1/x, a = 2$

3 $f(x) = \tan x$, $a = \pi/4$ **4** $f(x) = \sin x$, $a = \pi/2$

5 $f(x) = x \sin x$, $a = 2\pi$ **6** $f(x) = \sin^2 x$, $a = 0$

Compute 7-12 within .01 by deciding on $f(x)$, **choosing the basepoint** a, **and evaluating** $f(a) + f'(a)(x - a)$. **A calculator shows the error.**

7 $(2.001)^6$ **8** $\sin(.02)$

9 $\cos(.03)$ **10** $(15.99)^{1/4}$

11 $1/.98$ **12** $\sin(3.14)$

Calculate the numerical error in these linear approximations and compare with $\frac{1}{2}(\Delta x)^2 f''(x)$:

13 $(1.01)^3 \approx 1 + 3(.01)$ **14** $\cos(.01) \approx 1 + 0(.01)$

15 $(\sin .01)^2 \approx 0 + 0(.01)$ **16** $(1.01)^{-3} \approx 1 - 3(.01)$

17 $(1 + \frac{1}{10})^{10} \approx 2$ **18** $\sqrt{8.99} \approx 3 + \frac{1}{6}(-.01)$

Confirm the approximations 19-21 by computing $f'(0)$:

19 $\sqrt{1-x} \approx 1 - \frac{1}{2}x$

20 $1/\sqrt{1-x^2} \approx 1 + \frac{1}{2}x^2$ (use $f = 1/\sqrt{1-u}$, then put $u = x^2$)

21 $\sqrt{c^2 + x^2} \approx c + \frac{1}{2}\frac{x^2}{c}$ (use $f(u) = \sqrt{c^2 + u}$, then put $u = x^2$)

22 Write down the differentials df for $f(x) = \cos x$ and $(x + 1)/(x - 1)$ and $(x^2 + 1)^2$.

In 23-27 find the linear change dV **in the volume or** dA **in the surface area.**

23 dV if the sides of a cube change from 10 to 10.1.

24 dA if the sides of a cube change from x to $x + dx$.

25 dA if the radius of a sphere changes by dr.

26 dV if a circular cylinder with $r = 2$ changes height from 3 to 3.05 (recall $V = \pi r^2 h$).

27 dV if a cylinder of height 3 changes from $r = 2$ to $r = 1.9$. **Extra credit**: *What is* dV *if* r *and* h *both change* (dr *and* dh)?

28 In relativity the mass is $m_0/\sqrt{1 - (v/c)^2}$ at velocity v. By Problem 20 this is near $m_0 + \underline{\hspace{1cm}}$ for small v. Show that the kinetic energy $\frac{1}{2}mv^2$ and the change in mass satisfy Einstein's equation $e = (\Delta m)c^2$.

29 Enter 1.1 on your calculator. Press the square root key 5 times (slowly). What happens each time to the number after the decimal point? This is because $\sqrt{1 + x} \approx \underline{\hspace{1cm}}$.

30 In Problem 29 the numbers you see are less than 1.05, 1.025, ... The second derivative of $\sqrt{1 + x}$ is $\underline{\hspace{1cm}}$ so the linear approximation is higher than the curve.

31 Enter 0.9 on your calculator and press the square root key 4 times. Predict what will appear the fifth time and press again. You now have the $\underline{\hspace{1cm}}$ root of 0.9. How many decimals agree with $1 - \frac{1}{32}(0.1)$?

3.2 Maximum and Minimum Problems

Our goal is to learn about $f(x)$ from df/dx. We begin with two quick questions. If df/dx is positive, what does that say about f? If the slope is negative, how is that reflected in the function? Then the third question is the critical one:

How do you identify a **maximum** or **minimum**?   Normal answer: **The slope is zero**.

This may be the most important application of calculus, to reach $df/dx = 0$.

Take the easy questions first. Suppose df/dx is *positive* for every x between a and b. All tangent lines slope upward. *The function* $f(x)$ *is **increasing** as* x *goes from* a *to* b.

> **3B**   If $df/dx > 0$ then $f(x)$ is *increasing*.   If $df/dx < 0$ then $f(x)$ is *decreasing*.

To define increasing and decreasing, look at any two points $x < X$. "Increasing" requires $f(x) < f(X)$. "Decreasing" requires $f(x) > f(X)$. *A positive slope does not mean a positive function*. The function itself can be positive or negative.

EXAMPLE 1 $f(x) = x^2 - 2x$ has slope $2x - 2$. This slope is positive when $x > 1$ and negative when $x < 1$. The function increases after $x = 1$ and decreases before $x = 1$.
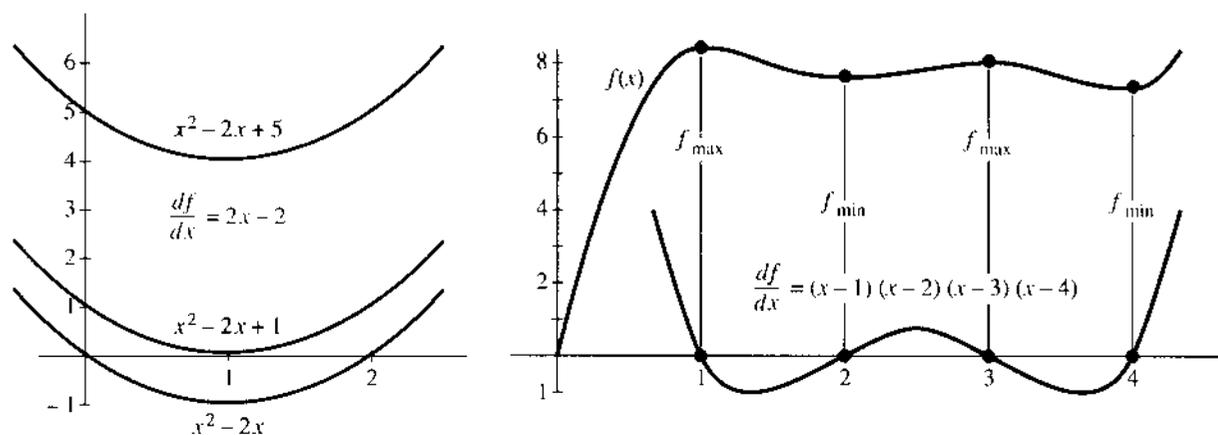
Fig. 3.3 Slopes are $- +$. Slope is $+ - + - +$ so f is up-down-up-down-up.

We say that without computing $f(x)$ at any point! The parabola in Figure 3.3 goes down to its minimum at $x = 1$ and up again.

EXAMPLE 2 $x^2 - 2x + 5$ has the same slope. Its graph is shifted up by 5, a number that disappears in df/dx. All functions with slope $2x - 2$ are parabolas $x^2 - 2x + C$, shifted up or down according to C. Some parabolas cross the x axis (those crossings are solutions to $f(x) = 0$). Other parabolas stay above the axis. The solutions to $x^2 - 2x + 5 = 0$ are complex numbers and we don't see them. The special parabola $x^2 - 2x + 1 = (x - 1)^2$ grazes the axis at $x = 1$. It has a "double zero," where $f(x) = df/dx = 0$.

EXAMPLE 3 Suppose $df/dx = (x - 1)(x - 2)(x - 3)(x - 4)$. This slope is positive beyond $x = 4$ and up to $x = 1$ ($df/dx = 24$ at $x = 0$). And df/dx is positive again between 2 and 3. At $x = 1, 2, 3, 4$, this slope is zero and $f(x)$ changes direction.

Here $f(x)$ is a fifth-degree polynomial, because $f'(x)$ is fourth-degree. The graph of f goes up-down-up-down-up. It might cross the x axis five times. *It must cross at least once* (like this one). When complex numbers are allowed, every fifth-degree polynomial has five roots.

You may feel that "*positive slope implies increasing function*" is obvious—perhaps it is. But there is still something delicate. Starting from $df/dx > 0$ at every *single* point, we have to deduce $f(X) > f(x)$ at *pairs* of points. That is a "local to global" question, to be handled by the Mean Value Theorem. It could also wait for the Fundamental Theorem of Calculus: *The difference $f(X) - f(x)$ equals the area under the graph of* df/dx. That area is positive, so $f(X)$ exceeds $f(x)$.

MAXIMA AND MINIMA

Which x makes $f(x)$ as large as possible? Where is the smallest $f(x)$? Without calculus we are reduced to computing values of $f(x)$ and comparing. With calculus, the information is in df/dx.

Suppose the maximum or minimum is at a particular point x. It is possible that the graph has a corner—and no derivative. *But if df/dx exists, it must be zero.* The tangent line is level. The parabolas in Figure 3.3 change from decreasing to increasing. The slope changes from negative to positive. At this crucial point *the slope is zero.*

> **3C** *Local Maximum or Minimum* Suppose the maximum or minimum occurs at a point x inside an interval where $f(x)$ and df/dx are defined. Then $f'(x) = 0$.

The word *"local"* allows the possibility that in other intervals, $f(x)$ goes higher or lower. *We only look near x*, and we use the definition of df/dx.

Start with $f(x + \Delta x) - f(x)$. If $f(x)$ is the maximum, this difference is negative or zero. The step Δx can be forward or backward:

$$\text{if } \Delta x > 0: \quad \frac{f(x + \Delta x) - f(x)}{\Delta x} = \frac{\text{negative}}{\text{positive}} \leqslant 0 \quad \text{and in the limit} \quad \frac{df}{dx} \leqslant 0.$$

$$\text{if } \Delta x < 0: \quad \frac{f(x + \Delta x) - f(x)}{\Delta x} = \frac{\text{negative}}{\text{negative}} \geqslant 0 \quad \text{and in the limit} \quad \frac{df}{dx} \geqslant 0.$$

Both arguments apply. Both conclusions $df/dx \leqslant 0$ and $df/dx \geqslant 0$ are correct. Thus $df/dx = 0$.

Maybe Richard Feynman said it best. He showed his friends a plastic curve that was made in a special way — *"no matter how you turn it, the tangent at the lowest point is horizontal."* They checked it out. It was true.

Surely You're Joking, Mr. Feynman! is a good book (but rough on mathematicians).

EXAMPLE 3 (continued) Look back at Figure 3.3b. The points that stand out are not the "ups" or "downs" but the "turns." Those are *stationary points*, where $df/dx = 0$. We see two maxima and two minima. None of them are absolute maxima or minima, because $f(x)$ starts at $-\infty$ and ends at $+\infty$.

EXAMPLE 4 $f(x) = 4x^3 - 3x^4$ has slope $12x^2 - 12x^3$. That derivative is zero when x^2 equals x^3, at the two points $x = 0$ and $x = 1$. To decide between minimum and maximum (local or absolute), the first step is to evaluate $f(x)$ at these *stationary points*. We find $f(0) = 0$ and $f(1) = 1$.

Now look at large x. The function goes down to $-\infty$ in both directions. (*You can mentally substitute $x = 1000$ and $x = -1000$*). For large x, $-3x^4$ dominates $4x^3$.

Conclusion $f = 1$ is an absolute maximum. $f = 0$ is not a maximum or minimum (local or absolute). We have to recognize this exceptional possibility, that a curve (or a car) can pause for an instant ($f' = 0$) and continue in the same direction. The reason is the "double zero" in $12x^2 - 12x^3$, from its double factor x^2.
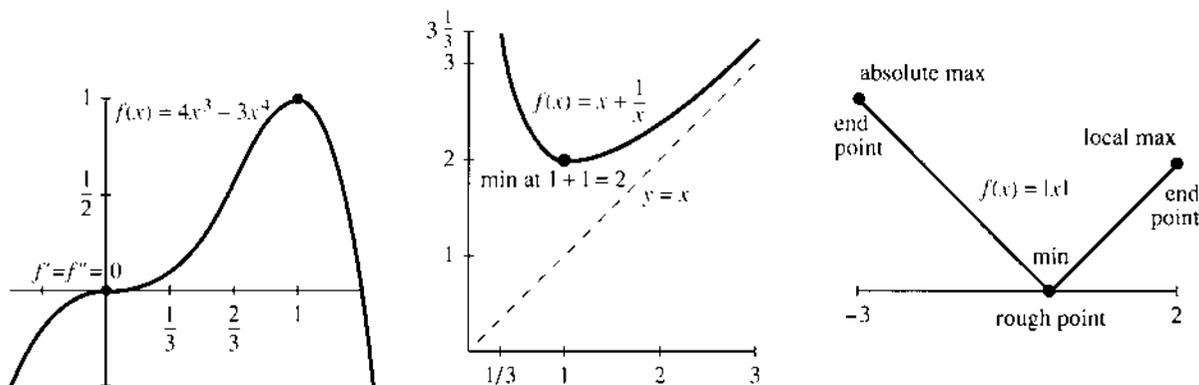

Fig. 3.4 The graphs of $4x^3 - 3x^4$ and $x - x^{-1}$. Check rough points and endpoints.

EXAMPLE 5 Define $f(x) = x + x^{-1}$ for $x > 0$. Its derivative $1 - 1/x^2$ is zero at $x = 1$. At that point $f(1) = 2$ is the minimum value. Every combination like $\frac{1}{3} + 3$ or $\frac{2}{3} + \frac{3}{2}$ is larger than $f_{\min} = 2$. Figure 3.4 shows that *the maximum of* $x + x^{-1}$ *is* $+\infty$.†

Important The maximum always occurs at a *stationary point* (where $df/dx = 0$) or a *rough point* (no derivative) or an *endpoint* of the domain. These are the three types of *critical points*. All maxima and minima occur at critical points! At every other point $df/dx > 0$ or $df/dx < 0$. Here is the procedure:

1. Solve $df/dx = 0$ to find the stationary points $f(x)$.
2. Compute $f(x)$ at every critical point—*stationary point, rough point, endpoint.*
3. Take the maximum and minimum of those critical values of $f(x)$.

EXAMPLE 6 (*Absolute value* $f(x) = |x|$) The minimum is zero at a rough point. The maximum is at an endpoint. There are no stationary points.

The derivative of $y = |x|$ is never zero. Figure 3.4 shows the maximum and minimum on the interval $[-3, 2]$. This is typical of piecewise linear functions.

Question Could the minimum be zero when the function never reaches $f(x) = 0$?
Answer *Yes*, $f(x) = 1/(1 + x)^2$ approaches but never reaches zero as $x \to \infty$.

Remark 1 $x \to \pm \infty$ and $f(x) \to \pm \infty$ are avoided when f is *continuous on a closed interval* $a \leqslant x \leqslant b$. Then $f(x)$ reaches its maximum and its minimum (*Extreme Value Theorem*). But $x \to \infty$ and $f(x) \to \infty$ are too important to rule out. You test $x \to \infty$ by considering large x. You recognize $f(x) \to \infty$ by going above every finite value.

Remark 2 Note the difference between critical *points* (specified by x) and critical *values* (specified by $f(x)$). The example $x + x^{-1}$ had the minimum *point* $x = 1$ and the minimum *value* $f(1) = 2$.

MAXIMUM AND MINIMUM IN APPLICATIONS

To find a maximum or minimum, solve $f'(x) = 0$. The slope is zero at the top and bottom of the graph. The idea is clear—and then check rough points and endpoints. But to be honest, that is not where the problem starts.

In a real application, the first step (often the hardest) is to choose the unknown and *find the function*. It is we ourselves who decide on x and $f(x)$. The equation $df/dx = 0$ comes in the middle of the problem, not at the beginning. I will start on a new example, with a question instead of a function.

EXAMPLE 7 Where should you get onto an expressway for minimum driving time, if the expressway speed is 60 mph and ordinary driving speed is 30 mph?

I know this problem well—it comes up every morning. The Mass Pike goes to MIT and I have to join it somewhere. There is an entrance near Route 128 and another entrance further in. I used to take the second one, now I take the first. Mathematics should decide which is faster—some mornings I think they are maxima.

Most models are simplified, to focus on the key idea. We will allow the expressway to be entered at *any point* x (Figure 3.5). Instead of two entrances (a discrete problem)

---

†A good word is *approach* when $f(x) \to \infty$. Infinity is not reached. But I still say "the maximum is ∞."

we have a continuous choice (a calculus problem). The trip has two parts, at speeds 30 and 60:

a distance $\sqrt{a^2 + x^2}$ up to the expressway, in $\sqrt{a^2 + x^2}/30$ hours

a distance $b - x$ on the expressway, in $(b - x)/60$ hours

Problem   Minimize $f(x) =$ total time $= \frac{1}{30}\sqrt{a^2 + x^2} + \frac{1}{60}(b - x)$.

We have the function $f(x)$. Now comes calculus. The first term uses the power rule: *The derivative of $u^{1/2}$ is $\frac{1}{2}u^{-1/2}du/dx$. Here $u = a^2 + x^2$ has $du/dx = 2x$:*

$$f'(x) = \frac{1}{30}\frac{1}{2}(a^2 + x^2)^{-1/2}(2x) - \frac{1}{60}. \tag{1}$$

To solve $f'(x) = 0$, multiply by 60 and square both sides:

$$(a^2 + x^2)^{-1/2}(2x) = 1 \quad \text{gives} \quad 2x = (a^2 + x^2)^{1/2} \quad \text{and} \quad 4x^2 = a^2 + x^2. \tag{2}$$

Thus $3x^2 = a^2$. This yields two candidates, $x = a/\sqrt{3}$ and $x = -a/\sqrt{3}$. But a negative x would mean useless driving on the expressway. In fact f' is *not zero* at $x = -a/\sqrt{3}$. That false root entered when we squared $2x$.
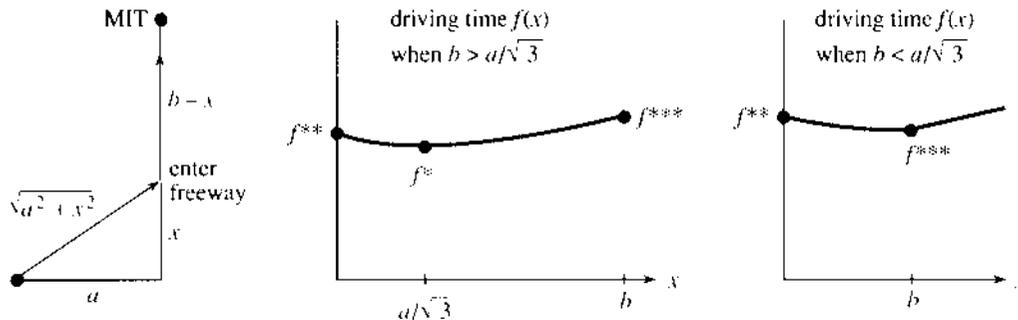

Fig. 3.5   Join the freeway at x minimize the driving time $f(x)$.

I notice something surprising. The stationary point $x = a/\sqrt{3}$ does not depend on b. The total time includes the constant $b/60$, which disappeared in df/dx. Somehow b must enter the answer. and this is a warning to go carefully. The minimum might occur at a rough point or an endpoint. Those are the other critical points of f, and our drawing may not be realistic. Certainly we expect $x \leqslant b$, or we are entering the expressway beyond MIT.

Continue with calculus. Compute the driving time $f(x)$ for an entrance at $x^* = a/\sqrt{3}$:

$$f(x) = \frac{1}{30}\sqrt{a^2 + (a^2/3)} + \frac{1}{60}\left(b - \frac{a}{\sqrt{3}}\right) = \frac{\sqrt{3}a}{60} + \frac{b}{60} = f^*.$$

The square root of $4a^2/3$ is $2a/\sqrt{3}$. We combined $2/30 - 1/60 = 3/60$ and divided by $\sqrt{3}$. **Is this stationary value f^* a minimum?** You must look also at *endpoints*:

enter at $x = 0$: travel time is $a/30 + b/60 = f^{**}$

enter at $x = b$: travel time is $\sqrt{a^2 + b^2}/30 = f^{***}$.

The comparison $f^* < f^{**}$ should be automatic. Entering at $x = 0$ was a candidate and calculus didn't choose it. The derivative is not zero at $x = 0$. It is not smart to go perpendicular to the expressway.

The second comparison has $x = b$. We drive directly to MIT at speed 30. This option has to be taken seriously. In fact it is optimal when b is small or a is large.

This choice $x = b$ can arise mathematically in two ways. If all entrances are between 0 and b, then b is an **endpoint**. If we can enter beyond MIT, then b is a **rough point**. The graph in Figure 3.5c has a corner at $x = b$, where the derivative jumps. The reason is that distance on the expressway is the *absolute value* $|b - x|$—never negative. Either way $x = b$ is a critical point. ***The optimal x is the smaller of $a/\sqrt{3}$ and b.***

if $a/\sqrt{3} \leqslant b$: stationary point wins, enter at $x = a/\sqrt{3}$, total time f^*

if $a/\sqrt{3} \geqslant b$: no stationary point, drive directly to MIT, time f^{***}

The heart of this subject is in "word problems." All the calculus is in a few lines, computing f' and solving $f'(x) = 0$. The formulation took longer. Step 1 usually does:

1. Express the quantity to be minimized or maximized as a function $f(x)$.
   The variable x has to be selected.
2. Compute $f'(x)$, solve $f'(x) = 0$, check critical points for f_{min} and f_{max}.

A picture of the problem (and the graph of $f(x)$) makes all the difference.

EXAMPLE 7 (continued)   Choose x as an *angle* instead of a distance. Figure 3.6 shows the triangle with angle x and side a. The driving distance to the expressway is $a \sec x$. The distance on the expressway is $b - a \tan x$. Dividing by the speeds 30 and 60, the driving time has a nice form:

$$f(x) = \text{total time} = \frac{a \sec x}{30} - \frac{b - a \tan x}{60}. \tag{3}$$

The derivatives of $\sec x$ and $\tan x$ go into df/dx:

$$\frac{df}{dx} = \frac{a}{30} \sec x \tan x - \frac{a}{60} \sec^2 x. \tag{4}$$

Now set $df/dx = 0$, divide by a, and multiply by $30 \cos^2 x$:

$$\sin x = \tfrac{1}{2}. \tag{5}$$

This answer is beautiful. The angle x is 30°! That optimal angle ($\pi/6$ radians) has $\sin x = \tfrac{1}{2}$. The triangle with side a and hypotenuse $a/\sqrt{3}$ is a 30 60 90 right triangle.

I don't know whether you prefer $\sqrt{a^2 + x^2}$ or trigonometry. The minimum is exactly as before—either at 30° or going directly to MIT.
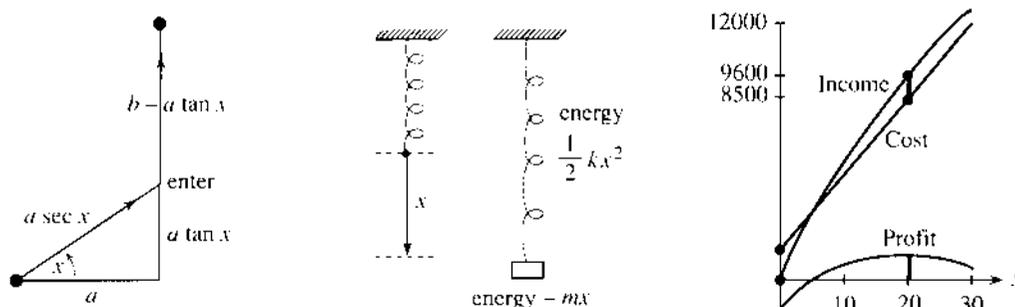

Fig. 3.6   (a) Driving at angle x. (b) Energies of spring and mass. (c) Profit = income — cost.

EXAMPLE 8 In mechanics, *nature chooses minimum energy*. A spring is pulled down by a mass, the energy is $f(x)$, and $df/dx = 0$ gives equilibrium. It is a philosophical question why so many laws of physics involve minimum energy or minimum time—which makes the mathematics easy.

The energy has two terms—for the spring and the mass. The spring energy is $\frac{1}{2}kx^2$—positive in stretching ($x > 0$ is downward) and also positive in compression ($x < 0$). The potential energy of the mass is taken as $-mx$—decreasing as the mass goes down. The balance is at the minimum of $f(x) = \frac{1}{2}kx^2 - mx$.

I apologize for giving you such a small problem, but it makes a crucial point. *When $f(x)$ is quadratic, the equilibrium equation $df/dx = 0$ is linear.*

$$df/dx = kx - m = 0.$$

Graphically, $x = m/k$ is at the bottom of the parabola. Physically, $kx = m$ is a balance of forces—the spring force against the weight. *Hooke's law* for the spring force is elastic constant k times displacement x.

EXAMPLE 9 *Derivative of cost = marginal cost* (our first management example).

The paper to print x copies of this book might cost $C = 1000 + 3x$ dollars. The derivative is $dC/dx = 3$. This is the *marginal cost* of paper for each additional book. If x increases by one book, the cost C increases by \$3. The marginal cost is like the velocity and the total cost is like the distance.

Marginal cost is in dollars per book. Total cost is in dollars. On the plus side, the income is $I(x)$ and the marginal income is dI/dx. To apply calculus, we overlook the restriction to whole numbers.

Suppose the number of books increases by dx.† The cost goes up by $(dC/dx)\, dx$. The income goes up by $(dI/dx)\, dx$. If we skip all other costs, then *profit* $P(x) =$ *income* $I(x) -$ *cost* $C(x)$. In most cases P increases to a maximum and falls back.

At the high point on the profit curve, *the marginal profit is zero*:

$$dP/dx = 0 \quad\text{or}\quad dI/dx = dC/dx. \tag{6}$$

Profit is maximized when marginal income I' equals marginal cost C'.

This basic rule of economics comes directly from calculus, and we give an example:

$C(x) = $ *cost of x advertisements* $= 900 + 400x - x^2$

 setup cost 900, print cost $400x$, volume savings x^2

$I(x) = $ *income due to x advertisements* $= 600x - 6x^2$

 sales 600 per advertisement, subtract $6x^2$ for diminishing returns

optimal decision $dC/dx = dI/dx$ or $400 - 2x = 600 - 12x$ or $x = 20$

 $profit = income - cost = 9600 - 8500 = 1100.$

The next section shows how to verify that this profit is a maximum not a minimum. The first exercises ask you to solve $df/dx = 0$. Later exercises also look for $f(x)$.

---

†Maybe dx is a differential calculus book. I apologize for that.

3.2 EXERCISES

Read-through questions

If $df/dx > 0$ in an interval then $f(x)$ is __a__. If a maximum or minimum occurs at x then $f'(x) =$ __b__. Points where $f'(x) = 0$ are called __c__ points. The function $f(x) = 3x^2 - x$ has a (minimum)(maximum) at $x =$ __d__. A stationary point that is not a maximum or minimum occurs for $f(x) =$ __e__.

Extreme values can also occur where __f__ is not defined or at the __g__ of the domain. The minima of $|x|$ and $5x$ for $-2 \leqslant x \leqslant 2$ are at $x =$ __h__ and $x =$ __i__, even though df/dx is not zero. x^* is an absolute __j__ when $f(x^*) \geqslant f(x)$ for all x. A __k__ minimum occurs when $f(x^*) \leqslant f(x)$ for all x near x^*.

The minimum of $\frac{1}{2}ax^2 - bx$ is __l__ at $x =$ __m__.

Find the stationary points and rough points and endpoints. Decide whether each point is a local or absolute minimum or maximum.

1 $f(x) = x^2 + 4x + 5, \ -\infty < x < \infty$

2 $f(x) = x^3 - 12x, \ -\infty < x < \infty$

3 $f(x) = x^2 + 3, \ -1 \leqslant x \leqslant 4$

4 $f(x) = x^2 + (2/x), \ 1 \leqslant x \leqslant 4$

5 $f(x) = (x - x^2)^2, \ -1 \leqslant x \leqslant 1$

6 $f(x) = 1/(x - x^2), \ 0 < x < 1$

7 $f(x) = 3x^4 + 8x^3 - 18x^2, \ -\infty < x < \infty$

8 $f(x) = \{x^2 - 4x \text{ for } 0 \leqslant x \leqslant 1, \ x^2 - 4 \text{ for } 1 \leqslant x \leqslant 2\}$

9 $f(x) = \sqrt{x-1} + \sqrt{9-x}, \ 1 \leqslant x \leqslant 9$

10 $f(x) = x + \sin x, \ 0 \leqslant x \leqslant 2\pi$

11 $f(x) = x^3(1 - x)^6, \ -\infty < x < \infty$

12 $f(x) = x/(1 + x), \ 0 \leqslant x \leqslant 100$

13 $f(x) = $ distance from $x \geqslant 0$ to nearest whole number

14 $f(x) = $ distance from $x \geqslant 0$ to nearest prime number

15 $f(x) = |x + 1| + |x - 1|, \ -3 \leqslant x \leqslant 2$

16 $f(x) = x\sqrt{1 - x^2}, \ 0 \leqslant x \leqslant 1$

17 $f(x) = x^{1/2} - x^{3/2}, \ 0 \leqslant x \leqslant 4$

18 $f(x) = \sin x + \cos x, \ 0 \leqslant x \leqslant 2\pi$

19 $f(x) = x + \sin x, \ 0 \leqslant x \leqslant 2\pi$

20 $f(\theta) = \cos^2\theta \sin \theta, \ -\pi \leqslant \theta \leqslant \pi$

21 $f(\theta) = 4 \sin \theta - 3 \cos \theta, \ 0 \leqslant \theta \leqslant 2\pi$

22 $f(x) = \{x^2 + 1 \text{ for } x \leqslant 1, \ x^2 - 4x + 5 \text{ for } x \geqslant 1\}$.

In applied problems, choose metric units if you prefer.

23 The airlines accept a box if length + width + height = $l + w + h \leqslant 62''$ or 158 cm. If h is fixed show that the maximum volume $(62-w-h)wh$ is $V = h(31 - \frac{1}{2}h)^2$. Choose h to maximize V. The box with greatest volume is a _____.

24 If a patient's pulse measures 70, then 80, then 120, what least squares value minimizes $(x - 70)^2 + (x - 80)^2 + (x - 120)^2$? If the patient got nervous, assign 120 a lower weight and minimize $(x - 70)^2 + (x - 80)^2 + \frac{1}{2}(x - 120)^2$.

25 At speed v, a truck uses $av + (b/v)$ gallons of fuel per mile. How many miles per gallon at speed v? Minimize the fuel consumption. Maximize the number of miles per gallon.

26 A limousine gets $(120 - 2v)/5$ miles per gallon. The chauffeur costs \$10/hour, the gas costs \$1/gallon.

   (a) Find the cost per mile at speed v.

   (b) Find the cheapest driving speed.

27 You should shoot a basketball at the angle θ requiring minimum speed. Avoid line drives and rainbows. Shooting from $(0, 0)$ with the basket at (a, b), minimize $f(\theta) = 1/(a \sin \theta \cos \theta - b \cos^2\theta)$.

   (a) If $b = 0$ you are level with the basket. Show that $\theta = 45°$ is best (Jabbar sky hook).

   (b) Reduce $df/d\theta = 0$ to $\tan 2\theta = -a/b$. Solve when $a = b$.

   (c) Estimate the best angle for a free throw.

The same angle allows the largest margin of error (*Sports Science* by Peter Brancazio). Section 12.2 gives the flight path.

28 On the longest and shortest days, in June and December, why does the length of day change the least?

29 Find the shortest **Y** connecting P, Q, and B in the figure. Originally B was a birdfeeder. The length of **Y** is $L(x) = (b - x) + 2\sqrt{a^2 + x^2}$.

   (a) Choose x to minimize L (not allowing $x > b$).

   (b) Show that the center of the **Y** has 120° angles.

   (c) The best **Y** becomes a **V** when $a/b =$ _____.
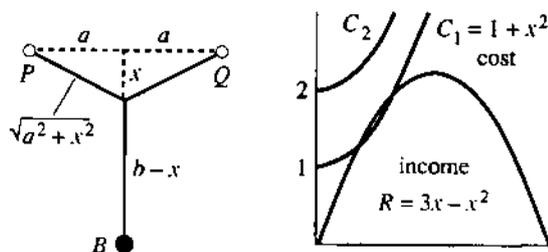
30 If the distance function is $f(t) = (1 + 3t)/(1 + 3t^2)$, when does the forward motion end? How far have you traveled? Extra credit: Graph $f(t)$ and df/dt.

In 31–34, we make and sell x pizzas. The income is $R(x) = ax + bx^2$ and the cost is $C(x) = c + dx + ex^2$.

31 The profit is $\Pi(x) =$ _____. The average profit per pizza is $=$ _____. The marginal profit per additional pizza is $d\Pi/dx =$ _____. We should maximize the (profit) (average profit) (marginal profit).

32 We receive $R(x) = ax + bx^2$ when the price per pizza is $p(x) =$ _____. In reverse: When the price is p we sell $x =$ _____ pizzas (a function of p). We expect $b < 0$ because _____.

33 Find x to maximize the profit $\Pi(x)$. At that x the marginal profit is $d\Pi/dx =$ _____.

34 Figure B shows $R(x) = 3x - x^2$ and $C_1(x) = 1 + x^2$ and $C_2(x) = 2 + x^2$. With cost C_1, which sales x makes a profit? Which x makes the most profit? With higher fixed cost in C_2, the best plan is _____.

The cookie box and popcorn box were created by Kay Dundas from a $12'' \times 12''$ square. A box with no top is a calculus classic.
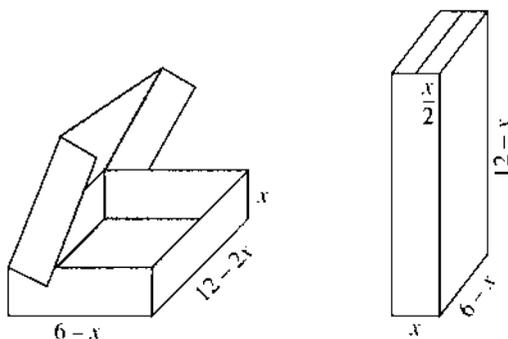
35 Choose x to find the maximum volume of the cookie box.

36 Choose x to maximize the volume of the popcorn box.

37 A high-class chocolate box adds a strip of width x down across the front of the cookie box. Find the new volume $V(x)$ and the x that maximizes it. Extra credit: Show that V_{max} is reduced by more than 20%.

38 For a box with no top, cut four squares of side x from the corners of the $12''$ square. Fold up the sides so the height is x. Maximize the volume.

Geometry provides many problems, more applied than they seem.

39 A wire four feet long is cut in two pieces. One piece forms a circle of radius r, the other forms a square of side x. Choose r to minimize the sum of their areas. Then choose r to maximize.

40 A fixed wall makes one side of a rectangle. We have 200 feet of fence for the other three sides. Maximize the area A in 4 steps:
 1 Draw a picture of the situation.
 2 Select one unknown quantity as x (but not A!).
 3 Find all other quantities in terms of x.
 4 Solve $dA/dx = 0$ and check endpoints.

41 With no fixed wall, the sides of the rectangle satisfy $2x + 2y = 200$. Maximize the area. Compare with the area of a circle using the same fencing.

42 Add 200 meters of fence to an existing straight 100-meter fence, to make a rectangle of maximum area (invented by Professor Klee).

43 How large a rectangle fits into the triangle with sides $x = 0$, $y = 0$, and $x/4 + y/6 = 1$? Find the point on this third side that maximizes the area xy.

44 The largest rectangle in Problem 43 may not sit straight up. Put one side along $x/4 + y/6 = 1$ and maximize the area.

45 The distance around the rectangle in Problem 43 is $P = 2x + 2y$. Substitute for y to find $P(x)$. Which rectangle has $P_{max} = 12$?

46 Find the right circular cylinder of largest volume that fits in a sphere of radius 1.

47 How large a cylinder fits in a cone that has base radius R and height H? For the cylinder, choose r and h on the sloping surface $r/R + h/H = 1$ to maximize the volume $V = \pi r^2 h$.

48 The cylinder in Problem 47 has side area $A = 2\pi rh$. Maximize A instead of V.

49 Including top and bottom, the cylinder has area

$$A = 2\pi rh + 2\pi r^2 = 2\pi rH(1 - (r/R)) + 2\pi r^2.$$

Maximize A when $H > R$. Maximize A when $R > H$.

*50 A wall 8 feet high is 1 foot from a house. Find the length L of the shortest ladder over the wall to the house. Draw a triangle with height y, base $1 + x$, and hypotenuse L.

51 Find the closed cylinder of volume $V = \pi r^2 h = 16\pi$ that has the least surface area.

52 Draw a kite that has a triangle with sides 1, 1, $2x$ next to a triangle with sides $2x$, 2, 2. Find the area A and the x that maximizes it. *Hint*: In dA/dx simplify $\sqrt{1 - x^2} - x^2/\sqrt{1 - x^2}$ to $(1 - 2x^2)/\sqrt{1 - x^2}$.

In 53–56, x and y are nonnegative numbers with $x + y = 10$. Maximize and minimize:

53 xy **54** $x^2 + y^2$ **55** $y - (1/x)$ **56** $\sin x \sin y$

57 Find the total distance $f(x)$ from A to X to C. Show that $df/dx = 0$ leads to $\sin a = \sin c$. Light reflects at an equal angle to minimize travel time.

reflection   refraction
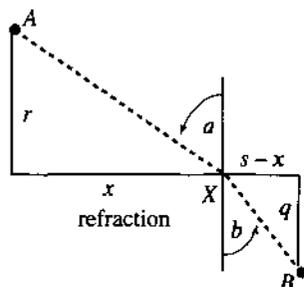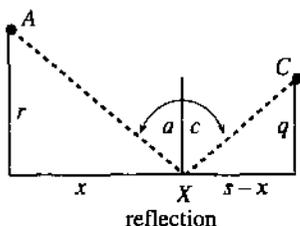
58 Fermat's principle says that light travels from A to B on the quickest path. Its velocity above the x axis is v and below the x axis is w.

(a) Find the time $T(x)$ from A to X to B. On AX, time = distance/velocity = $\sqrt{r^2 + x^2}/v$.

(b) Find the equation for the minimizing x.

(c) Deduce *Snell's law* $(\sin a)/v = (\sin b)/w$.

"Closest point problems" are models for many applications.

59 Where is the parabola $y = x^2$ closest to $x = 0$, $y = 2$?

60 Where is the line $y = 5 - 2x$ closest to $(0, 0)$?

61 What point on $y = -x^2$ is closest to what point on $y = 5 - 2x$? At the nearest points, the graphs have the same slope. Sketch the graphs.

62 Where is $y = x^2$ closest to $(0, \frac{1}{2})$? Minimizing $x^2 + (y - \frac{1}{2})^2 = y + (y - \frac{1}{2})^2$ gives $y < 0$. What went wrong?

63 Draw the line $y = mx$ passing near $(2, 3)$, $(1, 1)$, and $(-1, 1)$. For a least squares fit, minimize

$$(3 - 2m)^2 + (1 - m)^2 + (1 + m)^2.$$

64 A triangle has corners $(-1, 1)$, (x, x^2), and $(3, 9)$ on the parabola $y = x^2$. Find its maximum area for x between -1 and 3. *Hint*: The distance from (X, Y) to the line $y = mx + b$ is $|Y - mX - b|/\sqrt{1 + m^2}$.

65 Submarines are located at $(2, 0)$ and $(1, 1)$. Choose the slope m so the line $y = mx$ goes between the submarines but stays as far as possible from the nearest one.

Problems 66–72 go back to the theory.

66 To find where the graph of $y(x)$ has greatest slope, solve _____. For $y = 1/(1 + x^2)$ this point is _____.

67 When the difference between $f(x)$ and $g(x)$ is smallest, their slopes are _____. Show this point on the graphs of $f = 2 + x^2$ and $g = 2x - x^2$.

68 Suppose y is fixed. The minimum of $x^2 + xy - y^2$ (a function of x) is $m(y) =$ _____. Find the maximum of $m(y)$.

Now x is fixed. The maximum of $x^2 + xy - y^2$ (a function of y) is $M(x) =$ _____. Find the minimum of $M(x)$.

69 For each m the minimum value of $f(x) - mx$ occurs at $x = m$. What is $f(x)$?

70 $y = x + 2x^2 \sin(1/x)$ has slope 1 at $x = 0$. But show that y is not increasing on an *interval* around $x = 0$, by finding points where $dy/dx = 1 - 2\cos(1/x) + 4x\sin(1/x)$ is negative.

71 *True or false*, with a reason: Between two local minima of a smooth function $f(x)$ there is a local maximum.

72 Create a function $y(x)$ that has its maximum at a rough point and its minimum at an endpoint.

73 Draw a circular pool with a lifeguard on one side and a drowner on the opposite side. The lifeguard swims with velocity v and runs around the rest of the pool with velocity $w = 10v$. If the swim direction is at angle θ with the direct line, choose θ to minimize and maximize the arrival time.

3.3 Second Derivatives: Bending and Acceleration

When $f'(x)$ is positive, $f(x)$ is increasing. When dy/dx is negative, $y(x)$ is decreasing. That is clear, but what about the *second* derivative? From looking at the curve, can you decide the sign of $f''(x)$ or d^2y/dx^2? The answer is *yes* and the key is in the *bending*.

A straight line doesn't bend. The slope of $y = mx + b$ is m (a constant). The second derivative is zero. We have to go to curves, to see a changing slope. Changes in the derivative show up in $f''(x)$:

$f = x^2$ has $f' = 2x$ and $f'' = 2$ (this parabola bends *up*)

$y = \sin x$ has $dy/dx = \cos x$ and $d^2y/dx^2 = -\sin x$ (the sine bends *down*)

The slope $2x$ gets larger *even when the parabola is falling*. The sign of f or f' is not revealed by f''. The second derivative tells about *change in slope*.

A function with $f''(x) > 0$ is *concave up*. It bends upward as the slope increases. It is also called *convex*. A function with decreasing slope—this means $f''(x) < 0$—is *concave down*. Note how $\cos x$ and $1 + \cos x$ and even $1 + \frac{1}{2}x + \cos x$ change from concave down to concave up at $x = \pi/2$. At that point $f'' = -\cos x$ changes from negative to positive. The extra $1 + \frac{1}{2}x$ tilts the graph but the bending is the same.
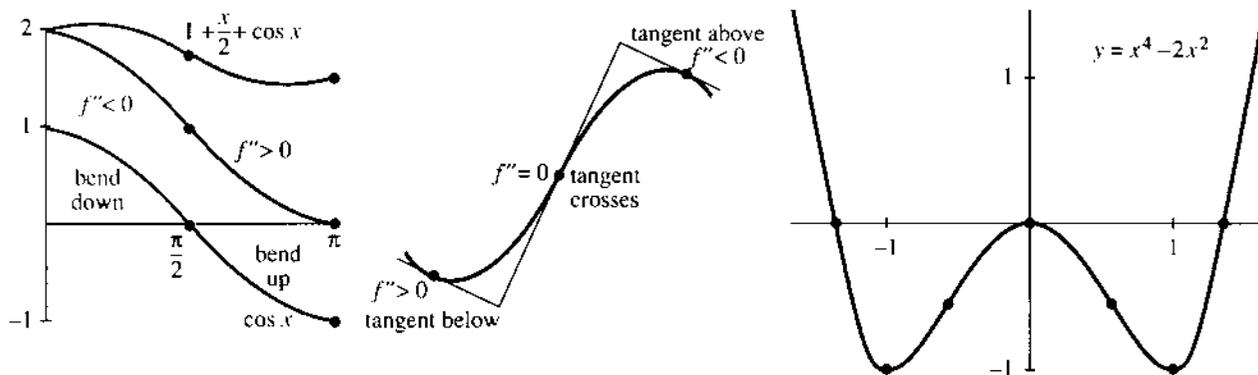

Fig. 3.7 Increasing slope = concave up $(f'' > 0)$. Concave down is $f'' < 0$. Inflection point $f'' = 0$.

Here is another way to see the sign of f''. *Watch the tangent lines*. When the curve is concave up, the tangent stays below it. A linear approximation is too low. This section computes a *quadratic* approximation—which includes the term with $f'' > 0$. When the curve bends down $(f'' < 0)$, the opposite happens—the tangent lines are above the curve. The linear approximation is too high, and f'' lowers it.

In physical motion, $f''(t)$ is the *acceleration*—in units of distance/(time)^2. Acceleration is rate of change of velocity. The oscillation $\sin 2t$ has $v = 2\cos 2t$ (maximum speed 2) and $a = -4\sin 2t$ (maximum acceleration 4).

An increasing population means $f' > 0$. *An increasing growth rate means $f'' > 0$*. Those are different. The rate can slow down while the growth continues.

MAXIMUM VS. MINIMUM

Remember that $f'(x) = 0$ locates a stationary point. That may be a *minimum* or a *maximum*. *The second derivative decides!* Instead of computing $f(x)$ at many points, we compute $f''(x)$ at one point—the stationary point. It is a minimum if $f''(x) > 0$.

3D When $f'(x) = 0$ and $f''(x) > 0$, there is a *local minimum* at x.
   When $f'(x) = 0$ and $f''(x) < 0$, there is a *local maximum* at x.

To the left of a minimum, the curve is falling. After the minimum, the curve rises. The slope has changed from negative to positive. The graph bends upward and $f''(x) > 0$.

At a maximum the slope drops from positive to negative. In the exceptional case, when $f'(x) = 0$ and also $f''(x) = 0$, anything can happen. An example is x^3, which pauses at $x = 0$ and continues up (its slope is $3x^2 \geqslant 0$). However x^4 pauses and goes down (with a very flat graph).

We emphasize that the information from $f'(x)$ and $f''(x)$ is only "*local*." To be certain of an *absolute* minimum or maximum, we need information over the whole domain.

EXAMPLE 1 $f(x) = x^3 - x^2$ has $f'(x) = 3x^2 - 2x$ and $f''(x) = 6x - 2$.

To find the maximum and/or minimum, solve $3x^2 - 2x = 0$. The stationary points are $x = 0$ and $x = \frac{2}{3}$. At those points we need the second derivative. It is $f''(0) = -2$ (local maximum) and $f''(\frac{2}{3}) = +2$ (local minimum).

Between the maximum and minimum is the *inflection point. That is where* $f''(x) = 0$. The curve changes from concave down to concave up. This example has $f''(x) = 6x - 2$, so the inflection point is at $x = \frac{1}{3}$.

INFLECTION POINTS

In mathematics it is a special event when a function passes through zero. When the function is f, its graph crosses the axis. When the function is f', the tangent line is horizontal. When f'' goes through zero, we have an *inflection point.*

The direction of bending changes at an inflection point. Your eye picks that out in a graph. For an instant the graph is straight (straight lines have $f'' = 0$). It is easy to see crossing points and stationary points and inflection points. Very few people can recognize where $f''' = 0$ or $f'''' = 0$. I am not sure if those points have names.

There is a genuine maximum or minimum when $f'(x)$ changes sign. Similarly, there is a genuine inflection point when $f''(x)$ changes sign. *The graph is concave down on one side of an inflection point and concave up on the other side.*† The tangents are above the curve on one side and below it on the other side. At an inflection point, *the tangent line crosses the curve* (Figure 3.7b).

Notice that a parabola $y = ax^2 + bx + c$ has no inflection points: y'' is constant. A cubic curve has one inflection point, because f'' is linear. A fourth-degree curve might or might not have inflection points—the quadratic $f''(x)$ might or might not cross the axis.

EXAMPLE 2 $x^4 - 2x^2$ is W-shaped, $4x^3 - 4x$ has two bumps, $12x^2 - 4$ is U-shaped. The table shows the signs at the important values of x:

x	$-\sqrt{2}$	-1	$-1/\sqrt{3}$	0	$1/\sqrt{3}$	1	$\sqrt{2}$
$f(x)$	0	$-$	$-$	$0,0$	$-$	$-$	0
$f'(x)$		0	$+$	0	$-$	0	
$f''(x)$			0	$-$	0		

Between zeros of $f(x)$ come zeros of $f'(x)$ (stationary points). Between zeros of $f'(x)$ come zeros of $f''(x)$ (inflection points). In this example $f(x)$ has a double zero at the origin, so a single zero of f' is caught there. It is a local maximum, since $f''(0) < 0$.

Inflection points are important—not just for mathematics. We know the world population will keep rising. We don't know if the *rate* of growth will slow down. Remember: *The rate of growth stops growing at the inflection point*. Here is the 1990 report of the UN Population Fund.

The next ten years will decide whether the world population trebles or merely doubles before it finally stops growing. This may decide the future of the earth as a habitation for humans. The population, now 5.3 billion, is increasing by a quarter of a million every day. Between 90 and 100 million people will be added every year

---

†That rules out $f(x) = x^4$, which has $f'' = 12x^2 > 0$ on both sides of zero. Its tangent line is the x axis. The line stays below the graph—so no inflection point.

during the 1990s; a billion people—a whole China—over the decade. The fastest growth will come in the poorest countries.

A few years ago it seemed as if the rate of population growth was slowing† everywhere except in Africa and parts of South Asia. The world's population seemed set to stabilize around 10.2 billion towards the end of the next century.

Today, the situation looks less promising. The world has overshot the marker points of the 1984 "most likely" medium projection. It is now on course for an eventual total that will be closer to 11 billion than to 10 billion.

If fertility reductions continue to be slower than projected, the mark could be missed again. In that case the world could be headed towards a total of up to 14 billion people.

Starting with a census, the UN follows each age group in each country. They estimate the death rate and fertility rate—the medium estimates are published. This report is saying that we are not on track with the estimate.

Section 6.5 will come back to population, with an equation that predicts 10 billion. It assumes we are now at the inflection point. But China's second census just started on July 1, 1990. When it's finished we will know if the inflection point is still ahead.

You now understand the meaning of $f''(x)$. Its sign gives the direction of bending—the change in the slope. *The rest of this section computes **how much** the curve bends—*using the *size* of f'' and not just its sign. We find quadratic approximations based on $f''(x)$. In some courses they are optional—the main points are highlighted.

CENTERED DIFFERENCES AND SECOND DIFFERENCES

Calculus begins with average velocities, computed on either side of x:

$$\frac{f(x + \Delta x) - f(x)}{\Delta x} \quad \text{and} \quad \frac{f(x) - f(x - \Delta x)}{\Delta x} \quad \text{are close to} \quad f'(x). \tag{1}$$

We never mentioned it, but a better approximation to $f'(x)$ comes from *averaging those two averages*. This produces a *centered difference*, which is based on $x + \Delta x$ and $x - \Delta x$. It divides by $2\Delta x$:

$$f'(x) \approx \frac{1}{2}\left[\frac{f(x + \Delta x) - f(x)}{\Delta x} + \frac{f(x) - f(x - \Delta x)}{\Delta x}\right] = \frac{f(x + \Delta x) - f(x - \Delta x)}{2\Delta x}. \tag{2}$$

We claim this is better. The test is to try it on powers of x.

For $f(x) = x$ these ratios all give $f' = 1$ (exactly). For $f(x) = x^2$, only the centered difference correctly gives $f' = 2x$. The one-sided ratio gave $2x + \Delta x$ (in Chapter 1 it was $2t + h$). It is only "first-order accurate." But centering leaves no error. We are averaging $2x + \Delta x$ with $2x - \Delta x$. Thus the centered difference is "second-order accurate."

I ask now: *What ratio converges to the second derivative?* One answer is to take differences of the first derivative. Certainly $\Delta f'/\Delta x$ approaches f''. But we want a ratio involving f itself. A natural idea is to take *differences of differences*, which brings us to "*second differences*":

$$\frac{\dfrac{f(x + \Delta x) - f(x)}{\Delta x} - \dfrac{f(x) - f(x - \Delta x)}{\Delta x}}{\Delta x} = \frac{f(x + \Delta x) - 2f(x) + f(x - \Delta x)}{(\Delta x)^2} \to \frac{d^2 f}{dx^2}. \tag{3}$$

---

†The United Nations watches the second derivative!

On the top, the difference of the difference is $\Delta(\Delta f) = \Delta^2 f$. It corresponds to $d^2 f$. On the bottom, $(\Delta x)^2$ corresponds to dx^2. This explains the way we place the 2's in $d^2 f/dx^2$. To say it differently: dx is squared, df is not squared—as in distance/(time)^2.

Note that $(\Delta x)^2$ becomes much smaller than Δx. If we divide Δf by $(\Delta x)^2$, the ratio blows up. It is the extra cancellation in the second difference $\Delta^2 f$ that allows the limit to exist. That limit is $f''(x)$.

Application   The great majority of differential equations can't be solved exactly. A typical case is $f''(x) = -\sin f(x)$ (the pendulum equation). To compute a solution, I would replace $f''(x)$ by the second difference in equation (3). Approximations at points spaced by Δx are a very large part of scientific computing.

To test the accuracy of these differences, here is an experiment on $f(x) = \sin x + \cos x$. The table shows the errors at $x = 0$ from formulas (1), (2), (3):

step length Δx	one-sided errors	centered errors	second difference errors
1/4	.1347	.0104	−.0052
1/8	.0650	.0026	−.0013
1/16	.0319	.0007	−.0003
1/32	.0158	.0002	−.0001

The one-sided errors are cut in half when Δx is cut in half. The other columns decrease like $(\Delta x)^2$. Each reduction divides those errors by 4. ***The errors from one-sided differences are $O(\Delta x)$ and the errors from centered differences are $O(\Delta x)^2$.***

The "big O" notation   When the errors are of order Δx, we write $E = O(\Delta x)$. This means that $E \leqslant C\Delta x$ for some constant C. We don't compute C—in fact we don't want to deal with it. The statement "one-sided errors are Oh of delta x" captures what is important. The main point of the other columns is $E = O(\Delta x)^2$.

LINEAR APPROXIMATION VS. QUADRATIC APPROXIMATION

The second derivative gives a tremendous improvement over linear approximation $f(a) + f'(a)(x - a)$. A tangent line starts out close to the curve, but *the line has no way to bend.* After a while it overshoots or undershoots the true function (see Figure 3.8). That is especially clear for the model $f(x) = x^2$, when the tangent is the x axis and the parabola curves upward.

You can almost guess the term with bending. *It should involve f'', and also $(\Delta x)^2$.* It might be exactly $f''(x)$ times $(\Delta x)^2$ but it is not. The model function x^2 has $f'' = 2$. There must be a factor $\frac{1}{2}$ to cancel that 2:

> **3E**   The *quadratic approximation* to a smooth function $f(x)$ near $x = a$ is
> $$f(x) \approx f(a) + f'(a)(x - a) + \tfrac{1}{2}f''(a)(x - a)^2. \qquad (4)$$

At the basepoint this is $f(a) = f(a)$. The derivatives also agree at $x = a$. Furthermore *the second derivatives agree.* On both sides of (4), the second derivative at $x = a$ is $f''(a)$.

The quadratic approximation bends with the function. It is not the absolutely final word, because there is a cubic term $\frac{1}{6}f'''(a)(x - a)^3$ and a fourth-degree term $\frac{1}{24}f''''(a)(x - a)^4$ and so on. The whole infinite sum is a "Taylor series." Equation (4) carries that series through the quadratic term—which for practical purposes gives a terrific approximation. You will see that in numerical experiments.

Two things to mention. First, equation (4) shows why $f'' > 0$ brings the curve above the tangent line. The linear part gives the line, while the quadratic part is positive and bends upward. Second, equation (4) comes from (2) and (3). Where one-sided differences give $f(x + \Delta x) \approx f(x) + f'(x)\Delta x$, centered differences give the quadratic:

$$\text{from (2):}\quad f(x + \Delta x) \approx f(x - \Delta x) + 2f'(x)\,\Delta x$$

$$\text{from (3):}\quad f(x + \Delta x) \approx 2f(x) - f(x - \Delta x) + f''(x)(\Delta x)^2.$$

Add and divide by 2. The result is $f(x + \Delta x) \approx f(x) + f'(x)\Delta x + \frac{1}{2}f''(x)(\Delta x)^2$. This is correct through $(\Delta x)^2$ and misses by $(\Delta x)^3$, as examples show:

EXAMPLE 3 $(x + \Delta x)^3 \approx (x^3) + (3x^2)(\Delta x) + \frac{1}{2}(6x)(\Delta x)^2 + \text{error }(\Delta x)^3$.

EXAMPLE 4 $(1 + x)^n \approx 1 + nx + \frac{1}{2}n(n - 1)x^2$.

The first derivative at $x = 0$ is n. The second derivative is $n(n - 1)$. The cubic term would be $\frac{1}{6}n(n - 1)(n - 2)x^3$. We are just producing the binomial expansion!

EXAMPLE 5 $\dfrac{1}{1 - x} \approx 1 + x + x^2 = \text{start of a geometric series.}$

$1/(1 - x)$ has derivative $1/(1 - x)^2$. Its second derivative is $2/(1 - x)^3$. At $x = 0$ those equal $1, 1, 2$. The factor $\frac{1}{2}$ cancels the 2, which leaves $1, 1, 1$. This explains $1 + x + x^2$.
The next terms are x^3 and x^4. The whole series is $1/(1 - x) = 1 + x + x^2 + x^3 + \cdots$.

Numerical experiment $1/\sqrt{1 + x} \approx 1 - \frac{1}{2}x + \frac{3}{8}x^2$ is tested for accuracy. Dividing x by 2 almost divides the error by 8. If we only keep the linear part $1 - \frac{1}{2}x$, the error is only divided by 4. Here are the errors at $x = \frac{1}{4}, \frac{1}{8}$, and $\frac{1}{16}$:

$$\text{linear approximation}\left(\text{error} \approx \frac{3}{8}x^2\right): \quad .0194 \quad .0053 \quad .0014$$

$$\text{quadratic approximation}\left(\text{error} \approx \frac{-5}{16}x^3\right): \quad -.00401 \quad -.00055 \quad -.00007$$

Fig. 3.8 (left margin): graph labeled $1 + x$ can't bend; $1 + x + x^2$ near $\frac{1}{1-x}$; axes from $-.5$ to $.5$, 2 at top.
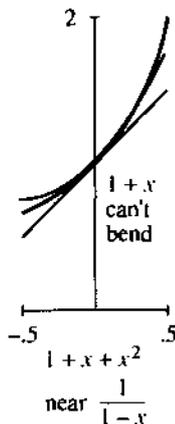

3.3 EXERCISES

Read-through questions

The direction of bending is given by the sign of __a__. If the second derivative is __b__ in an interval, the function is concave up (or convex). The graph bends __c__. The tangent lines are __d__ the graph. If $f''(x) < 0$ then the graph is concave __e__, and the slope is __f__.

At a point where $f'(x) = 0$ and $f''(x) > 0$, the function has a __g__. At a point where __h__, the function has a maximum. A point where $f''(x) = 0$ is an __i__ point, provided f'' changes sign. The tangent line __j__ the graph.

The centered approximation to $f'(x)$ is [__k__]$/2\Delta x$. The 3-point approximation to $f''(x)$ is [__l__]$/(\Delta x)^2$. The second-order approximation to $f(x + \Delta x)$ is $f(x) + f'(x)\Delta x + $__m__. Without that extra term this is just the __n__ approximation. With that term the error is $O($__o__$)$.

1 A graph that is concave upward is inaccurately said to "hold water." Sketch a graph with $f''(x) > 0$ that would not hold water.

2 Find a function that is concave down for $x < 0$ and concave up for $0 < x < 1$ and concave down for $x > 1$.

3 Can a function be always concave down and never cross zero? Can it be always concave down and positive? Explain.

4 Find a function with $f''(2) = 0$ and no other inflection point.

True or false, when $f(x)$ is a 9th degree polynomial with $f'(1) = 0$ and $f'(3) = 0$. Give (or draw) a reason.

5 $f(x) = 0$ somewhere between $x = 1$ and $x = 3$.

6 $f''(x) = 0$ somewhere between $x = 1$ and $x = 3$.

7 There is no absolute maximum at $x = 3$.

8 There are seven points of inflection.

9 If $f(x)$ has nine zeros, it has seven inflection points.

10 If $f(x)$ has seven inflection points, it has nine zeros.

In 11–16 decide which stationary points are maxima or minima.

11 $f(x) = x^2 - 6x$

12 $f(x) = x^3 - 6x^2$

13 $f(x) = x^4 - 6x^3$

14 $f(x) = x^{11} - 6x^{10}$

15 $f(x) = \sin x - \cos x$

16 $f(x) = x + \sin 2x$

Locate the inflection points and the regions where $f(x)$ is concave up or down.

17 $f(x) = x + x^2 - x^3$

18 $f(x) = \sin x + \tan x$

19 $f(x) = (x - 2)^2(x - 4)^2$

20 $f(x) = \sin x + (\sin x)^3$

21 If $f(x)$ is an even function, the centered difference $[f(\Delta x) - f(-\Delta x)]/2\Delta x$ exactly equals $f'(0) = 0$. Why?

22 If $f(x)$ is an odd function, the second difference $[f(\Delta x) - 2f(0) + f(-\Delta x)]/(\Delta x)^2$ exactly equals $f''(0) = 0$. Why?

Write down the quadratic $f(0) + f'(0)x + \frac{1}{2}f''(0)x^2$ in 23–26.

23 $f(x) = \cos x + \sin x$

24 $f(x) = \tan x$

25 $f(x) = (\sin x)/x$

26 $f(x) = 1 + x + x^2$

In 26, find $f(1) + f'(1)(x - 1) + \frac{1}{2}f''(1)(x - 1)^2$ around $a = 1$.

27 Find A and B in $\sqrt{1 - x} \approx 1 + Ax + Bx^2$.

28 Find A and B in $1/(1 - x)^2 \approx 1 + Ax + Bx^2$.

29 Substitute the quadratic approximation into $[f(x + \Delta x) - f(x)]/\Delta x$, to estimate the error in this one-sided approximation to $f'(x)$.

30 What is the quadratic approximation at $x = 0$ to $f(-\Delta x)$?

31 Substitute for $f(x + \Delta x)$ and $f(x - \Delta x)$ in the centered approximation $[f(x + \Delta x) - f(x - \Delta x)]/2\Delta x$, to get $f'(x) + $ error. Find the Δx and $(\Delta x)^2$ terms in this error. Test on $f(x) = x^3$ at $x = 0$.

32 Guess a third-order approximation $f(\Delta x) \approx f(0) + f'(0)\Delta x + \frac{1}{2}f''(0)(\Delta x)^2 + \underline{\hspace{1cm}}$. Test it on $f(x) = x^3$.

Construct a table as in the text, showing the actual errors at $x = 0$ in one-sided differences, centered differences, second differences, and quadratic approximations. By hand take two values of Δx, by calculator take three, by computer take four.

33 $f(x) = x^3 + x^4$

34 $f(x) = 1/(1 - x)$

35 $f(x) = x^2 + \sin x$

36 Example 5 was $1/(1 - x) \approx 1 + x + x^2$. What is the error at $x = 0.1$? What is the error at $x = 2$?

37 Substitute $x = .01$ and $x = -0.1$ in the geometric series $1/(1 - x) = 1 + x + x^2 + \cdots$ to find $1/.99$ and $1/1.1$—first to four decimals and then to all decimals.

38 Compute $\cos 1°$ by equation (4) with $a = 0$. OK to check on a calculator. Also compute $\cos 1$. Why so far off?

39 Why is $\sin x \approx x$ not only a linear approximation but also a quadratic approximation? $x = 0$ is an \underline{\hspace{1cm}} point.

40 If $f(x)$ is an even function, find its quadratic approximation at $x = 0$. What is the equation of the tangent line?

41 For $f(x) = x + x^2 + x^3$, what is the centered difference $[f(3) - f(1)]/2$, and what is the true slope $f'(2)$?

42 For $f(x) = x + x^2 + x^3$, what is the second difference $[f(3) - 2f(2) + f(1)]/1^2$, and what is the exact $f''(2)$?

43 The error in $f(a) + f'(a)(x - a)$ is approximately $\frac{1}{2}f''(a)(x - a)^2$. This error is positive when the function is \underline{\hspace{1cm}}. Then the tangent line is \underline{\hspace{1cm}} the curve.

44 Draw a piecewise linear $y(x)$ that is concave up. Define "concave up" without using the test $d^2y/dx^2 \geq 0$. If derivatives don't exist, a new definition is needed.

45 What do these sentences say about f or f' or f'' or f'''?

1. The population is growing more slowly.
2. The plane is landing smoothly.
3. The economy is picking up speed.
4. The tax rate is constant.
5. A bike accelerates faster but a car goes faster.
6. Stock prices have peaked.
7. The rate of acceleration is slowing down.
8. This course is going downhill.

46 (Recommended) Draw a curve that goes up-down-up. Below it draw its derivative. Then draw its second derivative. *Mark the same points on all curves*—the maximum, minimum, and inflection points of the first curve.

47 Repeat Problem 46 on a printout showing $y(x) = x^3 - 4x^2 + x + 2$ and dy/dx and d^2y/dx^2 on the same graph.

3.4 Graphs

Reading a graph is like appreciating a painting. Everything is there, but you have to know what to look for. One way to learn is by sketching graphs yourself, and in the past that was almost the only way. Now it is obsolete to spend weeks drawing curves—a computer or graphing calculator does it faster and better. That doesn't remove the need to appreciate a graph (or a painting), since a curve displays a tremendous amount of information.

This section combines two approaches. One is to study actual machine-produced graphs (especially electrocardiograms). The other is to understand the mathematics of graphs—slope, concavity, asymptotes, shifts, and scaling. We introduce the *centering transform* and *zoom transform*. These two approaches are like the rest of calculus, where special derivatives and integrals are done by hand and day-to-day applications are by computer. Both are essential—the machine can do experiments that we could never do. But without the mathematics our instructions miss the point. To create good graphs you have to know a few of them personally.

READING AN ELECTROCARDIOGRAM (ECG or EKG)

The graphs of an ECG show the electrical potential during a heartbeat. There are twelve graphs—six from leads attached to the chest, and six from leads to the arms and left leg. (It doesn't hurt, but everybody is nervous. You have to lie still, because contraction of other muscles will mask the reading from the heart.) The graphs record electrical impulses, as the cells depolarize and the heart contracts.

What can I explain in two pages? The graph shows the fundamental pattern of the ECG. *Note the P wave, the QRS complex, and the T wave.* Those patterns, seen differently in the twelve graphs, tell whether the heart is normal or out of rhythm— or suffering an infarction (a heart attack).
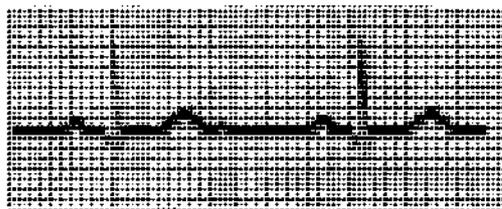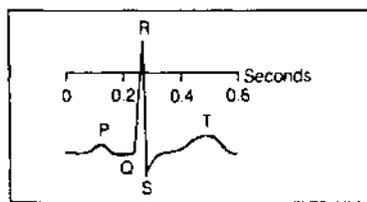
First of all the graphs show the *heart rate*. The dark vertical lines are by convention $\frac{1}{5}$ second apart. The light lines are $\frac{1}{25}$ second apart. If the heart beats every $\frac{1}{5}$ second (one dark line) the rate is 5 beats per second or 300 per minute. That is extreme *tachycardia*—not compatible with life. The normal rate is between three dark lines per beat ($\frac{3}{5}$ second, or 100 beats per minute) and five dark lines (one second between beats, 60 per minute). A baby has a faster rate, over 100 per minute. In this figure the rate is _____. A rate below 60 is *bradycardia*, not in itself dangerous. For a resting athlete that is normal.

Doctors memorize the six rates 300, 150, 100, 75, 60, 50. Those correspond to 1, 2, 3, 4, 5, 6 dark lines between heartbeats. The distance is easiest to measure between spikes (the peaks of the R wave). Many doctors put a printed scale next to the chart. One textbook emphasizes that "Where the next wave falls determines the rate. No mathematical computation is necessary." But you see where those numbers come from.

The next thing to look for is **heart rhythm**. The regular rhythm is set by the pacemaker, which produces the P wave. A constant distance between waves is good—and then each beat is examined. When there is a block in the pathway, it shows as a delay in the graph. Sometimes the pacemaker fires irregularly. Figure 3.10 shows *sinus arrythmia* (fairly normal). The time between peaks is changing. In disease or emergency, there are potential pacemakers in all parts of the heart.

I should have pointed out the main parts. We have four chambers, an atrium-ventricle pair on the left and right. The SA node should be the pacemaker. The stimulus spreads from the atria to the ventricles—from the small chambers that "prime the pump" to the powerful chambers that drive blood through the body. The P wave comes with contraction of the atria. There is a pause of $\frac{1}{10}$ second at the AV node. Then the big QRS wave starts contraction of the ventricles, and the T wave is when the ventricles relax. The cells switch back to negative charge and the heart cycle is complete.
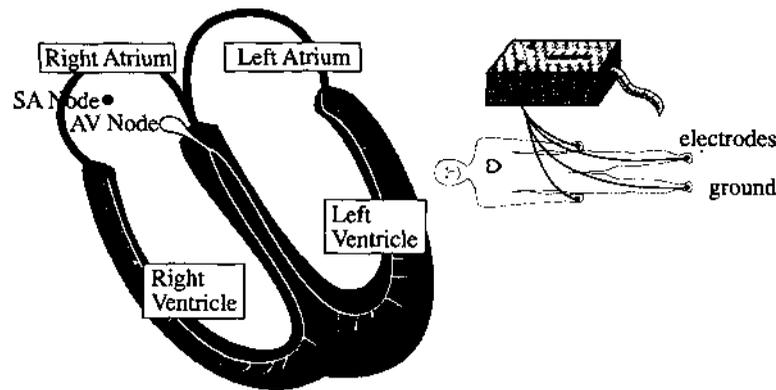

Fig. 3.9 Happy person with a heart and a normal electrocardiogram.

The ECG shows when the pacemaker goes wrong. Other pacemakers take over—the AV node will pace at 60/minute. An early firing in the ventricle can give a wide spike in the QRS complex, followed by a long pause. The impulses travel by a slow path. Also the pacemaker can suddenly speed up (paroxysmal tachycardia is 150–250/minute). But the most critical danger is *fibrillation*.

Figure 3.10h shows a dying heart. The ECG indicates irregular contractions—no normal PQRST sequence at all. What kind of heart would generate such a rhythm? The muscles are quivering or "fibrillating" independently. The pumping action is nearly gone, which means emergency care. The patient needs immediate CPR—someone to do the pumping that the heart can't do. Cardio-pulmonary resuscitation is a combination of chest pressure and air pressure (hand and mouth) to restart the rhythm. CPR can be done on the street. A hospital applies a defibrillator, which shocks the heart back to life. It depolarizes *all* the heart cells, so the timing can be reset. Then the charge spreads normally from SA node to atria to AV node to ventricles.

This discussion has not used all twelve graphs to locate the problem. That needs *vectors*. Look ahead at Section 11.1 for the heart vector, and especially at Section 11.2 for its *twelve projections*. Those readings distinguish between atrium and ventricle, left and right, forward and back. This information is of vital importance in the event of a heart attack. A "heart attack" is a *myocardial infarction* (MI).

An MI occurs when part of an artery to the heart is blocked (a coronary occlusion).

Fig. 3.10 Doubtful rhythm. Serious fibrillation. Signals of a heart attack.

An area is without blood supply—therefore without oxygen or glucose. Often the attack is in the thick left ventricle, which needs the most blood. The cells are first ischemic, then injured, and finally infarcted (dead). The classical ECG signals involve those three I's:

Ischemia: Reduced blood supply, upside-down T wave in the chest leads.
Injury: An elevated segment between S and T means a recent attack.
Infarction: The Q wave, normally a tiny dip or absent, is as wide as a small square ($\frac{1}{25}$ second). It may occupy a third of the entire QRS complex.

The Q wave gives the diagnosis. You can find all three I's in Figure 3.10c.
It is absolutely amazing how much a good graph can do.

THE MECHANICS OF GRAPHS

From the meaning of graphs we descend to the mechanics. A formula is now given for $f(x)$. The problem is *to create the graph*. It would be too old-fashioned to evaluate $f(x)$ by hand and draw a curve through a dozen points. A computer has a much better idea of a parabola than an artist (who tends to make it asymptotic to a straight line). There are some things a computer knows, and other things an artist knows, and still others that you and I know—because we understand derivatives.

Our job is to apply calculus. We extract information from f' and f'' as well as f. Small movements in the graph may go unnoticed, but the important properties come through. Here are the main tests:

1. The sign of $f(x)$ (above or below axis: $f = 0$ at *crossing point*)
2. The sign of $f'(x)$ (increasing or decreasing: $f' = 0$ at *stationary point*)
3. The sign of $f''(x)$ (concave up or down: $f'' = 0$ at *inflection point*)
4. The behavior of $f(x)$ as $x \to \infty$ and $x \to -\infty$
5. The points at which $f(x) \to \infty$ or $f(x) \to -\infty$
6. Even or odd?   Periodic?   Jumps in f or f'?   Endpoints? $f(0)$?

EXAMPLE 1 $f(x) = \dfrac{x^2}{1 - x^2}$ $f'(x) = \dfrac{2x}{(1 - x^2)^2}$ $f''(x) = \dfrac{2 + 6x^2}{(1 - x^2)^3}$

The sign of $f(x)$ depends on $1 - x^2$. Thus $f(x) > 0$ in the inner interval where $x^2 < 1$. The graph bends upwards ($f''(x) > 0$) in that same interval. There are no inflection points, since f'' is never zero. The stationary point where f' vanishes is $x = 0$. We have a *local minimum* at $x = 0$.

The guidelines (or *asymptotes*) meet the graph at infinity. For large x the important terms are x^2 and $-x^2$. Their ratio is $+x^2/-x^2 = -1$—which is the limit as $x \to \infty$ and $x \to -\infty$. *The horizontal asymptote is the line* $y = -1$.

The other infinities, where f blows up, occur when $1 - x^2$ is zero. That happens at $x = 1$ and $x = -1$. *The vertical asymptotes are the lines* $x = 1$ *and* $x = -1$. The graph

in Figure 3.11a approaches those lines.

if $f(x) \to b$ as $x \to +\infty$ *or* $-\infty$, the line $y = b$ is a *horizontal asymptote*

if $f(x) \to +\infty$ or $-\infty$ as $x \to a$, the line $x = a$ is a *vertical asymptote*

if $f(x) - (mx + b) \to 0$ as $x \to +\infty$ or $-\infty$, the line $y = mx + b$ is a *sloping asymptote*.

Finally comes the vital fact that this function is *even*: $f(x) = f(-x)$ because squaring x obliterates the sign. The graph is symmetric across the y axis.

To summarize the effect of dividing by $1 - x^2$: No effect near $x = 0$. Blowup at 1 and -1 from zero in the denominator. The function approaches -1 as $|x| \to \infty$.

EXAMPLE 2 $\qquad f(x) = \dfrac{x^2}{x-1} \qquad f'(x) = \dfrac{x^2 - 2x}{(x-1)^2} \qquad f''(x) = \dfrac{2}{(x-1)^3}$

This example divides by $x - 1$. Therefore $x = 1$ is a vertical asymptote, where $f(x)$ becomes infinite. Vertical asymptotes come mostly from *zero denominators*.

Look beyond $x = 1$. Both $f(x)$ and $f''(x)$ are positive for $x > 1$. The slope is zero at $x = 2$. That must be a local minimum.

What happens as $x \to \infty$? Dividing x^2 by $x - 1$, the leading term is x. The function becomes large. It grows linearly—we expect a *sloping asymptote*. To find it, do the division properly:

$$\frac{x^2}{x-1} = x + 1 + \frac{1}{x-1}. \tag{1}$$

The last term goes to zero. The function approaches $y = x + 1$ as the asymptote.

This function is not odd or even. Its graph is in Figure 3.11b. With *zoom out* you see the asymptotes. *Zoom in* for $f = 0$ or $f' = 0$ or $f'' = 0$.
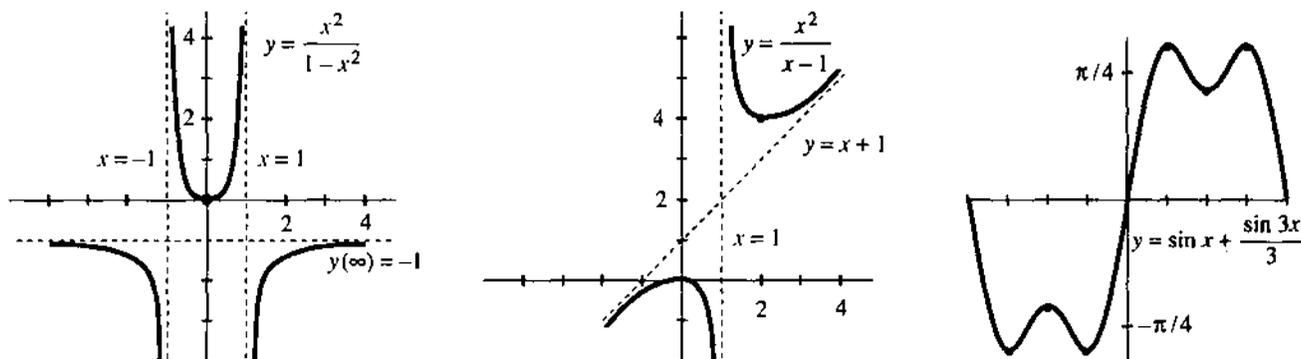

Fig. 3.11 The graphs of $x^2/(1 - x^2)$ and $x^2/(x - 1)$ and $\sin x + \frac{1}{3} \sin 3x$.

EXAMPLE 3 $\quad f(x) = \sin x + \frac{1}{3}\sin 3x \quad$ has the slope $\quad f'(x) = \cos x + \cos 3x$.

Above all these functions are *periodic*. If x increases by 2π, nothing changes. The graphs from 2π to 4π are repetitions of the graphs from 0 to 2π. Thus $f(x + 2\pi) = f(x)$ and the period is 2π. Any interval of length 2π will show a complete picture, and Figure 3.11c picks the interval from $-\pi$ to π.

The second outstanding property is that f is *odd*. The sine functions satisfy $f(-x) = -f(x)$. The graph is symmetric through the origin. By reflecting the right half through the origin, you get the left half. In contrast, the cosines in $f'(x)$ are even.

To find the zeros of $f(x)$ and $f'(x)$ and $f''(x)$, rewrite those functions as

$$f(x) = 2\sin x - \tfrac{4}{3}\sin^3 x \quad f'(x) = -2\cos x + 4\cos^3 x \quad f''(x) = -10\sin x + 12\sin^3 x.$$

We changed $\sin 3x$ to $3 \sin x - 4 \sin^3 x$. For the derivatives use $\sin^2 x = 1 - \cos^2 x$. Now find the zeros—the *crossing points, stationary points,* and *inflection points*:

$$f = 0 \quad 2 \sin x = \tfrac{4}{3} \sin^3 x \ \Rightarrow \ \sin x = 0 \ \text{or} \ \sin^2 x = \tfrac{3}{2} \ \Rightarrow \ x = 0, \ \pm \pi$$

$$f' = 0 \quad 2 \cos x = 4 \cos^3 x \ \Rightarrow \ \cos x = 0 \ \text{or} \ \cos^2 x = \tfrac{1}{2} \ \Rightarrow \ x = \pm \pi/4, \ \pm \pi/2, \ \pm 3\pi/4$$

$$f'' = 0 \quad 5 \sin x = 6 \sin^3 x \ \Rightarrow \ \sin x = 0 \ \text{or} \ \sin^2 x = \tfrac{5}{6} \ \Rightarrow \ x = 0, \ \pm 66°, \ \pm 114°, \ \pm \pi$$

That is more than enough information to sketch the graph. The stationary points $\pi/4$, $\pi/2$, $3\pi/4$ are evenly spaced. At those points $f(x)$ is $\sqrt{8}/3$ (maximum), $2/3$ (local minimum), $\sqrt{8}/3$ (maximum). Figure 3.11c shows the graph.

I would like to mention a beautiful continuation of this same pattern:

$$f(x) = \sin x + \tfrac{1}{3} \sin 3x + \tfrac{1}{5} \sin 5x + \cdots \qquad f'(x) = \cos x + \cos 3x + \cos 5x + \cdots$$

If we stop after ten terms, $f(x)$ is extremely close to a *step function*. If we don't stop, *the exact step function contains infinitely many sines.* It jumps from $-\pi/4$ to $+\pi/4$ as x goes past zero. More precisely it is a *"square wave,"* because the graph jumps back down at π and repeats. The slope $\cos x + \cos 3x + \cdots$ also has period 2π. *Infinitely many cosines add up to a delta function!* (The slope at the jump is an infinite spike.) These sums of sines and cosines are *Fourier series.*

GRAPHS BY COMPUTERS AND CALCULATORS

We have come to a topic of prime importance. If you have *graphing software* for a computer, or if you have a *graphing calculator*, you can bring calculus to life. A graph presents $y(x)$ in a new way—different from the formula. Information that is buried in the formula is clear on the graph. *But don't throw away $y(x)$ and dy/dx.* The derivative is far from obsolete.

These pages discuss how calculus and graphs go together. We work on a crucial problem of applied mathematics—to find where $y(x)$ reaches its minimum. There is no need to tell you a hundred applications. Begin with the formula. How do you find the point x^* where $y(x)$ is smallest?

First, draw the graph. That shows the main features. We should see (roughly) where x^* lies. There may be several minima, or possibly none. But what we see depends on a decision that is ours to make—*the range of x and y in the viewing window.*

If nothing is known about $y(x)$, the range is hard to choose. We can accept a default range, and zoom in or out. We can use the autoscaling program in Section 1.7. Somehow x^* can be observed on the screen. Then the problem is to compute it.

I would like to work with a specific example. We solved it by calculus—to find the best point x^* to enter an expressway. The speeds in Section 3.2 were 30 and 60. The length of the fast road will be $b = 6$. *The range of reasonable values for the entering point is $0 \leqslant x \leqslant 6$.* The distance to the road in Figure 3.12 is $a = 3$. We drive a distance $\sqrt{3^2 + x^2}$ at speed 30 and the remaining distance $6 - x$ at speed 60:

$$\text{driving time} \quad y(x) = \frac{1}{30} \sqrt{3^2 + x^2} + \frac{1}{60}(6 - x). \tag{2}$$

This is the function to be minimized. Its graph is extremely flat.

It may seem unusual for the graph to be so level. On the contrary, it is common. *A flat graph is the whole point of $dy/dx = 0$.*

The graph near the minimum looks like $y = Cx^2$. It is a parabola sitting on a horizontal tangent. At a distance of $\Delta x = .01$, we only go up by $C(\Delta x)^2 = .0001 C$. Unless C is a large number, this Δy can hardly be seen.
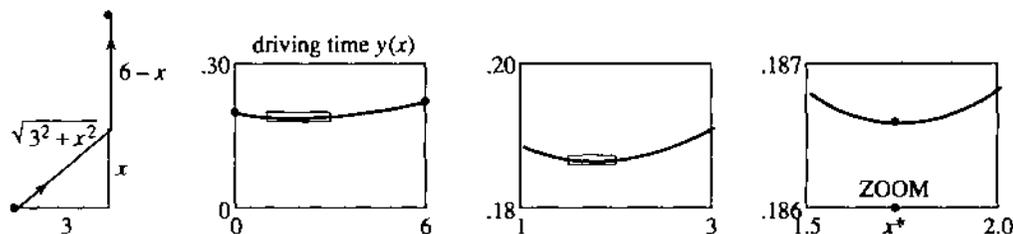
Fig. 3.12 Enter at x. The graph of driving time $y(x)$. Zoom boxes locate x^*.

*The solution is to change scale. **Zoom in on** x^*.* The tangent line stays flat, since dy/dx is still zero. But the bending from C is increased. Figure 3.12 shows the *zoom box* blown up into a new graph of $y(x)$.

A calculator has one or more ways to find x^*. With a TRACE mode, you direct a cursor along the graph. From the display of y values, read y_{max} and x^* to the nearest pixel. A zoom gives better accuracy, because it stretches the axes—each pixel represents a smaller Δx and Δy. The TI-81 stretches by 4 as default. Even better, let the whole process be graphical—*draw the actual* ZOOM BOX *on the screen.* Pick two opposite corners, press ENTER, and the box becomes the new viewing window (Figure 3.12).

The first zoom narrows the search for x^*. It lies between $x = 1$ and $x = 3$. We build a new ZOOM BOX and zoom in again. Now $1.5 \leqslant x^* \leqslant 2$. Reasonable accuracy comes quickly. High accuracy does not come quickly. It takes time to create the box and execute the zoom.

Question 1 What happens as we zoom in, if all boxes are square (equal scaling)?
Answer The picture gets flatter and flatter. We are zooming in to the tangent line. Changing x to $X/4$ and y to $Y/4$, the parabola $y = x^2$ flattens to $Y = X^2/4$. To see any bending, *we must use a long thin zoom box*.

I want to change to a totally different approach. Suppose we have a formula for dy/dx. *That derivative was produced by an infinite zoom!* The limit of $\Delta y/\Delta x$ came by brainpower alone:

$$\frac{dy}{dx} = \frac{x}{30\sqrt{3^2 + x^2}} - \frac{1}{60}. \qquad \text{Call this } f(x).$$

This function is zero at x^*. The computing problem is completely changed: Solve $f(x) = 0$. *It is easier to find a root of $f(x)$ than a minimum of $y(x)$.* The graph of $f(x)$ crosses the x axis. The graph of $y(x)$ goes flat—this is harder to pinpoint.

slope dy/dx

zero slope
at minimum

Fig. 3.13

Take the model function $y = x^2$ for $|x| < .01$. The slope $f = 2x$ changes from $-.02$ to $+.02$. The value of x^2 moves only by $.0001$—its minimum point is hard to see.

To repeat: Minimization is easier with dy/dx. The screen shows an order of magnitude improvement, when we trace or zoom on $f(x) = 0$. In calculus, we have been taking the derivative for granted. It is natural to get blasé about $dy/dx = 0$. We forget how intelligent it is, to work with the slope instead of the function.
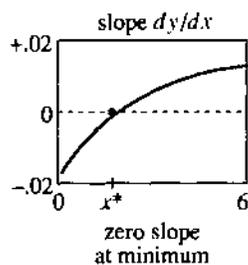
Question 2 How do you get another order of magnitude improvement?
Answer Use the next derivative! With a formula for df/dx, which is d^2y/dx^2, the convergence is even faster. In two steps the error goes from $.01$ to $.0001$ to $.00000001$. Another infinite zoom went into the formula for df/dx, and *Newton's method* takes account of it. Sections 3.6 and 3.7 study $f(x) = 0$.

The expressway example allows perfect accuracy. We can solve $dy/dx = 0$ by algebra. The equation simplifies to $60x = 30\sqrt{3^2 + x^2}$. Dividing by 30 and squaring yields $4x^2 = 3^2 + x^2$. Then $3x^2 = 3^2$. The exact solution is $x^* = \sqrt{3} = 1.73205...$

A model like this is a benchmark, to test competing methods. It also displays what we never appreciated—the extreme flatness of the graph. The difference in driving time between entering at $x^* = \sqrt{3}$ and $x = 2$ is *one second*.

THE CENTERING TRANSFORM AND ZOOM TRANSFORM

For a photograph we do two things—point the right way and stand at the right distance. Then take the picture. Those steps are the same for a graph. First we pick the new center point. The graph is *shifted*, to move that point from (a, b) to $(0, 0)$. Then we decide how far the graph should reach. It fits in a rectangle, just like the photograph. *Rescaling* to x/c and y/d puts the desired section of the curve into the rectangle.

A good photographer does more (like an artist). The subjects are placed and the camera is focused. For good graphs those are necessary too. But an everyday calculator or computer or camera is built to operate without an artist—just aim and shoot. I want to explain how to aim at $y = f(x)$.

We are doing exactly what a calculator does, with one big difference. *It doesn't change coordinates. We do.* When $x = 1$, $y = -2$ moves to the center of the viewing window, the calculator still shows that point as $(1, -2)$. When the *centering transform* acts on $y + 2 = m(x - 1)$, those numbers disappear. This will be confusing unless x and y also change. *The new coordinates are $X = x - 1$ and $Y = y + 2$. Then the new equation is $Y = mX$.*

The main point (for humans) is to make the algebra simpler. The computer has no preference for $Y = mX$ over $y - y_0 = m(x - x_0)$. It accepts $2x^2 - 4x$ as easily as x^2. But we do prefer $Y = mX$ and $y = x^2$, partly because their graphs go through $(0, 0)$. Ever since zero was invented, mathematicians have liked that number best.

3F A *centering transform* shifts left by a and down by b:

$$X = x - a \text{ and } Y = y - b \text{ change } y = f(x) \text{ into } Y + b = f(X + a).$$

EXAMPLE 4 The parabola $y = 2x^2 - 4x$ has its minimum when $dy/dx = 4x - 4 = 0$. Thus $x = 1$ and $y = -2$. Move this bottom point to the center: $y = 2x^2 - 4x$ is

$$Y + 2 = 2(X - 1)^2 - 4(X - 1) \qquad \text{or} \qquad Y = 2X^2.$$

The new parabola $Y = 2X^2$ has its bottom at $(0, 0)$. It is the same curve, shifted across and up. The only simpler parabola is $y = x^2$. This final step is the job of the zoom.

Next comes scaling. We may want more detail (zoom in to see the tangent line). We may want a big picture (zoom out to check asymptotes). We might stretch one axis more than the other, if the picture looks like a pancake or a skyscraper.

3G A *zoom transform* scales the X and Y axes by c and d:

$$\mathbf{x} = cX \quad \text{and} \quad \mathbf{y} = dY \quad \text{change} \quad Y = F(X) \quad \text{to} \quad \mathbf{y} = dF(\mathbf{x}/c).$$

The new $\mathbf{x}$ and $\mathbf{y}$ are boldface letters, and the graph is rescaled. Often $c = d$.

EXAMPLE 5 Start with $Y = 2X^2$. Apply a square zoom with $c = d$. In the new xy coordinates, the equation is $y/c = 2(x/c)^2$. The number 2 disappears if $c = d = 2$. With the right centering and the right zoom, every parabola that opens upward is $y = x^2$.

Question 3 What happens to the derivatives (*slope and bending*) after a zoom? **Answer** The slope (first derivative) is multiplied by d/c. Apply the chain rule to $y = dF(x/c)$. A square zoom has $d/c = 1$—*lines keep their slope*. The second derivative is multiplied by d/c^2, which changes the bending. A zoom out divides by small numbers $c = d$, so the big picture is more curved.

Combining the centering and zoom transforms, as we do in practice, gives y in terms of x:

$$y = f(x) \quad \text{becomes} \quad Y = f(X + a) - b \quad \text{and then} \quad y = d\left[f\left(\frac{x}{c} + a \right) - b \right].$$
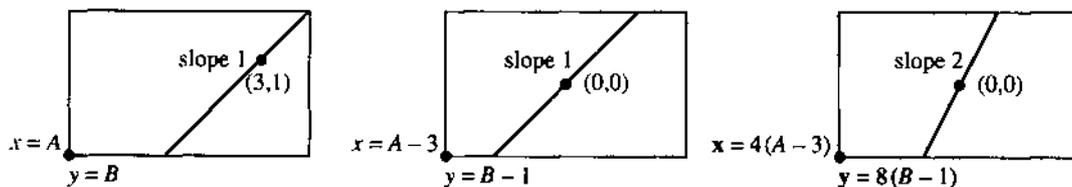
Fig. 3.14 Change of coordinates by centering and zoom. Calculators still show (x, y).

Question 4 Find x and y ranges after two transforms. Start between -1 and 1. **Answer** The window after centering is $-1 \leqslant x - a \leqslant 1$ and $-1 \leqslant y - b \leqslant 1$. The window after zoom is $-1 \leqslant c(x - a) \leqslant 1$ and $-1 \leqslant d(y - b) \leqslant 1$. The point (1, 1) was originally in the corner. The point $(c^{-1} + a, d^{-1} + b)$ is now in the corner.

The numbers a, b, c, d are chosen to produce a simpler function (like $y = x^2$). Or else—this is important in applied mathematics—they are chosen to make x and y "dimensionless." An example is $y = \frac{1}{2} \cos 8t$. The frequency 8 has dimension 1/time. The amplitude $\frac{1}{2}$ is a distance. With $d = 2$ cm and $c = 8$ sec, the units are removed and $y = \cos t$.

May I mention one transform that *does* change the slope? It is a **rotation**. The whole plane is turned. A photographer might use it—but normally people are supposed to be upright. You use rotation when you turn a map or straighten a picture. In the next section, an unrecognizable hyperbola is turned into $Y = 1/X$.

3.4 EXERCISES

Read-through questions

The position, slope, and bending of $y = f(x)$ are decided by __a__, __b__ and __c__. If $|f(x)| \to \infty$ as $x \to a$, the line $x = a$ is a vertical __d__. If $f(x) \to b$ for large x, then $y = b$ is a __e__. If $f(x) - mx \to b$ for large x, then $y = mx + b$ is a __f__. The asymptotes of $y = x^2/(x^2 - 4)$ are __g__. This function is even because $y(-x) = $ __h__. The function sin kx has period __i__.

Near a point where $dy/dx = 0$, the graph is extremely __j__. For the model $y = Cx^2$, $x = .1$ gives $y = $ __k__. A box around the graph looks long and __l__. We __m__ in to that box for another digit of x^*. But solving $dy/dx = 0$ is more accurate, because its graph __n__ the x axis. The slope of dy/dx is __o__. Each derivative is like an __p__ zoom.

To move (a, b) to $(0, 0)$, shift the variables to $X = $ __q__ and $Y = $ __r__. This __s__ transform changes $y = f(x)$ to $Y = $ __t__. The original slope at (a, b) equals the new slope at __u__. To stretch the axes by c and d, set $x = cX$ and __v__. The __w__ transform changes $Y = F(X)$ to $y = $ __x__. Slopes are multiplied by __y__. Second derivatives are multiplied by __z__.

1 Find the pulse rate when heartbeats are $\frac{1}{2}$ second or two dark lines or x seconds apart.

2 Another way to compute the heart rate uses marks for 6-second intervals. Doctors count the cycles in an interval.

(a) How many dark lines in 6 seconds?

(b) With 8 beats per interval, find the rate.

(c) Rule: Heart rate = cycles per interval times _____ .

Which functions in 3–18 are even or odd or periodic? Find all asymptotes: $y = b$ or $x = a$ or $y = mx + b$. **Draw roughly by hand or smoothly by computer.**

3 $f(x) = x - (9/x)$ **4** $f(x) = x^n$ (any integer n)

5 $f(x) = \dfrac{1}{1 - x^2}$ **6** $f(x) = \dfrac{x^3}{4 - x^2}$

7 $f(x) = \dfrac{x^2 + 3}{x^2 + 1}$ **8** $f(x) = \dfrac{x^2 + 3}{x + 1}$

9 $f(x) = (\sin x)(\sin 2x)$ **10** $f(x) = \cos x + \cos 3x + \cos 5x$

11 $f(x) = \dfrac{x \sin x}{x^2 - 1}$ **12** $f(x) = \dfrac{x}{\sin x}$

13 $f(x) = \dfrac{1}{x^3 + x^2}$ **14** $f(x) = \dfrac{1}{x - 1} - 2x$

15 $f(x) = \dfrac{x^3 + 1}{x^3 - 1}$ **16** $f(x) = \dfrac{\sin x + \cos x}{\sin x - \cos x}$

17 $f(x) = x - \sin x$ **18** $f(x) = (1/x) - \sqrt{x}$

In 19–24 construct $f(x)$ with exactly these asymptotes.

19 $x = 1$ and $y = 2$ **20** $x = 1, x = 2, y = 0$

21 $y = x$ and $x = 4$ **22** $y = 2x + 3$ and $x = 0$

23 $y = x \ (x \to \infty), \ y = -x(x \to -\infty)$

24 $x = 1, x = 3, y = x$

25 For $P(x)/Q(x)$ to have $y = 2$ as asymptote, the polynomials P and Q must be _____ .

26 For $P(x)/Q(x)$ to have a sloping asymptote, the degrees of P and Q must be _____ .

27 For $P(x)/Q(x)$ to have the asymptote $y = 0$, the degrees of P and Q must _____ . The graph of $x^4/(1 + x^2)$ has what asymptotes?

28 Both $1/(x - 1)$ and $1/(x - 1)^2$ have $x = 1$ *and* $y = 0$ as asymptotes. The most obvious difference in the graphs is _____ .

29 If $f'(x)$ has asymptotes $x = 1$ and $y = 3$ then $f(x)$ has asymptotes _____ .

30 True (with reason) or **false** (with example).

(a) Every ratio of polynomials has asymptotes

(b) If $f(x)$ is even so is $f''(x)$

(c) If $f''(x)$ is even so is $f(x)$

(d) Between vertical asymptotes, $f'(x)$ touches zero.

31 Construct an $f(x)$ that is "even around $x = 3$."

32 Construct $g(x)$ to be "odd around $x = \pi$."

Create graphs of 33–38 on a computer or calculator.

33 $y(x) = (1 + 1/x)^x, \ -3 \leqslant x \leqslant 3$

34 $y(x) = x^{1/x}, 0.1 \leqslant x \leqslant 2$

35 $y(x) = \sin(x/3) + \sin(x/5)$

36 $y(x) = (2 - x)/(2 + x), \ -3 \leqslant x \leqslant 3$

37 $y(x) = 2x^3 + 3x^2 - 12x + 5$ on $[-3, 3]$ and $[2.9, 3.1]$

38 $100[\sin(x + .1) - 2 \sin x + \sin(x - .1)]$

In 39–40 show the asymptotes on large-scale computer graphs.

39 (a) $y = \dfrac{x^3 + 8x - 15}{x^2 - 2}$ (b) $y = \dfrac{x^4 - 6x^3 + 1}{2x^4 + x^2}$

40 (a) $y = \dfrac{x^2 - 2}{x^3 + 8x - 15}$ (b) $y = \dfrac{x^2 - x + 2}{x^2 - 2x + 1}$

41 Rescale $y = \sin x$ so X is in degrees, not radians, and Y changes from meters to centimeters.

Problems 42–46 minimize the driving time $y(x)$ in the text. Some questions may not fit your software.

42 Trace along the graph of $y(x)$ to estimate x^*. Choose an xy range or use the default.

43 Zoom in by $c = d = 4$. How many zooms until you reach $x^* = 1.73205$ or 1.7320508?

44 Ask your program for the minimum of $y(x)$ and the solution of $dy/dx = 0$. Same answer?

45 What are the scaling factors c and d for the two zooms in Figure 3.12? They give the stretching of the x and y axes.

46 Show that $dy/dx = -1/60$ and $d^2y/dx^2 = 1/90$ at $x = 0$. Linear approximation gives $dy/dx \approx -1/60 + x/90$. So the slope is zero near $x =$ _____ . This is Newton's method, using the next derivative.

Change the function to $y(x) = \sqrt{15 + x^2}/30 + (10 - x)/60$.

47 Find x^* using only the graph of $y(x)$.

48 Find x^* using also the graph of dy/dx.

49 What are the xy and XY and xy equations for the line in Figure 3.14?

50 Define $f_n(x) = \sin x + \frac{1}{3} \sin 3x + \frac{1}{5} \sin 5x + \cdots$ (n terms). Graph f_5 and f_{10} from $-\pi$ to π. Zoom in and describe the *Gibbs phenomenon* at $x = 0$.

On the graphs of 51–56, zoom in to all maxima and minima (3 significant digits). *Estimate inflection points.*

51 $y = 2x^5 - 16x^4 + 5x^3 - 37x^2 + 21x + 683$

52 $y = x^5 - x^4 - \sqrt{3x + 1} - 2$

53 $y = x(x - 1)(x - 2)(x - 4)$

54 $y = 7 \sin 2x + 5 \cos 3x$

55 $y = (x^3 - 2x + 1)/(x^4 - 3x^2 - 15)$, $-3 \leqslant x \leqslant 5$

56 $y = x \sin (1/x)$, $0.1 \leqslant x \leqslant 1$

57 A 10-digit computer shows $y = 0$ and $dy/dx = .01$ at $x^* = 1$. This root should be correct to about (8 digits) (10 digits) (12 digits). *Hint:* Suppose $y = .01 (x - 1 + \text{error})$. What errors don't show in 10 digits of y?

58 Which is harder to compute accurately: Maximum point or inflection point? First derivative or second derivative?

3.5 Parabolas, Ellipses, and Hyperbolas

Here is a list of the most important curves in mathematics, so you can tell what is coming. It is not easy to rank the top four:

1. *straight lines*
2. *sines and cosines* (oscillation)
3. *exponentials* (growth and decay)
4. *parabolas, ellipses, and hyperbolas* (using $1, x, y, x^2, xy, y^2$).

The curves that I wrote last, the Greeks would have written first. It is so natural to go from linear equations to quadratic equations. Straight lines use $1, x, y$. Second degree curves include x^2, xy, y^2. If we go on to x^3 and y^3, the mathematics gets complicated. We now study equations of second degree, and the curves they produce.

It is quite important to see both the *equations* and the *curves*. This section connects two great parts of mathematics—*analysis* of the equation and *geometry* of the curve. Together they produce "*analytic geometry*." You already know about functions and graphs. Even more basic: Numbers correspond to points. We speak about "*the point* $(5, 2)$." Euclid might not have understood.

Where Euclid drew a 45° line through the origin, Descartes wrote down $y = x$. Analytic geometry has become central to mathematics—we now look at one part of it.
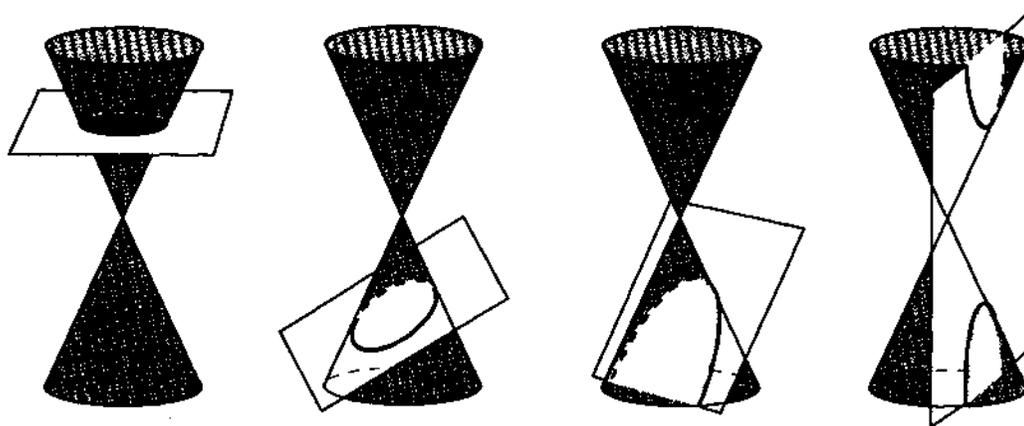

Fig. 3.15 The cutting plane gets steeper: circle to ellipse to parabola to hyperbola.

CONIC SECTIONS

The parabola and ellipse and hyperbola have absolutely remarkable properties. The Greeks discovered that all these curves come from *slicing a cone by a plane*. The curves are "conic sections." A level cut gives a *circle*, and a moderate angle produces an *ellipse*. A steep cut gives the two pieces of a *hyperbola* (Figure 3.15d). At the borderline, when the slicing angle matches the cone angle, the plane carves out a *parabola*. It has one branch like an ellipse, but it opens to infinity like a hyperbola.

Throughout mathematics, parabolas are on the border between ellipses and hyperbolas.

To repeat: We can slice through cones or we can look for equations. For a cone of light, we see an ellipse on the wall. (The wall cuts into the light cone.) For an equation $Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0$, we will work to make it simpler. The graph will be centered and rescaled (and rotated if necessary), aiming for an equation like $y = x^2$. Eccentricity and polar coordinates are left for Chapter 9.

THE PARABOLA $y = ax^2 + bx + c$

You knew this function long before calculus. The graph crosses the x axis when $y = 0$. The quadratic formula solves $y = 3x^2 - 4x + 1 = 0$, and so does factoring into $(x - 1)(3x - 1)$. The crossing points $x = 1$ and $x = \frac{1}{3}$ come from algebra.

The other important point is found by calculus. It is the *minimum* point, where $dy/dx = 6x - 4 = 0$. The x coordinate is $\frac{4}{6} = \frac{2}{3}$, halfway between the crossing points. The height is $y_{min} = -\frac{1}{3}$. This is the *vertex* V in Figure 3.16a—at the bottom of the parabola.

A parabola has no asymptotes. The slope $6x - 4$ doesn't approach a constant.

To center the vertex   Shift left by $\frac{2}{3}$ and up by $\frac{1}{3}$. So introduce the new variables $X = x - \frac{2}{3}$ and $Y = y + \frac{1}{3}$. Then $x = \frac{2}{3}$ and $y = -\frac{1}{3}$ correspond to $X = Y = 0$—which is the new vertex:

$$y = 3x^2 - 4x + 1 \quad \text{becomes} \quad Y = 3X^2. \tag{1}$$

Check the algebra. $Y = 3X^2$ is the same as $y + \frac{1}{3} = 3(x - \frac{2}{3})^2$. That simplifies to the original equation $y = 3x^2 - 4x + 1$. The second graph shows the centered parabola $Y = 3X^2$, with the vertex moved to the origin.

To zoom in on the vertex   Rescale X and Y by the zoom factor a:

$$Y = 3X^2 \quad \text{becomes} \quad y/a = 3(x/a)^2.$$

The final equation has x and y in boldface. With $a = 3$ we find $y = x^2$—the graph is magnified by 3. In two steps we have reached the model parabola opening upward.
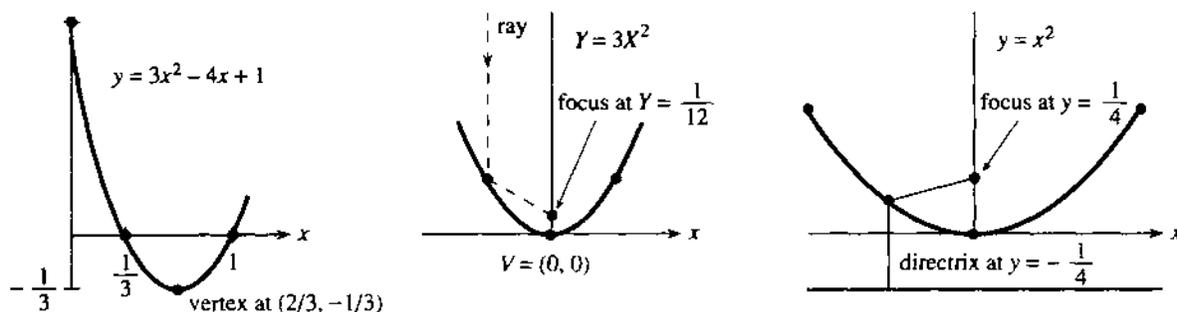

Fig. 3.16   Parabola with minimum at V. Rays reflect to focus. Centered in (b), rescaled in (c).

A parabola has another important point—the *focus*. Its distance from the vertex is called p. The special parabola $y = x^2$ has $p = 1/4$, and other parabolas $Y = aX^2$ have $p = 1/4a$. You magnify by a factor a to get $y = x^2$. The beautiful property of a parabola is that *every ray coming straight down is reflected to the focus*.

Problem 2.3.25 located the focus F—here we mention two applications. A solar collector and a TV dish are parabolic. They concentrate sun rays and TV signals onto a point—a heat cell or a receiver collects them at the focus. The 1982 *UMAP Journal* explains how radar and sonar use the same idea. Car headlights turn the idea around, and send the light outward.

Here is a classical fact about parabolas. *From each point on the curve, the distance to the focus equals the distance to the "directrix."* The directrix is the line $y = -p$ below the vertex (so the vertex is halfway between focus and directrix). With $p = \frac{1}{4}$, the distance down from any (x, y) is $y + \frac{1}{4}$. Match that with the distance to the focus at $(0, \frac{1}{4})$— this is the square root below. Out comes the special parabola $y = x^2$:

$$y + \tfrac{1}{4} = \sqrt{x^2 + (y - \tfrac{1}{4})^2} \quad \longrightarrow \quad \text{(square both sides)} \quad \longrightarrow \quad y = x^2. \quad (2)$$

The exercises give practice with all the steps we have taken—center the parabola to $Y = aX^2$, rescale it to $y = x^2$, locate the vertex and focus and directrix.

Summary for other parabolas $y = ax^2 + bx + c$ has its vertex where dy/dx is zero. Thus $2ax + b = 0$ and $x = -b/2a$. Shifting across to that point is "completing the square":

$$ax^2 + bx + c \quad \text{equals} \quad a\left(x + \frac{b}{2a}\right)^2 + C. \quad (3)$$

Here $C = c - (b^2/4a)$ is the height of the vertex. The centering transform $X = x + (b/2a)$, $Y = y - C$ produces $Y = aX^2$. It moves the vertex to $(0, 0)$, where it belongs.

For the ellipse and hyperbola, our plan of attack is the same:

1. Center the curve to remove any linear terms Dx and Ey.
2. Locate each focus and discover the reflection property.
3. Rotate to remove Bxy if the equation contains it.

$$\text{ELLIPSES} \quad \frac{x^2}{a^2} + \frac{y^2}{b^2} = 1 \quad \text{(CIRCLES HAVE } a = b\text{)}$$

This equation makes the ellipse symmetric about $(0, 0)$—the center. Changing x to $-x$ or y to $-y$ leaves the same equation. No extra centering or rotation is needed.

The equation also shows that x^2/a^2 and y^2/b^2 cannot exceed one. (They add to one and can't be negative.) Therefore $x^2 \leq a^2$, and x stays between $-a$ and a. Similarly y stays between b and $-b$. The ellipse is inside a rectangle.

By solving for y we get a function (or two functions!) of x:

$$\frac{y^2}{b^2} = 1 - \frac{x^2}{a^2} \quad \text{gives} \quad \frac{y}{b} = \pm\sqrt{1 - \frac{x^2}{a^2}} \quad \text{or} \quad y = \pm\frac{b}{a}\sqrt{a^2 - x^2}.$$

The graphs are the top half (+) and bottom half (−) of the ellipse. To draw the ellipse, plot them together. They meet when $y = 0$, at $x = a$ on the far right of Figure 3.17 and at $x = -a$ on the far left. The maximum $y = b$ and minimum $y = -b$ are at the top and bottom of the ellipse, where we bump into the enclosing rectangle.

A circle is a special case of an ellipse, when $a = b$. The circle equation $x^2 + y^2 = r^2$ is the ellipse equation with $a = b = r$. This circle is centered at $(0, 0)$; other circles are

centered at $x = h$, $y = k$. The circle is determined by its *radius r* and its *center* (h, k):

$$\text{Equation of circle: } (x - h)^2 + (y - k)^2 = r^2. \qquad (4)$$

In words, the distance from (x, y) on the circle to (h, k) at the center is r. The equation has linear terms $-2hx$ and $-2ky$—they disappear when the center is $(0, 0)$.

EXAMPLE 1 Find the circle that has a diameter from $(1, 7)$ to $(5, 7)$.

Solution The center is halfway at $(3, 7)$. So $r = 2$ and $(x - 3)^2 + (y - 7)^2 = 2^2$.

EXAMPLE 2 Find the center and radius of the circle $x^2 - 6x + y^2 - 14y = -54$.

Solution Complete $x^2 - 6x$ to the square $(x - 3)^2$ by adding 9. Complete $y^2 - 14y$ to $(y - 7)^2$ by adding 49. Adding 9 and 49 to both sides of the equation leaves $(x - 3)^2 + (y - 7)^2 = 4$—the same circle as in Example 1.

Quicker Solution Match the given equation with (4). Then $h = 3$, $k = 7$, and $r = 2$:

$$x^2 - 6x + y^2 - 14y = -54 \quad \text{must agree with} \quad x^2 - 2hx + h^2 + y^2 - 2ky + k^2 = r^2.$$

The change to $X = x - h$ and $Y = y - k$ moves the center of the circle from (h, k) to $(0, 0)$. This is equally true for an ellipse:

$$\text{The ellipse } \frac{(x - h)^2}{a^2} + \frac{(y - k)^2}{b^2} = 1 \quad \text{becomes} \quad \frac{X^2}{a^2} + \frac{Y^2}{b^2} = 1.$$

When we rescale by $x = X/a$ and $y = Y/b$, we get the unit circle $x^2 + y^2 = 1$.

The unit circle has area π. *The ellipse has area* πab (proved later in the book). The distance around the circle is 2π. The distance around an ellipse does not rescale—it has no simple formula.
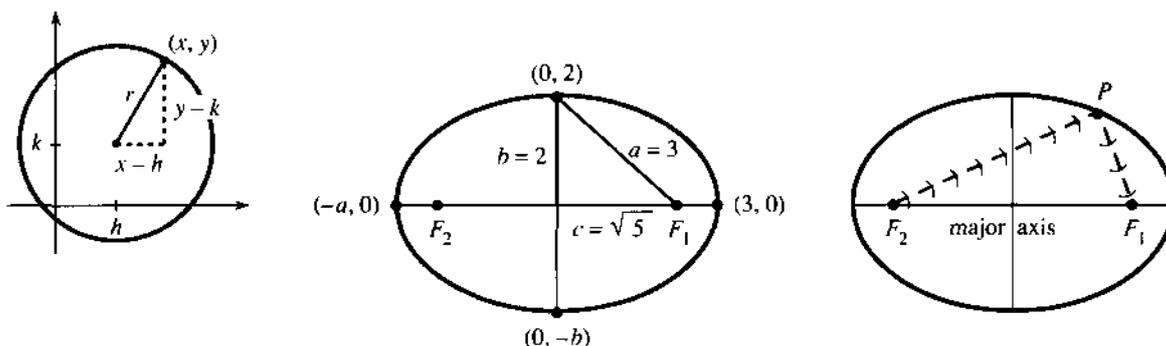

Fig. 3.17 Uncentered circle. Centered ellipse $x^2/3^2 + y^2/2^2 = 1$. *The distance from center to far right is also $a = 3$. All rays from F_2 reflect to F_1.*

Now we leave circles and concentrate on ellipses. They have *two foci* (pronounced *fo-sigh*). For a parabola, the second focus is at infinity. For a circle, both foci are at the center. The foci of an ellipse are on its longer axis (its *major* axis), one focus on each side of the center:

$$F_1 \text{ is at } x = c = \sqrt{a^2 - b^2} \quad \text{and} \quad F_2 \text{ is at } x = -c.$$

The right triangle in Figure 3.17 *has sides a, b, c.* From the top of the ellipse, the distance to each focus is a. From the endpoint at $x = a$, the distances to the foci are $a + c$ and $a - c$. Adding $(a + c) + (a - c)$ gives $2a$. *As you go around the ellipse, the distance to F_1 plus the distance to F_2 is constant* (always $2a$).

> **3H** At all points on the ellipse, the sum of distances from the foci is $2a$. This is another equation for the ellipse:
>
> from F_1 and F_2 to (x, y): $\quad \sqrt{(x-c)^2 + y^2} + \sqrt{(x+c)^2 + y^2} = 2a.$ $\quad$ (5)

To draw an ellipse, tie a string of length $2a$ to the foci. Keep the string taut and your moving pencil will create the ellipse. This description uses a and c—the other form uses a and b (remember $b^2 + c^2 = a^2$). Problem 24 asks you to simplify equation (5) until you reach $x^2/a^2 + y^2/b^2 = 1$.

The "whispering gallery" of the United States Senate is an ellipse. If you stand at one focus and speak quietly, you can be heard at the other focus (and nowhere else). Your voice is reflected off the walls to the other focus—following the path of the string. For a parabola the rays come in to the focus from infinity—where the second focus is.

A hospital uses this reflection property to split up kidney stones. The patient sits inside an ellipse with the kidney stone at one focus. At the other focus a *lithotripter* sends out hundreds of small shocks. You get a spinal anesthetic (I mean the patient) and the stones break into tiny pieces.

The most important focus is the Sun. The ellipse is the orbit of the Earth. See Section 12.4 for a terrible printing mistake by the Royal Mint, on England's last pound note. They put the Sun at the center.

Question 1 Why do the whispers (and shock waves) arrive together at the second focus?
Answer Whichever way they go, the distance is $2a$. Exception: straight path is $2c$.

Question 2 Locate the ellipse with equation $4x^2 + 9y^2 = 36$.
Answer Divide by 36 to change the constant to 1. Now identify a and b:

$$\frac{x^2}{9} + \frac{y^2}{4} = 1 \text{ so } a = \sqrt{9} \text{ and } b = \sqrt{4}. \text{ Foci at } \pm\sqrt{9-4} = \pm\sqrt{5}.$$

Question 3 Shift the center of that ellipse across and down to $x = 1$, $y = -5$.
Answer Change x to $x - 1$. Change y to $y + 5$. The equation becomes $(x-1)^2/9 + (y+5)^2/4 = 1$. In practice we start with this uncentered ellipse and go the other way to center it.

$$\text{HYPERBOLAS} \quad \frac{y^2}{a^2} - \frac{x^2}{b^2} = 1$$

Notice the minus sign for a hyperbola. That makes all the difference. Unlike an ellipse, x and y can both be large. The curve goes out to infinity. It is still symmetric, since x can change to $-x$ and y to $-y$.

The center is at $(0, 0)$. Solving for y again yields two functions $(+$ and $-)$:

$$\frac{y^2}{a^2} - \frac{x^2}{b^2} = 1 \quad \text{gives} \quad \frac{y}{a} = \pm\sqrt{1 + \frac{x^2}{b^2}} \quad \text{or} \quad y = \pm\frac{a}{b}\sqrt{b^2 + x^2}. \quad (6)$$

The hyperbola has two branches that never meet. The upper branch, with a plus sign, has $y \geqslant a$. The *vertex* V_1 is at $x = 0$, $y = a$—the lowest point on the branch. Much further out, when x is large, the hyperbola climbs up beside its *sloping asymptotes*:

$$\text{if} \quad \frac{x^2}{b^2} = 1000 \quad \text{then} \quad \frac{y^2}{a^2} = 1001. \quad \text{So } \frac{y}{a} \text{ is close to } \frac{x}{b} \text{ or } -\frac{x}{b}.$$
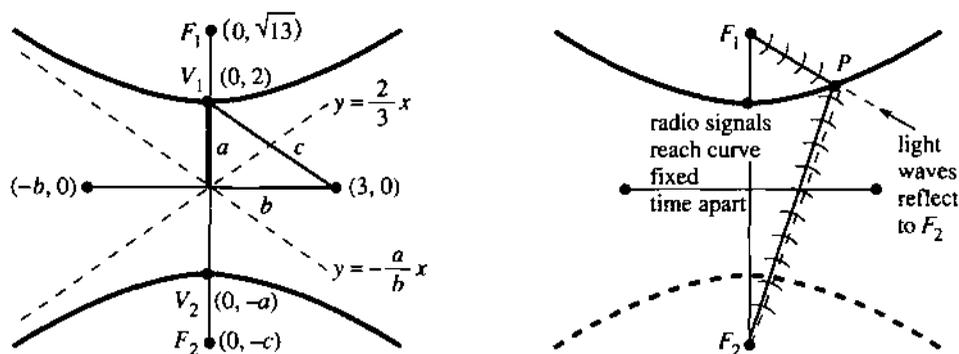
Fig. 3.18   The hyperbola $\frac{1}{4}y^2 - \frac{1}{9}x^2 = 1$ has $a = 2$, $b = 3$, $c = \sqrt{4+9}$. The distances to F_1 and F_2 differ by $2a = 4$.

The asymptotes are the lines $y/a = x/b$ and $y/a = -x/b$. Their slopes are a/b and $-a/b$. You can't miss them in Figure 3.18.

For a hyperbola, the foci are inside the two branches. Their distance from the center is still called c. But now $c = \sqrt{a^2 + b^2}$, which is larger than a and b. The vertex is a distance $c - a$ from one focus and $c + a$ from the other. The *difference* (not the sum) is $(c + a) - (c - a) = 2a$.

All points on the hyperbola have this property: *The difference between distances to the foci is constantly 2a*. A ray coming in to one focus is reflected toward the other. The reflection is on the *outside* of the hyperbola, and the *inside* of the ellipse.

Here is an application to navigation. Radio signals leave two fixed transmitters at the same time. A ship receives the signals a millisecond apart. Where is the ship? *Answer*: It is on a hyperbola with foci at the transmitters. Radio signals travel 186 miles in a millisecond, so $186 = 2a$. This determines the curve. In Long Range Navigation (LORAN) a third transmitter gives another hyperbola. Then the ship is located exactly.

Question 4   How do hyperbolas differ from parabolas, far from the center?
Answer   Hyperbolas have asymptotes. Parabolas don't.

The hyperbola has a natural rescaling. The appearance of x/b is a signal to change to X. Similarly y/a becomes Y. Then $Y = 1$ at the vertex, and we have a standard hyperbola:

$$y^2/a^2 - x^2/b^2 = 1 \quad \text{becomes} \quad Y^2 - X^2 = 1.$$

A 90° turn gives $X^2 - Y^2 = 1$—the hyperbola opens to the sides. A 45° turn produces $2XY = 1$. We show below how to recognize $x^2 + xy + y^2 = 1$ as an ellipse and $x^2 + 3xy + y^2 = 1$ as a hyperbola. (They are not circles because of the xy term.) When the xy coefficient increases past 2, $x^2 + y^2$ no longer indicates an ellipse.

Question 5   Locate the hyperbola with equation $9y^2 - 4x^2 = 36$.
Answer   Divide by 36. Then $y^2/4 - x^2/9 = 1$. Recognize $a = \sqrt{4}$ and $b = \sqrt{9}$.

Question 6   Locate the uncentered hyperbola $9y^2 - 18y - 4x^2 - 4x = 28$.
Answer   Complete $9y^2 - 18y$ to $9(y - 1)^2$ by adding 9. Complete $4x^2 + 4x$ to $4(x + \frac{1}{2})^2$ by adding $4(\frac{1}{2})^2 = 1$. The equation is rewritten as $9(y - 1)^2 - 4(x + \frac{1}{2})^2 = 28 + 9 - 1$. This is the hyperbola in Question 5 — except its center is $(-\frac{1}{2}, 1)$.

To summarize: Find the center by completing squares. Then read off a and b.

THE GENERAL EQUATION $Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0$

This equation is of second degree, containing any and all of $1, x, y, x^2, xy, y^2$. A plane is cutting through a cone. *Is the curve a parabola or ellipse or hyperbola?* Start with the most important case $Ax^2 + Bxy + Cy^2 = 1$.

3I The equation $Ax^2 + Bxy + Cy^2 = 1$ produces a hyperbola if $B^2 > 4AC$ and an ellipse if $B^2 < 4AC$. A parabola has $B^2 = 4AC$.

To recognize the curve, we remove Bxy by *rotating the plane*. This also changes A and C—but the combination $B^2 - 4AC$ is not changed (proof omitted). An example is $2xy = 1$, with $B^2 = 4$. It rotates to $y^2 - x^2 = 1$, with $-4AC = 4$. That positive number 4 signals a hyperbola—since $A = -1$ and $C = 1$ have opposite signs.

Another example is $x^2 + y^2 = 1$. It is a circle (a special ellipse). However we rotate, the equation stays the same. The combination $B^2 - 4AC = 0 - 4 \cdot 1 \cdot 1$ is negative, as predicted for ellipses.

To rotate by an angle α, change x and y to new variables x' and y':

$$x = x' \cos \alpha - y' \sin \alpha \qquad \text{and} \qquad x' = x \cos \alpha + y \sin \alpha$$
$$y = x' \sin \alpha + y' \cos \alpha \qquad\qquad y' = -y \sin \alpha + x \cos \alpha. \qquad (7)$$

Substituting for x and y changes $Ax^2 + Bxy + Cy^2 = 1$ to $A'x'^2 + B'x'y' + C'y'^2 = 1$. The formulas for A', B', C' are painful so I go to the key point:

$$B' \text{ is zero if the rotation angle } \alpha \text{ has } \tan 2\alpha = B/(A - C).$$

With $B' = 0$, the curve is easily recognized from $A'x'^2 + C'y'^2 = 1$. It is a hyperbola if A' and C' have opposite signs. Then $B'^2 - 4A'C'$ is positive. The original $B^2 - 4AC$ was also positive, because this special combination stays constant during rotation.

After the xy term is gone, we deal with x and y—by *centering*. To find the center, complete squares as in Questions 3 and 6. For total perfection, rescale to one of the model equations $y = x^2$ or $x^2 + y^2 = 1$ or $y^2 - x^2 = 1$.

The remaining question is about $F = 0$. What is the graph of $Ax^2 + Bxy + Cy^2 = 0$? The ellipse-hyperbola-parabola have disappeared. But if the Greeks were right, the cone is still cut by a plane. The degenerate case $F = 0$ occurs when the plane cuts *right through the sharp point of the cone*.

A level cut hits only that one point $(0, 0)$. The equation shrinks to $x^2 + y^2 = 0$, a circle with radius zero. A steep cut gives two lines. The hyperbola becomes $y^2 - x^2 = 0$, leaving only its asymptotes $y = \pm x$. A cut at the exact angle of the cone gives only one line, as in $x^2 = 0$. A *single point, two lines*, and *one line* are very extreme cases of an ellipse, hyperbola, and parabola.

All these "conic sections" come from planes and cones. The beauty of the geometry, which Archimedes saw, is matched by the importance of the equations. Galileo discovered that projectiles go along parabolas (Chapter 12). Kepler discovered that the Earth travels on an ellipse (also Chapter 12). Finally Einstein discovered that light travels on hyperbolas. That is in four dimensions, and not in Chapter 12.

	equation	vertices	foci
P	$y = ax^2 + bx + c$	$\left(-\dfrac{b}{2a},\, c - \dfrac{b^2}{4a}\right)$	$\dfrac{1}{4a}$ above vertex, also infinity
E	$\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} = 1,\ a > b$	$(a, 0)$ and $(-a, 0)$	$(c, 0)$ and $(-c, 0)$: $c = \sqrt{a^2 - b^2}$
H	$\dfrac{y^2}{a^2} - \dfrac{x^2}{b^2} = 1$	$(0, a)$ and $(0, -a)$	$(0, c)$ and $(0, -c)$: $c = \sqrt{a^2 + b^2}$

3.5 EXERCISES

Read-through questions

The graph of $y = x^2 + 2x + 5$ is a __a__. Its lowest point (the vertex) is $(x, y) = ($__b__$)$. Centering by $X = x + 1$ and $Y =$ __c__ moves the vertex to $(0, 0)$. The equation becomes $Y =$ __d__. The focus of this centered parabola is __e__. All rays coming straight down are __f__ to the focus.

The graph of $x^2 + 4y^2 = 16$ is an __g__. Dividing by __h__ leaves $x^2/a^2 + y^2/b^2 = 1$ with $a =$ __i__ and $b =$ __j__. The graph lies in the rectangle whose sides are __k__. The area is $\pi ab =$ __l__. The foci are at $x = \pm c =$ __m__. The sum of distances from the foci to a point on this ellipse is always __n__. If we rescale to $X = x/4$ and $Y = y/2$ the equation becomes __o__ and the graph becomes a __p__.

The graph of $y^2 - x^2 = 9$ is a __q__. Dividing by 9 leaves $y^2/a^2 - x^2/b^2 = 1$ with $a =$ __r__ and $b =$ __s__. On the upper branch $y \geqslant$ __t__. The asymptotes are the lines __u__. The foci are at $y = \pm c =$ __v__. The __w__ of distances from the foci to a point on this hyperbola is __x__.

All these curves are conic sections—the intersection of a __y__ and a __z__. A steep cutting angle yields a __A__. At the borderline angle we get a __B__. The general equation is $Ax^2 +$ __C__ $+ F = 0$. If $D = E = 0$ the center of the graph is at __D__. The equation $Ax^2 + Bxy + Cy^2 = 1$ gives an ellipse when __E__. The graph of $4x^2 + 5xy + 6y^2 = 1$ is a __F__.

1 The vertex of $y = ax^2 + bx + c$ is at $x = -b/2a$. What is special about this x? Show that it gives $y = c - (b^2/4a)$.

2 The parabola $y = 3x^2 - 12x$ has $x_{min} =$ _____. At this minimum, $3x^2$ is _____ as large as $12x$. Introducing $X = x - 2$ and $Y = y + 12$ centers the equation to _____.

Draw the curves 3–14 by hand or calculator or computer. Locate the vertices and foci.

3 $y = x^2 - 2x - 3$

4 $y = (x - 1)^2$

5 $4y = -x^2$

6 $4x = y^2$

7 $(x - 1)^2 + (y - 1)^2 = 1$

8 $x^2 + 9y^2 = 9$

9 $9x^2 + y^2 = 9$

10 $x^2/4 - (y - 1)^2 = 1$

11 $y^2 - 4x^2 = 1$

12 $(y - 1)^2 - 4x^2 = 1$

13 $y^2 - x^2 = 0$

14 $xy = 0$

Problems 15–20 are about parabolas, 21–34 are about ellipses, 35–41 are about hyperbolas.

15 Find the parabola $y = ax^2 + bx + c$ that goes through $(0, 0)$ and $(1, 1)$ and $(2, 12)$.

16 $y = x^2 - x$ has vertex at _____. To move the vertex to $(0, 0)$ set $X =$ _____ and $Y =$ _____. Then $Y = X^2$.

17 (a) In equation (2) change $\frac{1}{4}$ to p. Square and simplify.
(b) Locate the focus and directrix of $Y = 3X^2$. Which points are a distance 1 from the directrix and focus?

18 The parabola $y = 9 - x^2$ opens _____ with vertex at _____. Centering by $Y = y - 9$ yields $Y = -x^2$.

19 Find equations for all parabolas which
(a) open to the right with vertex at $(0, 0)$
(b) open upwards with focus at $(0, 0)$
(c) open downwards and go through $(0, 0)$ and $(1, 0)$.

20 A projectile is at $x = t$, $y = t - t^2$ at time t. Find dx/dt and dy/dt at the start, the maximum height, and an xy equation for the path.

21 Find the equation of the ellipse with extreme points at $(\pm 2, 0)$ and $(0, \pm 1)$. Then shift the center to $(1, 1)$ and find the new equation.

22 On the ellipse $x^2/a^2 + y^2/b^2 = 1$, solve for y when $x = c = \sqrt{a^2 - b^2}$. This height above the focus will be valuable in proving Kepler's third law.

23 Find equations for the ellipses with these properties:
(a) through $(5, 0)$ with foci at $(\pm 4, 0)$
(b) with sum of distances to $(1, 1)$ and $(5, 1)$ equal to 12
(c) with both foci at $(0, 0)$ and sum of distances = $2a = 10$.

24 Move a square root to the right side of equation (5) and square both sides. Then isolate the remaining square root and square again. Simplify to reach the equation of an ellipse.

25 Decide between circle-ellipse-parabola-hyperbola, based on the XY equation with $X = x - 1$ and $Y = y + 3$.

(a) $x^2 - 2x + y^2 + 6y = 6$

(b) $x^2 - 2x - y^2 - 6y = 6$

(c) $x^2 - 2x + 2y^2 + 12y = 6$

(d) $x^2 - 2x - y = 6$.

26 A tilted cylinder has equation $(x - 2y - 2z)^2 + (y - 2x - 2z)^2 = 1$. Show that the water surface at $z = 0$ is an ellipse. What is its equation and what is $B^2 - 4AC$?

27 $(4, 9/5)$ is above the focus on the ellipse $x^2/25 + y^2/9 = 1$. Find dy/dx at that point and the equation of the tangent line.

28 (a) Check that the line $xx_0 + yy_0 = r^2$ is tangent to the circle $x^2 + y^2 = r^2$ at (x_0, y_0).

(b) For the ellipse $x^2/a^2 + y^2/b^2 = 1$ show that the tangent equation is $xx_0/a^2 + yy_0/b^2 = 1$. (Check the slope.)
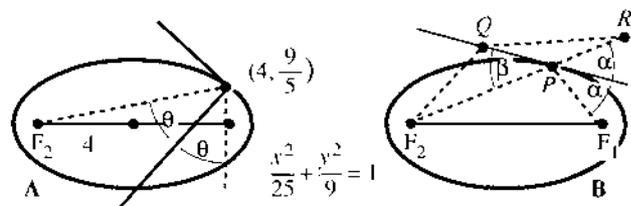
29 The slope of the normal line in Figure **A** is $s = -1/(\text{slope of tangent}) = \underline{\hspace{1cm}}$. The slope of the line from F_2 is $S = \underline{\hspace{1cm}}$. By the reflection property,

$$S = \cot 2\theta = \frac{1}{2}(\cot \theta - \tan \theta) = \frac{1}{2}\left(s - \frac{1}{s}\right).$$

Test your numbers s and S against this equation.

30 Figure **B** proves the reflecting property of an ellipse. R is the mirror image of F_1 in the tangent line; Q is any other point on the line. Deduce steps 2, 3, 4 from 1, 2, 3:

1. $PF_1 + PF_2 < QF_1 + QF_2$ (left side = 2a, Q is outside)
2. $PR + PF_2 < QR + QF_2$
3. P is on the straight line from F_2 to R
4. $\alpha = \beta$: the reflecting property is proved.

31 The ellipse $(x - 3)^2/4 + (y - 1)^2/4 = 1$ is really a $\underline{\hspace{1cm}}$ with center at $\underline{\hspace{1cm}}$ and radius $\underline{\hspace{1cm}}$. Choose X and Y to produce $X^2 + Y^2 = 1$.

32 Compute the area of a square that just fits inside the ellipse $x^2/a^2 + y^2/b^2 = 1$.

33 Rotate the axes of $x^2 + xy + y^2 = 1$ by using equation (7) with $\sin \alpha = \cos \alpha = 1/\sqrt{2}$. The $x'y'$ equation should show an ellipse.

34 What are a, b, c for the Earth's orbit around the sun?

35 Find an equation for the hyperbola with

(a) vertices $(0, \pm 1)$, foci $(0, \pm 2)$

(b) vertices $(0, \pm 3)$, asymptotes $y = \pm 2x$

(c) $(2, 3)$ on the curve, asymptotes $y = \pm x$

36 Find the slope of $y^2 - x^2 = 1$ at (x_0, y_0). Show that $yy_0 - xx_0 = 1$ goes through this point with the right slope (it has to be the tangent line).

37 If the distances from (x, y) to $(8, 0)$ and $(-8, 0)$ differ by 10, what hyperbola contains (x, y)?

38 If a cannon was heard by Napoleon and one second later by the Duke of Wellington, the cannon was somewhere on a $\underline{\hspace{1cm}}$ with foci at $\underline{\hspace{1cm}}$.

39 $y^2 - 4y$ is part of $(y - 2)^2 = \underline{\hspace{1cm}}$ and $2x^2 + 12x$ is part of $2(x + 3)^2 = \underline{\hspace{1cm}}$. Therefore $y^2 - 4y - 2x^2 - 12x = 0$ gives the hyperbola $(y - 2)^2 - 2(x + 3)^2 = \underline{\hspace{1cm}}$. Its center is $\underline{\hspace{1cm}}$ and it opens to the $\underline{\hspace{1cm}}$.

40 Following Problem 39 turn $y^2 + 2y = x^2 + 10x$ into $Y^2 = X^2 + C$ with X, Y, and C equal to $\underline{\hspace{1cm}}$.

41 Draw the hyperbola $x^2 - 4y^2 = 1$ and find its foci and asymptotes.

Problems 42–46 are about second-degree curves (conics).

42 For which A, C, F does $Ax^2 + Cy^2 + F = 0$ have no solution (empty graph)?

43 Show that $x^2 + 2xy + y^2 + 2x + 2y + 1 = 0$ is the equation (squared) of a single line.

44 Given any $\underline{\hspace{1cm}}$ points in the plane, a second-degree curve $Ax^2 + \cdots + F = 0$ goes through those points.

45 (a) When the plane $z = ax + by + c$ meets the cone $z^2 = x^2 + y^2$, eliminate z by squaring the plane equation. Rewrite in the form $Ax^2 + Bxy + Cy^2 - Dx + Ey + F = 0$.

(b) Compute $B^2 - 4AC$ in terms of a and b.

(c) Show that the plane meets the cone in an ellipse if $a^2 + b^2 < 1$ and a hyperbola if $a^2 + b^2 > 1$ (steeper).

46 The roots of $ax^2 + bx + c = 0$ also involve the special combination $b^2 - 4ac$. This quadratic equation has two real roots if $\underline{\hspace{1cm}}$ and no real roots if $\underline{\hspace{1cm}}$. The roots come together when $b^2 = 4ac$, which is the borderline case like a parabola.

3.6 Iterations $x_{n+1} = F(x_n)$

Iteration means repeating the same function. Suppose the function is $F(x) = \cos x$. Choose any starting value, say $x_0 = 1$. Take its cosine: $x_1 = \cos x_0 = .54$. *Then take the cosine of* x_1. That produces $x_2 = \cos .54 = .86$. *The iteration is* $x_{n+1} = \cos x_n$. I am in radian mode on a calculator, pressing "cos" each time. The early numbers are not important, what is important is the output after 12 or 30 or 100 steps:

EXAMPLE 1 $x_{12} = .75$, $x_{13} = .73$, $x_{14} = .74$, ..., $x_{29} = .7391$, $x_{30} = .7391$.

The goal is to explain why the x's approach $x^* = .739085$ ..... Every starting value x_0 leads to this same number x^*. *What is special about* $.7391$?

Note on iterations   Do $x_1 = \cos x_0$ and $x_2 = \cos x_1$ mean that $x_2 = \cos^2 x_0$? Absolutely not! Iteration creates a new and different function $\cos(\cos x)$. It uses the cos button, not the squaring button. The third step creates $F(F(F(x)))$. As soon as you can, iterate with $x_{n+1} = \frac{1}{2} \cos x_n$. What limit do the x's approach? Is it $\frac{1}{2}(.7931)$?

Let me slow down to understand these questions. *The central idea is expressed by the equation* $x_{n+1} = F(x_n)$. Substituting x_0 into F gives x_1. This output x_1 is the input that leads to x_2. In its turn, x_2 is the input and out comes $x_3 = F(x_2)$. This is *iteration*, and it produces the sequence $x_0, x_1, x_2, \dots$.

The x's may approach a limit x^*, depending on the function F. Sometimes x^* also depends on the starting value x_0. Sometimes there is *no* limit. Look at a second example, which does not need a calculator.

EXAMPLE 2 $x_{n+1} = F(x_n) = \frac{1}{2} x_n + 4$. Starting from $x_0 = 0$ the sequence is

$$x_1 = \frac{1}{2} \cdot 0 + 4 = 4, \quad x_2 = \frac{1}{2} \cdot 4 + 4 = 6, \quad x_3 = \frac{1}{2} \cdot 6 + 4 = 7, \quad x_4 = \frac{1}{2} \cdot 7 + 4 = 7\frac{1}{2}, \quad \dots$$

Those numbers $0, 4, 6, 7, 7\frac{1}{2}, \dots$ seem to be approaching $x^* = 8$. A computer would convince us. So will mathematics, when we see what is special about 8:

> When the x's approach x^*, the limit of $x_{n+1} = \frac{1}{2} x_n + 4$
>
> is $x^* = \frac{1}{2} x^* + 4$. This limiting equation yields $x^* = 8$.

8 is the "steady state" where *input equals output*: $8 = F(8)$. It is the *fixed point*.

If we start at $x_0 = 8$, the sequence is $8, 8, 8, \dots$. When we start at $x_0 = 12$, the sequence goes back toward 8:

$$x_1 = \frac{1}{2} \cdot 12 + 4 = 10, \quad x_2 = \frac{1}{2} \cdot 10 + 4 = 9, \quad x_3 = \frac{1}{2} \cdot 9 + 4 = 8.5, \quad \dots$$

Equation for limit: If the iterations $x_{n+1} = F(x_n)$ *converge to* x^*, *then* $x^* = F(x^*)$.

To repeat: 8 is special because it equals $\frac{1}{2} \cdot 8 + 4$. The number $.7391 \dots$ is special because it equals $\cos .7391 \dots$. *The graphs of* $y = x$ *and* $y = F(x)$ *intersect at* x^*. To explain *why* the x's converge (or why they don't) is the job of calculus.

EXAMPLE 3 $x_{n-1} = x_n^2$ has two fixed points: $0 = 0^2$ and $1 = 1^2$. Here $F(x) = x^2$.

Starting from $x_0 = \frac{1}{2}$ the sequence $\frac{1}{4}, \frac{1}{16}, \frac{1}{256}, \dots$ goes quickly to $x^* = 0$. The only approaches to $x^* = 1$ are from $x_0 = 1$ (of course) and from $x_0 = -1$. Starting from $x_0 = 2$ we get $4, 16, 256, \dots$ and *the sequence diverges to* $+\infty$.

Each limit x^* has a "*basin of attraction*." The basin contains all starting points x_0 that lead to x^*. For Examples 1 and 2, every x_0 led to $.7391$ and 8. The basins were

the whole line (that is still to be proved). Example 3 had three basins—the interval $-1 < x_0 < 1$, the two points $x_0 = \pm 1$, and all the rest. The outer basin $|x_0| > 1$ led to $\pm \infty$. I challenge you to find the limits and the basins of attraction (by calculator) for $F(x) = x - \tan x$.

In Example 3, $x^* = 0$ is **attracting**. Points near x^* move toward x^*. The fixed point $x^* = 1$ is **repelling**. Points near 1 move away. We now find the rule that decides whether x^* is attracting or repelling. *The key is the slope dF/dx at x^*.*

3J   Start from any x_0 near a fixed point $x^* = F(x^*)$:

x^* is **attracting** if $|dF/dx|$ is below 1 at x^*

x^* is **repelling** if $|dF/dx|$ is above 1 at x^*

First I will give a calculus proof. Then comes a picture of convergence, by "*cobwebs.*" Both methods throw light on this crucial test for attraction: $|dF/dx| < 1$.

First proof: Subtract $x^* = F(x^*)$ from $x_{n+1} = F(x_n)$. The difference $x_{n+1} - x^*$ is the same as $F(x_n) - F(x^*)$. This is ΔF. *The basic idea of calculus is that ΔF is close to $F'\Delta x$:*

$$x_{n+1} - x^* = F(x_n) - F(x^*) \approx F'(x^*)(x_n - x^*). \tag{1}$$

The "error" $x_n - x^*$ is multiplied by the slope dF/dx. The next error $x_{n+1} - x^*$ is smaller or larger, based on $|F'| < 1$ or $|F'| > 1$ at x^*. Every step multiplies approximately by $F'(x^*)$. *Its size controls the speed of convergence.*

In Example 1, $F(x)$ is $\cos x$ and $F'(x)$ is $-\sin x$. There is attraction to .7391 because $|\sin x^*| < 1$. In Example 2, F is $\frac{1}{2}x + 4$ and F' is $\frac{1}{2}$. There is attraction to 8. In Example 3, F is x^2 and F' is $2x$. There is superattraction to $x^* = 0$ (where $F' = 0$). There is repulsion from $x^* = 1$ (where $F' = 2$).

I admit one major difficulty. The approximation in equation (1) only holds *near* x^*. If x_0 is far away, does the sequence still approach x^*? When there are several attracting points, which x^* do we reach? This section starts with good iterations, which solve the equation $x^* = F(x^*)$ or $f(x) = 0$. At the end we discover *Newton's method*. The next section produces crazy but wonderful iterations, not converging and not blowing up. They lead to "*fractals*" and "*Cantor sets*" and "*chaos.*"

The mathematics of iterations is not finished. It may never be finished, but we are converging on the answers. Please choose a function and join in.

THE GRAPH OF AN ITERATION: COBWEBS

The iteration $x_{n+1} = F(x_n)$ involves two graphs at the same time. One is the graph of $y = F(x)$. The other is the graph of $y = x$ (the 45° line). The iteration jumps back and forth between these graphs. It is a very convenient way to see the whole process.

Example 1 was $x_{n+1} = \cos x_n$. Figure 3.19 shows the graph of $\cos x$ and the "*cobweb.*" Starting at (x_0, x_0) on the 45° line, the rule is based on $x_1 = F(x_0)$:

From (x_0, x_0) go up or down to (x_0, x_1) **on the curve.**

From (x_0, x_1) go across to (x_1, x_1) **on the 45° line.**

These steps are repeated forever. From x_1 go up to the curve at $F(x_1)$. That height is x_2. Now cross to the 45° line at (x_2, x_2). The iterations are aiming for $(x^*, x^*) = (.7391, .7391)$. This is the *crossing point* of the two graphs $y = F(x)$ and $y = x$.
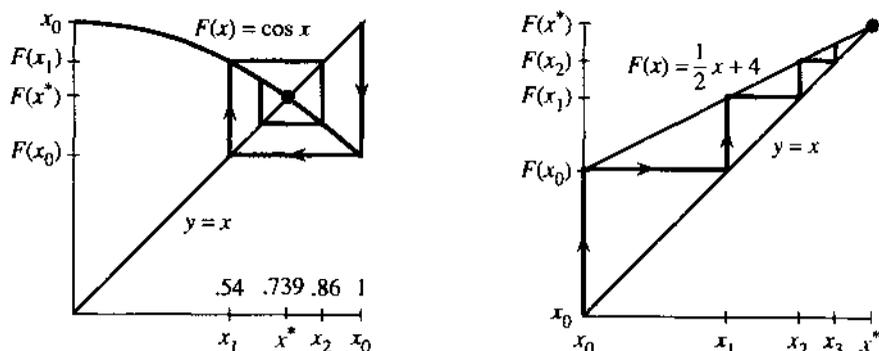
Fig. 3.19   Cobwebs go from (x_0, x_0) to (x_0, x_1) to (x_1, x_1)—line to curve to line.

Example 2 was $x_{n+1} = \frac{1}{2}x_n + 4$. Both graphs are straight lines. The cobweb is one-sided, from $(0, 0)$ to $(0, 4)$ to $(4, 4)$ to $(4, 6)$ to $(6, 6)$. Notice how y changes (vertical line) and then x changes (horizontal line). The slope of $F(x)$ is $\frac{1}{2}$, so the distance to 8 is multiplied by $\frac{1}{2}$ at every step.

Example 3 was $x_{n+1} = x_n^2$. The graph of $y = x^2$ crosses the $45°$ line at two fixed points: $0^2 = 0$ and $1^2 = 1$. Figure 3.20a starts the iteration close to 1, but it quickly goes away. This fixed point is repelling because $F'(1) = 2$. Distance from $x^* = 1$ is doubled (at the start). One path moves down to $x^* = 0$—which is *superattractive* because $F' = 0$. The path from $x_0 > 1$ diverges to infinity.

EXAMPLE 4 $F(x)$ has two attracting points x^* (a repelling x^* is always between).

Figure 3.20b shows two crossings with slope zero. The iterations and cobwebs converge quickly. In between, the graph of $F(x)$ must cross the $45°$ line from below. That requires a slope greater than one. Cobwebs diverge from this unstable point, which separates the basins of attraction. The fixed point $x = \pi$ is in a basin by itself!

Note 1   To draw cobwebs on a calculator, graph $y = F(x)$ on top of $y = x$. On a Casio, one way is to plot (x_0, x_0) and give the command L I N E : P L O T X , Y followed by E X E. Now move the cursor vertically to $y = F(x)$ and press E X E. Then move horizontally to $y = x$ and press E X E. Continue. Each step draws a line.
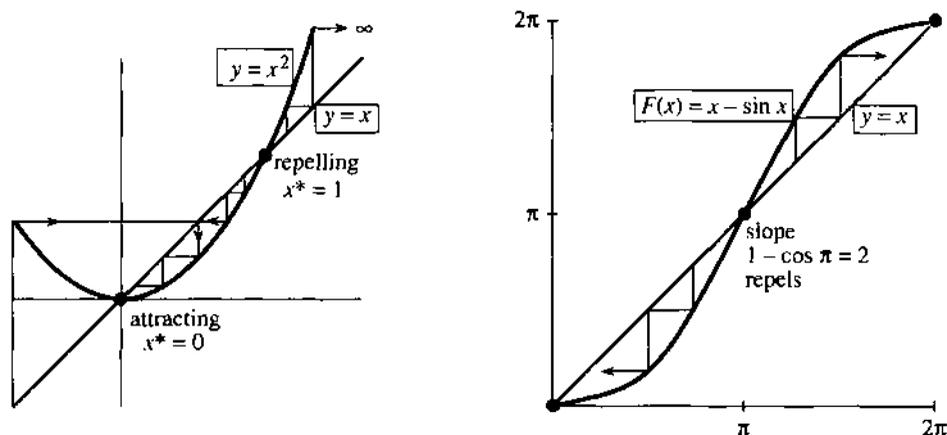
Fig. 3.20   Converging and diverging cobwebs: $F(x) = x^2$ and $F(x) = x - \sin x$.

For the TI-81 (and also the Casio) a short program produces a cobweb. Store $F(x)$ in the $Y =$ function slot Y_1. Set the range (square window or autoscaling). Run the program and answer the prompt with x_0:

```
PrgmC:COBWEB    :Disp "INITIAL X0"  :Input X  :All-Off
:Y1-On  :"X"→Y4  :Lbl 1  :X→S  :Y1→T  :Line(S,S,S,T)
:Line(S,T,T,T)  :T→X  :Pause  :Goto 1
```

Note 2 The x's approach x^* from one side when $0 < dF/dx < 1$.

Note 3 A basin of attraction can include faraway x_0's (basins can come in infinitely many pieces). This makes the problem interesting. If no fixed points are attracting, see Section 3.7 for "cycles" and "chaos."

THE ITERATION $x_{n+1} = x_n - cf(x_n)$

At this point we offer the reader a choice. One possibility is to jump ahead to the next section on "Newton's Method." That method is an iteration to solve $f(x) = 0$. The function $F(x)$ combines x_n and $f(x_n)$ and $f'(x_n)$ into an optimal formula for x_{n+1}. We will see how quickly Newton's method works (when it works). It is *the* outstanding algorithm to solve equations, and it is totally built on tangent approximations.

The other possibility is to understand (through calculus) a whole family of iterations. This family depends on a number c, which is at our disposal. *The best choice of c produces Newton's method.* I emphasize that iteration is by no means a new and peculiar idea. *It is a fundamental technique in scientific computing.*

We start by recognizing that there are many ways to reach $f(x^*) = 0$. (I write x^* for the solution.) A good algorithm may switch to Newton as it gets close. The iterations use $f(x_n)$ to decide on the next point x_{n+1}:

$$x_{n+1} = F(x_n) = x_n - cf(x_n). \tag{2}$$

Notice how F(x) is constructed from f(x)—they are different! We move f to the right side and multiply by a "preconditioner" c. *The choice of c* (or c_n, if it changes from step to step) *is absolutely critical*. The starting guess x_0 is also important—but its accuracy is not always under our control.

Suppose the x_n converge to x^*. Then the limit of equation (2) is

$$x^* = x^* - cf(x^*). \tag{3}$$

That gives $f(x^*) = 0$. If the x_n's have a limit, it solves the right equation. It is a fixed point of F (we can assume $c_n \to c \neq 0$ and $f(x_n) \to f(x^*)$). There are two key questions, and both of them are answered by the slope $F'(x^*)$:

1. How quickly does x_n approach x^* (or do the x_n diverge)?
2. What is a good choice of c (or c_n)?

EXAMPLE 5 $f(x) = ax - b$ is zero at $x^* = b/a$. The iteration $x_{n+1} = x_n - c(ax_n - b)$ intends to find b/a without actually dividing. (Early computers could not divide; they used iteration.) Subtracting x^* from both sides leaves an equation for the error:

$$x_{n+1} - x^* = x_n - x^* - c(ax_n - b).$$

Replace b by ax^*. The right side is $(1 - ca)(x_n - x^*)$. This "error equation" is

$$(\text{error})_{n+1} = (1 - ca)(\text{error})_n. \tag{4}$$

At every step the error is multiplied by $(1 - ca)$, which is F'. *The error goes to zero if* $|F'|$ *is less than* 1. The absolute value $|1 - ca|$ decides everything:

$$x_n \text{ converges to } x^* \text{ if and only if } -1 < 1 - ca < 1. \qquad (5)$$

The perfect choice (if we knew it) is $c = 1/a$, which turns the multiplier $1 - ca$ into zero. Then one iteration gives the exact answer: $x_1 = x_0 - (1/a)(ax_0 - b) = b/a$. That is the horizontal line in Figure 3.21a, converging in one step. But look at the other lines.

This example did not need calculus. Linear equations never do. The key idea is that *close to* x^* *the nonlinear equation* $f(x) = 0$ *is nearly linear*. We apply the tangent approximation. You are seeing how calculus is used, in a problem that doesn't start by asking for a derivative.

THE BEST CHOICE OF c

The immediate goal is to study the errors $x_n - x^*$. They go quickly to zero, if the multiplier is small. To understand $x_{n+1} = x_n - cf(x_n)$, subtract the equation $x^* = x^* - cf(x^*)$:

$$x_{n+1} - x^* = x_n - x^* - c(f(x_n) - f(x^*)). \qquad (6)$$

Now calculus enters. *When you see a difference of* f's *think of* df/dx. Replace $f(x_n) - f(x^*)$ by $A(x_n - x^*)$, where A stands for the slope df/dx at x^*:

$$x_{n+1} - x^* \approx (1 - cA)(x_n - x^*). \qquad (7)$$

This is the *error equation*. The new error at step $n + 1$ is approximately the old error multiplied by $m = 1 - cA$. This corresponds to $m = 1 - ca$ in the linear example. We keep returning to the basic test $|m| = |F'(x^*)| < 1$:

3K Starting near x^*, the errors $x_n - x^*$ go to zero if the multiplier has $|m| < 1$. The perfect choice is $c = 1/A = 1/f'(x^*)$. Then $m = 1 - cA = 0$.

There is only one difficulty: *We don't know* x^*. Therefore we don't know the perfect c. It depends on the slope $A = f'(x^*)$ at the unknown solution. However we can come close, by using the slope at x_n:

$$\text{Choose } c_n = 1/f'(x_n). \text{ Then } x_{n+1} = x_n - f(x_n)/f'(x_n) = F(x_n).$$

This is Newton's method. The multiplier $m = 1 - cA$ is as near to zero as we can make it. By building df/dx into $F(x)$, Newton speeded up the convergence of the iteration.
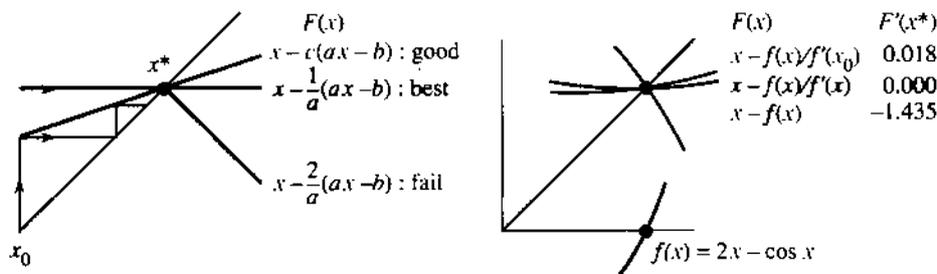
Fig. 3.21 The error multiplier is $m = 1 - cf'(x^*)$. Newton has $c = 1/f'(x_n)$ and $m \to 0$.

EXAMPLE 6 Solve $f(x) = 2x - \cos x = 0$ with different iterations (different c's).

The line $y = 2x$ crosses the cosine curve somewhere near $x = \frac{1}{2}$. The intersection point where $2x^* = \cos x^*$ has no simple formula. We start from $x_0 = \frac{1}{2}$ and iterate $x_{n+1} = x_n - c(2x_n - \cos x_n)$ with *three different choices* of c.

Take $c = 1$ or $c = 1/f'(x_0)$ or update c by Newton's rule $c_n = 1/f'(x_n)$:

	$c = 1$	$c = 1/f'(x_0)$	$c_n = 1/f'(x_n)$
$x_0 = .50$			
$x_1 =$	.38	.45063	.45062669
$x_2 =$	.55	.45019	.45018365
$x_3 =$	.30	.45018	.45018361...

The column with $c = 1$ is diverging (repelled from x^*). The second column shows convergence (attracted to x^*). The third column (Newton's method) approaches x^* so quickly that .4501836 *and seven more digits* are exact for x_3.

How does this convergence match the prediction? Note that $f'(x) = 2 + \sin x$ so $A = 2.435$. Look to see whether the actual errors $x_n - x^*$, going down each column, are multiplied by the predicted m below that column:

	$c = 1$	$c = 1/(2 + \sin \frac{1}{2})$	$c_n = 1/(2 + \sin x_n)$
$x_0 - x^* =$	0.05	$4.98 \cdot 10^{-2}$	$4.98 \cdot 10^{-2}$
$x_1 - x^* =$	-0.07	$4.43 \cdot 10^{-4}$	$4.43 \cdot 10^{-4}$
$x_2 - x^* =$	0.10	$7.88 \cdot 10^{-6}$	$3.63 \cdot 10^{-8}$
$x_3 - x^* =$	-0.15	$1.41 \cdot 10^{-7}$	$2.78 \cdot 10^{-16}$
multiplier	$m = -1.4$	$m = .018$	$m \to 0$ (Newton)

The first column shows a multiplier below -1. The errors grow at every step. Because m is negative the errors change sign—the cobweb goes outward.

The second column shows convergence with $m = .018$. It takes one genuine Newton step, then c is fixed. After n steps the error is closely proportional to $m^n = (.018)^n$— that is "*linear convergence*" with a good multiplier.

The third column shows the "*quadratic convergence*" of Newton's method. Multiplying the error by m is more attractive than ever, because $m \to 0$. In fact m itself is proportional to the error, so *at each step the error is squared*. Problem 3.8.31 will show that $(\text{error})_{n+1} \leqslant M(\text{error})_n^2$. This squaring carries us from 10^{-2} to 10^{-4} to 10^{-8} to "machine ε" in three steps. The number of correct digits is doubled at every step as Newton converges.

Note 1 The choice $c = 1$ produces $x_{n+1} = x_n - f(x_n)$. This is "successive substitution." The equation $f(x) = 0$ is rewritten as $x = x - f(x)$, and each x_n is substituted back to produce x_{n+1}. Iteration with $c = 1$ does not always fail!

Note 2 Newton's method is successive substitution for f/f', not f. Then $m \approx 0$.

Note 3 Edwards and Penney happened to choose the same example $2x = \cos x$. But they cleverly wrote it as $x_{n+1} = \frac{1}{2}\cos x_n$, which has $|F'| = |\frac{1}{2}\sin x| < 1$. This iteration fits into our family with $c = \frac{1}{2}$, and it succeeds. We asked earlier if its limit is $\frac{1}{2}(.7391)$. *No*, it is $x^* = .450....$

Note 4 The choice $c = 1/f'(x_0)$ is "*modified Newton.*" After one step of Newton's method, c is fixed. The steps are quicker, because they don't require a new $f'(x_n)$. But we need more steps. Millions of dollars are spent on Newton's method, so speed is important. In all its forms, $f(x) = 0$ is the central problem of computing.

3.6 EXERCISES

Read-through questions

$x_{n+1} = x_n^3$ describes an __a__. After one step $x_1 = $ __b__. After two steps $x_2 = F(x_1) = $ __c__. If it happens that input = output, or $x^* = $ __d__, then x^* is a __e__ point. $F = x^3$ has __f__ fixed points, at $x^* = $ __g__. Starting near a fixed point, the x_n will converge to it if __h__ < 1. That is because $x_{n+1} - x^* = F(x_n) - F(x^*) \approx$ __i__. The point is called __j__. The x_n are repelled if __k__. For $F = x^3$ the fixed points have $F' = $ __l__. The cobweb goes from (x_0, x_0) to (,) to (,) and converges to $(x^*, x^*) = $ __m__. This is an intersection of $y = x^3$ and $y = $ __n__, and it is super-attracting because __o__.

$f(x) = 0$ can be solved iteratively by $x_{n+1} = x_n - cf(x_n)$, in which case $F'(x^*) = $ __p__. Subtracting $x^* = x^* - cf(x^*)$, the error equation is $x_{n+1} - x^* \approx m($ __q__ $)$. The multiplier is $m = $ __r__. The errors approach zero if __s__. The choice $c_n = $ __t__ produces Newton's method. The choice $c = 1$ is "successive __u__" and $c = $ __v__ is modified Newton. Convergence to x^* is __w__ certain.

We have three ways to study iterations $x_{n+1} = F(x_n)$: (1) compute $x_1, x_2, ...$ from different x_0 (2) find the fixed points x^* and test $|dF/dx| < 1$ (3) draw cobwebs.

In Problems 1-8 start from $x_0 = .6$ **and** $x_0 = 2$. **Compute** $x_1, x_2, ...$ **to test convergence:**

1 $x_{n+1} = x_n^2 - \frac{1}{2}$
2 $x_{n+1} = 2x_n(1 - x_n)$
3 $x_{n+1} = \sqrt{x_n}$
4 $x_{n+1} = 1/\sqrt{x_n}$
5 $x_{n+1} = 3x_n(1 - x_n)$
6 $x_{n+1} = x_n^2 + x_n - 2$
7 $x_{n+1} = \frac{1}{2}x_n - 1$
8 $x_{n+1} = |x_n|$

9 Check dF/dx at all fixed points in Problems 1-6. Are they attracting or repelling?

10 From $x_0 = -1$ compute the sequence $x_{n+1} = -x_n^3$. Draw the cobweb with its "cycle." Two steps produce $x_{n+2} = x_n^9$, which has the fixed points _____.

11 Draw the cobwebs for $x_{n+1} = \frac{1}{2}x_n - 1$ and $x_{n+1} = 1 - \frac{1}{2}x_n$ starting from $x_0 = 2$. Rule: Cobwebs are two-sided when dF/dx is _____.

12 Draw the cobweb for $x_{n+1} = x_n^2 - 1$ starting from the periodic point $x_0 = 0$. Another periodic point is _____. Start nearby at $x_0 = .1$ to see if the iterations are attracted to 0, -1, 0, -1, ...

Solve equations 13-16 within 1% by iteration.

13 $x = \cos \frac{1}{2}x$
14 $x = \cos^2 x$
15 $x = \cos \sqrt{x}$
16 $x = 2x - 1$ (??)

17 For which numbers a does $x_{n+1} = a(x_n - x_n^2)$ converge to $x^* = 0$?

18 For which numbers a does $x_{n+1} = a(x_n - x_n^2)$ converge to $x^* = (a - 1)/a$?

19 Iterate $x_{n+1} = 4(x_n - x_n^2)$ to see chaos. Why don't the x_n approach $x^* = \frac{3}{4}$?

20 One fixed point of $F(x) = x^2 - \frac{1}{2}$ is attracting, the other is repelling. By experiment or cobwebs, find the basin of x_0's that go to the attractor.

21 (important) Find the fixed point for $F(x) = ax + s$. When is it attracting?

22 What happens in the linear case $x_{n+1} = ax_n + 4$ when $a = 1$ and when $a = -1$?

23 Starting with $1000, you spend half your money each year and a rich but foolish aunt gives you a new $1000. What is your steady state balance x^*? What is x^* if you start with a million dollars?

24 The US national debt was once $1 trillion. Inflation reduces its real value by 5% each year (so multiply by $a = .95$), but overspending adds another $100 billion. What is the steady state debt x^*?

25 $x_{n+1} = b/x_n$ has the fixed point $x^* = \sqrt{b}$. Show that $|dF/dx| = 1$ at that point--what is the sequence starting from x_0?

26 Show that both fixed points of $x_{n+1} = x_n^2 + x_n - 3$ are repelling. What do the iterations do?

27 A $5 calculator takes square roots but not cube roots. Explain why $x_{n+1} = \sqrt{2/x_n}$ converges to $\sqrt[3]{2}$.

28 Start the cobwebs for $x_{n+1} = \sin x_n$ and $x_{n+1} = \tan x_n$. In both cases $dF/dx = 1$ at $x^* = 0$. (a) Do the iterations converge? (b) Propose a theory based on F'' for cases when $F' = 1$.

Solve $f(x) = 0$ **in 29-32 by the iteration** $x_{n+1} = x_n - cf(x_n)$, **to find a** c **that succeeds and a** c **that fails.**

29 $f(x) = x^2 - 4$
30 $f(x) = x^2 - 4x + 3$
31 $f(x) = (x - 2)^9 - 1$
32 $f(x) = (1 - x)^{-1} - 3$

33 Newton's method computes a new $c = 1/f'(x_n)$ at each step. Write out the iteration formulas for $f(x) = x^3 - 2 = 0$ and $f(x) = \sin x - \frac{1}{2} = 0$.

34 Apply Problem 33 to find the first six decimals of $\sqrt[3]{2}$ and $\pi/6$.

35 By experiment find each x^* and its basin of attraction, when Newton's method is applied to $f(x) = x^2 - 5x + 4$.

36 Test Newton's method on $x^2 - 1 = 0$, starting far out at $x_0 = 10^6$. At first the error is reduced by about $m = \frac{1}{2}$. Near $x^* = 1$ the multiplier approaches $m = 0$.

37 Find the multiplier m at each fixed point of $x_{n+1} = x_n - c(x_n^2 - x_n)$. Predict the convergence for different c (to which x^*?).

38 Make a table of iterations for $c = 1$ and $c = 1/f'(x_0)$ and $c = 1/f'(x_n)$, when $f(x) = x^2 - \frac{1}{2}$ and $x_0 = 1$.

39 In the iteration for $x^2 - 2 = 0$, find dF/dx at x^*:

$$x_{n+1} = \frac{1}{2}\left(x_n + \frac{2}{x_n}\right).$$

(b) Newton's iteration has $F(x) = x - f(x)/f'(x)$. Show that $F' = 0$ when $f(x) = 0$. *The multiplier for Newton is $m = 0$.*

40 What are the solutions of $f(x) = x^2 + 2 = 0$ and why is Newton's method sure to fail? But carry out the iteration to see whether $x_n \to \infty$.

41 *Computer project* $F(x) = x - \tan x$ has fixed points where $\tan x^* = 0$. So x^* is any multiple of π. From $x_0 = 2.0$ and 1.8 and 1.9, which multiple do you reach? Test points in $1.7 < x_0 < 1.9$ to find basins of attraction to π, 2π, 3π, 4π.

Between any two basins there are basins for *every* multiple of π. And more basins between these (*a fractal*). Mark them on the line from 0 to π. Magnify the picture around $x_0 = 1.9$ (in color?).

42 Graph $\cos x$ and $\cos(\cos x)$ and $\cos(\cos(\cos x))$. Also $(\cos)^8 x$. What are these graphs approaching?

43 Graph $\sin x$ and $\sin(\sin x)$ and $(\sin)^8 x$. What are these graphs approaching? Why so slow?

3.7 Newton's Method (and Chaos)

The equation to be solved is $f(x) = 0$. Its solution x^* is the point where the graph crosses the x axis. Figure 3.22 shows x^* and a starting guess x_0. Our goal is to come as close as possible to x^*, *based on the information $f(x_0)$ and $f'(x_0)$.*

Section 3.6 reached Newton's formula for x_1 (the next guess). We now do that directly.

What do we see at x_0? The graph has height $f(x_0)$ and slope $f'(x_0)$. We know where we are, and which direction the curve is going. We don't know if the curve bends (we don't have f''). The best plan is *to follow the tangent line*, which uses all the information we have.

Newton replaces $f(x)$ by its linear approximation (= tangent approximation):

$$f(x) \approx f(x_0) + f'(x_0)(x - x_0). \tag{1}$$

We want the left side to be zero. The best we can do is to make the right side zero! The tangent line crosses the axis at x_1, while the curve crosses at x^*. The new guess x_1 comes from $f(x_0) + f'(x_0)(x_1 - x_0) = 0$. Dividing by $f'(x_0)$ and solving for x_1, this is step 1 of Newton's method:

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}. \tag{2}$$

At this new point, compute $f(x_1)$ and $f'(x_1)$—the height and slope at x_1. They give a new tangent line, which crosses at x_2. *At every step we want $f(x_{n+1}) = 0$ and we settle for* $f(x_n) + f'(x_n)(x_{n+1} - x_n) = 0$. After dividing by $f'(x_n)$, the formula for x_{n+1} is Newton's method.

> **3L**   The tangent line from x_n crosses the axis at x_{n+1}:
>
> $$\textbf{\textit{Newton's method}} \qquad x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}. \qquad (3)$$
>
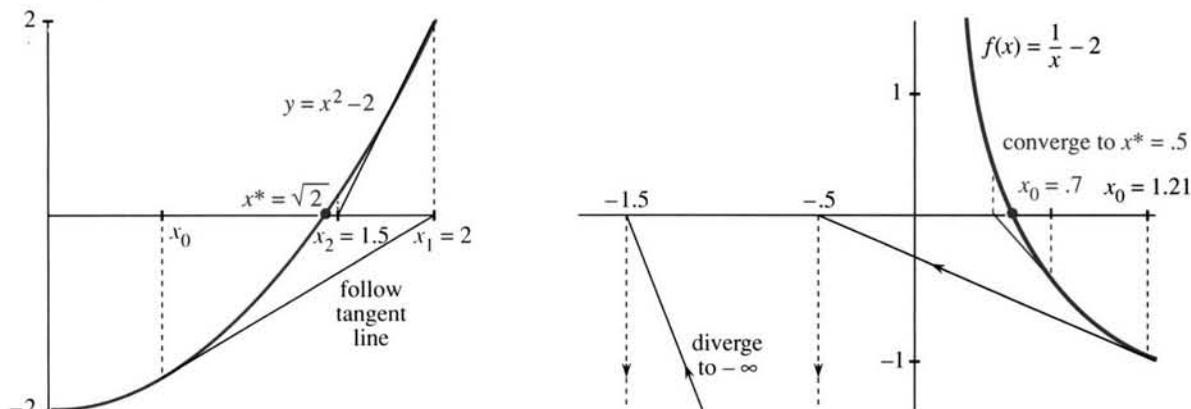> Usually this iteration $x_{n+1} = F(x_n)$ converges quickly to x^*.

Fig. 3.22   Newton's method along tangent lines from x_0 to x_1 to x_2.

Linear approximation involves three numbers. They are Δx (across) and Δf (up) and the slope $f'(x)$. If we know two of those numbers, we can estimate the third. It is remarkable to realize that calculus has now used all three calculations—they are the key to this subject:

    1. *Estimate the slope $f'(x)$ from $\Delta f / \Delta x$*   (Section 2.1)
    2. *Estimate the change Δf from $f'(x)\,\Delta x$*   (Section 3.1)
    3. *Estimate the change Δx from $\Delta f / f'(x)$*   (Newton's method)

The desired Δf is $-f(x_n)$. Formula (3) is exactly $\Delta x = -f(x_n)/f'(x_n)$.

EXAMPLE 1   (Square roots) $f(x) = x^2 - b$ is zero at $x^* = \sqrt{b}$ and also at $-\sqrt{b}$. Newton's method is a quick way to find square roots—probably built into your calculator. The slope is $f'(x_n) = 2x_n$, and formula (3) for the new guess becomes

$$x_{n+1} = x_n - \frac{x_n^2 - b}{2x_n} = x_n - \frac{1}{2}x_n + \frac{b}{2x_n}. \qquad (4)$$

This simplifies to $x_{n+1} = \frac{1}{2}(x_n + b/x_n)$. ***Guess the square root, divide into b, and average the two numbers.*** The ancient Babylonians had this same idea, without knowing functions or slopes. They iterated $x_{n+1} = F(x_n)$:

$$F(x) = \frac{1}{2}\left(x + \frac{b}{x}\right) \qquad \text{and} \qquad F'(x) = \frac{1}{2}\left(1 - \frac{b}{x^2}\right). \qquad (5)$$

The Babylonians did exactly the right thing. The slope F' is zero *at the solution*, when $x^2 = b$. That makes Newton's method converge at high speed. The convergence test is $|F'(x^*)| < 1$. Newton achieves $F'(x^*) = 0$—which is *superconvergence*.

To find $\sqrt{4}$, start the iteration $x_{n+1} = \frac{1}{2}(x_n + 4/x_n)$ at $x_0 = 1$. Then $x_1 = \frac{1}{2}(1 + 4)$:

$$x_1 = 2.5 \quad x_2 = 2.05 \quad x_3 = 2.0006 \quad x_4 = 2.000000009.$$

The wrong decimal is twice as far out at each step. *The error is squared.* Subtracting $x^* = 2$ from both sides of $x_{n+1} = F(x_n)$ gives an *error equation* which displays that square:

$$x_{n+1} - 2 = \frac{1}{2}\left(x_n + \frac{4}{x_n}\right) - 2 = \frac{1}{2x_n}(x_n - 2)^2. \tag{6}$$

This is $(error)_{n+1} \approx \frac{1}{4}(error)_n^2$. It explains the speed of Newton's method.

Remark 1   You can't start this iteration at $x_0 = 0$. The first step computes $4/0$ and blows up. Figure 3.22a shows why—the tangent line at zero is horizontal. It will never cross the axis.

Remark 2   Starting at $x_0 = -1$, Newton converges to $-\sqrt{2}$ instead of $+\sqrt{2}$. That is the other x^*. Often it is difficult to predict which x^* Newton's method will choose. Around every solution is a "basin of attraction," but other parts of the basin may be far away. Numerical experiments are needed, with many starts x_0. Finding basins of attraction was one of the problems that led to fractals.

EXAMPLE 2   *Solve* $\frac{1}{x} - a = 0$ *to find* $x^* = \frac{1}{a}$ *without dividing by a.*

Here $f(x) = (1/x) - a$. Newton uses $f'(x) = -1/x^2$. Surprisingly, we don't divide:

$$x_{n+1} = x_n - \frac{(1/x_n) - a}{-1/x_n^2} = x_n + x_n - ax_n^2. \tag{7}$$

Do these iterations converge? I will take $a = 2$ and aim for $x^* = \frac{1}{2}$. Subtracting $\frac{1}{2}$ from both sides of (7) changes the iteration into the error equation:

$$x_{n+1} = 2x_n - 2x_n^2 \quad \text{becomes} \quad x_{n+1} - \frac{1}{2} = -2(x_n - \frac{1}{2})^2. \tag{8}$$

At each step the error is squared. This is terrific if (and only if) you are close to $x^* = \frac{1}{2}$. Otherwise squaring a large error and multiplying by -2 is not good:

$$x_0 = .70 \quad x_1 = .42 \quad x_2 = .487 \quad x_3 = .4997 \quad x_4 = .4999998$$

$$x_0 = 1.21 \quad x_1 = -.5 \quad x_2 = -1.5 \quad x_3 = -7.5 \quad x_4 = -127.5$$

The algebra in Problem 18 confirms those experiments. There is fast convergence if $0 < x_0 < 1$. There is divergence if x_0 is negative or $x_0 > 1$. The tangent line goes to a negative x_1. After that Figure 3.22 shows a long trip backwards.

In the previous section we drew $F(x)$. The iteration $x_{n+1} = F(x_n)$ converged to the 45° line, where $x^* = F(x^*)$. In this section we are drawing $f(x)$. Now x^* is the point on the axis where $f(x^*) = 0$.

To repeat: It is $f(x^*) = 0$ that we aim for. But it is the slope $F'(x^*)$ that decides whether we get there. Example 2 has $F(x) = 2x - 2x^2$. The fixed points are $x^* = \frac{1}{2}$ (our solution) and $x^* = 0$ (not attractive). The slopes $F'(x^*)$ are zero (typical Newton) and 2 (typical repeller). *The key to Newton's method is* $F' = 0$ *at the solution*:

The slope of $F(x) = x - \dfrac{f(x)}{f'(x)}$ *is* $\dfrac{f(x)f''(x)}{(f'(x))^2}$. *Then* $F'(x) = 0$ *when* $f(x) = 0$.
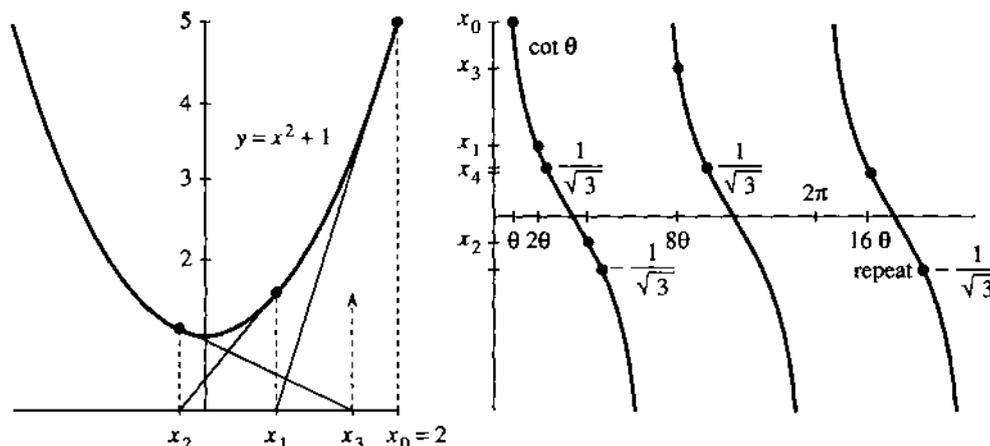
Fig. 3.23 Newton's method for $x^2 + 1 = 0$. Iteration gives $x_n = \cot 2^n \theta$.

Example B The sequence $1/\sqrt{3}$, $-1/\sqrt{3}$, $1/\sqrt{3}$ matches the cotangents of $\pi/3$, $2\pi/3$, and $4\pi/3$. This sequence *cycles forever* because $x_0 = x_2 = x_4 = \ldots$.

Example C Start with a large x_0 (a small θ). Then x_1 is about half as large (at 2θ). Eventually one of the angles 4θ, 8θ, $\ldots$ hits on a large cotangent, and the x's go far out again. *This is typical.* Examples A and B were special, when θ/π was $\frac{1}{4}$ or $\frac{1}{3}$.

What we have here is *chaos*. The x's can't converge. They are strongly repelled by all points. They are also extremely sensitive to the value of θ. After ten steps θ is multiplied by $2^{10} = 1024$. The starting angles $60°$ and $61°$ look close, but now they are different by $1024°$. If that were a multiple of $180°$, the cotangents would still be close. In fact the x_{10}'s are 0.6 and 14.

This chaos in mathematics is also seen in nature. The most familiar example is the weather, which is much more delicate than you might think. The headline "Forecasting Pushed Too Far" appeared in *Science* (1989). The article said that the snowballing of small errors destroys the forecast after six days. We can't follow the weather equations for a month—the flight of a plane can change everything. This is a revolutionary idea, that a simple rule can lead to answers that are too sensitive to compute.

We are accustomed to complicated formulas (or no formulas). We are not accustomed to innocent-looking formulas like $\cot 2^n \theta$, which are absolutely hopeless after 100 steps.

CHAOS FROM A PARABOLA

Now I get to tell you about new mathematics. First I will change the iteration $x_{n+1} = \frac{1}{2}(x_n - 1/x_n)$ into one that is even simpler. By switching from x to $z = 1/(1 + x^2)$, each new z turns out to involve only the old z and z^2:

$$z_{n+1} = 4z_n - 4z_n^2. \qquad (10)$$

This is the most famous quadratic iteration in the world. There are books about it, and Problem 28 shows where it comes from. Our formula for x_n leads to z_n:

$$z_n = \frac{1}{1 + x_n^2} = \frac{1}{1 + (\cot 2^n\theta)^2} = (\sin 2^n\theta)^2. \qquad (11)$$

The sine is just as unpredictable as the cotangent, when $2^n\theta$ gets large. The new thing is to locate this quadratic as the last member (when $a = 4$) of the family

$$z_{n+1} = az_n - az_n^2, \quad 0 \leqslant a \leqslant 4. \tag{12}$$

Example 2 happened to be the middle member $a = 2$, converging to $\frac{1}{2}$. I would like to give a brief and very optional report on this iteration, for different a's.

The general principle is to start with a number z_0 between 0 and 1, and compute $z_1, z_2, z_3, \ldots$. It is fascinating to watch the behavior change as a increases. *You can see it on your own computer.* Here we describe some things to look for. All numbers stay between 0 and 1 and they may approach a limit. That happens when a is small:

for $0 \leqslant a \leqslant 1$ the z_n approach $z^* = 0$
for $1 \leqslant a \leqslant 3$ the z_n approach $z^* = (a - 1)/a$

Those limit points are the solutions of $z = F(z)$. They are the fixed points where $z^* = az^* - a(z^*)^2$. But remember the test for approaching a limit: *The slope at z^* cannot be larger than one.* Here $F = az - az^2$ has $F' = a - 2az$. It is easy to check $|F'| \leqslant 1$ at the limits predicted above. The hard problem—sometimes impossible— is to predict what happens above $a = 3$. Our case is $a = 4$.

The z's cannot approach a limit when $|F'(z^*)| > 1$. Something has to happen, and there are at least three possibilities:

The z_n's can cycle or fill the whole interval (0, 1) *or approach a Cantor set.*

I start with a random number z_0, take 100 steps, and write down steps 101 to 105:

	$a = 3.4$	$a = 3.5$	$a = 3.8$	$a = 4.0$
$z_{101} =$	.842	.875	.336	.169
$z_{102} =$	.452	.383	.848	.562
$z_{103} =$	.842	.827	.491	.985
$z_{104} =$	.452	.501	.950	.060
$z_{105} =$	.842	.875	.182	.225

The first column is converging to a "2-cycle." It alternates between $x = .842$ and $y = .452$. Those satisfy $y = F(x)$ and $x = F(y) = F(F(x))$. If we look at a *double step* when $a = 3.4$, x and y are fixed points of the double iteration $z_{n+2} = F(F(z_n))$. When a increases past 3.45, this cycle becomes unstable.

At that point the period doubles from 2 *to* 4. With $a = 3.5$ you see a "4-cycle" in the table—it repeats after four steps. The sequence bounces from .875 to .383 to .827 to .501 *and back to* .875. This cycle must be attractive or we would not see it. But it also becomes unstable as a increases. Next comes an 8-cycle, which is stable in a little window (you could compute it) around $a = 3.55$. *The cycles are stable for shorter and shorter intervals of a's.* Those stability windows are reduced by the *Feigenbaum shrinking factor* 4.6692.... Cycles of length 16 and 32 and 64 can be seen in physical experiments, but they are all unstable before $a = 3.57$. What happens then?

The new and unexpected behavior is between 3.57 and 4. Down each line of Figure 3.24, the computer has plotted the values of z_{1001} to z_{2000}—omitting the first thousand points to let a stable period (or chaos) become established. No points appeared in the big white wedge. I don't know why. In the window for period 3, you
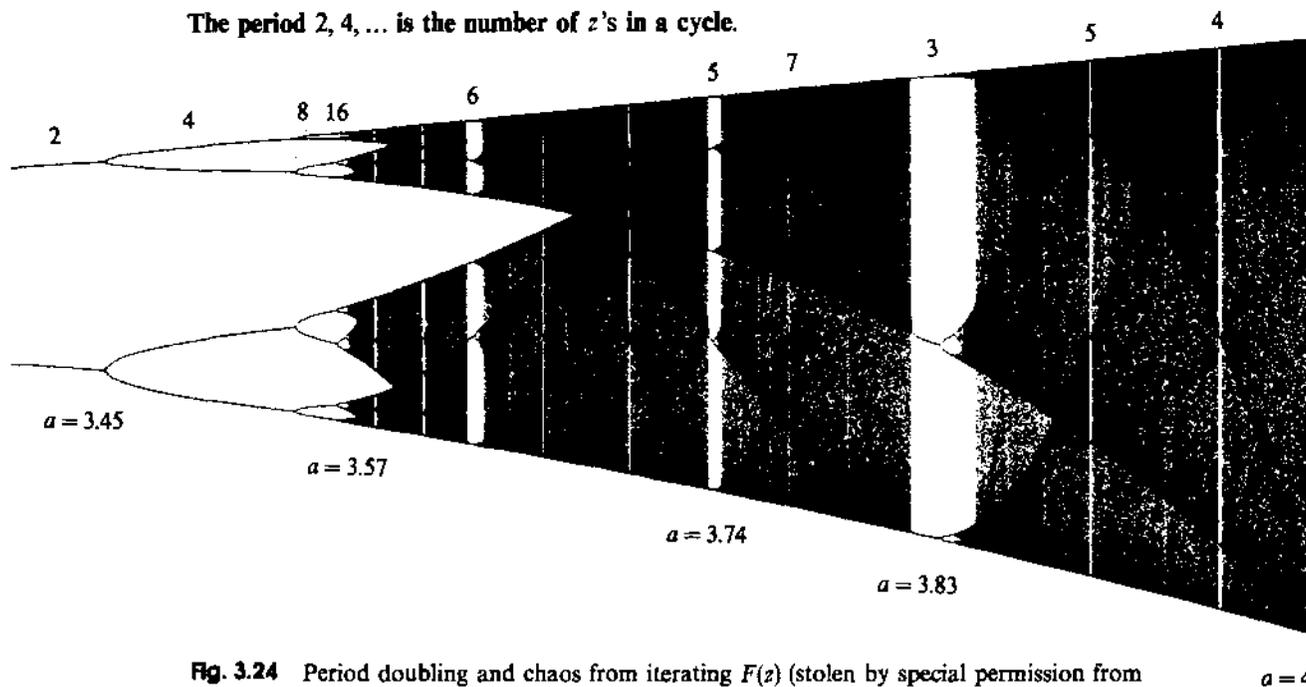
Fig. 3.24 Period doubling and chaos from iterating $F(z)$ (stolen by special permission from *Introduction to Applied Mathematics* by Gilbert Strang, Wellesley-Cambridge Press).

see only three z's. Period 3 is followed by 6, 12, 24, ... There is *period doubling* at the end of every window (including all the windows that are too small to see). You can reproduce this figure by iterating $z_{n+1} = az_n - az_n^2$ from any z_0 and plotting the results.

CANTOR SETS AND FRACTALS

I can't tell what happens at $a = 3.8$. There may be a stable cycle of some long period. The z's may come close to every point between 0 and 1. A third possibility is to approach a very thin limit set, which looks like the famous *Cantor set*:

To construct the Cantor set, divide $[0, 1]$ into three pieces and remove the open interval $(\frac{1}{3}, \frac{2}{3})$. Then remove $(\frac{1}{9}, \frac{2}{9})$ and $(\frac{7}{9}, \frac{8}{9})$ from what remains. At each step *take out the middle thirds*. The points that are left form the Cantor set.

All the endpoints $\frac{1}{3}, \frac{2}{3}, \frac{1}{9}, \frac{2}{9}, \ldots$ are in the set. So is $\frac{3}{4}$ (Problem 42). Nevertheless the lengths of the removed intervals add to 1 and the Cantor set has "measure zero." What is especially striking is its *self-similarity*: Between 0 and $\frac{1}{3}$ you see the same Cantor set three times smaller. From 0 to $\frac{1}{9}$ the Cantor set is there again, scaled down by 9. Every section, when blown up, copies the larger picture.

Fractals That self-similarity is typical of a *fractal*. There is an infinite sequence of scales. A mathematical snowflake starts with a triangle and adds a bump in the middle of each side. At every step the bumps lengthen the sides by 4/3. The final boundary is self-similar, like an infinitely long coastline.

The word "fractal" comes from *fractional dimension*. The snowflake boundary has dimension larger than 1 and smaller than 2. The Cantor set has dimension larger than 0 and smaller than 1. Covering an ordinary line segment with circles of radius r would take c/r circles. For fractals it takes c/r^D circles—and D is the dimension.
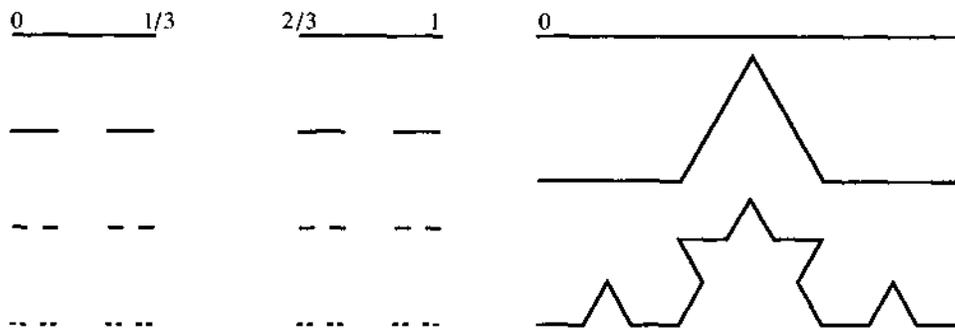
Fig. 3.25 Cantor set (middle thirds removed). Fractal snowflake (infinite boundary).

Our iteration $z_{n+1} = 4z_n - 4z_n^2$ has $a = 4$, at the end of Figure 3.24. The sequence $z_0, z_1, \ldots$ goes everywhere and nowhere. Its behavior is chaotic, and statistical tests find no pattern. For all practical purposes the numbers are random.

Think what this means in an experiment (or the stock market). If simple rules produce chaos, there is *absolutely no way* to predict the results. No measurement can ever be sufficiently accurate. The newspapers report that Pluto's orbit is chaotic—even though it obeys the law of gravity. The motion is totally unpredictable over long times. I don't know what that does for astronomy (or astrology).

The most readable book on this subject is Gleick's best-seller *Chaos: Making a New Science*. The most dazzling books are *The Beauty of Fractals* and *The Science of Fractal Images*, in which Peitgen and Richter and Saupe show photographs that have been in art museums around the world. The most original books are Mandelbrot's *Fractals* and *Fractal Geometry*. Our cover has a fractal from Figure 13.11.

We return to friendlier problems in which calculus is not helpless.

NEWTON'S METHOD VS. SECANT METHOD: CALCULATOR PROGRAMS

The hard part of Newton's method is to find df/dx. We need it for the slope of the tangent line. But calculus can approximate by $\Delta f/\Delta x$—using the values of $f(x)$ already computed at x_n and x_{n-1}.

The *secant method* follows the secant line instead of the tangent line:

$$\textbf{Secant:} \qquad x_{n+1} = x_n - \frac{f(x_n)}{(\Delta f/\Delta x)_n} \quad \text{where} \quad \left(\frac{\Delta f}{\Delta x}\right)_n = \frac{f(x_n) - f(x_{n-1})}{x_n - x_{n-1}}. \qquad (13)$$

The secant line connects the two latest points on the graph of $f(x)$. Its equation is $y - f(x_n) = (\Delta f/\Delta x)(x - x_n)$. Set $y = 0$ to find equation (13) for the new $x = x_{n+1}$, where the line crosses the axis.

Prediction: *Three* secant steps are about as good as *two* Newton steps. Both should give four times as many correct decimals: $(error) \to (error)^4$. Probably the secant method is also chaotic for $x^2 + 1 = 0$.

These Newton and secant programs are for the TI-81. Place the formula for $f(x)$ in slot Y1 and the formula for $f'(x)$ in slot Y2 on the Y = function edit screen. Answer the prompt with the initial $x_0 = X0$. The programs pause to display each approximation x_n, the value $f(x_n)$, and the difference $x_n - x_{n-1}$. Press ENTER to continue or press ON and select item 2 : Quit to break. If $f(x_n) = 0$, the programs display ROOT AT and the root x_n.

```
PrgmN:NEWTON    :Disp "ENTER FOR MORE"    PrgmS:SECANT      :Y→T
:Disp "X0="     :Disp "ON 2 TO BREAK"     :Disp "X0="       :Y1→Y
:Input X        :Disp " "                 :Input X          :Disp "ENTER FOR MORE"
:X→S            :Disp "XN FXN XN-XNM1"    :X→S              :Disp "XN FXN XN-XNM1"
:Y1→Y           :Disp X                   :Y1→T             :Disp X
:Lbl 1          :Disp Y                   :Disp "X1="       :Disp Y
:X-Y/Y2→X       :Disp D                   :Input X          :Disp D
:X-S→D          :Pause                    :Y1→Y             :Pause
:X→S            :If Y≠0                   :Lbl 1            :If Y≠0
:Y1→Y           :Goto 1                   :X-S→D            :Goto 1
                :Disp "ROOT AT"           :X→S              :Disp "ROOT AT"
                :Disp X                   :X-YD/(Y-T)→X     :Disp X
```

3.7 EXERCISES

Read-through questions

When $f(x) = 0$ is linearized to $f(x_n) + f'(x_n)(x - x_n) = 0$, the solution $x =$ __a__ is Newton's x_{n+1}. The __b__ to the curve crosses the axis at x_{n+1}, while the __c__ crosses at x^*. The errors at x_n and x_{n+1} are normally related by $(\text{error})_{n+1} \approx M$ __d__. This is __e__ convergence. The number of correct decimals __f__ at every step.

For $f(x) = x^2 - b$, Newton's iteration is $x_{n+1} =$ __g__. The x_n converge to __h__ if $x_0 > 0$ and to __i__ if $x_0 < 0$. For $f(x) = x^2 + 1$, the iteration becomes $x_{n+1} =$ __j__. This cannot converge to __k__. Instead it leads to chaos. Changing to $z = 1/(x^2 + 1)$ yields the parabolic iteration $z_{n+1} =$ __l__.

For $a \leqslant 3$, $z_{n+1} = az_n - az_n^2$ converges to a single __m__. After $a = 3$ the limit is a 2-cycle, which means __n__. Later the limit is a Cantor set, which is a one-dimensional example of a __o__. The Cantor set is self-__p__.

1 To solve $f(x) = x^3 - b = 0$, what iteration comes from Newton's method?

2 For $f(x) = (x - 1)/(x + 1)$ Newton's formula is $x_{n+1} = F(x_n) =$ _____. Solve $x^* = F(x^*)$ and find $F'(x^*)$. What limit do the x_n's approach?

3 I believe that Newton only applied his method in public to one equation $x^3 - 2x - 5 = 0$. Raphson carried the idea forward but got partial credit at best. After two steps from $x_0 = 2$, how many decimals in $x^* = 2.09455148$ are correct?

4 Show that Newton's method for $f(x) = x^{1/3}$ gives the strange formula $x_{n+1} = -2x_n$. Draw a graph to show the iterations.

5 Find x_1 if (a) $f(x_0) = 0$; (b) $f'(x_0) = 0$.

6 Graph $f(x) = x^3 - 3x - 1$ and estimate its roots x^*. Run Newton's method starting from 0, 1, $-\frac{1}{2}$, and 1.1. Experiment to decide which x_0 converge to which root.

7 Solve $x^2 - 6x + 5 = 0$ by Newton's method with $x_0 = 2.5$ and 3. Draw a graph to show which x_0 lead to which root.

8 If $f(x)$ is increasing and concave up $(f' > 0$ and $f'' > 0)$ show by a graph that Newton's method converges. From which side?

Solve 9–17 to four decimal places by Newton's method with a computer or calculator. Choose any x_0 except x^*.

9 $x^2 - 10 = 0$

10 $x^4 - 100 = 0$ (faster or slower than Problem 9?)

11 $x^2 - x = 0$ (which x_0 to which root?)

12 $x^3 - x = 0$ (which x_0 to which root?)

13 $x + 5 \cos x = 0$ (this has three roots)

14 $x + \tan x = 0$ (find two roots) (are there more?)

15 $1/(1 - x) = 2$

16 $1 + x + x^2 + x^3 + x^4 = 2$

17 $x^3 + (x + 1)^3 = 10^3$

18 (a) Show that $x_{n+1} = 2x_n - 2x_n^2$ in Example 2 is the same as $(1 - 2x_{n+1}) = (1 - 2x_n)^2$.

(b) Prove divergence if $|1 - 2x_0| > 1$. Prove convergence if $|1 - 2x_0| < 1$ or $0 < x_0 < 1$.

19 With $a = 3$ in Example 2, experiment with the Newton iteration $x_{n+1} = 2x_n - 3x_n^2$ to decide which x_0 lead to $x^* = \frac{1}{3}$.

20 Rewrite $x_{n+1} = 2x_n - ax_n^2$ as $(1 - ax_{n+1}) = (1 - ax_n)^2$. For which x_0 does the sequence $1 - ax_n$ approach zero (so $x_n \to 1/a$)?

21 What is Newton's method to find the kth root of 7? Calculate $\sqrt[7]{7}$ to 7 places.

22 Find all solutions of $x^3 = 4x - 1$ (5 decimals).

Problems 23–29 are about $x^2 + 1 = 0$ and chaos.

23 For $\theta = \pi/16$ when does $x_n = \cot 2^n\theta$ blow up? For $\theta = \pi/7$ when does $\cot 2^n\theta = \cot \theta$? (The angles $2^n\theta$ and θ differ by a multiple of π.)

24 For $\theta = \pi/9$ follow the sequence until $x_n = x_0$.

25 For $\theta = 1$, x_n never returns to $x_0 = \cot 1$. The angles 2^n and 1 never differ by a multiple of π because _____.

26 If z_0 equals $\sin^2\theta$, show that $z_1 = 4z_0 - 4z_0^2$ equals $\sin^2 2\theta$.

27 If $y = x^2 + 1$, each new y is

$$y_{n+1} = x_{n+1}^2 + 1 = \frac{1}{4}\left(x_n - \frac{1}{x_n}\right)^2 + 1.$$

Show that this equals $y_n^2/4(y_n - 1)$.

28 Turn Problem 27 upside down, $1/y_{n+1} = 4(y_n - 1)/y_n^2$, to find the quadratic iteration (10) for $z_n = 1/y_n = 1/(1 + x_n^2)$.

29 If $F(z) = 4z - 4z^2$ what is $F(F(z))$? How many solutions to $z = F(F(z))$? How many are not solutions to $z = F(z)$?

30 Apply Newton's method to $x^3 - .64x - .36 = 0$ to find the basin of attraction for $x^* = 1$. Also find a pair of points for which $y = F(z)$ and $z = F(y)$. In this example Newton does not always find a root.

31 Newton's method solves $x/(1 - x) = 0$ by $x_{n+1} = $ _____. From which x_0 does it converge? The distance to $x^* = 0$ is exactly squared.

Problems 33–41 are about competitors of Newton.

32 At a double root, Newton only converges linearly. What is the iteration to solve $x^2 = 0$?

33 To speed up Newton's method, find the step Δx from $f(x_n) + \Delta x f'(x_n) + \frac{1}{2}(\Delta x)^2 f''(x_n) = 0$. Test on $f(x) = x^2 - 1$ from $x_0 = 0$ and explain.

34 Halley's method uses $f_n + \Delta x f_n' + \frac{1}{2}\Delta x(-f_n/f_n')f_n'' = 0$. For $f(x) = x^2 - 1$ and $x_0 = 1 + \varepsilon$, show that $x_1 = 1 + O(\varepsilon^3)$ — which is *cubic* convergence.

35 Apply the secant method to $f(x) = x^2 - 4 = 0$, starting from $x_0 = 1$ and $x_1 = 2.5$. Find $\Delta f/\Delta x$ and the next point x_2 by hand. Newton uses $f'(x_1) = 5$ to reach $x_2 = 2.05$. Which is closer to $x^* = 2$?

36 Draw a graph of $f(x) = x^2 - 4$ to show the secant line in Problem 35 and the point x_2 where it crosses the axis.

Bisection method If $f(x)$ changes sign between x_0 and x_1, find its sign at the midpoint $x_2 = \frac{1}{2}(x_0 + x_1)$. Decide whether $f(x)$ changes sign between x_0 and x_2 or x_2 and x_1. Repeat on that half-length (bisected) interval. Continue. Switch to a faster method when the interval is small enough.

37 $f(x) = x^2 - 4$ is negative at $x = 1$, positive at $x = 2.5$, and negative at the midpoint $x = 1.75$. So x^* lies in what interval? Take a second step to cut the interval in half again.

38 Write a code for the bisection method. At each step print out an interval that contains x^*. The inputs are x_0 and x_1; the code calls $f(x)$. Stop if $f(x_0)$ and $f(x_1)$ have the same sign.

39 Three bisection steps reduce the interval by what factor? Starting from $x_0 = 0$ and $x_1 = 8$, take three steps for $f(x) = x^2 - 10$.

40 A direct method is to *zoom in* where the graph crosses the axis. Solve $10x^3 - 8.3x^2 + 2.295x - .21141 = 0$ by several zooms.

41 If the zoom factor is 10, then the number of correct decimals _____ for every zoom. Compare with Newton.

42 The number $\frac{3}{4}$ equals $\frac{2}{3}(1 + \frac{1}{9} + \frac{1}{81} + \cdots)$. Show that it is in the Cantor set. It survives when middle thirds are removed.

43 The solution to $f(x) = (x - 1.9)/(x - 2.0) = 0$ is $x^* = 1.9$. Try Newton's method from $x_0 = 1.5$, 2.1, and 1.95. Extra credit: Which x_0's give convergence?

44 Apply the secant method to solve $\cos x = 0$ from $x_0 = .308$.

45 Try Newton's method on $\cos x = 0$ from $x_0 = .308$. If $\cot x_0$ is exactly π, show that $x_1 = x_0 + \pi$ (and $x_2 = x_1 + \pi$). From $x_0 = .308169071$ does Newton's method ever stop?

46 Use the Newton and secant programs to solve $x^3 - 10x^2 + 22x + 6 = 0$ from $x_0 = 2$ and 1.39.

47 Newton's method for $\sin x = 0$ is $x_{n+1} = x_n - \tan x_n$. Graph $\sin x$ and three iterations from $x_0 = 2$ and $x_0 = 1.8$. Predict the result for $x_0 = 1.9$ and test. This leads to the *computer project* in Problem 3.6.41, which finds fractals.

48 Graph $Y_1(x) = 3.4(x - x^2)$ and $Y_2(x) = Y_1(Y_1(x))$ in the square window $(0, 0) \leqslant (x, y) \leqslant (1, 1)$. Then graph $Y_3(x) = Y_2(Y_1(x))$ and $Y_4, \ldots, Y_9$. The cycle is from .842 to .452.

49 Repeat Problem 48 with 3.4 changed to 2 or 3.5 or 4.

████████   **3.8 The Mean Value Theorem and l'Hôpital's Rule**   ████████

Now comes one of the cornerstones of calculus: the *Mean Value Theorem*. It connects the local picture (slope at a point) to the global picture (average slope across an interval). In other words it relates df/dx to $\Delta f/\Delta x$. Calculus depends on this connec-
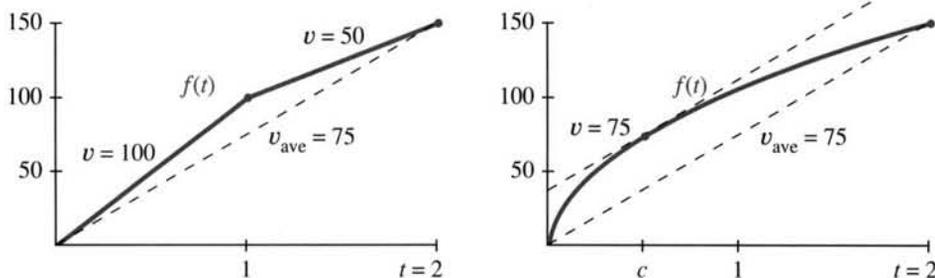
Fig. 3.26 (a) v jumps over v_{average}. (b) v equals v_{average}.

tion, which we saw first for velocities. If the average velocity is 75, is there a moment when the instantaneous velocity is 75?

Without more information, the answer to that question is *no*. The velocity could be 100 and then 50—averaging 75 but never equal to 75. If we allow a jump in velocity, it can jump right over its average. At that moment the velocity does not exist. (The distance function in Figure 3.26a has no derivative at $x = 1$.) We will take away this cheap escape by requiring a derivative at all points inside the interval.

In Figure 3.26b the distance increases by 150 when t increases by 2. There is a derivative df/dt at all interior points (but an infinite slope at $t = 0$). The average velocity is

$$\frac{\Delta f}{\Delta t} = \frac{f(2) - f(0)}{2 - 0} = \frac{150}{2} = 75.$$

The conclusion of the theorem is that $df/dt = 75$ *at some point inside the interval.* There is at least one point where $f'(c) = 75$.

This is not a constructive theorem. The value of c is not known. We don't find c, we just claim (with proof) that such a point exists.

3M Mean Value Theorem Suppose $f(x)$ is continuous in the closed interval $a \leqslant x \leqslant b$ and has a derivative everywhere in the open interval $a < x < b$. Then

$$\frac{f(b) - f(a)}{b - a} = f'(c) \text{ at some point } a < c < b. \tag{1}$$

The left side is the average slope $\Delta f / \Delta x$. It equals df/dx at c. The notation for a closed interval [with endpoints] is $[a, b]$. For an open interval (without endpoints) we write (a, b). Thus f' is defined in (a, b), and f remains continuous at a and b. A derivative is allowed at those endpoints too—but the theorem doesn't require it.

The proof is based on a special case—when $f(a) = 0$ and $f(b) = 0$. *Suppose the function starts at zero and returns to zero.* The average slope or velocity is zero. We have to prove that $f'(c) = 0$ at a point in between. This special case (keeping the assumptions on $f(x)$) is called *Rolle's theorem*.

Geometrically, if f goes away from zero and comes back, then $f' = 0$ *at the turn*.

3N Rolle's theorem Suppose $f(a) = f(b) = 0$ (zero at the ends). Then $f'(c) = 0$ at some point with $a < c < b$.

Proof At a point inside the interval where $f(x)$ reaches its maximum or minimum, df/dx must be zero. That is an acceptable point c. Figure 3.27a shows the difference between $f = 0$ (assumed at a and b) and $f' = 0$ (proved at c).

Small problem: The maximum could be reached at the ends a and b, if $f(x) < 0$ in between. At those endpoints df/dx might not be zero. But in that case the *minimum* is reached at an interior point c, which is equally acceptable. The key to our proof is that *a continuous function on* $[a, b]$ *reaches its maximum and minimum*. This is the *Extreme Value Theorem*.†

It is ironic that Rolle himself did not believe the logic behind calculus. He may not have believed his own theorem! Probably he didn't know what it meant—the language of "evanescent quantities" (Newton) and "infinitesimals" (Leibniz) was exciting but frustrating. Limits were close but never reached. Curves had infinitely many flat sides. Rolle didn't accept that reasoning, and what was really serious, he didn't accept the conclusions. The Académie des Sciences had to stop his battles (he fought against ordinary mathematicians, not Newton and Leibniz). So he went back to number theory, but his special case when $f(a) = f(b) = 0$ leads directly to the big one.
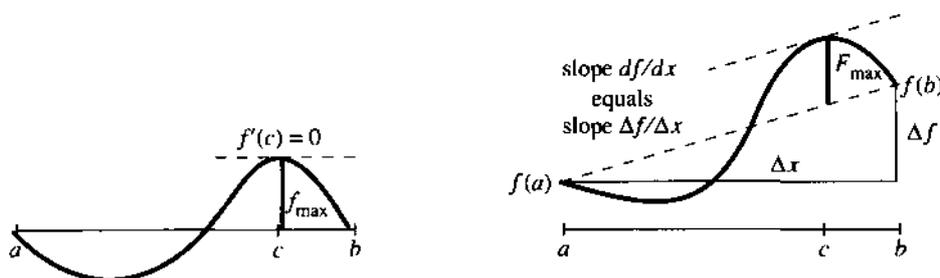

Fig. 3.27   Rolle's theorem is when $f(a) = f(b) = 0$ in the Mean Value Theorem.

Proof of the Mean Value Theorem   We are looking for a point where df/dx equals $\Delta f/\Delta x$. The idea is *to tilt the graph back to Rolle's special case* (when Δf was zero). In Figure 3.27b, the distance $F(x)$ between the curve and the dotted secant line comes from subtraction:

$$F(x) = f(x) - \left[f(a) + \frac{\Delta f}{\Delta x}(x - a) \right]. \tag{2}$$

At a and b, this distance is $F(a) = F(b) = 0$. *Rolle's theorem applies to* $F(x)$. There is an interior point where $F'(c) = 0$. At that point take the derivative of equation (2): $0 = f'(c) - (\Delta f/\Delta x)$. The desired point c is found, proving the theorem.

EXAMPLE 1   The function $f(x) = \sqrt{x}$ goes from zero at $x = 0$ to ten at $x = 100$. Its average slope is $\Delta f/\Delta x = 10/100$. The derivative $f'(x) = 1/2\sqrt{x}$ exists in the open interval $(0, 100)$, even though it blows up at the end $x = 0$. By the Mean Value Theorem there must be a point where $10/100 = f'(c) = 1/2\sqrt{c}$. That point is $c = 25$.

The truth is that nobody cares about the exact value of c. Its existence is what matters. Notice how it affects the linear approximation $f(x) \approx f(a) + f'(a)(x - a)$, which was basic to this chapter. Close becomes exact ($\approx$ becomes $=$) when f' is computed at c instead of a:

---

†If $f(x)$ doesn't reach its maximum M, then $1/(M - f(x))$ would be continuous but also approach infinity. Essential fact: *A continuous function on* $[a, b]$ *cannot approach infinity*.

---

30   The derivative at c gives an exact prediction of $f(x)$:

$$f(x) = f(a) + f'(c)(x - a). \qquad (3)$$

The Mean Value Theorem is rewritten here as $\Delta f = f'(c) \Delta x$. Now $a < c < x$.

EXAMPLE 2   The function $f(x) = \sin x$ starts from $f(0) = 0$. The linear prediction (tangent line) uses the slope $\cos 0 = 1$. The exact prediction uses the slope $\cos c$ at an unknown point between 0 and x:

$$(approximate) \ \sin x \approx x \qquad (exact) \ \sin x = (\cos c)x. \qquad (4)$$

The approximation is useful, because everything is computed at $x = a = 0$. The exact formula is interesting, because $\cos c \leqslant 1$ proves again that $\sin x \leqslant x$. The slope is below 1, so the sine graph stays below the $45°$ line.

EXAMPLE 3   *If $f'(c) = 0$ at all points in an interval then $f(x)$ is constant.*

Proof   When f' is everywhere zero, the theorem gives $\Delta f = 0$. Every pair of points has $f(b) = f(a)$. The graph is a horizontal line. That deceptively simple case is a key to the Fundamental Theorem of Calculus.

    Most applications of $\Delta f = f'(c) \Delta x$ do not end up with a number. They end up with another theorem (like this one). The goal is to connect derivatives (local) to differences (global). But the next application—*l'Hôpital's Rule*—manages to produce a number out of $0/0$.

L'HÔPITAL'S RULE

When $f(x)$ and $g(x)$ both approach zero, what happens to their ratio $f(x)/g(x)$?

$$\frac{f(x)}{g(x)} = \frac{x^2}{x} \quad \text{or} \quad \frac{\sin x}{x} \quad \text{or} \quad \frac{x - \sin x}{1 - \cos x} \quad \text{all become} \quad \frac{0}{0} \quad \text{at } x = 0.$$

Since $0/0$ is meaningless, we cannot work separately with $f(x)$ and $g(x)$. This is a "*race toward zero,*" in which two functions become small while their ratio might do anything. The problem is to find the limit of $f(x)/g(x)$.

    One such limit is already studied. *It is the derivative!* $\Delta f / \Delta x$ automatically builds in a race toward zero, whose limit is df/dx:

$$\begin{matrix} f(x) - f(a) \to 0 \\ x \ - \ a \ \to 0 \end{matrix} \quad \text{but} \quad \lim_{x \to a} \frac{f(x) - f(a)}{x - a} = f'(a). \qquad (5)$$

The idea of l'Hôpital is to use f'/g' to handle f/g. The derivative is the special case $g(x) = x - a$, with $g' = 1$. The Rule is followed by examples and proofs.

3P *l'Hôpital's Rule*   Suppose $f(x)$ and $g(x)$ both approach zero as $x \to a$. Then $f(x)/g(x)$ approaches the same limit as $f'(x)/g'(x)$, if that second limit exists:

$$\lim_{x \to a} \frac{f(x)}{g(x)} = \lim_{x \to a} \frac{f'(x)}{g'(x)}. \qquad \text{Normally this limit is} \quad \frac{f'(a)}{g'(a)}. \qquad (6)$$

This is not the quotient rule! The derivatives of $f(x)$ and $g(x)$ are taken separately. Geometrically, l'Hôpital is saying that *when functions go to zero their slopes control their size.* An easy case is $f = 6(x - a)$ and $g = 2(x - a)$. The ratio f/g is exactly $6/2$,
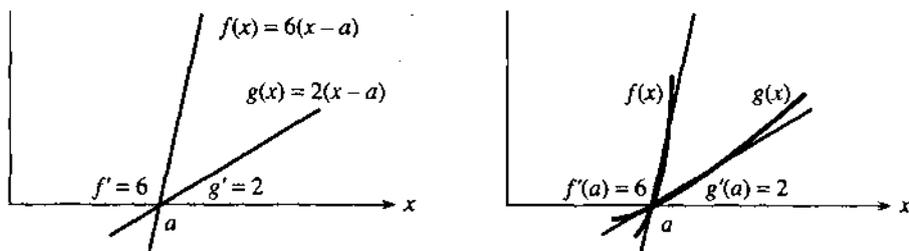
Fig. 3.28 (a) $\dfrac{f(x)}{g(x)}$ is exactly $\dfrac{f'(a)}{g'(a)} = 3$. (b) $\dfrac{f(x)}{g(x)}$ approaches $\dfrac{f'(a)}{g'(a)} = 3$.

the ratio of their slopes. Figure 3.28 shows these straight lines dropping to zero, controlled by 6 and 2.

The next figure shows the same limit 6/2, when the curves are *tangent* to the lines. That picture is the key to l'Hôpital's rule.

Generally the limit of f/g can be a finite number L or $+\infty$ or $-\infty$. (Also the limit point $x = a$ can represent a finite number or $+\infty$ or $-\infty$. We keep it finite.) The one absolute requirement is that $f(x)$ and $g(x)$ must separately approach zero—we insist on $0/0$. Otherwise there is no reason why equation (6) should be true. With $f(x) = x$ and $g(x) = x - 1$, *don't* use l'Hôpital:

$$\frac{f(x)}{g(x)} \to \frac{a}{a-1} \qquad \text{but} \qquad \frac{f'(x)}{g'(x)} = \frac{1}{1}.$$

Ordinary ratios approach $\lim f(x)$ divided by $\lim g(x)$. l'Hôpital enters only for $0/0$.

EXAMPLE 4 (an old friend) $\displaystyle \lim_{x \to 0} \frac{1 - \cos x}{x}$ equals $\displaystyle \lim_{x \to 0} \frac{\sin x}{1}$. This equals zero.

EXAMPLE 5 $\dfrac{f}{g} = \dfrac{\tan x}{\sin x}$ leads to $\dfrac{f'}{g'} = \dfrac{\sec^2 x}{\cos x}$. At $x = 0$ the limit is $\dfrac{1}{1}$.

EXAMPLE 6 $\dfrac{f}{g} = \dfrac{x - \sin x}{1 - \cos x}$ leads to $\dfrac{f'}{g'} = \dfrac{1 - \cos x}{\sin x}$. At $x = 0$ this is still $\dfrac{0}{0}$.

Solution *Apply the Rule to* f'/g'. It has the same limit as f''/g'':

$$\text{if } \frac{f}{g} \to \frac{0}{0} \text{ and } \frac{f'}{g'} \to \frac{0}{0} \text{ then compute } \frac{f''(x)}{g''(x)} = \frac{\sin x}{\cos x} \to \frac{0}{1} = 0.$$

The reason behind l'Hôpital's Rule is that the following fractions are the same:

$$\frac{f(x)}{g(x)} = \frac{f(x) - f(a)}{x - a} \bigg/ \frac{g(x) - g(a)}{x - a}. \tag{7}$$

That is just algebra; the limit hasn't happened yet. The factors $x - a$ cancel, and the numbers $f(a)$ and $g(a)$ are zero by assumption. Now take the limit on the right side of (7) as x approaches a.

What normally happens is that one part approaches f' at $x = a$. The other part approaches $g'(a)$. We hope $g'(a)$ is not zero. In this case we can divide one limit by

the other limit. That gives the "normal" answer

$$\lim_{x \to a} \frac{f(x)}{g(x)} = \text{limit of (7)} = \frac{f'(a)}{g'(a)}. \tag{8}$$

This is also l'Hôpital's answer. When $f'(x) \to f'(a)$ and separately $g'(x) \to g'(a)$, his overall limit is $f'(a)/g'(a)$. He published this rule in the first textbook ever written on differential calculus. (That was in 1696—the limit was actually discovered by his teacher Bernoulli.) Three hundred years later we apply his name to other cases permitted in (6), when f'/g' might approach a limit even if the separate parts do not.

To prove this more general form of l'Hôpital's Rule, we need a more general Mean Value Theorem. *I regard the discussion below as optional in a calculus course* (but required in a calculus book). The important idea already came in equation (8).

Remark **The basic "*indeterminate*" is** $\infty - \infty$. If $f(x)$ and $g(x)$ approach infinity, anything is possible for $f(x) - g(x)$. We could have $x^2 - x$ or $x - x^2$ or $(x + 2) - x$. Their limits are ∞ and $-\infty$ and 2.

At the next level are $0/0$ and ∞/∞ and $0 \cdot \infty$. To find the limit in these cases, try l'Hôpital's Rule. See Problem 24 when $f(x)/g(x)$ approaches ∞/∞. When $f(x) \to 0$ and $g(x) \to \infty$, apply the $0/0$ rule to $f(x)/(1/g(x))$.

The next level has 0^0 and 1^∞ and ∞^0. Those come from limits of $f(x)^{g(x)}$. If $f(x)$ approaches 0, 1, or ∞ while $g(x)$ approaches 0, ∞, or 0, we need more information. A really curious example is $x^{1/\ln x}$, which shows all three possibilities 0^0 and 1^∞ and ∞^0. This function is actually a constant! It equals e.

To go back down a level, take logarithms. Then $g(x) \ln f(x)$ returns to $0/0$ and $0 \cdot \infty$ and l'Hôpital's Rule. But logarithms and e have to wait for Chapter 6.

THE GENERALIZED MEAN VALUE THEOREM

The MVT can be extended to *two functions*. The extension is due to Cauchy, who cleared up the whole idea of limits. You will recognize the special case $g = x$ as the ordinary Mean Value Theorem.

3Q ***Generalized*** **MVT** If $f(x)$ and $g(x)$ are continuous on $[a, b]$ and differentiable on (a, b), there is a point $a < c < b$ where

$$[f(b) - f(a)]g'(c) = [g(b) - g(a)]f'(c). \tag{9}$$

The proof comes by constructing a new function that has $F(a) = F(b)$:

$$F(x) = [f(b) - f(a)]g(x) - [g(b) - g(a)]f(x).$$

The ordinary Mean Value Theorem leads to $F'(c) = 0$—which is equation (9).

Application 1 (Proof of l'Hôpital's Rule) The rule deals with $f(a)/g(a) = 0/0$. Inserting those zeros into equation (9) leaves $f(b)g'(c) = g(b)f'(c)$. Therefore

$$\frac{f(b)}{g(b)} = \frac{f'(c)}{g'(c)}. \tag{10}$$

As b approaches a, so does c. The point c is squeezed between a and b. The limit of equation (10) as $b \to a$ and $c \to a$ is l'Hôpital's Rule.

Application 2 (Error in linear approximation) Section 3.2 stated that the distance between a curve and its tangent line grows like $(x - a)^2$. Now we can prove this, and find out more. Linear approximation is

$$f(x) = f(a) + f'(a)(x - a) + \text{error } e(x). \qquad (11)$$

The pattern suggests an error involving $f''(x)$ and $(x - a)^2$. The key example $f = x^2$ shows the need for a factor $\frac{1}{2}$ (to cancel $f'' = 2$). **The error in linear approximation is**

$$e(x) = \tfrac{1}{2} f''(c)(x - a)^2 \quad \text{with} \quad a < c < x. \qquad (12)$$

Key idea Compare the error $e(x)$ to $(x - a)^2$. Both are zero at $x = a$:

$$e = f(x) - f(a) - f'(a)(x - a) \qquad e' = f'(x) - f'(a) \qquad e'' = f''(x)$$
$$g = (x - a)^2 \qquad\qquad\qquad g' = 2(x - a) \qquad\quad g'' = 2$$

The Generalized Mean Value Theorem finds a point C between a and x where $e(x)/g(x) = e'(C)/g'(C)$. This is equation (10) with different letters. After checking $e'(a) = g'(a) = 0$, apply the same theorem to $e'(x)$ and $g'(x)$. It produces a point c between a and C—certainly between a and x—where

$$\frac{e'(C)}{g'(C)} = \frac{e''(c)}{g''(c)} \quad \text{and therefore} \quad \frac{e(x)}{g(x)} = \frac{e''(c)}{g''(c)}.$$

With $g = (x - a)^2$ and $g'' = 2$ and $e'' = f''$, the equation on the right is $e(x) = \frac{1}{2} f''(c)(x - a)^2$. The error formula is proved. A very good approximation is $\frac{1}{2} f''(a)(x - a)^2$.

EXAMPLE 7 $f(x) = \sqrt{x}$ near $a = 100$: $\sqrt{102} \approx 10 + \left(\dfrac{1}{20}\right) 2 + \dfrac{1}{2}\left(\dfrac{-1}{4000}\right) 2^2.$

That last term predicts $e = -.0005$. The actual error is $\sqrt{102} - 10.1 = -.000496$.

3.8   EXERCISES

Read-through questions

The Mean Value Theorem equates the average slope $\Delta f / \Delta x$ over an __a__ $[a, b]$ to the slope df/dx at an unknown __b__. The statement is __c__. It requires $f(x)$ to be __d__ on the __e__ interval $[a, b]$, with a __f__ on the open interval (a, b). Rolle's theorem is the special case when $f(a) = f(b) = 0$, and the point c satisfies __g__. The proof chooses c as the point where f reaches its __h__.

Consequences of the Mean Value Theorem include: If $f'(x) = 0$ everywhere in an interval then $f(x) = $ __i__. The prediction $f(x) = f(a) + $ __j__ $(x - a)$ is exact for some c between a and x. The quadratic prediction $f(x) = f(a) + f'(a)(x - a) + $ __k__ $(x - a)^2$ is exact for another c. The error in $f(a) + f'(a)(x - a)$ is less than $\frac{1}{2} M(x - a)^2$ where M is the maximum of __l__.

A chief consequence is l'Hôpital's Rule, which applies when $f(x)$ and $g(x) \to $ __m__ as $x \to a$. In that case the limit of $f(x)/g(x)$ equals the limit of __n__, provided this limit exists. Normally this limit is $f'(a)/g'(a)$. If this is also $0/0$, go on to the limit of __o__.

Find all points $0 < c < 2$ where $f(2) - f(0) = f'(c)(2 - 0)$.

1 $f(x) = x^3$ 2 $f(x) = \sin \pi x$

3 $f(x) = \tan 2\pi x$ 4 $f(x) = 1 + x + x^2$

5 $f(x) = (x - 1)^{10}$ 6 $f(x) = (x - 1)^9$

In 7–10 show that no point c yields $f(1) - f(-1) = f'(c)(2)$. Explain why the Mean Value Theorem fails to apply.

7 $f(x) = |x - \frac{1}{2}|$ 8 $f(x) = $ unit step function

9 $f(x) = |x|^{1/2}$ 10 $f(x) = 1/x^2$

11 Show that $\sec^2 x$ and $\tan^2 x$ have the same derivative, and draw a conclusion about $f(x) = \sec^2 x - \tan^2 x$.

12 Show that $\csc^2 x$ and $\cot^2 x$ have the same derivative and find $f(x) = \csc^2 x - \cot^2 x$.

Evaluate the limits in 13–22 by l'Hôpital's Rule.

13 $\displaystyle\lim_{x \to 3} \frac{x^2 - 9}{x - 3}$ 14 $\displaystyle\lim_{x \to 3} \frac{x^2 - 9}{x + 3}$

15 $\lim\limits_{x\to 0} \dfrac{(1+x)^{-2}-1}{x}$

16 $\lim\limits_{x\to 0} \dfrac{\sqrt{1-\cos x}}{x}$

17 $\lim\limits_{x\to \pi} \dfrac{x-\pi}{\sin x}$

18 $\lim\limits_{x\to 1} \dfrac{x-1}{\sin x}$

19 $\lim\limits_{x\to 0} \dfrac{(1+x)^r-1}{x}$

20 $\lim\limits_{x\to 0} \dfrac{(1+x)^r-1-nx}{x^2}$

21 $\lim\limits_{x\to 0} \dfrac{\sin x-\tan x}{x^3}$

22 $\lim\limits_{x\to 0} \dfrac{\sqrt{1+x}-\sqrt{1-x}}{x}$

23 For $f = x^2 - 4$ and $g = x + 2$, the ratio f'/g' approaches 4 as $x \to 2$. What is the limit of $f(x)/g(x)$? What goes wrong in l'Hôpital's Rule?

24 *l'Hôpital's Rule still holds for* $f(x)/g(x) \to \infty/\infty$: L **is**

$$\lim \frac{f(x)}{g(x)} = \lim \frac{1/g(x)}{1/f(x)} = \lim \frac{g'(x)/g^2(x)}{f'(x)/f^2(x)} = L^2 \lim \frac{g'(x)}{f'(x)}.$$

Then L equals $\lim [f'(x)/g'(x)]$ if this limit exists. Where did we use the rule for $0/0$? What other limit rule was used?

25 Compute $\lim\limits_{x\to 0} \dfrac{1+(1/x)}{1-(1/x)}$.

26 Compute $\lim\limits_{x\to \infty} \dfrac{x^2+x}{2x^2}$.

27 Compute $\lim\limits_{x\to \infty} \dfrac{x+\cos x}{x+\sin x}$ by common sense. Show that l'Hôpital gives no answer.

28 Compute $\lim\limits_{x\to 0} \dfrac{\csc x}{\cot x}$ by common sense or trickery.

29 The Mean Value Theorem applied to $f(x) = x^3$ guarantees that some number c between 1 and 4 has a certain property. Say what the property is and find c.

30 If $|df/dx| \leqslant 1$ at all points, prove this fact:
$$|f(x) - f(y)| \leqslant |x - y| \text{ at all } x \text{ and } y.$$

31 The error in Newton's method is squared at each step: $|x_{n+1} - x^*| \leqslant M|x_n - x^*|^2$. The proof starts from $0 = f(x^*) = f(x_n) + f'(x_n)(x^* - x_n) + \frac{1}{2}f''(c)(x^* - x_n)^2$. Divide by $f'(x_n)$, recognize x_{n+1}, and estimate M.

32 (Rolle's theorem backward) Suppose $f'(c) = 0$. Are there necessarily two points around c where $f(a) = f(b)$?

33 Suppose $f(0) = 0$. If $f(x)/x$ has a limit as $x \to 0$, that limit is better known to us as _____. L'Hôpital's Rule looks instead at the limit of _____.

Conclusion from l'Hôpital: The limit of $f'(x)$, if it exists, agrees with $f'(0)$. Thus $f'(x)$ cannot have a "removable _____."

34 It is possible that $f'(x)/g'(x)$ has *no* limit but $f(x)/g(x) \to L$. This is why l'Hôpital included an "if."

(a) Find L as $x \to 0$ when $f(x) = x^2 \cos(1/x)$ and $g(x) = x$. Remember that cosines are below 1.

(b) From the formula $f'(x) = \sin(1/x) + 2x\cos(1/x)$ show that f'/g' has no limit as $x \to 0$.

35 Stein's calculus book asks for the limiting ratio of $f(x) =$ triangular area ABC to $g(x) =$ curved area ABC.
(a) Guess the limit of f/g as the angle x goes to zero.
(b) Explain why $f(x)$ is $\frac{1}{2}(\sin x - \sin x \cos x)$ and $g(x)$ is $\frac{1}{2}(x - \sin x \cos x)$. (c) Compute the true limit of $f(x)/g(x)$.
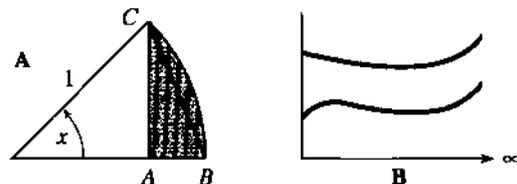
36 If you drive 3000 miles from New York to L.A. in 100 hours (sleeping and eating and going backwards are allowed) then at some moment your speed is _____.

37 As $x \to \infty$ l'Hôpital's Rule still applies. The limit of $f(x)/g(x)$ equals the limit of $f'(x)/g'(x)$, if that limit exists. What is the limit as the graphs become parallel in Figure B?

38 Prove that $f(x)$ is increasing when its slope is positive: *If $f'(c) > 0$ at all points c, then $f(b) > f(a)$ at all pairs of points $b > a$.*