CALCULUS

Gilbert Strang
Massachusetts Institute of Technology

WELLESLEY-CAMBRIDGE PRESS
Box 82-279 Wellesley MA 02181

Contents

Contents

CHAPTER 1

Introduction to Calculus

1.1 Velocity and Distance

The right way to begin a calculus book is with calculus. This chapter will jump directly into the two problems that the subject was invented to solve. You will see what the questions are, and you will see an important part of the answer. There are plenty of good things left for the other chapters, so why not get started?

The book begins with an example that is familiar to everybody who drives a car. It is calculus in action—the driver sees it happening. The example is the relation between the *speedometer* and the *odometer*. One measures the speed (or ***velocity***); the other measures the ***distance traveled***. We will write v for the velocity, and f for how far the car has gone. The two instruments sit together on the dashboard:

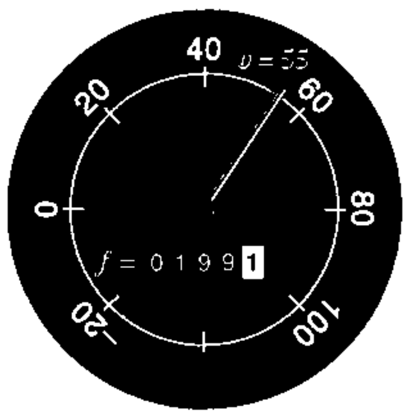


Fig. 1.1 Velocity v and total distance f (at one instant of time).

Notice that the units of measurement are different for v and f. The distance f is measured in kilometers or miles (it is easier to say miles). The velocity v is measured in km/hr or *miles per hour*. A unit of *time* enters the velocity but not the distance. Every formula to compute v from f will have f divided by time.

The central question of calculus is the relation between v and f.

Can you find v if you know f, and vice versa, and how? If we know the velocity over the whole history of the car, we should be able to compute the total distance traveled. In other words, if the speedometer record is complete but the odometer is missing, its information could be recovered. One way to do it (without calculus) is to put in a new odometer and drive the car all over again at the right speeds. That seems like a hard way; calculus may be easier. But the point is that *the information is there.* If we know everything about v, there must be a method to find f.

What happens in the opposite direction, when f is known? If you have a complete record of distance, could you recover the complete velocity? In principle you could drive the car, repeat the history, and read off the speed. Again there must be a better way.

The whole subject of calculus is built on the relation between v and f. The question we are raising here is not some kind of joke, after which the book will get serious and the mathematics will get started. On the contrary, *I am serious now*—and the mathematics has already started. We need to know how to find the velocity from a record of the distance. (That is called ***differentiation***, and it is the central idea of ***differential calculus***.) We also want to compute the distance from a history of the velocity. (That is ***integration***, and it is the goal of ***integral calculus***.)

Differentiation goes from f to v; integration goes from v to f. We look first at examples in which these pairs can be computed and understood.

CONSTANT VELOCITY

Suppose the velocity is fixed at $v = 60$ (miles per hour). Then f increases at this constant rate. After two hours the distance is $f = 120$ (miles). After four hours $f = 240$ and after t hours $f = 60t$. We say that f increases ***linearly*** with time—its graph is a straight line.

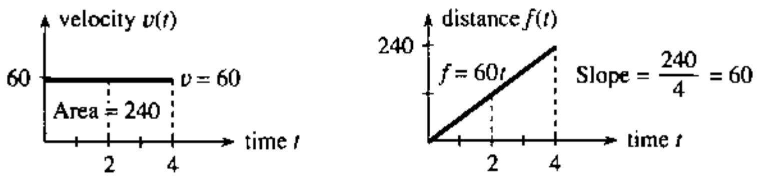


Fig. 1.2 Constant velocity $v = 60$ and linearly increasing distance $f = 60t$.

Notice that this example starts the car at full velocity. No time is spent picking up speed. (The velocity is a "step function.") Notice also that the distance starts at zero; the car is new. Those decisions make the graphs of v and f as neat as possible. One is the horizontal line $v = 60$. The other is the sloping line $f = 60t$. This v, f, t relation needs algebra but not calculus:

if v is constant and f starts at zero then $f = vt$.

The opposite is also true. When f increases linearly, v is constant. *The division by time gives the slope.* The distance is $f_1 = 120$ miles when the time is $t_1 = 2$ hours. Later $f_2 = 240$ at $t_2 = 4$. At both points, the ratio f/t is 60 miles/hour. Geometrically, ***the velocity is the slope of the distance graph***:

$$\text{slope} = \frac{\text{change in distance}}{\text{change in time}} = \frac{vt}{t} = v.$$

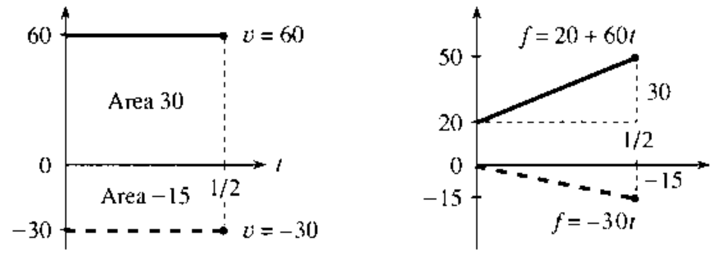


Fig. 1.3 Straight lines $f = 20 + 60t$ (slope 60) and $f = -30t$ (slope -30).

The slope of the f-graph gives the v-graph. Figure 1.3 shows two more possibilities:

1. The distance starts at 20 instead of 0. The distance formula changes from $60t$ to $20 + 60t$. The number 20 cancels when we compute *change* in distance—so the slope is still 60.
2. When v is *negative*, the graph of f goes *downward*. The car goes backward and the slope of $f = -30t$ is $v = -30$.

I don't think speedometers go below zero. But driving backwards, it's not that safe to watch. If you go fast enough, Toyota says they measure "absolute values"—the speedometer reads $+30$ when the velocity is -30. For the odometer, as far as I know it just stops. It should go backward.†

VELOCITY *vs.* DISTANCE: SLOPE *vs.* AREA

How do you compute f from v? The point of the question is to see *$f = vt$ on the graphs.* We want to start with the graph of v and discover the graph of f. Amazingly, the opposite of slope is *area.*

The distance f is the area under the v-graph. When v is constant, the region under the graph is a rectangle. Its height is v, its width is t, and its area is v times t. This is *integration*, to go from v to f by computing the area. We are glimpsing two of the central facts of calculus.

> **1A** *The slope of the f-graph gives the velocity v. The area under the v-graph gives the distance f.*

That is certainly not obvious, and I hesitated a long time before I wrote it down in this first section. The best way to understand it is to look first at more examples. The whole point of calculus is to deal with velocities that are *not* constant, and from now on v has several values.

EXAMPLE (***Forward and back***) There is a motion that you will understand right away. The car goes forward with velocity V, and comes back at the same speed. To say it more correctly, the *velocity in the second part is* $-V$. If the forward part lasts until $t = 3$, and the backward part continues to $t = 6$, ***the car will come back where it started.*** The total distance after both parts will be $f = 0$.

†This actually happened in *Ferris Bueller's Day Off*, when the hero borrowed his father's sports car and ran up the mileage. At home he raised the car and drove in reverse. I forget if it worked.

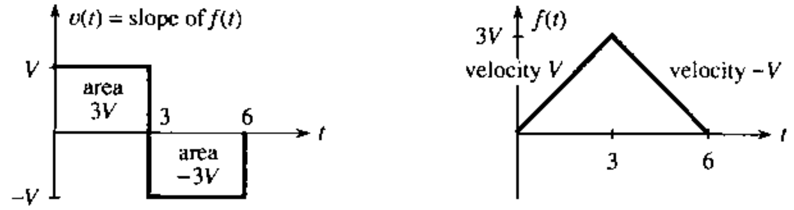


Fig. 1.4 Velocities $+V$ and $-V$ give motion forward and back, ending at $f(6)=0$.

The v-graph shows velocities $+V$ and $-V$. The distance starts up with slope $+V$ and reaches $f=3V$. Then the car starts backward. The distance goes down with slope $-V$ and returns to $f=0$ at $t=6$.

Notice what that means. The total area "under" the v-graph is zero! A negative velocity makes the distance graph go *downward* (*negative slope*). The car is moving backward. ***Area below the axis in the v-graph is counted as negative.***

FUNCTIONS

This forward-back example gives practice with a crucially important idea—the concept of a "***function.***" We seize this golden opportunity to explain functions:

The number $v(t)$ is the value of the function v at the time t.

The time t is the ***input*** to the function. The velocity $v(t)$ at that time is the ***output.*** Most people say "*v of t*" when they read $v(t)$. The number "*v of* 2" is the velocity when $t=2$. The forward-back example has $v(2)=+V$ and $v(4)=-V$. The function contains the whole history, like a memory bank that has a record of v at each t.

It is simple to convert forward-back motion into a formula. Here is $v(t)$:

$$v(t)=\begin{cases} +V & \text{if} \quad 0<t<3 \\ \;\;? & \text{if} \quad t=3 \\ -V & \text{if} \quad 3<t<6 \end{cases}$$

The right side contains the instructions for finding $v(t)$. The input t is converted into the output $+V$ or $-V$. The velocity $v(t)$ depends on t. In this case the function is "discontinuous," because the needle jumps at $t=3$. *The velocity is not defined at that instant*. There is no $v(3)$. (You might argue that v is zero at the jump, but that leads to trouble.) The graph of f has a corner, and we can't give its slope.

The problem also involves a second function, namely the distance. The principle behind $f(t)$ is the same: ***$f(t)$ is the distance at time t.*** It is the net distance forward, and again the instructions change at $t=3$. In the forward motion, $f(t)$ equals Vt as before. In the backward half, a calculation is built into the formula for $f(t)$:

$$f(t)=\begin{cases} Vt & \text{if} \quad 0\leqslant t\leqslant 3 \\ V(6-t) & \text{if} \quad 3\leqslant t\leqslant 6 \end{cases}$$

At the switching time the right side gives two instructions (one on each line). This would be bad except that they agree: $f(3)=3V$.† The distance function is "*con-*

†A function is only allowed *one value* $f(t)$ or $v(t)$ at each time t.

tinuous." There is no jump in f, even when there is a jump in v. After $t = 3$ the distance decreases because of $-Vt$. At $t = 6$ the second instruction correctly gives $f(6) = 0$.

Notice something more. The functions were given by graphs before they were given by formulas. The graphs tell you f and v at every time t—sometimes more clearly than the formulas. The values $f(t)$ and $v(t)$ can also be given by tables or equations or a set of instructions. (In some way all functions are instructions—the function tells how to find f at time t.) Part of knowing f is knowing all its inputs and outputs—its ***domain*** and ***range***:

The domain of a function is the set of inputs. The range is the set of outputs.

The domain of f consists of all times $0 \leqslant t \leqslant 6$. The range consists of all distances $0 \leqslant f(t) \leqslant 3V$. (The range of v contains only the two velocities $+V$ and $-V$.) We mention now, and repeat later, that every "linear" function has a formula $f(t) = vt + C$. Its graph is a line and v is the slope. The constant C moves the line up and down. It adjusts the line to go through any desired starting point.

SUMMARY: MORE ABOUT FUNCTIONS

May I collect together the ideas brought out by this example? We had two functions v and f. One was *velocity*, the other was *distance*. Each function had a *domain*, and a *range*, and most important a *graph*. For the f-graph we studied the slope (which agreed with v). For the v-graph we studied the area (which agreed with f). Calculus produces functions in pairs, and the best thing a book can do early is to show you more of them.

in	***input*** t	$\rightarrow$	***function*** f	$\rightarrow$	***output*** $f(t)$	***in***
the	***input*** 2	$\rightarrow$	***function*** v	$\rightarrow$	***output*** $v(2)$	***the***
domain	***input*** 7	$\rightarrow$	$f(t) = 2t + 6$	$\rightarrow$	$f(7) = 20$	***range***

Note about the definition of a function. The idea behind the symbol $f(t)$ is absolutely crucial to mathematics. Words don't do it justice! By definition, a function is a "rule" that assigns one member of the range to each member of the domain. Or, a function is a set of pairs $(t, f(t))$ with no t appearing twice. (These are "ordered pairs" because we write t before $f(t)$.) Both of those definitions are correct—but somehow they are too passive.

In practice what matters is the active part. The number $f(t)$ is produced from the number t. We read a graph, plug into a formula, solve an equation, run a computer program. The input t is "mapped" to the output $f(t)$, which changes as t changes. Calculus is about the ***rate of change***. This rate is our other function v.

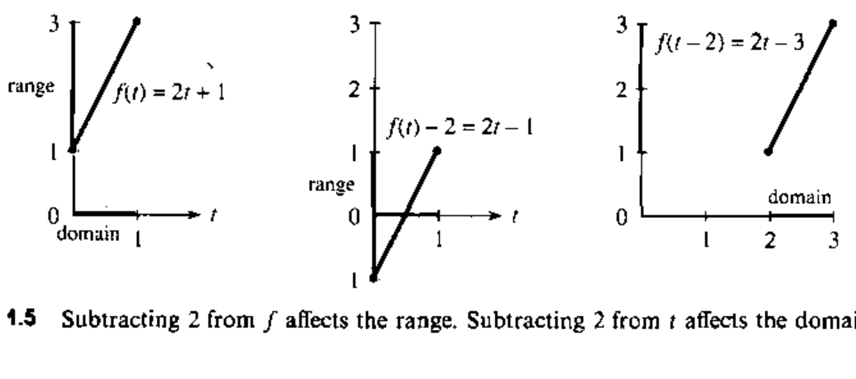


Fig. 1.5 Subtracting 2 from f affects the range. Subtracting 2 from t affects the domain.

It is quite hard at the beginning, and not automatic, to see the difference between $f(t)-2$ and $f(t-2)$. Those are both new functions, created out of the original $f(t)$. In $f(t)-2$, we subtract 2 from all the distances. That moves the whole graph *down*. In $f(t-2)$, we subtract 2 from the time. That moves the graph over *to the right*. Figure 1.5 shows both movements, starting from $f(t)=2t+1$. The formula to find $f(t-2)$ is $2(t-2)+1$, which is $2t-3$.

A graphing calculator also moves the graph, when you change the viewing window. You can pick any rectangle $A \leqslant t \leqslant B$, $C \leqslant f(t) \leqslant D$. The screen shows that part of the graph. But on the calculator, *the function $f(t)$ remains the same*. It is the axes that get renumbered. In our figures the axes stay the same and the function is changed.

There are two more basic ways to change a function. (We are always creating new functions—that is what mathematics is all about.) Instead of subtracting or adding, we can *multiply* the distance by 2. Figure 1.6 shows $2f(t)$. And instead of shifting the time, we can *speed it up*. The function becomes $f(2t)$. Everything happens twice as fast (and takes half as long). On the calculator those changes correspond to a "***zoom***"—on the f axis or the t axis. We soon come back to zooms.

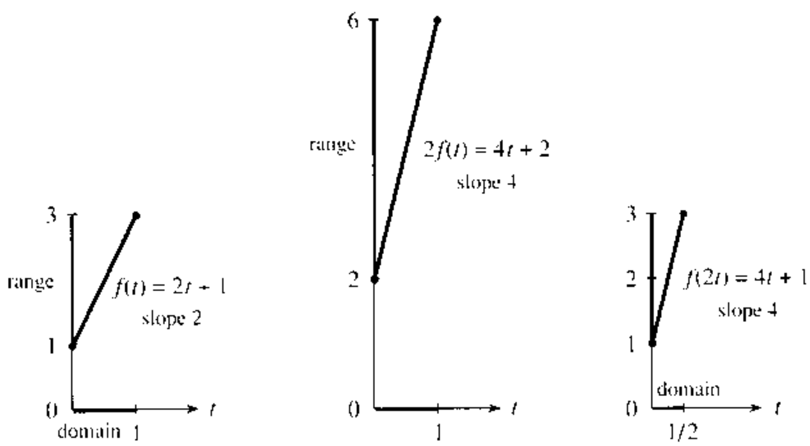


Fig. 1.6 Doubling the distance or speeding up the time doubles the slope.

1.1 EXERCISES

Each section of the book contains read-through questions. They allow you to outline the section yourself—more actively than reading a summary. This is probably the best way to remember the important ideas.

Starting from $f(0)=0$ at constant velocity v, the distance function is $f(t)=$ __a__. When $f(t)=55t$ the velocity is $v=$ __b__. When $f(t)=55t+1000$ the velocity is still __c__ and the starting value is $f(0)=$ __d__. In each case v is the __e__ of the graph of f. When __f__ is negative, the graph of __g__ goes downward. In that case area in the v-graph counts as __h__.

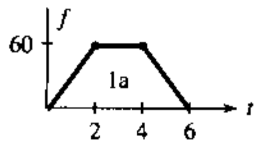
Forward motion from $f(0)=0$ to $f(2)=10$ has $v=$ __i__. Then backward motion to $f(4)=0$ has $v=$ __j__. The distance function is $f(t)=5t$ for $0 \leqslant t \leqslant 2$ and then $f(t)=$ __k__ (not $-5t$). The slopes are __l__ and __m__. The distance $f(3)=$ __n__. The area under the v-graph up to time 1.5 is __o__. The domain of f is the time interval __p__, and the range is the distance interval __q__. The range of $v(t)$ is only __r__.

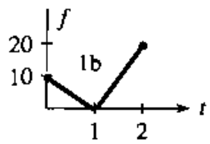
The value of $f(t)=3t+1$ at $t=2$ is $f(2)=$ __s__. The value 19 equals $f($__t__$)$. The difference $f(4)-f(1)=$ __u__. That is the change in distance, when $4-1$ is the change in __v__. The ratio of those changes equals __w__, which is the __x__ of the graph. The formula for $f(t)+2$ is $3t+3$ whereas $f(t+2)$ equals __y__. Those functions have the same __z__ as f: the graph of $f(t)+2$ is shifted __A__ and $f(t+2)$ is shifted __B__. The formula for $f(5t)$ is __C__. The formula for $5f(t)$ is __D__. The slope has jumped from 3 to __E__.

The set of inputs to a function is its __F__. The set of outputs is its __G__. The functions $f(t) = 7 + 3(t-2)$ and $f(t) = vt + C$ are __H__. Their graphs are __I__ with slopes equal to __J__ and __K__. They are the same function, if $v =$ __L__ and $C =$ __M__.

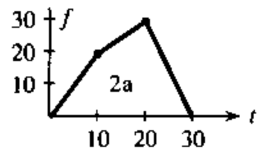
Draw the velocity graph that goes with each distance graph.

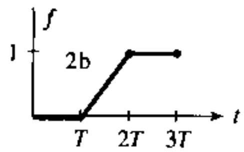
1





2





3 Write down three-part formulas for the velocities $v(t)$ in Problem 2, starting from $v(t) = 2$ for $0 < t < 10$.

4 The distance in 1b starts with $f(t) = 10 - 10t$ for $0 \leqslant t \leqslant 1$. Give a formula for the second part.

5 In the middle of graph 2a find $f(15)$ and $f(12)$ and $f(t)$.

6 In graph 2b find $f(1.4T)$. If $T = 3$ what is $f(4)$?

7 Find the *average speed* between $t = 0$ and $t = 5$ in graph 1a. What is the speed at $t = 5$?

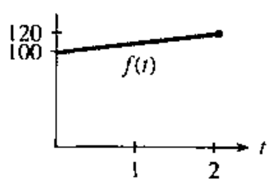
8 What is the average speed between $t = 0$ and $t = 2$ in graph 1b? The average speed is zero between $t = \frac{1}{2}$ and $t =$ ________.

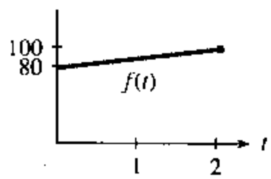
9 (recommended) A car goes at speed $v = 20$ into a brick wall at distance $f = 4$. Give two-part formulas for $v(t)$ and $f(t)$ (before and after), and draw the graphs.

10 Draw any reasonable graphs of $v(t)$ and $f(t)$ when

(a) the driver backs up, stops to shift gear, then goes fast;
(b) the driver slows to 55 for a police car;
(c) in a rough gear change, the car accelerates in jumps;
(d) the driver waits for a light that turns green.

11 Your bank account earns simple interest on the opening balance $f(0)$. What are the interest rates per year?

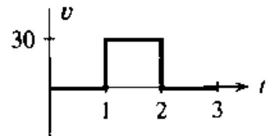




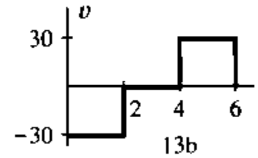
12 The earth's population is growing at $v = 100$ million a year, starting from $f = 5.2$ billion in 1990. Graph $f(t)$ and find $f(2000)$.

Draw the distance graph that goes with each velocity graph. Start from $f = 0$ at $t = 0$ and mark the distance.

13

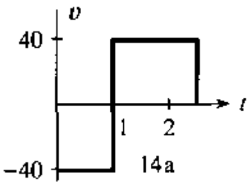


13a

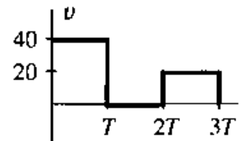


13b

14



14a



14b

15 Write down formulas for $v(t)$ in Problem 14, starting with $v = -40$ for $0 < t < 1$. Find the average velocities to $t = 2.5$ and $t = 3T$.

16 Give 3-part formulas for the areas $f(t)$ under $v(t)$ in 13.

17 The distance in 14a starts with $f(t) = -40t$ for $0 \leqslant t \leqslant 1$. Find $f(t)$ in the other part, which passes through $f = 0$ at $t = 2$.

18 Draw the velocity and distance graphs if $v(t) = 8$ for $0 < t < 2$, $f(t) = 20 + t$ for $2 \leqslant t \leqslant 3$.

19 Draw rough graphs of $y = \sqrt{x}$ and $y = \sqrt{x-4}$ and $y = \sqrt{x} - 4$. They are "half-parabolas" with infinite slope at the start.

20 What is the break-even point if x yearbooks cost \$1200 + 30x to produce and the income is $40x$? The slope of the cost line is ________ (cost per additional book). If it goes above ________ you can't break even.

21 What are the domains and ranges of the distance functions in 14a and 14b—all values of t and $f(t)$ if $f(0) = 0$?

22 What is the range of $v(t)$ in 14b? Why is $t = 1$ not in the domain of $v(t)$ in 14a?

Problems 23–28 involve *linear functions* $f(t) = vt + C$. Find the constants v and C.

23 What linear function has $f(0) = 3$ and $f(2) = -11$?

24 Find *two* linear functions whose domain is $0 \leqslant t \leqslant 2$ and whose range is $1 \leqslant f(t) \leqslant 9$.

25 Find the linear function with $f(1) = 4$ and slope 6.

26 What functions have $f(t+1) = f(t) + 2$?

27 Find the linear function with $f(t+2) = f(t) + 6$ and $f(1) = 10$.

28 Find the only $f = vt$ that has $f(2t) = 4f(t)$. Show that every $f = \frac{1}{2}at^2$ has this property. To go ________ times as far in twice the time, you must accelerate.

29 Sketch the graph of $f(t)=|5-2t|$ (absolute value) for $|t|\leqslant 2$ and find its slopes and range.

30 Sketch the graph of $f(t)=4-t-|4-t|$ for $2\leqslant t\leqslant 5$ and find its slope and range.

31 Suppose $v=8$ up to time T, and after that $v=-2$. Starting from zero, when does f return to zero? Give formulas for $v(t)$ and $f(t)$.

32 Suppose $v=3$ up to time $T=4$. What new velocity will lead to $f(7)=30$ if $f(0)=0$? Give formulas for $v(t)$ and $f(t)$.

33 What function F(C) converts Celsius temperature C to Fahrenheit temperature F? The slope is ________, which is the number of Fahrenheit degrees equivalent to 1°C.

34 What function C(F) converts Fahrenheit to Celsius (or Centigrade), and what is its slope?

35 What function converts the weight w in grams to the weight $f(w)$ in kilograms? Interpret the slope of $f(w)$.

36 (Newspaper of March 1989) Ten hours after the accident the alcohol reading was .061. Blood alcohol is eliminated at .015 per hour. What was the reading at the time of the accident? How much later would it drop to .04 (the maximum set by the Coast Guard)? The usual limit on drivers is .10 percent.

Which points between $t=0$ and $t=5$ can be in the domain of $f(t)$? With this domain find the range in 37–42.

37 $f(t)=\sqrt{t-1}$ **38** $f(t)=1/\sqrt{t-1}$

39 $f(t)=|t-4|$ (absolute value) **40** $f(t)=1/(t-4)^2$

41 $f(t)=2^t$ **42** $f(t)=2^{-t}$

43 (a) Draw the graph of $f(t)=\frac{1}{2}t+3$ with domain $0\leqslant t\leqslant 2$. Then give a formula and graph for
(b) $f(t)+1$ (c) $f(t+1)$
(d) $4f(t)$ (e) $f(4t)$.

44 (a) Draw the graph of $U(t)=$ step function $=\{0$ for $t<0$, 1 for $t\geqslant 0\}$. Then draw
(b) $U(t)+2$ (c) $U(t+2)$
(d) $3U(t)$ (e) $U(3t)$.

45 (a) Draw the graph of $f(t)=t+1$ for $-1\leqslant t\leqslant 1$. Find the domain, range, slope, and formula for
(b) $2f(t)$ (c) $f(t-3)$ (d) $-f(t)$ (e) $f(-t)$.

46 If $f(t)=t-1$ what are $2f(3t)$ and $f(1-t)$ and $f(t-1)$?

47 In the forward-back example find $f(\frac{1}{2}T)$ and $f(\frac{3}{2}T)$. Verify that those agree with the areas "under" the v-graph in Figure 1.4.

48 Find formulas for the outputs $f_1(t)$ and $f_2(t)$ which come from the input t:

(1) inside = input∗3 (2) inside ← input + 6
output = inside + 3 output ← inside∗3

Note BASIC and FORTRAN (and calculus itself) use = instead of ←. But the symbol ← or := is in some ways better. The instruction $t\leftarrow t+6$ produces a new t equal to the old t plus six. The equation $t=t+6$ is not intended.

49 Your computer can add and multiply. Starting with the number 1 and the input called t, give a list of instructions to lead to these outputs:

$$f_1(t)=t^2+t \quad f_2(t)=f_1(f_1(t)) \quad f_3(t)=f_1(t+1).$$

50 In fifty words or less explain what a *function* is.

The last questions are challenging but possible.

51 If $f(t)=3t-1$ for $0\leqslant t\leqslant 2$ give formulas (with domain) and find the slopes of these six functions:
(a) $f(t+2)$ (b) $f(t)+2$ (c) $2f(t)$
(d) $f(2t)$ (e) $f(-t)$ (f) $f(f(t))$.

52 For $f(t)=vt+C$ find the formulas and slopes of
(a) $3f(t)+1$ (b) $f(3t+1)$ (c) $2f(4t)$
(d) $f(-t)$ (e) $f(t)-f(0)$ (f) $f(f(t))$.

53 (hardest) The forward-back function is $f(t)=2t$ for $0\leqslant t\leqslant 3$, $f(t)=12-2t$ for $3\leqslant t\leqslant 6$. Graph $f(f(t))$ and find its *four-part* formula. First try $t=1.5$ and 3.

54 (a) Why is the letter **X** not the graph of a function?
(b) Which capital letters are the graphs of functions?
(c) Draw graphs of their slopes.

1.2 Calculus Without Limits

The next page is going to reveal one of the key ideas behind calculus. The discussion is just about numbers—functions and slopes can wait. The numbers are not even special, they can be any numbers. The crucial point is to look at their differences:

Suppose the numbers are $f=$ 0 2 6 7 4 9
Their differences are $v=$ 2 4 1 −3 5

The differences are printed in between, to show $2-0=2$ and $6-2=4$ and $7-6=1$.

Notice how $4-7$ gives a negative answer -3. The numbers in f can go up or down, the differences in v can be positive or negative. The idea behind calculus comes when you ***add up those differences***:

$$\mathbf{2+4+1-3+5=9}$$

The sum of differences is 9. This is the last number on the top line (in f). Is this an accident, or is this always true? If we stop earlier, after $2+4+1$, we get the 7 in f. Test any prediction on a second example:

Suppose the numbers are $f=$	1	3	7	8	5	10
Their differences are $v=$		2	4	1	-3	5

The f's are increased by 1. ***The differences are exactly the same***—no change. The sum of differences is still 9. But the last f is now 10. That prediction is not right, we don't always get the last f.

The first f is now 1. The answer 9 (the sum of differences) is $10-1$, ***the last f minus the first f***. What happens when we change the f's in the middle?

Suppose the numbers are $f=$	1	5	12	7	10
Their differences are $v=$		4	7	-5	3

The differences add to $4+7-5+3=9$. This is still $10-1$. No matter what f's we choose or how many, the sum of differences is controlled by the first f and last f. If this is always true, there must be a clear reason why *the middle f's cancel out*.

The sum of differences is $(5-1)+(12-5)+(7-12)+(10-7)=10-1$.

The 5's cancel, the 12's cancel, and the 7's cancel. It is only $10-1$ that doesn't cancel. This is the key to calculus!

1B The differences of the f's add up to $(f_{\text{last}}-f_{\text{first}})$.

EXAMPLE 1 The numbers grow linearly: $f=$ 2 3 4 5 6 7
Their differences are constant: $v=$ 1 1 1 1 1

The sum of differences is certainly 5. This agrees with $7-2=f_{\text{last}}-f_{\text{first}}$. The numbers in v remind us of constant velocity. The numbers in f remind us of a straight line $f=vt+C$. This example has $v=1$ and the f's start at 2. The straight line would come from $f=t+2$.

EXAMPLE 2 The numbers are squares: $f=$ 0 1 4 9 16
Their differences grow linearly: $v=$ 1 3 5 7

$1+3+5+7$ agrees with $4^2=16$. It is a beautiful fact that the first j odd numbers always add up to j^2. The v's are the odd numbers, the f's are perfect squares.

Note The letter j is sometimes useful to tell which number in f we are looking at. For this example the zeroth number is $f_0=0$ and the jth number is $f_j=j^2$. This is a part of algebra, to give a formula for the f's instead of a list of numbers. We can also use j to tell which difference we are looking at. The first v is the first odd number $v_1=1$. The jth difference is the jth odd number $v_j=2j-1$. (Thus v_4 is $8-1=7$.) It is better to start the differences with $j=1$, since there is no zeroth odd number v_0.

With this notation the jth ***difference is*** $v_j=f_j-f_{j-1}$. Sooner or later you will get comfortable with subscripts like j and $j-1$, but it can be later. The important point is that the sum of the v's equals $f_{\text{last}}-f_{\text{first}}$. We now connect the v's to slopes and the f's to areas.

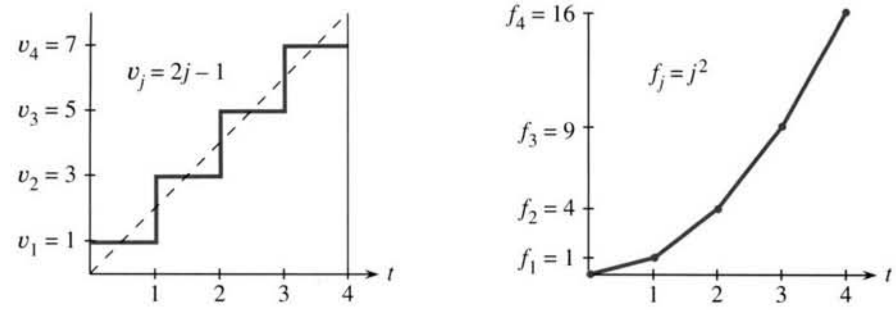


Fig. 1.7 Linear increase in $v = 1, 3, 5, 7$. Squares in the distances $f = 0, 1, 4, 9, 16$.

Figure 1.7 shows a natural way to graph Example 2, with the odd numbers in v and the squares in f. Notice an important difference between the v-graph and the f-graph. The graph of f is "***piecewise linear***." We plotted the numbers in f and connected them by straight lines. The graph of v is "***piecewise constant***." We plotted the differences as constant over each piece. This reminds us of the distance-velocity graphs, when the distance $f(t)$ is a straight line and the velocity $v(t)$ is a horizontal line.

Now make the connection to slopes:

$$\text{The slope of the } f\text{-graph is } \frac{\text{distance up}}{\text{distance across}} = \frac{\text{change in } f}{\text{change in } t} = v.$$

Over each piece, the change in t (across) is 1. The change in f (upward) is the difference that we are calling v. The ratio is the slope $v/1$ or just v. The slope makes a sudden change at the breakpoints $t = 1, 2, 3, \ldots$. At those special points the slope of the f-graph is not defined—we connected the v's by vertical lines but this is very debatable. ***The main idea is that between the breakpoints, the slope of*** $f(t)$ ***is*** $v(t)$.

Now make the connection to areas:

The total area under the v***-graph is*** $f_{\text{last}} - f_{\text{first}}$.

This area, underneath the staircase in Figure 1.7, is composed of rectangles. The base of every rectangle is 1. The heights of the rectangles are the v's. So the areas also equal the v's, and the total area is the sum of the v's. This area is $f_{\text{last}} - f_{\text{first}}$.

Even more is true. We could start at any time and end at any later time —not necessarily at the special times $t = 0, 1, 2, 3, 4$. Suppose we stop at $t = 3.5$. Only half of the last rectangular area (under $v = 7$) will be counted. The total area is $1 + 3 + 5 + \frac{1}{2}(7) = 12.5$. This still agrees with $f_{\text{last}} - f_{\text{first}} = 12.5 - 0$. At this new ending time $t = 3.5$, we are only halfway up the last step in the f-graph. Halfway between 9 and 16 is 12.5.

1C ***The*** v***'s are slopes of*** $f(t)$***. The area under the*** v***-graph is*** $f(t_{\text{end}}) - f(t_{\text{start}})$.

This is nothing less than the Fundamental Theorem of Calculus. But we have only used algebra (no curved graphs and no calculations involving limits). For now the Theorem is restricted to piecewise linear $f(t)$ and piecewise constant $v(t)$. In Chapter 5 that restriction will be overcome.

Notice that a proof of $1 + 3 + 5 + 7 = 4^2$ is suggested by Figure 1.7a. The triangle under the dotted line has the same area as the four rectangles under the staircase. The area of the triangle is $\frac{1}{2} \cdot \text{base} \cdot \text{height} = \frac{1}{2} \cdot 4 \cdot 8$, which is the perfect square 4^2. When there are j rectangles instead of 4, we get $\frac{1}{2} \cdot j \cdot 2j = j^2$ for the area.

The next examples show other patterns, where f and v increase exponentially or oscillate around zero. I hope you like them but I don't think you have to learn them. They are like the special functions 2^t and $\sin t$ and $\cos t$—except they go in steps. You get a first look at the important functions of calculus, but you only need algebra. ***Calculus is needed for a steadily changing velocity, when the graph of f is curved.***

The last example will be *income tax*—which really does go in steps. Then Section 1.3 will introduce the slope of a curve. The crucial step for curves is working with ***limits***. That will take us from algebra to calculus.

EXPONENTIAL VELOCITY AND DISTANCE

Start with the numbers $f = 1, 2, 4, 8, 16$. These are "powers of 2." They start with the zeroth power, which is $2^0 = 1$. *The exponential starts at* 1 *and not* 0. After j steps there are j factors of 2, and f_j equals 2^j. ***Please recognize the difference between $2j$ and j^2 and 2^j.*** The numbers $2j$ grow linearly, the numbers j^2 grow quadratically, the numbers 2^j grow exponentially. At $j = 10$ these are 20 and 100 and 1024. The exponential 2^j quickly becomes much larger than the others.

The differences of $f = 1, 2, 4, 8, 16$ are exactly $v = 1, 2, 4, 8$. We get the same beautiful numbers. ***When the f's are powers of 2, so are the v's.*** The formula $v_j = 2^{j-1}$ is slightly different from $f_j = 2^j$, because the first v is numbered v_1. (Then $v_1 = 2^0 = 1$. The zeroth power of every number is 1, except that 0^0 is meaningless.) The two graphs in Figure 1.8 use the same numbers but they look different, because f is piecewise linear and v is piecewise constant.

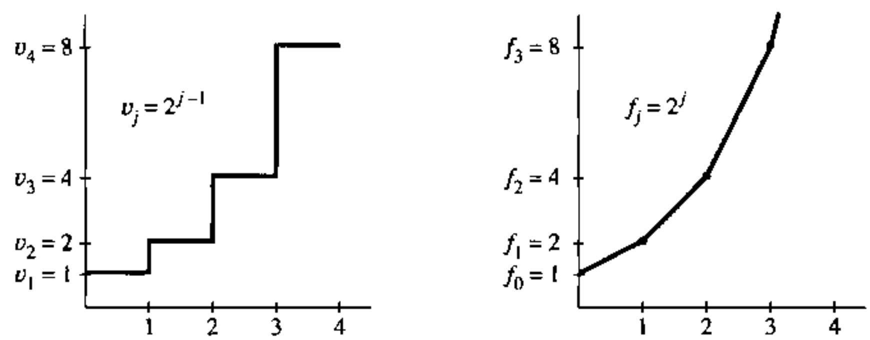


Fig. 1.8 The velocity and distance grow exponentially (powers of 2).

Where will calculus come in? It works with the smooth curve $f(t) = 2^t$. This exponential growth is critically important for population and money in a bank and the national debt. You can spot it by the following test: $v(t)$ ***is proportional to*** $f(t)$.

Remark The function 2^t is trickier than t^2. For $f = t^2$ the slope is $v = 2t$. It is proportional to t and not t^2. For $f = 2^t$ the slope is $v = c2^t$, and we won't find the constant $c = .693\ldots$ until Chapter 6. (The number c is the natural logarithm of 2.) Problem 37 estimates c with a calculator—the important thing is that it's constant.

OSCILLATING VELOCITY AND DISTANCE

We have seen a forward-back motion, velocity V followed by $-V$. That is oscillation of the simplest kind. The graph of f goes linearly up and linearly down. Figure 1.9 shows another oscillation that returns to zero, but the path is more interesting.

The numbers in f are now $0, 1, 1, 0, -1, -1, 0$. Since $f_6 = 0$ the motion brings us back to the start. The whole oscillation can be repeated.

The differences in v are $1, 0, -1, -1, 0, 1$. They add up to zero, which agrees with $f_{\text{last}} - f_{\text{first}}$. It is the same oscillation as in f (and also repeatable), but shifted in time.

The f-graph resembles (roughly) a ***sine curve***. The v-graph resembles (even more roughly) a ***cosine curve***. The waveforms in nature are smooth curves, while these are "digitized"—the way a digital watch goes forward in jumps. You recognize that the change from analog to digital brought the computer revolution. The same revolution is coming in CD players. Digital signals (off or on, 0 or 1) seem to win every time.

The piecewise v and f start again at $t = 6$. The ordinary sine and cosine repeat at $t = 2\pi$. A repeating motion is ***periodic***—here the "period" is 6 or 2π. (With t in degrees the period is 360—a full circle. The period becomes 2π when angles are measured in ***radians***. We virtually always use radians—which are degrees times $2\pi/360$.) A watch has a period of 12 hours. If the dial shows AM and PM, the period is _______.

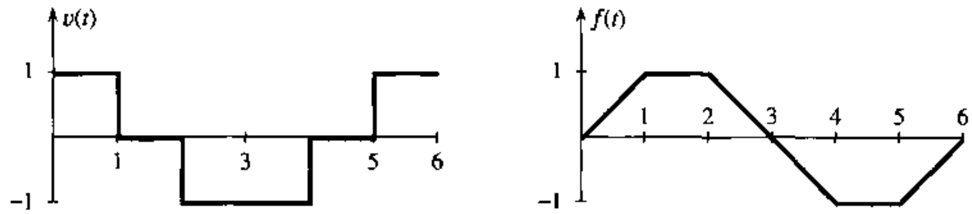


Fig. 1.9 Piecewise constant "cosine" and piecewise linear "sine." They both repeat.

A SHORT BURST OF SPEED

The next example is a car that is driven fast for a short time. The speed is V until the distance reaches $f = 1$, when the car suddenly stops. The graph of f goes up linearly with slope V, and then across with slope zero:

$$v(t) = \begin{cases} V & \text{up to} \quad t = T \\ 0 & \text{after} \quad t = T \end{cases} \qquad f(t) = \begin{cases} Vt & \text{up to} \quad t = T \\ 1 & \text{after} \quad t = T \end{cases}$$

This is another example of "function notation." Notice the general time t and the particular stopping time T. The distance is $f(t)$. The domain of f (the inputs) includes all times $t \geq 0$. The range of f (the outputs) includes all distances $0 \leq f \leq 1$.

Figure 1.10 allows us to compare three cars—a Jeep and a Corvette and a Maserati. They have different speeds but they all reach $f = 1$. So the areas under the v-graphs are all 1. The rectangles have height V and base $T = 1/V$.

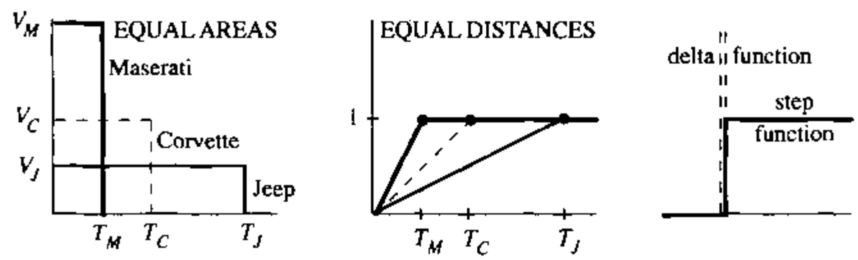


Fig. 1.10 Bursts of speed with $V_M T_M = V_C T_C = V_J T_J = 1$. Step function has infinite slope.

Optional remark It is natural to think about faster and faster speeds, which means steeper slopes. The f-graph reaches 1 in shorter times. The extreme case is a ***step function***, when the graph of f goes straight up. This is the unit step $U(t)$, which is zero up to $t = 0$ and jumps immediately to $U = 1$ for $t > 0$.

What is the slope of the step function? It is zero except at the jump. At that moment, which is $t = 0$, the slope is *infinite*. We don't have an ordinary velocity $v(t)$—instead we have an impulse that makes the car jump. The graph is a spike over the single point $t = 0$, and it is often denoted by δ—so the slope of the step function is called a "***delta function***." The area under the infinite spike is 1.

You are absolutely not responsible for the theory of delta functions! Calculus is about curves, not jumps.

Our last example is a real-world application of slopes and rates—to explain "how taxes work." Note especially the difference between tax rates and tax brackets and total tax. The rates are v, the brackets are on x, the total tax is f.

EXAMPLE 3 ***Income tax is piecewise linear. The slopes are the tax rates*** .15, .28, .31.

Suppose you are single with taxable income of x dollars (Form 1040, line 37—after all deductions). These are the 1991 instructions from the Internal Revenue Service:

If x is not over \$20,350, the tax is 15% of x.

If $\$20{,}350 \leqslant x \leqslant \$49{,}300$, the tax is \$3052.50 + 28% of the amount over \$20,350.

If x is over \$49,300, the tax is \$11,158.50 + 31% of the amount over \$49,300.

The first bracket is $0 \leqslant x \leqslant \$20{,}350$. (The IRS never uses this symbol $\leqslant$, but I think it is OK here. We know what it means.) The second bracket is $\$20{,}350 \leqslant x \leqslant \$49{,}300$. The top bracket $x \geqslant \$49{,}300$ pays tax at the top rate of 31%. But only the income ***in that bracket*** is taxed at that rate.

Figure 1.11 shows the rates and the brackets and the tax due. Those are not ***average*** rates, they are ***marginal*** rates. Total tax divided by total income would be the average rate. The marginal rate of .28 or .31 gives the tax on each *additional* dollar of income—it is the slope at the point x. *Tax* is like *area* or *distance*—it adds up. ***Tax rate*** is like ***slope*** or ***velocity***—it depends where you are. This is often unclear in the news media.

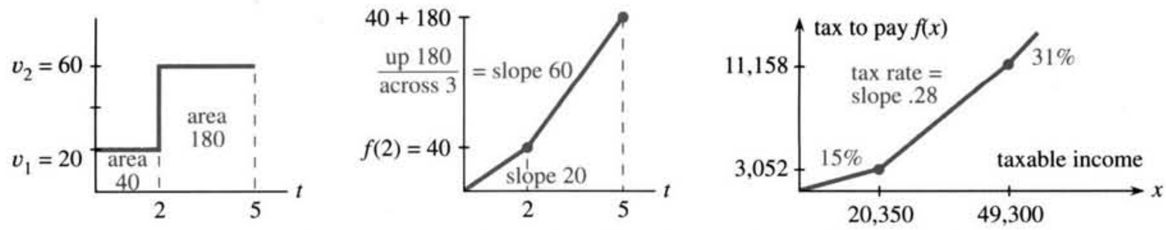


Fig. 1.11 The tax rate is v, the total tax is f. Tax brackets end at breakpoints.

Question What is the equation for the straight line in the top bracket?
Answer The bracket begins at $x = \$49{,}300$ when the tax is $f(x) = \$11{,}158.50$. The slope of the line is the tax rate .31. When we know a point on the line and the slope, we know the equation. This is important enough to be highlighted.

> **1D** For x in the top bracket the tax is $f(x) = \$11{,}158.50 + .31\ (x - \$49{,}300)$. This is the tax on \$49,300 plus the extra tax on extra income.

Section 2.3 presents this "point-slope equation" for any straight line. Here you see it for one specific example. Where does the number \$11,158.50 come from? It is the tax at the *end* of the middle bracket, so it is the tax at the *start* of the top bracket.

Figure 1.11 also shows a distance-velocity example. The distance at $t=2$ is $f(2)=40$ miles. After that time the velocity is 60 miles per hour. So the line with slope 60 on the f-graph has the equation

$$f(t) = \text{starting distance} + \text{extra distance} = 40 + 60(t-2).$$

The starting point is $(2, 40)$. The new speed 60 multiplies the extra time $t-2$. The point-slope equation makes sense. ***We now review this section, with comments.***

Central idea Start with any numbers in f. Their differences go in v. Then the sum of those differences is $f_{\text{last}} - f_{\text{first}}$.

Subscript notation The numbers are $f_0, f_1, \ldots$ and the first difference is $v_1 = f_1 - f_0$. A typical number is f_j and the jth difference is $v_j = f_j - f_{j-1}$. When those differences are added, all f's in the middle (like f_1) cancel out:

$$v_1 + v_2 + \cdots + v_j = (f_1 - f_0) + (f_2 - f_1) + \cdots + (f_j - f_{j-1}) = f_j - f_0.$$

Examples $f_j = j$ or j^2 or 2^j. Then $v_j = 1$ (constant) or $2j-1$ (odd numbers) or 2^{j-1}.

Functions Connect the f's to be piecewise linear. Then the slope v is piecewise constant. The area under the v-graph from any t_{start} to any t_{end} equals $f(t_{\text{end}}) - f(t_{\text{start}})$.

Units Distance in miles and velocity in miles per hour. Tax in dollars and tax rate in (dollars paid)/(dollars earned). Tax rate is a percentage like .28, with no units.

1.2 EXERCISES

Read-through questions

Start with the numbers $f = 1, 6, 2, 5$. Their differences are $v =$ __a__. The sum of those differences is __b__. This is equal to f_{last} minus __c__. The numbers 6 and 2 have no effect on this answer, because in $(6-1)+(2-6)+(5-2)$ the numbers 6 and 2 __d__. The slope of the line between $f(0)=1$ and $f(1)=6$ is __e__. The equation of that line is $f(t) =$ __f__.

With distances 1, 5, 25 at unit times, the velocities are __g__. These are the __h__ of the f-graph. The slope of the tax graph is the tax __i__. If $f(t)$ is the postage cost for t ounces or t grams, the slope is the __j__ per __k__. For distances 0, 1, 4, 9 the velocities are __l__. The sum of the first j odd numbers is $f_j =$ __m__. Then f_{10} is __n__ and the velocity v_{10} is __o__.

The piecewise linear sine has slopes __p__. Those form a piecewise __q__ cosine. Both functions have __r__ equal to 6, which means that $f(t+6) =$ __s__ for every t. The velocities $v = 1, 2, 4, 8, \ldots$ have $v_j =$ __t__. In that case $f_0 = 1$ and $f_j =$ __u__. The sum of 1, 2, 4, 8, 16 is __v__. The difference $2^j - 2^{j-1}$ equals __w__. After a burst of speed V to time T, the distance is __x__. If $f(T) = 1$ and V increases, the burst lasts only to $T =$ __y__. When V approaches infinity, $f(t)$ approaches a __z__ function. The velocities approach a __A__ function, which is concentrated at $t = 0$ but has area __B__ under its graph. The slope of a step function is __C__.

Problems 1–4 are about numbers f and differences v.

1 From the numbers $f = 0, 2, 7, 10$ find the differences v and the sum of the three v's. Write down another f that leads to the same v's. For $f = 0, 3, 12, 10$ the sum of the v's is still ______.

2 Starting from $f = 1, 3, 2, 4$ draw the f-graph (linear pieces) and the v-graph. What are the areas "under" the v-graph that add to $4 - 1$? If the next number in f is 11, what is the area under the next v?

3 From $v = 1, 2, 1, 0, -1$ find the f's starting at $f_0 = 3$. Graph v and f. The maximum value of f occurs when $v =$ ______. Where is the maximum f when $v = 1, 2, 1, -1$?

4 For $f = 1, b, c, 7$ find the differences v_1, v_2, v_3 and add them up. Do the same for $f = a, b, c, 7$. Do the same for $f = a, b, c, d$.

Problems 5–11 are about linear functions and constant slopes.

5 Write down the slopes of these linear functions:

(a) $f(t) = 1.1t$ (b) $f(t) = 1 - 2t$ (c) $f(t) = 4 + 5(t-6)$.

Compute $f(6)$ and $f(7)$ for each function and confirm that $f(7) - f(6)$ equals the slope.

6 If $f(t) = 5 + 3(t-1)$ and $g(t) = 1.5 + 2.5(t-1)$ what is $h(t) = f(t) - g(t)$? Find the slopes of f, g, and h.

7 Suppose $v(t) = 2$ for $t < 5$ and $v(t) = 3$ for $t > 5$.

(a) If $f(0) = 0$ find a two-part formula for $f(t)$.

(b) Check that $f(10)$ equals the area under the graph of $v(t)$ (two rectangles) up to $t = 10$.

8 Suppose $v(t) = 10$ for $t < 1/10$, $v(t) = 0$ for $t > 1/10$. Starting from $f(0) = 1$ find $f(t)$ in two pieces.

9 Suppose $g(t) = 2t + 1$ and $f(t) = 4t$. Find $g(3)$ and $f(g(3))$ and $f(g(t))$. How is the slope of $f(g(t))$ related to the slopes of f and g?

10 For the same functions, what are $f(3)$ and $g(f(3))$ and $g(f(t))$? When t is changed to $4t$, distance increases ________ times as fast and the velocity is multiplied by ________.

11 Compute $f(6)$ and $f(8)$ for the functions in Problem 5. Confirm that the slopes v agree with

$$\text{slope} = \frac{f(8) - f(6)}{8 - 6} = \frac{\text{change in } f}{\text{change in } t}.$$

Problems 12–18 are based on Example 3 about income taxes.

12 What are the income taxes on $x = \$10{,}000$ and $x = \$30{,}000$ and $x = \$50{,}000$?

13 What is the equation for income tax $f(x)$ in the second bracket $\$20{,}350 \leqslant x \leqslant \$49{,}300$? How is the number 11,158.50 connected with the other numbers in the tax instructions?

14 Write the tax function $F(x)$ for a married couple if the IRS treats them as two single taxpayers each with taxable income $x/2$. (This is not done.)

15 In the 15% bracket, with 5% state tax as a deduction, the combined rate is not 20% but ________. Think about the tax on an extra \$100.

16 A piecewise linear function is ***continuous*** when $f(t)$ at the end of each interval equals $f(t)$ at the start of the following interval. If $f(t) = 5t$ up to $t = 1$ and $v(t) = 2$ for $t > 1$, define f beyond $t = 1$ so it is (a) continuous (b) discontinuous. (c) Define a tax function $f(x)$ with rates .15 and .28 so you would lose by earning an extra dollar beyond the breakpoint.

17 The difference between a tax *credit* and a *deduction* from income is the difference between $f(x) - c$ and $f(x - d)$. Which is more desirable, a credit of $c = \$1000$ or a deduction of $d = \$1000$, and why? Sketch the tax graphs when $f(x) = .15x$.

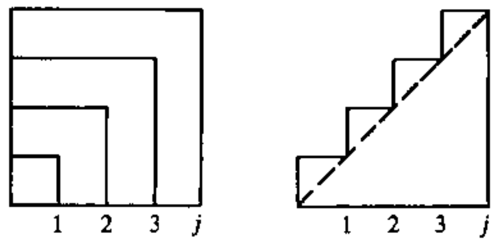
18 The average tax rate on the taxable income x is $a(x) = f(x)/x$. This is the slope between $(0, 0)$ and the point $(x, f(x))$. Draw a rough graph of $a(x)$. The average rate a is below the marginal rate v because ________.

Problems 19–30 involve numbers $f_0, f_1, f_2, \ldots$ and their differences $v_j = f_j - f_{j-1}$. They give practice with subscripts $0, \ldots, j$.

19 Find the velocities v_1, v_2, v_3 and formulas for v_j and f_j:
(a) $f = 1, 3, 5, 7 \ldots$ (b) $f = 0, 1, 0, 1, \ldots$ (c) $f = 0, \frac{1}{2}, \frac{3}{4}, \frac{7}{8}, \ldots$

20 Find f_1, f_2, f_3 and a formula for f_j with $f_0 = 0$:
(a) $v = 1, 2, 4, 8, \ldots$ (b) $v = -1, 1, -1, 1, \ldots$

21 The areas of these nested squares are $1^2, 2^2, 3^2, \ldots$. What are the areas of the L-shaped bands (the differences between squares)? How does the figure show that $1 + 3 + 5 + 7 = 4^2$?



22 From the area under the staircase (by rectangles and then by triangles) show that the first j whole numbers 1 to j add up to $\frac{1}{2}j^2 + \frac{1}{2}j$. Find $1 + 2 + \cdots + 100$.

23 If $v = 1, 3, 5, \ldots$ then $f_j = j^2$. If $v = 1, 1, 1, \ldots$ then $f_j =$ ________. Add those to find the sum of $2, 4, 6, \ldots, 2j$. Divide by 2 to find the sum of $1, 2, 3, \ldots, j$. (Compare Problem 22.)

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

(a) When the f's are increasing so are the v's.

(b) When the v's are increasing so are the f's.

(c) When the f's are periodic so are the v's.

(d) When the v's are periodic so are the f's.

25 If $f(t) = t^2$, compute $f(99)$ and $f(101)$. Between those times, what is the increase in f divided by the increase in t?

26 If $f(t) = t^2 + t$, compute $f(99)$ and $f(101)$. Between those times, what is the increase in f divided by the increase in t?

27 If $f_j = j^2 + j + 1$ find a formula for v_j.

28 Suppose the v's increase by 4 at every step. Show by example and then by algebra that the "second difference" $f_{j+1} - 2f_j + f_{j-1}$ equals 4.

29 Suppose $f_0 = 0$ and the v's are $1, \frac{1}{2}, \frac{1}{2}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \ldots$. For which j does $f_j = 5$?

30 Show that $a_j = f_{j+1} - 2f_j + f_{j-1}$ always equals $v_{j+1} - v_j$. If v is velocity then a stands for ________.

Problems 31–34 involve periodic f's and v's (like $\sin t$ and $\cos t$).

31 For the discrete sine $f = 0, 1, 1, 0, -1, -1, 0$ find the second differences $a_1 = f_2 - 2f_1 + f_0$ and $a_2 = f_3 - 2f_2 + f_1$ and a_3. Compare a_j with f_j.

32 If the sequence $v_1, v_2, \ldots$ has period 6 and $w_1, w_2, \ldots$ has period 10, what is the period of $v_1 + w_1, v_2 + w_2, \ldots$?

33 Draw the graph of $f(t)$ starting from $f_0 = 0$ when $v = 1, -1, -1, 1$. If v has period 4 find $f(12)$, $f(13)$, $f(100.1)$.

34 Graph $f(t)$ from $f_0 = 0$ to $f_4 = 4$ when $v = 1, 2, 1, 0$. If v has period 4, find $f(12)$ and $f(14)$ and $f(16)$. *Why doesn't f have period* 4?

Problems 35–42 are about exponential v's and f's.

35 Find the v's for $f = 1, 3, 9, 27$. Predict v_4 and v_j. Algebra gives $3^j - 3^{j-1} = (3-1)3^{j-1}$.

36 Find $1 + 2 + 4 + \cdots + 32$ and also $1 + \frac{1}{2} + \frac{1}{4} + \cdots + \frac{1}{32}$.

37 Estimate the slope of $f(t) = 2^t$ at $t = 0$. Use a calculator to compute (increase in f)/(increase in t) when t is small:

$$\frac{f(t) - f(0)}{t} = \frac{2 - 1}{1} \text{ and } \frac{2^{.1} - 1}{.1} \text{ and } \frac{2^{.01} - 1}{.01} \text{ and } \frac{2^{.001} - 1}{.001}.$$

38 Suppose $f_0 = 1$ and $v_j = 2f_{j-1}$. Find f_4.

39 (a) From $f = 1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}$ find v_1, v_2, v_3 and predict v_j.
(b) Check $f_3 - f_0 = v_1 + v_2 + v_3$ and $f_j - f_{j-1} = v_j$.

40 Suppose $v_j = r^j$. Show that $f_j = (r^{j+1} - 1)/(r - 1)$ starts from $f_0 = 1$ and has $f_j - f_{j-1} = v_j$. (Then this is the correct $f_j = 1 + r + \cdots + r^j =$ sum of a geometric series.)

41 From $f_j = (-1)^j$ compute v_j. What is $v_1 + v_2 + \cdots + v_j$?

42 Estimate the slope of $f(t) = e^t$ at $t = 0$. Use a calculator that knows e (or else take $e = 2.78$) to compute

$$\frac{f(t) - f(0)}{t} = \frac{e - 1}{1} \text{ and } \frac{e^{.1} - 1}{.1} \text{ and } \frac{e^{.01} - 1}{.01}.$$

Problems 43–47 are about $U(t) =$ step from 0 to 1 at $t = 0$.

43 Graph the four functions $U(t-1)$ and $U(t) - 2$ and $U(3t)$ and $4U(t)$. Then graph $f(t) = 4U(3t - 1) - 2$.

44 Graph the square wave $U(t) - U(t-1)$. If this is the velocity $v(t)$, graph the distance $f(t)$. If this is the distance $f(t)$, graph the velocity.

45 Two bursts of speed lead to the same distance $f = 10$:

$v =$ ______ to $t = .001$ $\quad v = V$ to $t =$ ______.

As $V \to \infty$ the limit of the $f(t)$'s is ______.

46 Draw the staircase function $U(t) + U(t-1) + U(t-2)$. Its slope is a sum of three ______ functions.

47 Which capital letters like **L** are the graphs of functions when steps are allowed? The slope of **L** is minus a delta function. Graph the slopes of the others.

48 Write a subroutine FINDV whose input is a sequence $f_0, f_1, \ldots, f_N$ and whose output is $v_1, v_2, \ldots, v_N$. Include graphical output if possible. Test on $f_j = 2j$ and j^2 and 2^j.

49 Write a subroutine FINDF whose input is $v_1, \ldots, v_N$ and f_0, and whose output is $f_0, f_1, \ldots, f_N$. The default value of f_0 is zero. Include graphical output if possible. Test $v_j = j$.

50 If FINDV is applied to the output of FINDF, what sequence is returned? If FINDF is applied to the output of FINDV, what sequence is returned? Watch f_0.

51 Arrange $2j$ and j^2 and 2^j and $\sqrt{j}$ in increasing order
(a) when j is large: $j = 9$ (b) when j is small: $j = \frac{1}{9}$.

52 The average age of your family since 1970 is a piecewise linear function $A(t)$. Is it continuous or does it jump? What is its slope? Graph it the best you can.

1.3 The Velocity at an Instant

We have arrived at the central problems that calculus was invented to solve. There are two questions, in opposite directions, and I hope you could see them coming.

1. If the velocity is changing, *how can you compute the distance traveled*?
2. If the graph of $f(t)$ is not a straight line, *what is its slope*?

Find the distance from the velocity, find the velocity from the distance. Our goal is to do both—but not in one section. Calculus may be a good course, but it is not magic. The first step is to let the velocity change in the steadiest possible way.

Question 1 ***Suppose the velocity at each time t is*** $v(t) = 2t$. Find $f(t)$.

With $v = 2t$, a physicist would say that the acceleration is constant (it equals 2). The driver steps on the gas, the car accelerates, and the speedometer goes steadily up. The distance goes up too—faster and faster. If we measure t in seconds and v in feet per second, the distance f comes out in feet. After 10 seconds the speed is 20 feet per second. After 44 seconds the speed is 88 feet/second (which is 60 miles/hour). The acceleration is clear, ***but how far has the car gone***?

Question 2 *The distance traveled by time t is* $f(t) = t^2$. Find the velocity $v(t)$.

The graph of $f(t) = t^2$ is on the right of Figure 1.12. It is a *parabola*. The curve starts at zero, when the car is new. At $t = 5$ the distance is $f = 25$. By $t = 10$, f reaches 100.

Velocity is distance divided by time, but what happens when the speed is changing? Dividing $f = 100$ by $t = 10$ gives $v = 10$—the *average velocity* over the first ten seconds. Dividing $f = 121$ by $t = 11$ gives the average speed over 11 seconds. But how do we find the ***instantaneous velocity***—the reading on the speedometer at the exact instant when $t = 10$?

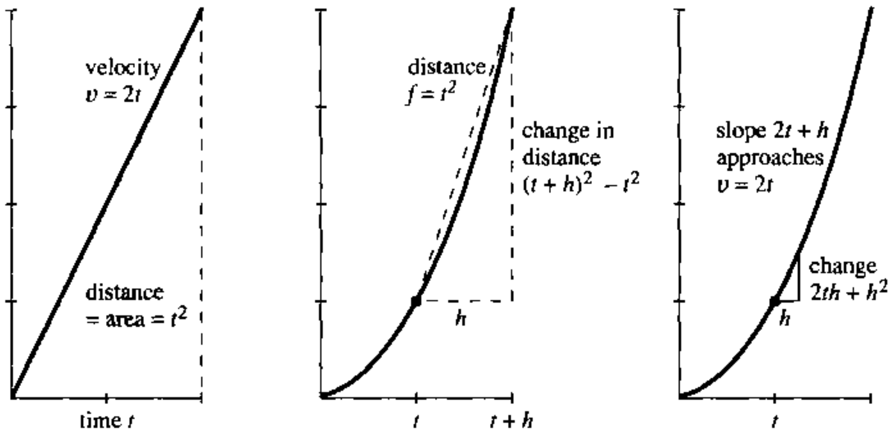


Fig. 1.12 The velocity $v = 2t$ is linear. The distance $f = t^2$ is quadratic.

I hope you see the problem. As the car goes faster, the graph of t^2 gets steeper—because more distance is covered in each second. The average velocity between $t = 10$ and $t = 11$ is a good approximation—but only an approximation—to the speed at the moment $t = 10$. Averages are easy to find:

distance at $t = 10$ is $f(10) = 10^2 = 100$ distance at $t = 11$ is $f(11) = 11^2 = 121$

$$\textit{average velocity is } \frac{f(11) - f(10)}{11 - 10} = \frac{121 - 100}{1} = 21.$$

The car covered 21 feet in that 1 second. Its average speed was 21 feet/second. Since it was gaining speed, the velocity at the beginning of that second was below 21.

Geometrically, what is the average? It is a slope, but not the slope of the curve. ***The average velocity is the slope of a straight line.*** The line goes between two points on the curve in Figure 1.12. When we compute an average, we pretend the velocity is constant—so we go back to the easiest case. It only requires a division of distance by time:

$$\text{average velocity} = \frac{\text{change in } f}{\text{change in } t}. \tag{1}$$

Calculus and the Law You enter a highway at 1:00. If you exit 150 miles away at 3:00, your average speed is 75 miles per hour. I'm not sure if the police can give you a ticket. You could say to the judge, "When was I doing 75?" The police would have to admit that they have no idea—but they would have a definite feeling that you must have been doing 75 sometime.†

We return to the central problem—computing $v(10)$ at the instant $t = 10$. The average velocity over the next second is 21. We can also find the average over the *half-second* between $t = 10.0$ and $t = 10.5$. Divide the change in distance by the change in time:

$$\frac{f(10.5) - f(10.0)}{10.5 - 10.0} = \frac{(10.5)^2 - (10.0)^2}{.5} = \frac{110.25 - 100}{.5} = 20.5.$$

That average of 20.5 is closer to the speed at $t = 10$. It is still not exact.

The way to find $v(10)$ is to *keep reducing the time interval*. This is the basis for Chapter 2, and the key to differential calculus. ***Find the slope between points that are closer and closer on the curve***. The "limit" is the slope at a single point.

Algebra gives the average velocity between $t = 10$ and any later time $t = 10 + h$. The distance increases from 10^2 to $(10 + h)^2$. The change in time is h. So divide:

$$v_{\text{average}} = \frac{(10 + h)^2 - 10^2}{h} = \frac{100 + 20h + h^2 - 100}{h} = 20 + h. \tag{2}$$

This formula fits our previous calculations. The interval from $t = 10$ to $t = 11$ had $h = 1$, and the average was $20 + h = 21$. When the time step was $h = \frac{1}{2}$, the average was $20 + \frac{1}{2} = 20.5$. Over a millionth of a second the average will be 20 plus 1/1,000,000—which is very near 20.

Conclusion: ***The velocity at*** $t = 10$ ***is*** $v = 20$. *That is the slope of the curve*. It agrees with the v-graph on the left side of Figure 1.12, which also has $v(10) = 20$.

We now show that the two graphs match at all times. If $f(t) = t^2$ then $v(t) = 2t$. You are seeing the key computation of calculus, and we can put it into words before equations. Compute the distance at time $t + h$, *subtract* the distance at time t, and *divide* by h. That gives the average velocity:

$$v_{\text{ave}} = \frac{f(t + h) - f(t)}{h} = \frac{(t + h)^2 - t^2}{h} = \frac{t^2 + 2th + h^2 - t^2}{h} = 2t + h. \tag{3}$$

This fits the previous calculation, where t was 10. The average was $20 + h$. Now the average is $2t + h$. It depends on the time step h, because the velocity is changing. But we can see what happens *as h approaches zero*. The average is closer and closer to the speedometer reading of $2t$, at the exact moment when the clock shows time t:

1E As h approaches zero, the average velocity $2t + h$ approaches $v(t) = 2t$.

Note The computation (3) shows how calculus needs algebra. If we want the whole v-graph, we have to let time be a "*variable*." It is represented by the letter t. Numbers are enough at the specific time $t = 10$ and the specific step $h = 1$—but algebra gets beyond that. The average between any t and any $t + h$ is $2t + h$. Please don't hesitate to put back numbers for the letters—that checks the algebra.

†This is our first encounter with the much despised "Mean Value Theorem." If the judge can prove the theorem, you are dead. A few v-graphs and f-graphs will confuse the situation (possibly also a delta function).

There is also a step beyond algebra! Calculus requires the ***limit of the average***. As h shrinks to zero, the points on the graph come closer. "Average over an interval" becomes "velocity at an instant." The general theory of limits is not particularly simple, but here we don't need it. (It isn't particularly hard either.) In this example *the limiting value is easy to identify*. The average $2t + h$ approaches $2t$, as $h \to 0$.

What remains to do in this section? We answered Question 2—to find velocity from distance. We have not answered Question 1. If $v(t) = 2t$ increases linearly with time, what is the distance? This goes in the opposite direction (it is *integration*).

The Fundamental Theorem of Calculus says that no new work is necessary. ***If the slope of $f(t)$ leads to $v(t)$, then the area under that v-graph leads back to the f-graph.*** The odometer readings $f = t^2$ produced speedometer readings $v = 2t$. By the Fundamental Theorem, the area under $2t$ should be t^2. But we have certainly not proved any fundamental theorems, so it is better to be safe—by actually computing the area.

Fortunately, it is the area of a triangle. The base of the triangle is t and the height is $v = 2t$. The area agrees with $f(t)$:

$$\text{area} = \tfrac{1}{2}(\text{base})(\text{height}) = \tfrac{1}{2}(t)(2t) = t^2. \tag{4}$$

EXAMPLE 1 The graphs are *shifted in time*. The car doesn't start until $t = 1$. Therefore $v = 0$ and $f = 0$ up to that time. After the car starts we have $v = 2(t - 1)$ and $f = (t - 1)^2$. You see how the time delay of 1 enters the formulas. Figure 1.13 shows how it affects the graphs.

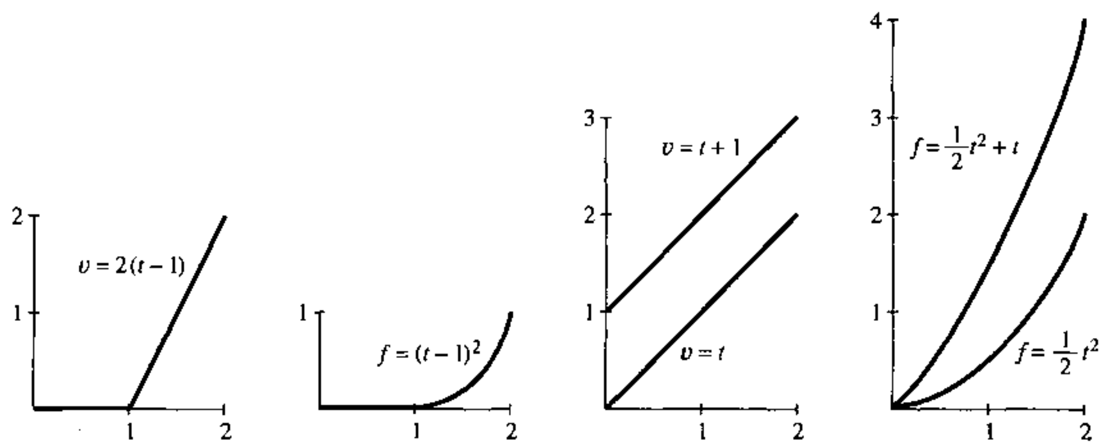


Fig. 1.13 Delayed velocity and distance. The pairs $v = at + b$ and $f = \frac{1}{2}at^2 + bt$.

EXAMPLE 2 The acceleration changes from 2 to another constant a. The velocity changes from $v = 2t$ to $v = at$. ***The acceleration is the slope of the velocity curve!*** The distance is also proportional to a, but notice the factor $\frac{1}{2}$:

$$\text{acceleration } a \quad \Leftrightarrow \quad \text{velocity } v = at \quad \Leftrightarrow \quad \text{distance } f = \tfrac{1}{2}at^2.$$

If a equals 1, then $v = t$ and $f = \frac{1}{2}t^2$. That is one of the most famous pairs in calculus. If a equals the gravitational constant g, then $v = gt$ is the velocity of a falling body. The speed doesn't depend on the mass (tested by Galileo at the Leaning Tower of Pisa). Maybe he saw the distance $f = \frac{1}{2}gt^2$ more easily than the speed $v = gt$. Anyway, this is the most famous pair in physics.

EXAMPLE 3 Suppose $f(t) = 3t + t^2$. The average velocity from t to $t + h$ is

$$v_{\text{ave}} = \frac{f(t+h) - f(t)}{h} = \frac{3(t+h) + (t+h)^2 - 3t - t^2}{h}.$$

The change in distance has an extra $3h$ (coming from $3(t + h)$ minus $3t$). The velocity contains an additional 3 (coming from $3h$ divided by h). When $3t$ is added to the distance, 3 is added to the velocity. If Galileo had thrown a weight instead of dropping it, the starting velocity v_0 would have added $v_0 t$ to the distance.

FUNCTIONS ACROSS TIME

The idea of slope is not difficult—for one straight line. Divide the change in f by the change in t. In Chapter 2, divide the change in y by the change in x. Experience shows that the hard part is to see what happens to the slope as the line moves.

Figure 1.14a shows the line between points A and B on the curve. This is a "secant line." Its slope is an *average* velocity. What calculus does is to bring that point B ***down the curve toward*** A.

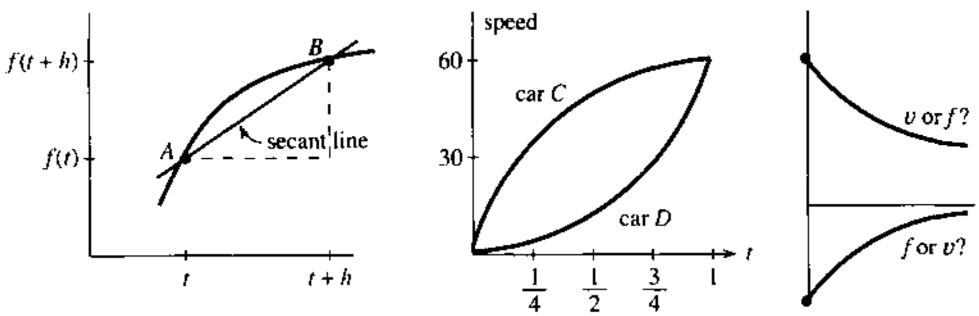


Fig. 1.14 Slope of line, slope of curve. Two velocity graphs. *Which is which?*

Question 1 What happens to the "change in f"—the height of B above A?
Answer The change in f decreases to zero. So does the change in t.

Question 2 As B approaches A, does the slope of the line increase or decrease?
Answer I am not going to answer that question. It is too important. Draw another secant line with B closer to A. Compare the slopes.

This question was created by Steve Monk at the University of Washington—where 57% of the class gave the right answer. Probably 97% would have found the right slope from a formula. Figure 1.14b shows the opposite problem. We know the velocity, not the distance. But calculus answers questions about both functions.

Question 3 Which car is going faster at time $t = 3/4$?
Answer Car C has higher speed. Car D has greater acceleration.

Question 4 If the cars start together, is D catching up to C at the end? Between $t = \frac{1}{2}$ and $t = 1$, do the cars get closer or further apart?
Answer This time more than half the class got it wrong. You won't but you can see why they did. You have to look at the speed graph and imagine the distance graph. When car C is going faster, the distance between them ______.

To repeat: The cars start together, but they don't finish together. They reach the same speed at $t = 1$, not the same distance. Car C went faster. You really should draw their distance graphs, to see how they bend.

These problems help to emphasize one more point. Finding the speed (or slope) is entirely different from finding the distance (or area):

1. To find the ***slope*** of the f-graph at a particular time t, you *don't* have to know the whole history.
2. To find the ***area*** under the v-graph up to a particular time t, you *do* have to know the whole history.

A short record of distance is enough to recover $v(t)$. Point B moves toward point A. The problem of slope is *local*—the speed is completely decided hy $f(t)$ near point A.

In contrast, a short record of speed is ***not enough*** to recover the total distance. We have to know what the mileage was earlier. Otherwise we can only know the *increase* in mileage, not the total.

1.3 EXERCISES

Read-through questions

Between the distances $f(2) = 100$ and $f(6) = 200$, the average velocity is __a__. If $f(t) = \frac{1}{4}t^2$ then $f(6) =$ __b__ and $f(8) =$ __c__. The average velocity in between is __d__. The instantaneous velocities at $t = 6$ and $t = 8$ are __e__ and __f__.

The average velocity is computed from $f(t)$ and $f(t+h)$ by $v_{ave} =$ __g__. If $f(t) = t^2$ then $v_{ave} =$ __h__. From $t = 1$ to $t = 1.1$ the average is __i__. The instantaneous velocity is the __j__ of v_{ave}. If the distance is $f(t) = \frac{1}{2}at^2$ then the velocity is $v(t) =$ __k__ and the acceleration is __l__.

On the graph of $f(t)$, the average velocity between A and B is the slope of __m__. The velocity at A is found by __n__. The velocity at B is found by __o__. When the velocity is positive, the distance is __p__. When the velocity is increasing, the car is __q__.

1 Compute the average velocity between $t = 5$ and $t = 8$:

(a) $f(t) = 6t$ (b) $f(t) = 6t + 2$
(c) $f(t) = \frac{1}{2}at^2$ (d) $f(t) = t - t^2$
(e) $f(t) = 6$ (f) $v(t) = 2t$

2 For the same functions compute $[f(t+h) - f(t)]/h$. This depends on t and h. Find the limit as $h \to 0$.

3 If the odometer reads $f(t) = t^2 + t$ (f in miles or kilometers, t in hours), find the average speed between

(a) $t = 1$ and $t = 2$
(b) $t = 1$ and $t = 1.1$
(c) $t = 1$ and $t = 1 + h$
(d) $t = 1$ and $t = .9$ (note $h = -.1$)

4 For the same $f(t) = t^2 + t$, find the average speed between (a) $t = 0$ and 1 (b) $t = 0$ and $\frac{1}{2}$ (c) $t = 0$ and h.

5 In the answer to 3(c), find the limit as $h \to 0$. What does that limit tell us?

6 Set $h = 0$ in your answer to 4(c). Draw the graph of $f(t) = t^2 + t$ and show its slope at $t = 0$.

7 Draw the graph of $v(t) = 1 + 2t$. From geometry find the area under it from 0 to t. Find the slope of that area function $f(t)$.

8 Draw the graphs of $v(t) = 3 - 2t$ and the area $f(t)$.

9 ***True or false***

(a) If the distance $f(t)$ is positive, so is $v(t)$.
(b) If the distance $f(t)$ is increasing, so is $v(t)$.
(c) If $f(t)$ is positive, $v(t)$ is increasing.
(d) If $v(t)$ is positive, $f(t)$ is increasing.

10 If $f(t) = 6t^2$ find the slope of the f-graph and also the v-graph. The slope of the v-graph is the ________.

11 If $f(t) = t^2$ what is the average velocity between $t = .9$ and $t = 1.1$? What is the average between $t - h$ and $t + h$?

12 (a) Show that for $f(t) = \frac{1}{2}at^2$ the average velocity between $t - h$ and $t + h$ is exactly the velocity at t.

(b) The area under $v(t) = at$ from $t - h$ to $t + h$ is exactly the base $2h$ times ________.

13 Find $f(t)$ from $v(t) = 20t$ if $f(0) = 12$. Also if $f(1) = 12$.

14 ***True or false***, for any distance curves.

(a) The slope of the line from A to B is the average velocity between those points.

(b) Secant lines have smaller slopes than the curve.

(c) If $f(t)$ and $F(t)$ start together and finish together, the average velocities are equal.

(d) If $v(t)$ and $V(t)$ start together and finish together, the increases in distance are equal.

15 When you jump up and fall back your height is $y = 2t - t^2$ in the right units.

(a) Graph this parabola and its slope.

(b) Find the time in the air and maximum height.

(c) *Prove: Half the time you are above* $y = \frac{3}{4}$.

Basketball players "hang" in the air partly because of (c).

16 Graph $f(t) = t^2$ and $g(t) = f(t) - 2$ and $h(t) = f(2t)$, all from $t = 0$ to $t = 1$. Find the velocities.

17 (Recommended) An up and down velocity is $v(t) = 2t$ for $t \leqslant 3$, $v(t) = 12 - 2t$ for $t \geqslant 3$. Draw the piecewise parabola $f(t)$. Check that $f(6) =$ area under the graph of $v(t)$.

18 Suppose $v(t) = t$ for $t \leqslant 2$ and $v(t) = 2$ for $t \geqslant 2$. Draw the graph of $f(t)$ out to $t = 3$.

19 Draw $f(t)$ up to $t = 4$ when $v(t)$ increases linearly from

(a) 0 to 2 (b) -1 to 1 (c) -2 to 0.

20 (Recommended) Suppose $v(t)$ is the piecewise linear sine function of Section 1.2. (In Figure 1.8 it was the distance.) Find the area under $v(t)$ between $t = 0$ and $t = 1, 2, 3, 4, 5, 6$. Plot those points $f(1), \ldots, f(6)$ and draw the complete piecewise parabola $f(t)$.

21 Draw the graph of $f(t) = |1 - t^2|$ for $0 \leqslant t \leqslant 2$. Find a three-part formula for $v(t)$.

22 Draw the graphs of $f(t)$ for these velocities (to $t = 2$):

(a) $v(t) = 1 - t$

(b) $v(t) = |1 - t|$

(c) $v(t) = (1 - t) + |1 - t|$.

23 When does $f(t) = t^2 - 3t$ reach 10? Find the average velocity up to that time and the instantaneous velocity at that time.

24 If $f(t) = \frac{1}{2}at^2 + bt + c$, what is $v(t)$? What is the slope of $v(t)$? When does $f(t)$ equal 41, if $a = b = c = 1$?

25 If $f(t) = t^2$ then $v(t) = 2t$. Does the speeded-up function $f(4t)$ have velocity $v(4t)$ or $4v(t)$ or $4v(4t)$?

26 If $f(t) = t - t^2$ find $v(t)$ and $f(3t)$. Does the slope of $f(3t)$ equal $v(3t)$ or $3v(t)$ or $3v(3t)$?

27 For $f(t) = t^2$ find $v_{ave}(t)$ between 0 and t. Graph $v_{ave}(t)$ and $v(t)$.

28 If you know the average velocity $v_{ave}(t)$, how can you find the distance $f(t)$? Start from $f(0) = 0$.

1.4 Circular Motion

This section introduces completely new distances and velocities—***the sines and cosines from trigonometry***. As I write that last word, I ask myself how much trigonometry it is essential to know. There will be the basic picture of a right triangle, with sides cos t and sin t and 1. There will also be the crucial equation $(\cos t)^2 + (\sin t)^2 = 1$, which is Pythagoras' law $a^2 + b^2 = c^2$. The squares of two sides add to the square of the hypotenuse (and the 1 is really 1^2). Nothing else is needed immediately. If you don't know trigonometry, don't stop—an important part can be learned now.

You will recognize the wavy graphs of the sine and cosine. *We intend to find the slopes of those graphs.* That can be done without using the formulas for $\sin(x + y)$ and $\cos(x + y)$—which later give the same slopes in a more algebraic way. Here it is only basic things that are needed.† And anyway, how complicated can a triangle be?

Remark You might think trigonometry is only for surveyors and navigators (people with triangles). Not at all! By far the biggest applications are to ***rotation*** and ***vibration*** and ***oscillation***. It is fantastic that sines and cosines are so perfect for "repeating motion"—around a circle or up and down.

†Sines and cosines are so important that I added a review of trigonometry in Section 1.5. But the concepts in this section can be more valuable than formulas.

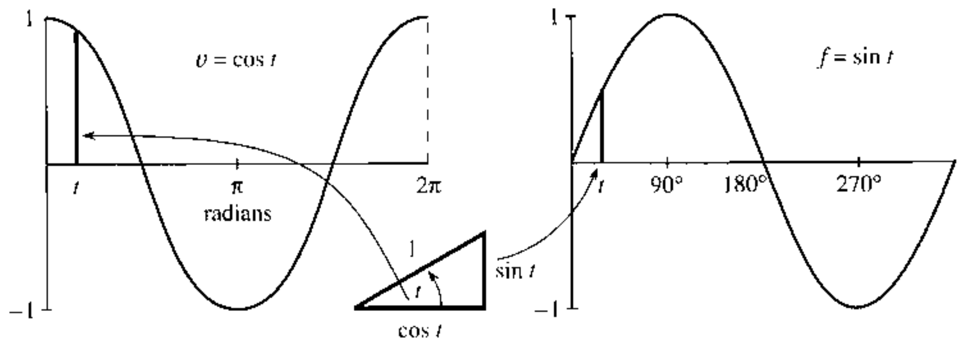


Fig. 1.15 As the angle t changes, the graphs show the sides of the right triangle.

Our underlying goal is to offer one more example in which the velocity can be computed by common sense. Calculus is mainly an extension of common sense, but here that extension is not needed. We will find the slope of the sine curve. The straight line $f = vt$ was easy and the parabola $f = \frac{1}{2}at^2$ was harder. The new example also involves realistic motion, seen every day. We start with ***circular motion***, in which the position is given and the velocity will be found.

A ball goes around a circle of radius one. The center is at $x = 0$, $y = 0$ (the origin). The x and y coordinates satisfy $x^2 + y^2 = 1^2$, to keep the ball on the circle. We specify its position in Figure 1.16a by giving its angle with the horizontal. And we make the ball travel with constant speed, by requiring that ***the angle is equal to the time*** t. The ball goes counterclockwise. At time 1 it reaches the point where the angle equals 1. The angle is measured in ***radians*** rather than degrees, so a full circle is completed at $t = 2\pi$ instead of $t = 360$.

The ball starts on the x axis, where the angle is zero. Now find it at time t:

The ball is at the point where $x = \cos t$ ***and*** $y = \sin t$.

This is where trigonometry is useful. The cosine oscillates between 1 and -1, as the ball goes from far right to far left and back again. The sine also oscillates between 1 and -1, starting from $\sin 0 = 0$. At time $\pi/2$ the sine (the height) increases to one. The cosine is zero and the ball reaches the top point $x = 0$, $y = 1$. At time π the cosine is -1 and the sine is back to zero—the coordinates are $(-1, 0)$. At $t = 2\pi$ the circle is complete (the angle is also 2π), and $x = \cos 2\pi = 1$, $y = \sin 2\pi = 0$.

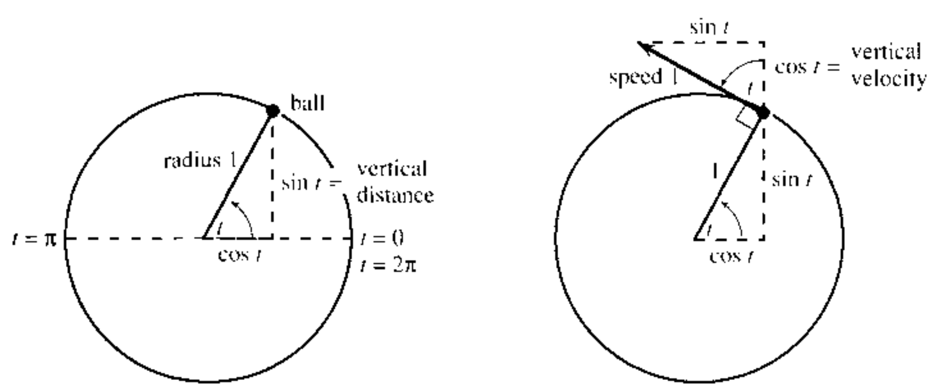


Fig. 1.16 Circular motion with speed 1, angle t, height $\sin t$, upward velocity $\cos t$.

Important point: The distance around the circle (its circumference) is $2\pi r = 2\pi$, because the radius is 1. The ball travels a distance 2π in a time 2π. ***The speed equals*** 1. It remains to find the velocity, which involves not only speed but *direction.*

Degrees vs. radians A full circle is 360 degrees and 2π radians. Therefore

$$1 \text{ radian} = 360/2\pi \text{ degrees} \approx 57.3 \text{ degrees}$$

$$1 \text{ degree} = 2\pi/360 \text{ radians} \approx .01745 \text{ radians}$$

Radians were invented to avoid those numbers! The speed is exactly 1, reaching t radians at time t. The speed would be .01745, if the ball only reached t degrees. The ball would complete the circle at time $T = 360$. We cannot accept the division of the circle into 360 pieces (by whom?), which produces these numbers.

To check degree mode vs. radian mode, verify that $\sin 1° \approx .017$ and $\sin 1 \approx .84$.

VELOCITY OF THE BALL

At time t, which direction is the ball going? Calculus watches the motion between t and $t + h$. For a ball on a string, we don't need calculus—just let go. ***The direction of motion is tangent to the circle.*** With no force to keep it on the circle, *the ball goes off on a tangent.* If the ball is the moon, the force is gravity. If it is a hammer swinging around on a chain, the force is from the center. When the thrower lets go, the hammer takes off—and it is an art to pick the right moment. (I once saw a friend hit by a hammer at MIT. He survived, but the thrower quit track.) Calculus will find that same tangent direction, when the points at t and $t + h$ come close.

The "***velocity triangle***" is in Figure 1.16b. It is the same as the position triangle, but rotated through 90°. The hypotenuse is tangent to the circle, in the direction the ball is moving. Its length equals 1 (the speed). The angle t still appears, but now it is the angle with the vertical. ***The upward component of velocity is*** $\cos t$, ***when the upward component of position is*** $\sin t$. That is our common sense calculation, based on a figure rather than a formula. The rest of this section depends on it—and we check $v = \cos t$ at special points.

At the starting time $t = 0$, the movement is all upward. The height is $\sin 0 = 0$ and the upward velocity is $\cos 0 = 1$. At time $\pi/2$, the ball reaches the top. The height is $\sin \pi/2 = 1$ and the upward velocity is $\cos \pi/2 = 0$. At that instant the ball is not moving up or down.

The horizontal velocity contains a minus sign. At first the ball travels to the *left.* The value of x is $\cos t$, but *the speed in the x direction is* $-\sin t$. Half of trigonometry is in that figure (the good half), and you see how $\sin^2 t + \cos^2 t = 1$ is so basic. That equation applies to position and velocity, at every time.

Application of plane geometry: The right triangles in Figure 1.16 are the same size and shape. They look congruent and they are—the angle t above the ball equals the angle t at the center. That is because the three angles at the ball add to 180°.

OSCILLATION: UP AND DOWN MOTION

We now use circular motion to study *straight-line motion.* That line will be the y axis. Instead of a ball going around a circle, a mass will move up and down. It oscillates between $y = 1$ and $y = -1$. ***The mass is the "shadow of the ball,"*** as we explain in a moment.

There is a jumpy oscillation that we do not want, with $v = 1$ and $v = -1$. That "bang-bang" velocity is like a billiard ball, bouncing between two walls without slowing down. If the distance between the walls is 2, then at $t = 4$ the ball is back to the start. The distance graph is a zigzag (or sawtooth) from Section 1.2.

We prefer a smoother motion. Instead of velocities that jump between $+1$ and -1, a real oscillation *slows down to zero* and gradually builds up speed again. The mass is on a spring, whicb pulls it back. The velocity drops to zero as the spring is fully stretched. Then v is negative, as the mass goes the same distance in the opposite direction. ***Simple harmonic motion*** is the most important back and forth motion, while $f = vt$ and $f = \frac{1}{2}at^2$ are the most important one-way motions.

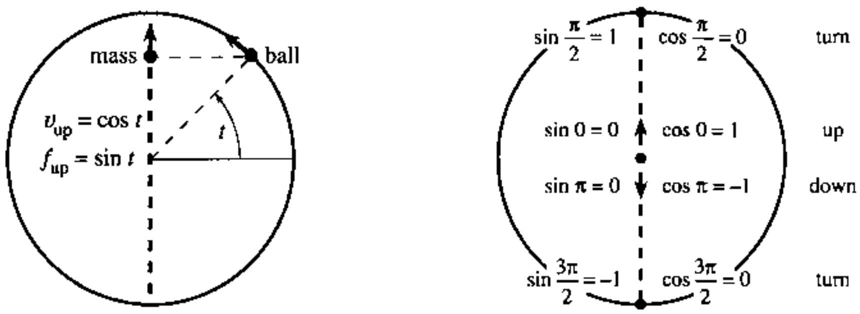


Fig. 1.17 Circular motion of the ball and harmonic motion of the mass (its shadow).

How do we describe this oscillation? The best way is to match it with the ball on the circle. ***The height of the ball will be the height of the mass.*** The "shadow of the hall" goes up and down, level with the ball. As the ball passes the top of the circle, the mass stops at the top and starts down. As the ball goes around the bottom, the mass stops and turns back up the y axis. Halfway up (or down), the speed is 1.

Figure 1.17a shows the mass at a typical time t. The height is $y = f(t) = \sin t$, level with the ball. This height oscillates between $f = 1$ and $f = -1$. But the mass does not move with constant speed. ***The speed of the mass is changing although the speed of the ball is always*** 1. The time for a full cycle is still 2π, but within that cycle the mass speeds up and slows down. The problem is to find the changing velocity v. Since the distance is $f = \sin t$, the velocity will be the *slope of the sine curve.*

THE SLOPE OF THE SINE CURVE

At the top and bottom ($t = \pi/2$ and $t = 3\pi/2$) the ball changes direction and $v = 0$. *The slope at the top and bottom of the sine curve is zero.*† At time zero, when the ball is going straight up, the slope of the sine curve is $v = 1$. At $t = \pi$, when the ball and mass and f-graph are going down, the velocity is $v = -1$. The mass goes fastest at the center. The mass goes slowest (in fact it stops) when the height reaches a maximum or minimum. The velocity triangle yields v at every time t.

To find the upward velocity of the mass, look at the upward velocity of the ball. Those velocities are the same! The mass and ball stay level, and we know v from circular motion: ***The upward velocity is*** $v = \cos t$.

†That looks easy but you will see later that it is extremely important. ***At a maximum or minimum the slope is zero.*** The curve levels off.

Figure 1.18 shows the result we want. On the right, $f = \sin t$ gives the height. On the left is the velocity $v = \cos t$. That velocity is the slope of the f-curve. The height and velocity (red lines) are oscillating together, but they are out of phase—just as the position triangle and velocity triangle were at right angles. This is absolutely fantastic, that in calculus the two most famous functions of trigonometry form a pair: ***The slope of the sine curve is given by the cosine curve.***

When the distance is $f(t) = \sin t$, ***the velocity is*** $v(t) = \cos t$.

Admission of guilt: The slope of $\sin t$ was not computed in the standard way. Previously we compared $(t+h)^2$ with t^2, and divided that distance by h. This average velocity approached the slope $2t$ as h became small. ***For*** $\sin t$ ***we could have done the same***:

$$\text{average velocity} = \frac{\text{change in } \sin t}{\text{change in } t} = \frac{\sin(t+h) - \sin t}{h}. \tag{1}$$

This is where we need the formula for $\sin(t+h)$, coming soon. Somehow the ratio in (1) should approach $\cos t$ as $h \to 0$. (It does.) The sine and cosine fit the same pattern as t^2 and $2t$—our shortcut was to watch the shadow of motion around a circle.

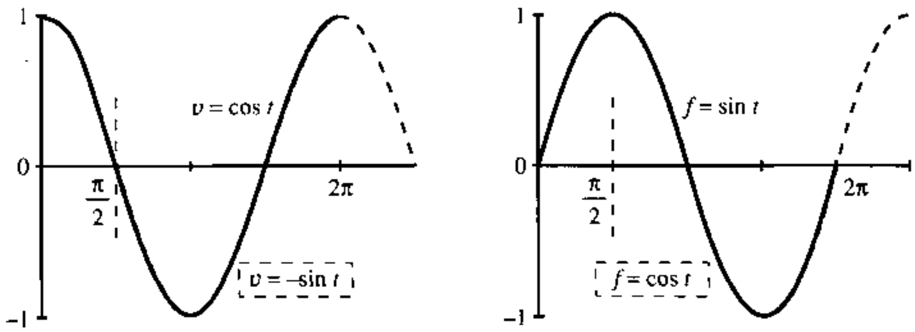


Fig. 1.18 $v = \cos t$ when $f = \sin t$ (red); $v = -\sin t$ when $f = \cos t$ (black).

Question 1 ***What if the ball goes twice as fast, to reach angle $2t$ at time t?***

Answer The speed is now 2. The time for a full circle is only π. The ball's position is $x = \cos 2t$ and $y = \sin 2t$. The velocity is still tangent to the circle—but the tangent is at angle $2t$ where the ball is. Therefore $\cos 2t$ enters the upward velocity and $-\sin 2t$ enters the horizontal velocity. The difference is that ***the velocity triangle is twice as big***. The upward velocity is not $\cos 2t$ but $2\cos 2t$. The horizontal velocity is $-2\sin 2t$. Notice these 2's!

Question 2 ***What is the area under the cosine curve from $t = 0$ to $t = \pi/2$?***

You can answer that, if you accept the Fundamental Theorem of Calculus—***computing areas is the opposite of computing slopes***. The slope of $\sin t$ is $\cos t$, so the area under $\cos t$ is the increase in $\sin t$. No reason to believe that yet, but we use it anyway.

From $\sin 0 = 0$ to $\sin \pi/2 = 1$, the increase is 1. Please realize the power of calculus. No other method could compute the area under a cosine curve so fast.

THE SLOPE OF THE COSINE CURVE

I cannot resist uncovering another distance and velocity (another f-v pair) with no extra work. This time f is the cosine. The time clock starts *at the top of the circle.* The old time $t = \pi/2$ is now $t = 0$. The dotted lines in Figure 1.18 show the new start. But the shadow has exactly the same motion—the hall keeps going around the circle, and the mass follows it up and down. The f-graph and v-graph are still correct, both with a time shift of $\pi/2$.

The new f-graph is the cosine. The new v-graph is ***minus the sine***. The slope of the cosine curve follows the *negative* of the sine curve. That is another famous pair, twins of the first:

When the distance is $f(t) = \cos t$, ***the velocity is*** $v(t) = -\sin t$.

You could see that coming, by watching the ball go left and right (instead of up and down). Its distance across is $f = \cos t$. Its velocity across is $v = -\sin t$. That twin pair completes the calculus in Chapter 1 (trigonometry to come). We review the ideas:

v is the ***velocity***

the ***slope*** of the distance curve

the ***limit*** of average velocity over a short time

the ***derivative*** of f.

f is the ***distance***

the ***area*** under the velocity curve

the ***limit*** of total distance over many short times

the ***integral*** of v.

Differential calculus: *Compute v from f.* ***Integral calculus***: *Compute f from v.*

With constant velocity, f equals vt. With constant acceleration, $v = at$ and $f = \frac{1}{2}at^2$. In harmonic motion, $v = \cos t$ and $f = \sin t$. One part of our goal is to extend that list—for which we need the tools of calculus. Another and more important part is to put these ideas to use.

Before the chapter ends, may I add a note about the book and the course? The book is more personal than usual, and I hope readers will approve. What I write is very close to what I would say, if you were in this room. The sentences are spoken before they are written.† Calculus is alive and moving forward—it needs to be taught that way.

One new part of the subject has come with the computer. It works with a finite step h, not an "infinitesimal" limit. What it can do, it does quickly—even if it cannot find exact slopes or areas. The result is an overwhelming growth in the range of problems that can be solved. We landed on the moon because f and v were so accurate. (The moon's orbit has sines and cosines, the spacecraft starts with $v = at$ and $f = \frac{1}{2}at^2$. Only the computer can account for the atmosphere and the sun's gravity and the changing mass of the spacecraft.) ***Modern mathematics is a combination of exact formulas and approximate computations.*** Neither part can be ignored, and I hope you will see numerically what we derive algebraically. The exercises are to help you master both parts.

†On television you know immediately when the words are live. The same with writing.

The course has made a quick start—not with an abstract discussion of sets or functions or limits, but with the concrete questions that led to those ideas. You have seen a distance function f and a limit v of average velocities. We will meet more functions and more limits (and their definitions!) but it is crucial to study important examples early. There is a lot to do, but the course has definitely begun.

1.4 EXERCISES

Read-through questions

A ball at angle t on the unit circle has coordinates $x =$ __a__ and $y =$ __b__. It completes a full circle at $t =$ __c__. Its speed is __d__. Its velocity points in the direction of the __e__, which is __f__ to the radius coming out from the center. The upward velocity is __g__ and the horizontal velocity is __h__.

A mass going up and down level with the ball has height $f(t) =$ __i__. This is called simple __j__ motion. The velocity is $v(t) =$ __k__. When $t = \pi/2$ the height is $f =$ __l__ and the velocity is $v =$ __m__. If a speeded-up mass reaches $f = \sin 2t$ at time t, its velocity is $v =$ __n__. A shadow traveling *under* the ball has $f = \cos t$ and $v =$ __o__. When f is distance = area = integral, v is __p__ = __q__ = __r__.

1 For a ball going around a unit circle with speed 1,

(a) how long does it take for 5 revolutions?

(b) at time $t = 3\pi/2$ where is the ball?

(c) at $t = 22$ where is the ball (approximately)?

2 For the same motion find the exact x and y coordinates at $t = 2\pi/3$. At what time would the ball hit the x axis, if it goes off on the tangent at $t = 2\pi/3$?

3 A ball goes around a circle of radius 4. At time t (when it reaches angle t) find

(a) its x and y coordinates

(b) the speed and the distance traveled

(c) the vertical and horizontal velocity.

4 On a circle of radius R find the x and y coordinates at time t (and angle t). Draw the velocity triangle and find the x and y velocities.

5 A ball travels around a unit circle (radius 1) with speed 3, starting from angle zero. At time t,

(a) what angle does it reach?

(b) what are its x and y coordinates?

(c) what are its x and y velocities? This part is harder.

6 If another ball stays $\pi/2$ radians ahead of the ball with speed 3, find its angle, its x and y coordinates, and its vertical velocity at time t.

7 A mass moves on the x axis under or over the original ball (on the unit circle with speed 1). What is the position $x = f(t)$? Find x and v at $t = \pi/4$. Plot x and v up to $t = \pi$.

8 Does the new mass (under or over the ball) meet the old mass (level with the ball)? What is the distance between the masses at time t?

9 Draw graphs of $f(t) = \cos 3t$ and $\cos 2\pi t$ and $2\pi \cos t$, marking the time axes. How long until each f repeats?

10 Draw graphs of $f = \sin(t + \pi)$ and $v = \cos(t + \pi)$. This oscillation stays level with what ball?

11 Draw graphs of $f = \sin(\pi/2 - t)$ and $v = -\cos(\pi/2 - t)$. This oscillation stays level with a ball going which way starting where?

12 Draw a graph of $f(t) = \sin t + \cos t$. Estimate its greatest height (maximum f) and the time it reaches that height. By computing f^2 check your estimate.

13 How fast should you run across the circle to meet the ball again? It travels at speed 1.

14 A mass falls from the top of the unit circle when the ball of speed 1 passes by. What acceleration a is necessary to meet the ball at the bottom?

Find the area under $v = \cos t$ from the change in $f = \sin t$:

15 from $t = 0$ to $t = \pi$

16 from $t = 0$ to $t = \pi/6$

17 from $t = 0$ to $t = 2\pi$

18 from $t = \pi/2$ to $t = 3\pi/2$.

19 The distance curve $f = \sin 4t$ yields the velocity curve $v = 4 \cos 4t$. Explain both 4's.

20 The distance curve $f = 2 \cos 3t$ yields the velocity curve $v = -6 \sin 3t$. Explain the -6.

21 The velocity curve $v = \cos 4t$ yields the distance curve $f = \frac{1}{4} \sin 4t$. Explain the $\frac{1}{4}$.

22 The velocity $v = 5 \sin 5t$ yields what distance?

23 Find the slope of the sine curve at $t=\pi/3$ from $v=\cos t$. Then find an average slope by dividing $\sin\pi/2-\sin\pi/3$ by the time difference $\pi/2-\pi/3$.

24 The slope of $f=\sin t$ at $t=0$ is $\cos 0=1$. Compute average slopes $(\sin t)/t$ for $t=1, .1, .01, .001$.

The ball at $x=\cos t$, $y=\sin t$ circles (1) counterclockwise (2) with radius 1 (3) starting from $x=1$, $y=0$ (4) at speed 1. Find (1)(2)(3)(4) for the motions 25–30.

25 $x=\cos 3t$, $y=-\sin 3t$

26 $x=3\cos 4t$, $y=3\sin 4t$

27 $x=5\sin 2t$, $y=5\cos 2t$

28 $x=1+\cos t$, $y=\sin t$

29 $x=\cos(t+1)$, $y=\sin(t+1)$

30 $x=\cos(-t)$, $y=\sin(-t)$

The oscillation $x=0$, $y=\sin t$ goes (1) up and down (2) between -1 and 1 (3) starting from $x=0$, $y=0$ (4) at velocity $v=\cos t$. Find (1)(2)(3)(4) for the oscillations 31–36.

31 $x=\cos t$, $y=0$

32 $x=0$, $y=\sin 5t$

33 $x=0$, $y=2\sin(t+\theta)$

34 $x=\cos t$, $y=\cos t$

35 $x=0$, $y=-2\cos\frac{1}{2}t$

36 $x=\cos^2 t$, $y=\sin^2 t$

37 If the ball on the unit circle reaches t *degrees* at time t, find its position and speed and upward velocity.

38 Choose the number k so that $x=\cos kt$, $y=\sin kt$ completes a rotation at $t=1$. Find the speed and upward velocity.

39 If a pitcher doesn't pause before starting to throw, a balk is called. The American League decided mathematically that there is always a stop between backward and forward motion, even if the time is too short to see it. (Therefore no balk.) Is that true?

1.5 A Review of Trigonometry

Trigonometry begins with a right triangle. The size of the triangle is not as important as the angles. We focus on one particular angle—call it θ—and on the *ratios* between the three sides x, y, r. The ratios don't change if the triangle is scaled to another size. Three sides give six ratios, which are the basic functions of trigonometry:

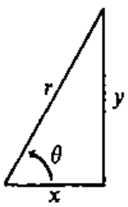


Fig. 1.19

$$\cos\theta=\frac{x}{r}=\frac{\text{near side}}{\text{hypotenuse}}\qquad \sec\theta=\frac{r}{x}=\frac{1}{\cos\theta}$$

$$\sin\theta=\frac{y}{r}=\frac{\text{opposite side}}{\text{hypotenuse}}\qquad \csc\theta=\frac{r}{y}=\frac{1}{\sin\theta}$$

$$\tan\theta=\frac{y}{x}=\frac{\text{opposite side}}{\text{near side}}\qquad \cot\theta=\frac{x}{y}=\frac{1}{\tan\theta}$$

Of course those six ratios are not independent. The three on the right come directly from the three on the left. And the tangent is the sine divided by the cosine:

$$\tan\theta=\frac{\sin\theta}{\cos\theta}=\frac{y/r}{x/r}=\frac{y}{x}.$$

Note that "tangent of an angle" and "tangent to a circle" and "tangent line to a graph" are different uses of the same word. As the cosine of θ goes to zero, the tangent of θ goes to infinity. The side x becomes zero, θ approaches 90°, and the triangle is infinitely steep. The sine of 90° is $y/r=1$.

Triangles have a serious limitation. They are excellent for angles up to 90°, and they are OK up to 180°, but after that they fail. We cannot put a 240° angle into a triangle. Therefore we change now to a circle.

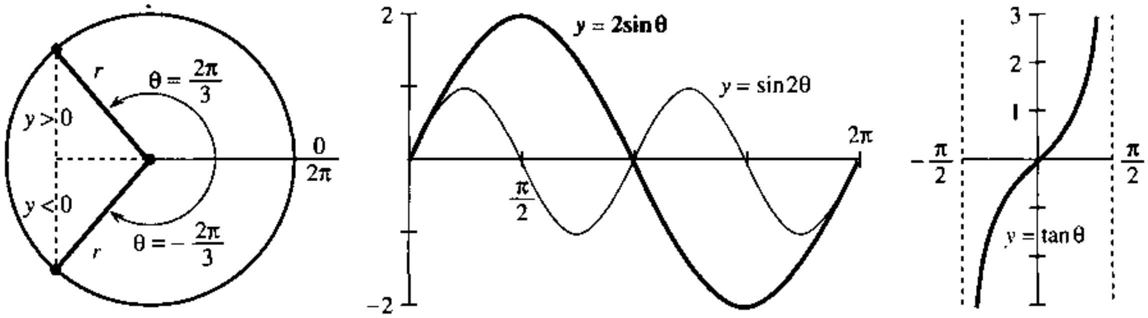


Fig. 1.20 Trigonometry on a circle. Compare $2 \sin \theta$ with $\sin 2\theta$ and $\tan \theta$ (periods 2π, π, π).

Angles are measured from the positive x axis (counterclockwise). Thus 90° is straight up, 180° is to the left, and 360° is in the same direction as 0°. (Then 450° is the same as 90°.) Each angle yields a point on the circle of radius r. The coordinates x and y of that point can be negative (*but never* r). As the point goes around the circle, the six ratios $\cos\theta$, $\sin\theta$, $\tan\theta$, ... trace out six graphs. The cosine waveform is the same as the sine waveform—just shifted by 90°.

One more change comes with the move to a circle. Degrees are out. Radians are in. The distance around the whole circle is $2\pi r$. The distance around to other points is θr. ***We measure the angle by that multiple*** θ. For a half-circle the distance is πr, so the angle is π radians—which is 180°. A quarter-circle is $\pi/2$ radians or 90°. ***The distance around to angle*** θ ***is*** r ***times*** θ.

When $r = 1$ this is the ultimate in simplicity: *The distance is* θ. A 45° angle is $\frac{1}{8}$ of a circle and $2\pi/8$ radians—and the length of the circular arc is $2\pi/8$. Similarly for 1°:

$$360° = 2\pi \text{ radians} \qquad 1° = 2\pi/360 \text{ radians} \qquad 1 \text{ radian} = 360/2\pi \text{ degrees}.$$

An angle going clockwise is *negative*. The angle $-\pi/3$ is $-60°$ and takes us $\frac{1}{6}$ of the *wrong* way around the circle. What is the effect on the six functions?

Certainly the radius r is not changed when we go to $-\theta$. Also x is not changed (see Figure 1.20a). But y reverses sign, because $-\theta$ is below the axis when $+\theta$ is above. This change in y affects y/r and y/x but not x/r:

$$\cos(-\theta) = \cos\theta \qquad \sin(-\theta) = -\sin\theta \qquad \tan(-\theta) = -\tan\theta.$$

The cosine is *even* (no change). The sine and tangent are *odd* (change sign).

The same point is $\frac{5}{6}$ of the *right* way around. Therefore $\frac{5}{6}$ of 2π radians (or 300°) gives the same direction as $-\pi/3$ radians or $-60°$. ***A difference of*** 2π ***makes no difference to*** x, y, r. Thus $\sin\theta$ and $\cos\theta$ and the other four functions have period 2π. We can go five times or a hundred times around the circle, adding 10π or 200π to the angle, and the six functions repeat themselves.

EXAMPLE Evaluate the six trigonometric functions at $\theta = 2\pi/3$ (or $\theta = -4\pi/3$).

This angle is shown in Figure 1.20a (where $r = 1$). The ratios are

$$\cos\theta = x/r = -1/2 \qquad \sin\theta = y/r = \sqrt{3}/2 \qquad \tan\theta = y/x = -\sqrt{3}$$

$$\sec\theta = -2 \qquad \csc\theta = 2/\sqrt{3} \qquad \cot\theta = -1/\sqrt{3}$$

Those numbers illustrate basic facts about the sizes of four functions:

$$|\cos\theta| \leqslant 1 \qquad |\sin\theta| \leqslant 1 \qquad |\sec\theta| \geqslant 1 \qquad |\csc\theta| \geqslant 1.$$

The tangent and cotangent can fall anywhere, as long as $\cot\theta = 1/\tan\theta$.

The numbers reveal more. The tangent $-\sqrt{3}$ is the ratio of sine to cosine. The secant -2 is $1/\cos\theta$. Their squares are 3 and 4 (differing by 1). That may not seem remarkable, but it is. There are three relationships in the *squares* of those six numbers, and they are the key identities of trigonometry:

$$\cos^2\theta + \sin^2\theta = 1 \qquad 1 + \tan^2\theta = \sec^2\theta \qquad \cot^2\theta + 1 = \csc^2\theta$$

Everything flows from the Pythagoras formula $x^2 + y^2 = r^2$. Dividing by r^2 gives $(x/r)^2 + (y/r)^2 = 1$. That is $\cos^2\theta + \sin^2\theta = 1$. Dividing by x^2 gives the second identity, which is $1 + (y/x)^2 = (r/x)^2$. Dividing by y^2 gives the third. All three will be needed throughout the book—and the first one has to be unforgettable.

DISTANCES AND ADDITION FORMULAS

To compute the distance between points we stay with Pythagoras. The points are in Figure 1.21a. They are known by their x and y coordinates, and d is the distance between them. The third point completes a right triangle.

For the x distance along the bottom we don't need help. It is $x_2 - x_1$ (or $|x_2 - x_1|$ since distances can't be negative). The distance up the side is $|y_2 - y_1|$. Pythagoras immediately gives the distance d:

$$\textit{distance between points} = d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}. \qquad (1)$$

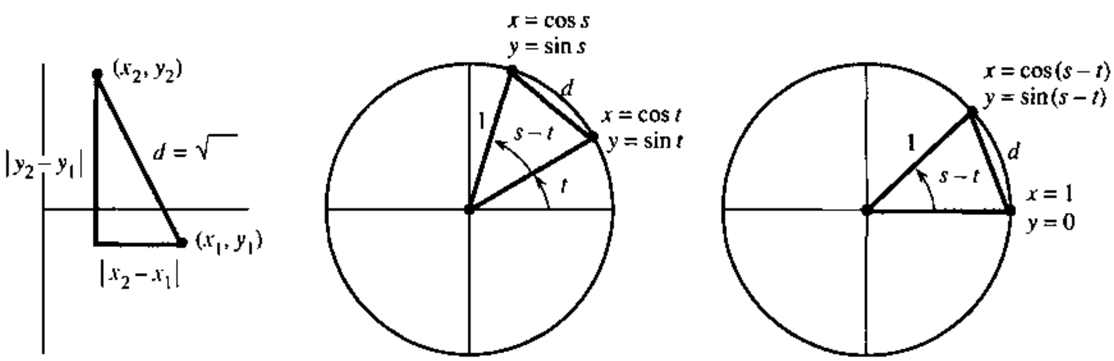


Fig. 1.21 Distance between points and equal distances in two circles.

By applying this distance formula in two identical circles, we discover the cosine of $s - t$. (Subtracting angles is important.) In Figure 1.21b, the distance squared is

$$\begin{aligned} d^2 &= (\text{change in } x)^2 + (\text{change in } y)^2 \\ &= (\cos s - \cos t)^2 + (\sin s - \sin t)^2. \qquad (2) \end{aligned}$$

Figure 1.21c shows the same circle and triangle (but rotated). The same distance squared is

$$d^2 = (\cos(s - t) - 1)^2 + (\sin(s - t))^2. \qquad (3)$$

Now multiply out the squares in equations (2) and (3). Whenever $(\text{cosine})^2 + (\text{sine})^2$ appears, replace it by 1. The distances are the same, so (2) = (3):

$$(2) = 1 + 1 - 2\cos s\cos t - 2\sin s\sin t$$

$$(3) = 1 + 1 - 2\cos(s - t).$$

After canceling $1+1$ and then -2, we have the "***addition formula***" for $\cos(s-t)$:

$$\text{The cosine of } s-t \text{ equals } \cos s \cos t + \sin s \sin t. \tag{4}$$

$$\text{The cosine of } s+t \text{ equals } \cos s \cos t - \sin s \sin t. \tag{5}$$

The easiest is $t=0$. Then $\cos t = 1$ and $\sin t = 0$. The equations reduce to $\cos s = \cos s$.

To go from (4) to (5) in all cases, replace t by $-t$. No change in $\cos t$, but a "minus" appears with the sine. In the special case $s=t$, we have $\cos(t+t) = (\cos t)(\cos t) - (\sin t)(\sin t)$. This is a much-used formula for $\cos 2t$:

$$\textbf{\textit{Double angle}}\text{: } \cos 2t = \cos^2 t - \sin^2 t = 2\cos^2 t - 1 = 1 - 2\sin^2 t. \tag{6}$$

I am constantly using $\cos^2 t + \sin^2 t = 1$, to switch between sines and cosines.

We also need addition formulas and double-angle formulas for the *sine* of $s-t$ and $s+t$ and $2t$. For that we connect sine to cosine, rather than $(\text{sine})^2$ to $(\text{cosine})^2$. The connection goes back to the ratio y/r in our original triangle. This is the sine of the angle θ and also the cosine of the *complementary angle* $\pi/2-\theta$:

$$\sin\theta = \cos(\pi/2-\theta) \qquad \text{and} \qquad \cos\theta = \sin(\pi/2-\theta). \tag{7}$$

The complementary angle is $\pi/2-\theta$ because the two angles add to $\pi/2$ (a right angle). By making this connection in Problem 19, formulas (4–5–6) move from cosines to sines:

$$\sin(s-t) = \sin s \cos t - \cos s \sin t \tag{8}$$

$$\sin(s+t) = \sin s \cos t + \cos s \sin t \tag{9}$$

$$\sin 2t = \sin(t+t) = 2 \sin t \cos t \tag{10}$$

I want to stop with these ten formulas, even if more are possible. Trigonometry is full of identities that connect its six functions—basically because all those functions come from a single right triangle. The x, y, r ratios and the equation $x^2+y^2=r^2$ can be rewritten in many ways. But you have now seen the formulas that are needed by calculus.† They give derivatives in Chapter 2 and integrals in Chapter 5. And it is typical of our subject to add something of its own—a limit in which an angle approaches zero. *The essence of calculus is in that limit.*

Review of the ten formulas Figure 1.22 shows $d^2 = (0-\frac{1}{2})^2 + (1-\sqrt{3}/2)^2$.

$$\cos\frac{\pi}{6} = \cos\frac{\pi}{2}\cos\frac{\pi}{3} + \sin\frac{\pi}{2}\sin\frac{\pi}{3} \qquad (s-t) \qquad \sin\frac{\pi}{6} = \sin\frac{\pi}{2}\cos\frac{\pi}{3} - \cos\frac{\pi}{2}\sin\frac{\pi}{3}$$

$$\cos\frac{5\pi}{6} = \cos\frac{\pi}{2}\cos\frac{\pi}{3} - \sin\frac{\pi}{2}\sin\frac{\pi}{3} \qquad (s+t) \qquad \sin\frac{5\pi}{6} = \sin\frac{\pi}{2}\cos\frac{\pi}{3} + \cos\frac{\pi}{2}\sin\frac{\pi}{3}$$

$$\cos 2\frac{\pi}{3} = \cos^2\frac{\pi}{3} - \sin^2\frac{\pi}{3} \qquad (2t) \qquad \sin 2\frac{\pi}{3} = 2\sin\frac{\pi}{3}\cos\frac{\pi}{3}$$

$$\cos\frac{\pi}{6} = \sin\frac{\pi}{3} = \sqrt{3}/2 \qquad \left(\frac{\pi}{2}-\theta\right) \qquad \sin\frac{\pi}{6} = \cos\frac{\pi}{3} = 1/2$$

†Calculus turns (6) around to $\cos^2 t = \frac{1}{2}(1+\cos 2t)$ and $\sin^2 t = \frac{1}{2}(1-\cos 2t)$.

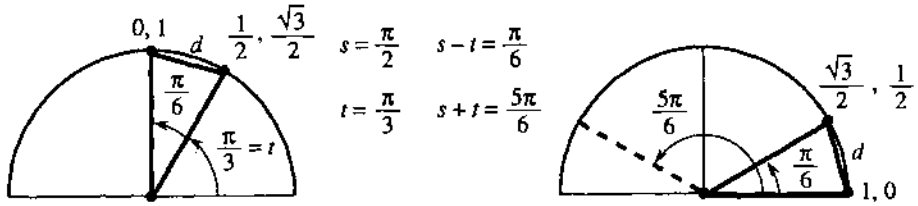


Fig. 1.22

1.5 EXERCISES

Read-through questions

Starting with a __a__ triangle, the six basic functions are the __b__ of the sides. Two ratios (the cosine x/r and the __c__) are below 1. Two ratios (the secant r/x and the __d__) are above 1. Two ratios (the __e__ and the __f__) can take any value. The six functions are defined for all angles θ, by changing from a triangle to a __g__.

The angle θ is measured in __h__. A full circle is $\theta =$ __i__, when the distance around is $2\pi r$. The distance to angle θ is __j__. All six functions have period __k__. Going clockwise changes the sign of θ and __l__ and __m__. Since $\cos(-\theta) = \cos\theta$, the cosine is __n__.

Coming from $x^2 + y^2 = r^2$ are the three identities $\sin^2\theta + \cos^2\theta = 1$ and __o__ and __p__. (Divide by r^2 and __q__ and __r__.) The distance from (2, 5) to (3, 4) is $d =$ __s__. The distance from (1, 0) to $(\cos(s-t), \sin(s-t))$ leads to the addition formula $\cos(s-t) =$ __t__. Changing the sign of t gives $\cos(s+t) =$ __u__. Choosing $s = t$ gives $\cos 2t =$ __v__ or __w__. Therefore $\frac{1}{2}(1 + \cos 2t) =$ __x__, a formula needed in calculus.

1 In a 60–60–60 triangle show why $\sin 30° = \frac{1}{2}$.

2 Convert π, 3π, $-\pi/4$ to degrees and 60°, 90°, 270° to radians. What angles between 0 and 2π correspond to $\theta = 480°$ and $\theta = -1°$?

3 Draw graphs of $\tan\theta$ and $\cot\theta$ from 0 to 2π. What is their (shortest) period?

4 Show that $\cos 2\theta$ and $\cos^2\theta$ have period π and draw them on the same graph.

5 At $\theta = 3\pi/2$ compute the six basic functions and check $\cos^2\theta + \sin^2\theta$, $\sec^2\theta - \tan^2\theta$, $\csc^2\theta - \cot^2\theta$.

6 Prepare a table showing the values of the six basic functions at $\theta = 0, \pi/4, \pi/3, \pi/2, \pi$.

7 The area of a circle is πr^2. What is the area of the sector that has angle θ? It is a fraction ______ of the whole area.

8 Find the distance from (1, 0) to (0, 1) along (a) a straight line (b) a quarter-circle (c) a semicircle centered at $(\frac{1}{2}, \frac{1}{2})$.

9 Find the distance d from (1, 0) to $(\frac{1}{2}, \sqrt{3}/2)$ and show on a circle why $6d$ is less than 2π.

10 In Figure 1.22 compute d^2 and (with calculator) $12d$. Why is $12d$ close to and below 2π?

11 Decide whether these equations are true or false:

(a) $\dfrac{\sin\theta}{1 - \cos\theta} = \dfrac{1 + \cos\theta}{\sin\theta}$

(b) $\dfrac{\sec\theta + \csc\theta}{\tan\theta + \cot\theta} = \sin\theta + \cos\theta$

(c) $\cos\theta - \sec\theta = \sin\theta\tan\theta$

(d) $\sin(2\pi - \theta) = \sin\theta$

12 Simplify $\sin(\pi - \theta)$, $\cos(\pi - \theta)$, $\sin(\pi/2 + \theta)$, $\cos(\pi/2 + \theta)$.

13 From the formula for $\cos(2t + t)$ find $\cos 3t$ in terms of $\cos t$.

14 From the formula for $\sin(2t + t)$ find $\sin 3t$ in terms of $\sin t$.

15 By averaging $\cos(s - t)$ and $\cos(s + t)$ in (4–5) find a formula for $\cos s \cos t$. Find a similar formula for $\sin s \sin t$.

16 Show that $(\cos t + i\sin t)^2 = \cos 2t + i\sin 2t$, if $i^2 = -1$.

17 Draw $\cos\theta$ and $\sec\theta$ on the same graph. Find all points where $\cos\theta = \sec\theta$.

18 Find all angles s and t between 0 and 2π where $\sin(s + t) = \sin s + \sin t$.

19 Complementary angles have $\sin\theta = \cos(\pi/2 - \theta)$. Write $\sin(s + t)$ as $\cos(\pi/2 - s - t)$ and apply formula (4) with $\pi/2 - s$ instead of s. In this way derive the addition formula (9).

20 If formula (9) is true, how do you prove (8)?

21 Check the addition formulas (4–5) and (8–9) for $s = t = \pi/4$.

22 Use (5) and (9) to find a formula for $\tan(s + t)$.

In 23–28 find *every* θ that satisfies the equation.

23 $\sin\theta = -1$

24 $\sec\theta = -2$

25 $\sin\theta = \cos\theta$

26 $\sin\theta = \theta$

27 $\sec^2\theta + \csc^2\theta = 1$

28 $\tan\theta = 0$

29 Rewrite $\cos\theta + \sin\theta$ as $\sqrt{2}\sin(\theta+\phi)$ by choosing the correct "phase angle" ϕ. (Make the equation correct at $\theta = 0$. Square both sides to check.)

30 Match $a\sin x + b\cos x$ with $A\sin(x+\phi)$. From equation (9) show that $a = A\cos\phi$ and $b = A\sin\phi$. Square and add to find $A =$ ________. Divide to find $\tan\phi = b/a$.

31 Draw the base of a triangle from the origin $O = (0, 0)$ to $P = (a, 0)$. The third corner is at $Q = (b\cos\theta, b\sin\theta)$. What are the side lengths OP and OQ? From the distance formula (1) show that the side PQ has length

$$d^2 = a^2 + b^2 - 2ab\cos\theta \quad \text{(law of cosines).}$$

32 Extend the same triangle to a parallelogram with its fourth corner at $R = (a + b\cos\theta, b\sin\theta)$. Find the length squared of the other diagonal OR.

Draw graphs for equations 33–36, and mark three points.

33 $y = \sin 2x$

34 $y = 2\sin \pi x$

35 $y = \frac{1}{2}\cos 2\pi x$

36 $y = \sin x + \cos x$

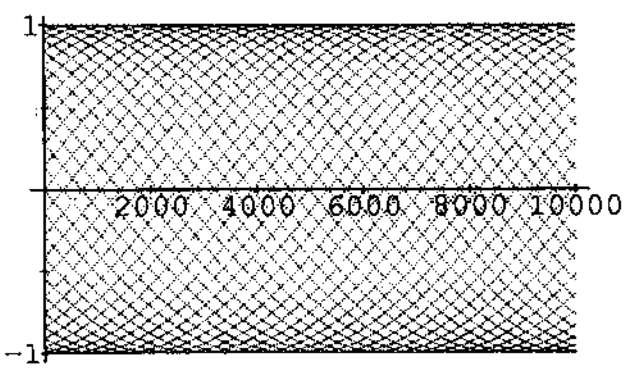
37 Which of the six trigonometric functions are infinite at what angles?

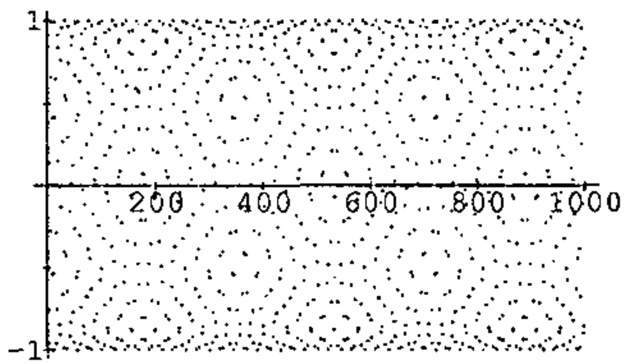
38 Draw rough graphs or computer graphs of $t\sin t$ and $\sin 4t \sin t$ from 0 to 2π.

1.6 A Thousand Points of Light

The graphs on the back cover of the book show $y = \sin n$. This is very different from $y = \sin x$. The graph of $\sin x$ is one continuous curve. By the time it reaches $x = 10{,}000$, the curve has gone up and down $10{,}000/2\pi$ times. Those 1591 oscillations would be so crowded that you couldn't see anything. The graph of $\sin n$ has picked 10,000 points from the curve—and for some reason those points seem to lie on more than 40 separate sine curves.

The second graph shows the first 1000 points. They *don't* seem to lie on sine curves. Most people see hexagons. *But they are the same thousand points*! It is hard to believe that the graphs are the same, but I have learned what to do. ***Tilt the second graph and look from the side at a narrow angle.*** Now the first graph appears. You see "diamonds." The narrow angle compresses the x axis—back to the scale of the first graph.





The effect of scale is something we don't think of. We understand it for maps. Computers can zoom in or zoom out—those are changes of scale. What our eyes see depends on what is "close." We think we see sine curves in the 10,000 point graph, and they raise several questions:

1. Which points are near (0, 0)?
2. How many sine curves are there?
3. Where does the middle curve, going upward from (0, 0), come back to zero?

A point near (0, 0) really means that $\sin n$ is close to zero. That is certainly not true of sin 1 (1 is one radian!). In fact sin 1 is up the axis at .84, at the start of the seventh sine curve. Similarly sin 2 is .91 and sin 3 is .14. (The numbers 3 and .14 make us think of π. The sine of 3 equals the sine of $\pi - 3$. Then sin .14 is near .14.) Similarly sin 4, sin 5, ..., sin 21 are not especially close to zero.

The first point to come close is sin 22. This is because 22/7 is near π. Then 22 is close to 7π, whose sine is zero:

$$\sin 22 = \sin(7\pi - 22) \approx \sin(-.01) \approx -.01.$$

That is the first point to the right of (0, 0) and slightly below. You can see it on graph 1, and more clearly on graph 2. It begins a curve downward.

The next point to come close is sin 44. This is because 44 is just past 14π.

$$44 \approx 14\pi + .02 \qquad \text{so} \qquad \sin 44 \approx \sin .02 \approx .02.$$

This point (44, sin 44) ***starts the middle sine curve***. Next is (88, sin 88).

Now we know something. ***There are*** 44 ***curves***. They begin near the heights sin 0, sin 1, ..., sin 43. Of these 44 curves, 22 start upward and 22 start downward. I was confused at first, because I could only find 42 curves. The reason is that sin 11 equals −0.99999 and sin 33 equals .9999. Those are so close to the bottom and top that you can't see their curves. The sine of 11 is near −1 because sin 22 is near zero. It is almost impossible to follow a single curve past the top—coming back down it is not the curve you think it is.

The points on the middle curve are at $n = 0$ and 44 and 88 and every number $44N$. Where does that curve come back to zero? In other words, when does $44N$ come *very close* to a multiple of π? We know that 44 is $14\pi + .02$. More exactly 44 is $14\pi + .0177$. So we multiply .0177 until we reach π:

$$\text{if} \quad N = \pi/.0177 \quad \text{then} \quad 44N = (14\pi + .0177)N = 14\pi N + \pi.$$

This gives $N = 177.5$. At that point $44N = 7810$. *This is half the period of the sine curve*. The sine of 7810 is very near zero.

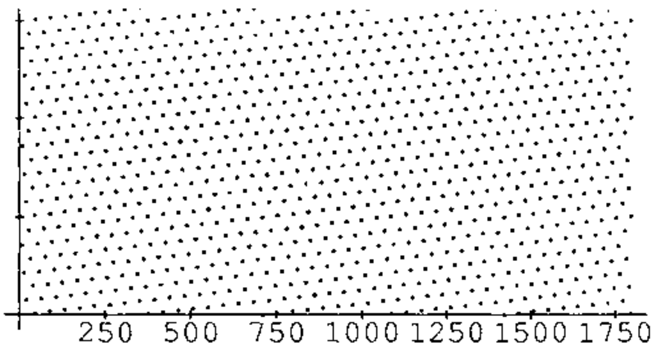
If you follow the middle sine curve, you will see it come back to zero above 7810. The actual points on that curve have $n = 44 \cdot 177$ and $n = 44 \cdot 178$, with sines just above and below zero. Halfway between is $n = 7810$. ***The equation for the middle sine curve is*** $y = \sin(\pi x/7810)$. Its period is 15,620—beyond our graph.

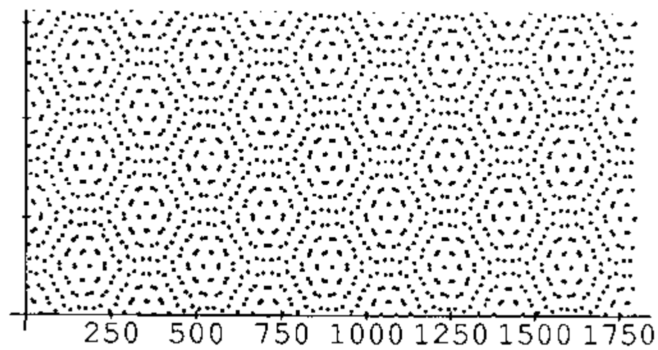
Question The fourth point on that middle curve looks the same as the fourth point coming down from sin 3. What is this "double point?"

Answer 4 times 44 is 176. On the curve going up, the point is (176, sin 176). On the curve coming down it is (179, sin 179). ***The sines of*** 176 ***and*** 179 ***differ only by*** .00003.

The second graph spreads out this double point. Look above 176 and 179, at the center of a hexagon. You can follow the sine curve all the way across graph 2.

Only a little question remains. Why does graph 2 have hexagons? *I don't know*. The problem is with your eyes. To understand the hexagons, Doug Hardin plotted points on straight lines as well as sine curves. Graph 3 shows y = fractional part of $n/2\pi$. Then he made a second copy, turned it over, and placed it on top. That produced graph 4—with hexagons. Graphs 3 and 4 are on the next page.





This is called a ***Moiré pattern***. If you can get a transparent copy of graph 3, and turn it slowly over the original, you will see fantastic hexagons. They come from interference between periodic patterns—in our case 44/7 and 25/4 and 19/3 are near 2π. This interference is an enemy of printers, when color screens don't line up. It can cause vertical lines on a TV. Also in making cloth, operators get dizzy from seeing Moiré patterns move. There are good applications in engineering and optics—but we have to get back to calculus.

1.7 Computing in Calculus

Software is available for calculus courses—a lot of it. The packages keep getting better. Which program to use (if any) depends on cost and convenience and purpose. *How* to use it is a much harder question. These pages identify some of the goals, and also particular packages and calculators. Then we make a beginning (this is still Chapter 1) on the connection of computing to calculus.

The discussion will be informal. It makes no sense to copy the manual. Our aim is to support, with examples and information, the effort to use computing to help learning.

For calculus, ***the greatest advantage of the computer is to offer graphics.*** You see the function, not just the formula. As you watch, $f(x)$ reaches a maximum or a minimum or zero. A separate graph shows its derivative. Those statements are not 100% true, as everybody learns right away—as soon as a few functions are typed in. But the power to *see this subject* is enormous, because it is adjustable. If we don't like the picture we change to a new viewing window.

This is computer-based graphics. It combines ***numerical*** computation with ***graphical*** computation. You get pictures as well as numbers—a powerful combination. The computer offers the experience of actually working with a function. The domain and range are not just abstract ideas. *You choose them.* May I give a few examples.

EXAMPLE 1 Certainly x^3 equals 3^x when $x = 3$. ***Do those graphs ever meet again?*** At this point we don't know the full meaning of 3^x, except when x is a nice number. (Neither does the computer.) Checking at $x = 2$ and 4, the function x^3 is smaller both times: 2^3 is below 3^2 and $4^3 = 64$ is below $3^4 = 81$. If x^3 is always less than 3^x we ought to know—these are among the basic functions of mathematics.

The computer will answer numerically or graphically. At our command, it solves $x^3 = 3^x$. At another command, it plots both functions—this shows more. The screen proves a point of logic (or mathematics) that escaped us. If the graphs cross once, they must cross again—because 3^x is higher at 2 and 4. A crossing point near 2.5 is seen by zooming in. I am less interested in the exact number than its position—it comes before $x = 3$ rather than after.

A few conclusions from such a basic example:

1. A supercomputer is not necessary.
2. High-level programming is not necessary.
3. We can do mathematics without completely understanding it.

The third point doesn't sound so good. Write it differently: ***We can learn mathematics while doing it.*** The hardest part of teaching calculus is to turn it from a spectator sport into a workout. The computer makes that possible.

EXAMPLE 2 (mental computer) Compare x^2 with 2^x. The functions meet at $x = 2$. Where do they meet again? Is it before or after 2?

That is mental computing because the answer happens to be a whole number (4). Now we are on a different track. Does an accident like $2^4 = 4^2$ ever happen again? Can the machine tell us about integers? Perhaps it can plot the solutions of $x^b = b^x$. I asked *Mathematica* for a formula, hoping to discover x as a function of b—but the program just gave back the equation. For once the machine typed HELP instead of the user.

Well, mathematics is not helpless. I am proud of calculus. There is a new exercise at the end of Section 6.4, to show that we never see whole numbers again.

EXAMPLE 3 Find the number b for which $x^b = b^x$ has only ***one*** solution (at $x = b$).

When b is 3, the second solution is below 3. When b is 2, the second solution (4) is above 2. If we move b from 2 to 3, there must be a special "double point"—where the graphs barely touch but don't cross. For that particular b—and only for that one value—the curve x^b never goes above b^x.

This special point b can be found with computer-based graphics. In many ways it is the "***center point of calculus.***" Since the curves touch but don't cross, they are tangent. They have the same slope at the double point. Calculus was created to work with slopes, and we already know the slope of x^2. Soon comes x^b. Eventually we discover the slope of b^x, and identify the most important number in calculus.

The point is that this number can be discovered first by experiment.

EXAMPLE 4 Graph $y(x) = e^x - x^e$. Locate its minimum.

The next example was proposed by Don Small. Solve $x^4 - 11x^3 + 5x - 2 = 0$. The first tool is algebra—try to factor the polynomial. That succeeds for quadratics, and then gets extremely hard. Even if the computer can do algebra better than we can, factoring is seldom the way to go. In reality we have two good choices:

1. (***Mathematics***) Use the derivative. Solve by Newton's method.
2. (***Graphics***) *Plot the function and zoom in.*

Both will be done by the computer. Both have potential problems! Newton's method is fast, but that means it can fail fast. (It is usually terrific.) Plotting the graph is also fast—but solutions can be outside the viewing window. This particular function is zero only once, in the standard window from -10 to 10. The graph seems to be leaving zero, but mathematics again predicts a second crossing point. So we zoom out before we zoom in.

The use of the zoom is the best part of graphing. Not only do we *choose* the domain and range, we *change* them. The viewing window is controlled hy four numbers. They can be the limits $A \leqslant x \leqslant B$ and $C \leqslant y \leqslant D$. They can be the coordinates of two opposite corners: (A, C) and (B, D). They can be the center position (a, b) and the scale factors c and d. Clicking on opposite corners of the zoom box is the fastest way, unless the center is unchanged and we only need to give scale factors. (Even faster: Use the default factors.) Section 3.4 discusses the ***centering transform*** and ***zoom transform***—a change of picture on the screen and a change of variable within the function.

EXAMPLE 5 Find all real solutions to $x^4 - 11x^3 + 5x - 2 = 0$.

EXAMPLE 6 Zoom out and in on the graphs of $y = \cos 40x$ and $y = x \sin(1/x)$. Describe what you see.

EXAMPLE 7 What does $y = (\tan x - \sin x)/x^3$ become at $x = 0$? For small x the machine eventually can't separate $\tan x$ from $\sin x$. It may give $y = 0$. Can you get close enough to see the limit of y?

For these examples, and for most computer exercises in this book, a ***menu-driven*** system is entirely adequate. There is a list of commands to choose from. The user provides a formula for $y(x)$, and many functions are built in. A calculus supplement can be very useful—*MicroCalc* or *True* BASIC or *Exploring Calculus* or *MPP* (in the public domain). Specific to graphics are *Surface Plotter* and *Master Grapher* and *Gyrographics* (animated). The best software for linear algebra is *MATLAB*.

Powerful packages are increasing in convenience and decreasing in cost. They are capable of ***symbolic*** computation—which opens up a third avenue of computing in calculus.

SYMBOLIC COMPUTATION

In symbolic computation, answers can be *formulas* as well as numbers and graphs. The derivative of $y = x^2$ is seen as "$2x$." The derivative of $\sin t$ is "$\cos t$." The slope of b^x is known to the program. The computer does more than substitute numbers into formulas—it operates directly on the formulas. We need to think where this fits with learning calculus.

In a way, symbolic computing is close to what we ourselves do. Maybe too close—there is some danger that symbolic manipulation is *all* we do. With a higher-level language and enough power, a computer can print the derivative of $\sin(x^2)$. So why learn the chain rule? Because mathematics goes deeper than "algebra with formulas." We deal with *ideas*.

I want to say clearly: Mathematics is not formulas, or computations, or even proofs, but ideas. The symbols and pictures are the language. The book and the professor and the computer can join in teaching it. The computer should be non-threatening (like this book and your professor)—you can work at your own pace. Your part is to learn by doing.

EXAMPLE 8 A computer algebra system quickly finds 100 factorial. This is $100! = (100)(99)(98) \ldots (1)$. The number has 158 digits (not written out here). The last 24 digits are zeros. For $10! = 3628800$ there are seven digits and two zeros. Between 10 and 100, and beyond, are simple questions that need ideas:

1. How many digits (approximately) are in the number $N!$?
2. How many zeros (exactly) are at the end of $N!$?

For Question 1, the computer shows more than N digits when $N = 100$. It will never show more than N^2 digits, because none of the N terms can have more than N digits. A much tighter bound would be $2N$, but is it true? *Does $N!$ always have fewer than $2N$ digits*?

For Question 2, the zeros in 10! can be explained. One comes from 10, the other from 5 times 2. (10 is also 5 times 2.) Can you explain the 24 zeros in 100!? An idea from the card game blackjack applies here too: *Count the fives.*

Hard question: How many zeros at the end of 200!?

The outstanding package for full-scale symbolic computation is *Mathematica*. It was used to draw graphs for this book, including $y = \sin n$ on the back cover. *The complete command was* ListPlot[Table [Sin[n], {n, 10000}]]. This system has rewards and also drawbacks, including the price. Its original purpose, like *MathCAD* and *MACSYMA* and *REDUCE*, was not to teach calculus—but it can. The computer algebra system *MAPLE* is good.

As I write in 1990, DERIVE is becoming well established for the PC. For the Macintosh, *Calculus T/L* is a "sleeper" that deserves to be widely known. It builds on *MAPLE* and is much more accessible for calculus. An important alternative is *Theorist*. These are menu-driven (therefore easier at the start) and not expensive.

I strongly recommend that students share terminals and work together. Two at a terminal and 3–5 in a working group seems to be optimal. Mathematics can be learned by *talking* and *writing*—it is a human activity. Our goal is not to test but to teach and learn.

Writing in Calculus May I emphasize the importance of writing? We totally miss it, when the answer is just a number. A one-page report is harder on instructors as well as students—but much more valuable. A word processor keeps it neat. You can't write sentences without being forced to organize ideas—and part of yourself goes into it.

I will propose a writing exercise with options. If you have computer-based graphing, follow through on Examples 1–4 above and report. Without a computer, pick a paragraph from this book that should be clearer and *make it clearer*. Rewrite it with examples. Identify the key idea at the start, explain it, and come back to express it differently at the end. Ideas are like surfaces—they can be seen many ways.

Every reader will understand that in software there is no last word. New packages keep coming (*Analyzer* and *EPIC* among them). The biggest challenges at this moment are three-dimensional graphics and calculus workbooks. In $3D$, the problem is the position of the eye—since the screen is only $2D$. In workbooks, the problem is to get past symbol manipulation and reach ideas. Every teacher, including this one, knows how hard that is and hopes to help.

GRAPHING CALCULATORS

The most valuable feature for calculus—*computer-based graphing*—is available on hand calculators. With trace and zoom their graphs are quite readable. By creating the graphs you subconsciously learn about functions. These are genuinely *personal* computers, and the following pages aim to support and encourage their use.

Programs for a hand-held machine tend to be simple and short. We don't count the zeros in 100 factorial (probably we could). A calculator finds crossing points and maximum points to good accuracy. Most of all it allows you to explore calculus by yourself. You set the viewing window and define the function. Then you see it.

There is a choice of calculators—*which one to buy*? For this book there was also a choice—*which one to describe*? To provide you with listings for useful programs, we had to choose. Fortunately the logic is so clear that you can translate the instructions into any language—for a computer as well as a calculator. The programs given here are the "greatest common denominator" of computing in calculus.

The range of choices starts with the Casio *fx* 7000G—the first and simplest, with very limited memory but a good screen. The Casio 7500, 8000, and 8500 have increasing memory and extra features. The Sharp *EL*-5200 (or 9000 in Canada and Europe) is comparable to the Casio 8000. These machines have *algebraic entry*—the normal order as in $y = x + 3$. They are inexpensive and good. More expensive and much more powerful are the Hewlett-Packard calculators—the *HP*-28*S* and *HP*-48*SX*. They have large memories and extensive menus (and symbolic algebra). They use *reverse Polish notation*—numbers first in the stack, then commands. They require extra time and effort, and other books do justice to their amazing capabilities. It is estimated that those calculators could get 95 on a typical calculus exam.

While this book was being written, Texas Instruments produced a new graphing calculator: the *TI*-81. It is closer to the Casio and Sharp (emphasis on graphing, easy to learn, no symbolic algebra, moderate price). With earlier machines as a starting point, many improvements were added. There is some risk in a choice that is available only Δt before this textbook is published, and we hope that the experts we asked are right. Anyway, *our programs are for the TI-81*. It is impressive.

These few pages are no substitute for the manual that comes with a calculator. A valuable supplement is a guide directed especially at calculus—my absolute favorites are *Calculus Activities for Graphic Calculators* by Dennis Pence (PWS-Kent, 1990 for the Casio and Sharp and *HP*-28*S*, 1991 for the *TI*-81). A series of *Calculator Enhancements*, using *HP*'s, is being published by Harcourt Brace Jovanovich. What follows is an introduction to one part of a calculus laboratory. Later in the book, we supply *TI*-81 programs close to the mathematics and the exercises that they are prepared for.

A few words to start: To select from a menu, press the item number and ENTER. Edit a command line using DEL(ete) and INS(ert). Every line ends with ENTER. For calculus select radians on the MODE screen. For powers use ^. For special powers choose x^2, x^{-1}, $\sqrt{x}$. Multiplication has priority, so $(-)3 + 2 \times 2$ produces 1. Use keys for SIN, IF, IS, ... When you press letters, I multiplies S.

If a program says 3 → C, type 3 STO C ENTER. Storage locations are A to Z or Greek θ.

Functions A graphing calculator helps you (forces you?) to understand the concept of a function. It also helps you to understand specific functions—especially when changing the viewing window.

To evaluate $y = x^2 - 2x$ *just once*, use the home screen. To define $y(x)$ for *repeated use*, move to the function edit screen: Press MODE, choose Function, and press Y=. Then type in the formula. ***Important tip***: for X on the TI-81, the key X|T is faster than two steps Alpha X. The Y= edit screen is the same place where the formula is needed for graphing.

Example $Y_1 = X^2 - 2X$ ENTER on the Y= screen. 4 STO X ENTER on the home screen. Y_1 ENTER on the Y-VARS screen. The screen shows 8, which is $Y(4)$. The formula remains when the calculator is off.

Graphing You specify the X range and Y range. (We should say X domain but we don't.) The screen is a grid of 96 × 64 little rectangles called "*pixels*." The first column of pixels represents Xmin and the last column is Xmax. Press RANGE to reset. With Xres = 1 the function is evaluated 96 times as it is graphed. Xscl and Yscl give the spaces between ticks on the axes.

The ZOOM menu is a fast way to set ranges. ZOOM Standard gives the default $-10 \leqslant x \leqslant 10$, $-10 \leqslant y \leqslant 10$. ZOOM Trig gives $-2\pi \leqslant x \leqslant 2\pi$, $-3 \leqslant y \leqslant 3$.

The keystroke GRAPH shows the graphing screen with the current functions.

Example Set the ranges (−)2 ⩽ X ⩽ 3 and (−)150 ⩽ Y ⩽ 50. Press Y= and store Y_1 = X (in MATH)$^3 - 28X^2 + 15X + 36$ ENTER. Press GRAPH. You won't see much of the graph! Press RANGE and reset (−)10 ⩽ X ⩽ 30, (−)4000 ⩽ Y ⩽ 2300. Press GRAPH. See a cubic polynomial.

"Smart Graph" recalls the graph instantly without redrawing it, if no settings have changed. The DRAW menu is for points, lines, and shaded regions. This is perfect for our piecewise linear functions—just connect the breakpoints with lines. In Section 3.6 the lines show an iteration by its "cobweb."

Programming This book contains programs that you can type in once and save. We chose ***Autoscaling***, ***Newton's Method***, ***Secant Method***, ***Cobweb Iteration***, and ***Numerical Integration***. You will create others—to do calculations or to add features that are not available as single keystrokes. The calculator is like a computer, with a fairly small set of instructions. *One difference*: Memory is too precious to store comments with the code. You have to see the logic by rereading the program.

To enter the world of programming, press PRGM. Each PRGM submenu lists all programs by *name*—a digit, a letter, or θ (37 names). The program *title* has up to eight characters. Select the EDIT submenu and press G for the edit screen. Type the title GRAPHS and press ENTER. Practice on this one:

```
:"X²+X" STO (Y-VARS) Y1 ENTER
:"X-1" STO (Y-VARS) Y2 ENTER
:(PRGM)(I/O) DispGraph
```

The menus to call are in parentheses. Leave the edit screen with QUIT (not CLEAR—that erases the line with the cursor). Set the default window by ZOOM Standard.

To execute, press PRGM (EXEC) G ENTER. The program draws the graphs. It leaves Y_1 and Y_2 on the Y= screen. To erase the program from the home screen, press (PRGM)(ERASE)G. Practice again by creating Prgm2:FUNC. Type :√ X STO Y and :(PRGM)(I/O)Disp Y. Move to the home screen, store X by 4 STO X ENTER, and execute by (PRGM)(EXEC)2 ENTER. Also try $X = -1$. When it fails to imagine i, select 1:Goto Error.

Piecewise functions and ***Input*** (to a running program). The definition of a piecewise function includes the *domain of each piece*. Logical tests like "IF X ⩾ 7" determine which domain the input value X falls into. An IF statement only affects the following line—which is executed when TEST = 1 (meaning *true*) and skipped when TEST = 0 (meaning *false*). IF commands are in the PRGM (CTL) submenu; TEST calls the menu of inequalities.

An input value $X = 4$ need not be stored in advance. Program P stops while running to request input. Execute with P ENTER after selecting the PRGM (EXEC) menu. Answer ? with 4 and ENTER. After completion, rerun by pressing ENTER again. The function is $y = 14 - x$ if $x < 7$, $y = x$ *if* $x \geqslant 7$.

PrgmP: PIECES	
:Disp "X="	PGRM (I/O) Ask for input
:Input X	PGRM (I/O) Screen ? ENTER X
:14-X → Y	First formula for all X
:If 7≤X	PRGM (CTL) TEST
:X → Y	Overwrite if TEST = 1
:Disp Y	Display $Y(X)$

Overwriting is faster than checking both ends $A \leqslant X \leqslant B$ for each piece. Even faster: a whole formula $(14 - X)(X < 7) + (X)(7 \leqslant X)$ can go on a single line using 1 and 0 from the tests. Compute-store-display $Y(X)$ as above, or define Y1 on the edit screen.

Exercise Define a third piece $Y = 8 + X$ if $X < 3$. Rewrite P using Y1 =. A product of tests (3≤X) (X<7) evaluates to 1 if *all true* and to 0 if *any false*.

TRACE *and* ZOOM The best feature is graphing. But a whole graph can be like a whole book—too much at once. You want to focus on one part. A computer or calculator will trace along the graph, stop at a point, and zoom in.

There is also ZOOM OUT, to widen the ranges and see more. Our eyes work the same way—they put together information on different scales. Looking around the room uses an amazingly large part of the human brain. With a big enough computer we can try to imitate the eyes—this is a key problem in artificial intelligence. With a small computer and a zoom feature, we can *use* our eyes to understand functions.

Press TRACE to locate a point on the graph. A blinking cursor appears. Move left or right—the cursor stays on the graph. Its coordinates appear at the bottom of the screen. When x changes by a pixel, the calculator evaluates $y(x)$. To solve $y(x) = 0$, read off x at the point when y is nearest to zero. To minimize or maximize $y(x)$, read off the smallest and largest y. In all these problems, zoom in for more accuracy.

To blow up a figure we can choose new ranges. The fast way is to use a ZOOM command. For a preset range, use ZOOM Standard or ZOOM Trig. To shrink or stretch by XFact or YFact (default values 4), use ZOOM In or ZOOM Out. Choose the center point and press ENTER. The new graph appears. Change those scaling factors with ZOOM Set Factors. Best of all, ***create your own viewing window***. Press ZOOM Box.

To draw the box, move the cursor to one corner. Press ENTER and this point is a small square. The same keys move a second (blinking) square to the opposite corner—the box grows as you move. Press ENTER, and the box is the new viewing window. The graphs show the same function with a change of scale. Section 3.4 will discuss the mathematics—here we concentrate on the graphics.

EXAMPLE 9 Place :Y1=X sin (1/X) in the Y= edit screen. Press ZOOM Trig for a first graph. Set XFact = 1 and YFact = 2.5. Press ZOOM In with center at (0, 0). To see a larger picture, use XFact = 10 and YFact = 1. Then Zoom Out again. As X gets large, the function $X \sin(1/X)$ approaches _______.

Now return to ZOOM Trig. Zoom In with the factors set to 4 (default). Zoom again by pressing ENTER. With the center and the factors fixed, this is faster than drawing a zoom box.

EXAMPLE 10 Repeat for the more erratic function $Y = \sin(1/X)$. After `ZOOM Trig`, create a box to see this function near $X = .01$. The Y range is now ______.

Scaling is crucial. For a new function it can be tedious. A formula for $y(x)$ does not easily reveal the range of y's, when $A \leqslant x \leqslant B$ is given. The following program is often more convenient than zooms. It samples the function $L = 19$ times across the x-range (every 5 pixels). The inputs *Xmin*, *Xmax*, Y_1 are previously stored on other screens. After sampling, the program sets the y-range from $C = Ymin$ to $D = Ymax$ and draws the graph.

Notice the *loop* with counter K. The loop ends with the command `IS>(K,L)`, which increases K by 1 and skips a line if the new K exceeds L. Otherwise the command `Goto 1` restarts the loop. The screen shows the short form on the left.

Example: `Y1` $= x^3 + 10x^2 - 7x + 42$ with range `Xmin` $= -12$ and `Xmax` $= 10$. Set tick spacing `Xscl` $= 4$ and `Yscl` $= 250$. Execute with `PRGM (EXEC) A ENTER`. For this program we also list menu locations and comments.

```
PrgmA :AUTOSCL      Menu (Submenu) Comment
:All-Off            Y-VARS (OFF) Turn off functions
:Xmin → A           VARS (RNG) Store Xmin using STO
:19 → L             Store number of evaluations (19)
:(Xmax-A)/L → H     Spacing between evaluations
:A → X              Start at x = A
:Y1 → C             Y-VARS (Y) Evaluate the function
:C → D              Start C and D with this value
:1 → K              Initialize counter K = 1
:Lbl 1              PRGM (CTL) Mark loop start
:A+KH → X           Calculate next x
:Y1 → Y             Evaluate function at x
:IF Y<C             PGRM (CTL) New minimum?
:Y → C              Update C
:IF D<Y             PRGM (CTL) New maximum?
:Y → D              Update D
:IS>(K,L)           PRGM (CTL) Add 1 to K, skip Goto if >L
:Goto 1             PRGM (CTL) Loop return to Lbl 1
:Y1-On              Y-VARS (ON) Turn on Y1
:C → Ymin           STO VARS (RNG) Set Ymin = C
:D → Ymax           STO VARS (RNG) Set Ymax = D
:DispGraph          PRGM (I/O) Generate graph
```

Contents

CHAPTER 2

Derivatives

2.1 The Derivative of a Function

This chapter begins with the definition of the derivative. Two examples were in Chapter 1. When the distance is t^2, the velocity is $2t$. When $f(t) = \sin t$ we found $v(t) = \cos t$. *The velocity is now called the* ***derivative*** *of* $f(t)$. As we move to a more formal definition and new examples, we use new symbols f' and df/dt for the derivative.

> **2A** At time t, the *derivative* $f'(t)$ or df/dt or $v(t)$ is
>
> $$f'(t) = \lim_{\Delta t \to 0} \frac{f(t + \Delta t) - f(t)}{\Delta t}. \qquad (1)$$

The ratio on the right is the average velocity over a short time Δt. The derivative, on the left side, is its limit as the step Δt (*delta t*) approaches zero.

Go slowly and look at each piece. The distance at time $t + \Delta t$ is $f(t + \Delta t)$. The distance at time t is $f(t)$. Subtraction gives the ***change in distance***, between those times. We often write Δf for this difference: $\Delta f = f(t + \Delta t) - f(t)$. ***The average velocity is the ratio*** $\Delta f/\Delta t$—change in distance divided by change in time.

The limit of the average velocity is the derivative, if this limit exists:

$$\frac{df}{dt} = \lim_{\Delta t \to 0} \frac{\Delta f}{\Delta t}. \qquad (2)$$

This is the neat notation that Leibniz invented: $\Delta f/\Delta t$ *approaches* df/dt. Behind the innocent word "*limit*" is a process that this course will help you understand.

Note that Δf is not Δ times f! ***It is the change in*** f. Similarly Δt is not Δ times t. It is the time step, positive or negative and eventually small. To have a one-letter symbol we replace Δt by h.

The right sides of (1) and (2) contain average speeds. On the graph of $f(t)$, the distance *up* is divided by the distance *across*. That gives the average slope $\Delta f/\Delta t$.

The left sides of (1) and (2) are ***instantaneous*** speeds df/dt. They give the slope at the instant t. This is the derivative df/dt (when Δt and Δf shrink to zero). Look again at the calculation for $f(t) = t^2$:

$$\frac{\Delta f}{\Delta t} = \frac{f(t+\Delta t) - f(t)}{\Delta t} = \frac{t^2 + 2t\,\Delta t + (\Delta t)^2 - t^2}{\Delta t} = 2t + \Delta t. \tag{3}$$

Important point: Those steps are taken before Δt goes to zero. ***If we set $\Delta t = 0$ too soon, we learn nothing.*** The ratio $\Delta f/\Delta t$ becomes 0/0 (which is meaningless). The numbers Δf and Δt must approach zero together, not separately. Here their ratio is $2t + \Delta t$, the average speed.

To repeat: Success came by writing out $(t + \Delta t)^2$ and subtracting t^2 and dividing by Δt. Then and only then can we approach $\Delta t = 0$. The limit is the derivative $2t$.

There are several new things in formulas (1) and (2). Some are easy but important, others are more profound. The idea of a function we will come back to, and the definition of a limit. But the notations can be discussed right away. They are used constantly and you also need to know how to read them aloud:

$f(t)$ = "f of t" = the value of the function f at time t

Δt = "delta t" = the time step forward or backward from t

$f(t + \Delta t)$ = "f of t plus delta t" = the value of f at time $t + \Delta t$

Δf = "delta f" = the change $f(t + \Delta t) - f(t)$

$\Delta f/\Delta t$ = "delta f over delta t" = the average velocity

$f'(t)$ = "f prime of t" = the value of the derivative at time t

df/dt = "$d\,f\,d\,t$" = the same as f' (the instantaneous velocity)

$\lim_{\Delta t \to 0}$ = "limit as delta t goes to zero" = the process that starts with numbers $\Delta f/\Delta t$ and produces the number df/dt.

From those last words you see what lies behind the notation df/dt. The symbol Δt indicates a nonzero (usually short) length of time. The symbol dt indicates an infinitesimal (even shorter) length of time. Some mathematicians work separately with df and dt, and df/dt is their ratio. For us df/dt is a single notation (don't cancel d and don't cancel Δ). The derivative df/dt is the limit of $\Delta f/\Delta t$. *When that notation df/dt is awkward, use f' or v.*

Remark The notation hides one thing we should mention. The time step can be *negative* just as easily as positive. We can compute the average $\Delta f/\Delta t$ over a time interval *before* the time t, instead of after. This ratio also approaches df/dt.

The notation also hides another thing: ***The derivative might not exist.*** The averages $\Delta f/\Delta t$ might not approach a limit (it has to be the same limit going forward and backward from time t). In that case $f'(t)$ is not defined. At that instant there is no clear reading on the speedometer. This will happen in Example 2.

EXAMPLE 1 (Constant velocity $V = 2$) The distance f is V times t. The distance at time $t + \Delta t$ is V times $t + \Delta t$. ***The difference Δf is V times Δt:***

$$\frac{\Delta f}{\Delta t} = \frac{V\Delta t}{\Delta t} = V \quad \text{so the limit is} \quad \frac{df}{dt} = V.$$

The derivative of Vt is V. The derivative of $2t$ is 2. The averages $\Delta f/\Delta t$ are always $V = 2$, in this exceptional case of a constant velocity.

EXAMPLE 2 Constant velocity 2 up to time $t=3$, ***then stop.***

For small times we still have $f(t)=2t$. But after the stopping time, the distance is fixed at $f(t)=6$. The graph is flat beyond time 3. Then $f(t+\Delta t)=f(t)$ and $\Delta f=0$ and ***the derivative of a constant function is zero***:

$$t>3: \quad f'(t)=\lim_{\Delta t\to 0}\frac{f(t+\Delta t)-f(t)}{\Delta t}=\lim_{\Delta t\to 0}\frac{0}{\Delta t}=0. \tag{4}$$

In this example ***the derivative is not defined at the instant when*** $t=3$. The velocity falls suddenly from 2 to zero. The ratio $\Delta f/\Delta t$ depends, at that special moment, on whether Δt is positive or negative. The average velocity *after* time $t=3$ is zero. The average velocity *before* that time is 2. When the graph of f has a corner, the graph of v has a ***jump***. It is a ***step function.***

One new part of that example is the notation (df/dt or f' instead of v). Please look also at the third figure. It shows how the function takes t (on the left) to $f(t)$. Especially it shows Δt and Δf. At the start, $\Delta f/\Delta t$ is 2. After the stop at $t=3$, all t's go to the same $f(t)=6$. So $\Delta f=0$ and $df/dt=0$.

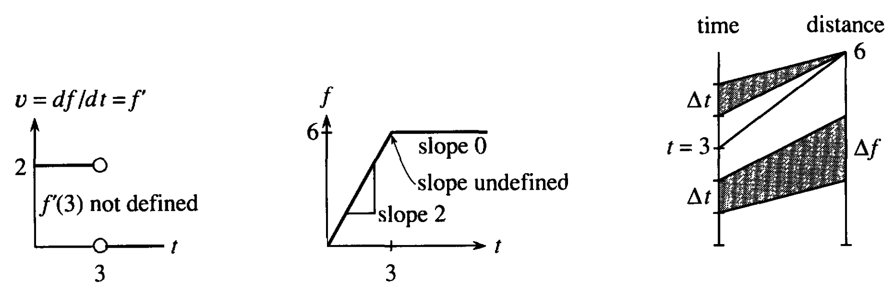


Fig. 2.1 The derivative is 2 then 0. It does not exist at $t=3$.

THE DERIVATIVE OF $1/t$

Here is a completely different slope, for the "demand function" $f(t)=1/t$. The demand is $1/t$ when the price is t. A high price t means a low demand $1/t$. Increasing the price reduces the demand. The calculus question is: ***How quickly does*** $1/t$ ***change when*** t ***changes***? The "marginal demand" is the slope of the demand curve.

The big thing is to find the derivative of $1/t$ once and for all. It is $-1/t^2$.

EXAMPLE 3 $f(t)=\dfrac{1}{t}$ has $\Delta f=\dfrac{1}{t+\Delta t}-\dfrac{1}{t}$. This equals $\dfrac{t-(t+\Delta t)}{t(t+\Delta t)}=\dfrac{-\Delta t}{t(t+\Delta t)}$.

Divide by Δt and let $\Delta t\to 0$: $\dfrac{\Delta f}{\Delta t}=\dfrac{-1}{t(t+\Delta t)}$ approaches $\dfrac{df}{dt}=\dfrac{-1}{t^2}$.

Line 1 is algebra, line 2 is calculus. The first step in line 1 subtracts $f(t)$ from $f(t+\Delta t)$. The difference is $1/(t+\Delta t)$ minus $1/t$. The common denominator is t times $t+\Delta t$—this makes the algebra possible. We can't set $\Delta t=0$ in line 2, until we have divided by Δt.

The average is $\Delta f/\Delta t=-1/t(t+\Delta t)$. Now set $\Delta t=0$. The derivative is $-1/t^2$. Section 2.4 will discuss the first of many cases when substituting $\Delta t=0$ is not possible, and the idea of a limit has to be made clearer.

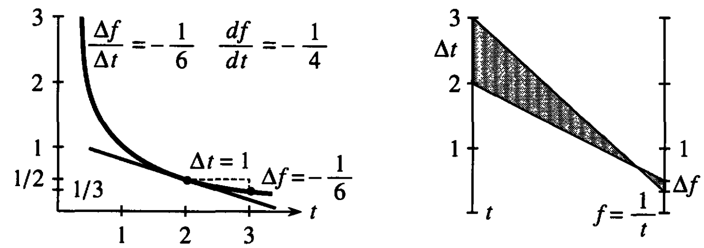


Fig. 2.2 Average slope is $-\frac{1}{6}$, true slope is $-\frac{1}{4}$. Increase in t produces decrease in f.

Check the algebra at $t = 2$ and $t + \Delta t = 3$. The demand $1/t$ drops from 1/2 to 1/3. The difference is $\Delta f = -1/6$, which agrees with $-1/(2)(3)$ in line 1. As the steps Δf and Δt get smaller, their ratio approaches $-1/(2)(2) = -1/4$.

This derivative is negative. The function $1/t$ is *decreasing*, and Δf is below zero. The graph is going *downward* in Figure 2.2, and its slope is negative:

An increasing $f(t)$ has positive slope. A decreasing $f(t)$ has negative slope.

The slope $-1/t^2$ is very negative for small t. A price increase severely cuts demand.

The next figure makes a small but important point. There is nothing sacred about t. Other letters can be used—especially x. A quantity can depend on ***position instead of time***. The height changes as we go west. The area of a square changes as the side changes. Those are not affected by the passage of time, and there is no reason to use t. You will often see $y = f(x)$, with x across and y up—connected by a function f.

Similarly, f is not the only possibility. Not every function is named f! That letter is useful because it stands for the word function—but we are perfectly entitled to write $y(x)$ or $y(t)$ instead of $f(x)$ or $f(t)$. The distance up is a function of the distance across. This relationship "y of x" is all-important to mathematics.

The slope is also a function. Calculus is about two functions, $y(x)$ and dy/dx.

Question If we add 1 to $y(x)$, what happens to the slope? *Answer* Nothing.

Question If we add 1 to the slope, what happens to the height? *Answer* ________.

The symbols t and x represent ***independent variables***—they take any value they want to (in the domain). Once they are set, $f(t)$ and $y(x)$ are determined. Thus f and y represent ***dependent variables***—they *depend* on t and x. A change Δt produces a

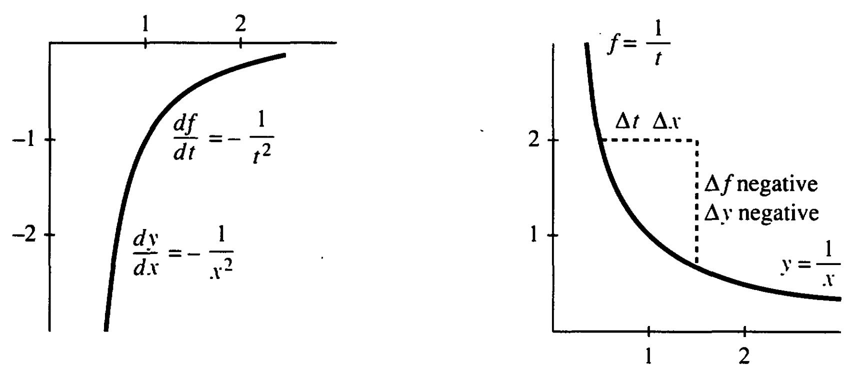


Fig. 2.3 The derivative of $1/t$ is $-1/t^2$. The slope of $1/x$ is $-1/x^2$.

change Δf. A change Δx produces Δy. The *in*dependent variable goes *in*side the parentheses in $f(t)$ and $y(x)$. It is not the letter that matters, it is the idea:

independent variable t or x

dependent variable f or g or y or z or u

derivative df/dt or df/dx or dy/dx or $\cdots$

The derivative dy/dx comes from [change in y] divided by [change in x]. The time step becomes a space step, forward or backward. The slope is the rate at which y changes with x. ***The derivative of a function is its "rate of change."***

I mention that physics books use $x(t)$ for distance. Darn it.

To emphasize the definition of a derivative, here it is again with y and x:

$$\frac{\Delta y}{\Delta x}=\frac{y(x+\Delta x)-y(x)}{\Delta x}=\frac{\text{distance up}}{\text{distance across}} \qquad \frac{dy}{dx}=\lim_{\Delta x\to 0}\frac{\Delta y}{\Delta x}=y'(x).$$

The notation $y'(x)$ pins down the point x where the slope is computed. In dy/dx that extra precision is omitted. This book will try for a reasonable compromise between logical perfection and ordinary simplicity. The notation $dy/dx(x)$ is not good; $y'(x)$ is better; when x is understood it need not be written in parentheses.

You are allowed to say that the function is $y=x^2$ and the derivative is $y'=2x$—even if the strict notation requires $y(x)=x^2$ and $y'(x)=2x$. You can even say that the function is x^2 and its derivative is $2x$ and its ***second derivative*** is 2—provided everybody knows what you mean.

Here is an example. It is a little early and optional but terrific. You get excellent practice with letters and symbols, and out come new derivatives.

EXAMPLE 4 If $u(x)$ has slope du/dx, what is the slope of $f(x)=(u(x))^2$?

From the derivative of x^2 this will give the derivative of x^4. In that case $u=x^2$ and $f=x^4$. First point: ***The derivative of u^2 is not*** $(du/dx)^2$. We do not square the derivative $2x$. To find the "square rule" we start as we have to—with $\Delta f=f(x+\Delta x)-f(x)$:

$$\Delta f=(u(x+\Delta x))^2-(u(x))^2=[u(x+\Delta x)+u(x)][u(x+\Delta x)-u(x)].$$

This algebra puts Δf in a convenient form. We factored a^2-b^2 into $[a+b]$ times $[a-b]$. Notice that we don't have $(\Delta u)^2$. We have Δf, the change in u^2. Now divide by Δx and take the limit:

$$\frac{\Delta f}{\Delta x}=[u(x+\Delta x)+u(x)]\left[\frac{u(x+\Delta x)-u(x)}{\Delta x}\right] \text{ approaches } 2u(x)\frac{du}{dx}. \qquad (5)$$

This is the *square rule*: ***The derivative of*** $(u(x))^2$ ***is*** $2u(x)$ ***times*** du/dx. From the derivatives of x^2 and $1/x$ and $\sin x$ (all known) the examples give new derivatives.

EXAMPLE 5 $(u=x^2)$ The derivative of x^4 is $2u\,du/dx=2(x^2)(2x)=4x^3$.

EXAMPLE 6 $(u=1/x)$ The derivative of $1/x^2$ is $2u\,du/dx=(2/x)(-1/x^2)=-2/x^3$.

EXAMPLE 7 $(u=\sin x,\ du/dx=\cos x)$ The derivative of $u^2=\sin^2 x$ is $2\sin x\cos x$.

Mathematics is really about ideas. The notation is created to express those ideas. Newton and Leibniz invented calculus independently, and Newton's friends spent a lot of time proving that he was first. He was, but it was Leibniz who thought of writing dy/dx—which caught on. It is the perfect way to suggest the limit of $\Delta y/\Delta x$. Newton was one of the great scientists of all time, and calculus was one of the great inventions of all time—but the notation must help. You now can write and speak about the derivative. What is needed is a longer list of functions and derivatives.

2.1 EXERCISES

Read-through questions

The derivative is the __a__ of $\Delta f/\Delta t$ as Δt approaches __b__. Here Δf equals __c__. The step Δt can be positive or __d__. The derivative is written v or __e__ or __f__. If $f(x)=2x+3$ and $\Delta x=4$ then $\Delta f=$ __g__. If $\Delta x=-1$ then $\Delta f=$ __h__. If $\Delta x=0$ then $\Delta f=$ __i__. The slope is not 0/0 but $df/dx=$ __j__.

The derivative does not exist where $f(t)$ has a __k__ and $v(t)$ has a __l__. For $f(t)=1/t$ the derivative is __m__. The slope of $y=4/x$ is $dy/dx=$ __n__. A decreasing function has a __o__ derivative. The __p__ variable is t or x and the __q__ variable is f or y. The slope of y^2 (is) (is not) $(dy/dx)^2$. The slope of $(u(x))^2$ is __r__ by the square rule. The slope of $(2x+3)^2$ is __s__.

1 Which of the following numbers (*as is*) gives df/dt at time t? If in doubt test on $f(t)=t^2$.

(a) $\dfrac{f(t+\Delta t)-f(t)}{\Delta t}$ (b) $\lim\limits_{h\to 0}\dfrac{f(t+2h)-f(t)}{2h}$

(c) $\lim\limits_{\Delta t\to 0}\dfrac{f(t-\Delta t)-f(t)}{-\Delta t}$ (d) $\lim\limits_{t\to 0}\dfrac{f(t+\Delta t)-f(t)}{\Delta t}$

2 Suppose $f(x)=x^2$. Compute each ratio and set $h=0$:

(a) $\dfrac{f(x+h)-f(x)}{h}$ (b) $\dfrac{f(x+5h)-f(x)}{5h}$

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

3 For $f(x)=3x$ and $g(x)=1+3x$, find $f(4+h)$ and $g(4+h)$ and $f'(4)$ and $g'(4)$. Sketch the graphs of f and g—why do they have the same slope?

4 Find three functions with the same slope as $f(x)=x^2$.

5 For $f(x)=1/x$, sketch the graphs of $f(x)+1$ and $f(x+1)$. Which one has the derivative $-1/x^2$?

6 Choose c so that the line $y=x$ is tangent to the parabola $y=x^2+c$. They have the same slope where they touch.

7 Sketch the curve $y(x)=1-x^2$ and compute its slope at $x=3$.

8 If $f(t)=1/t$, what is the average velocity between $t=\frac{1}{2}$ and $t=2$? What is the average between $t=\frac{1}{2}$ and $t=1$? What is the average (to one decimal place) between $t=\frac{1}{2}$ and $t=101/200$?

9 Find $\Delta y/\Delta x$ for $y(x)=x+x^2$. Then find dy/dx.

10 Find $\Delta y/\Delta x$ and dy/dx for $y(x)=1+2x+3x^2$.

11 When $f(t)=4/t$, simplify the difference $f(t+\Delta t)-f(t)$, divide by Δt, and set $\Delta t=0$. The result is $f'(t)$.

12 Find the derivative of $1/t^2$ from $\Delta f(t)=1/(t+\Delta t)^2-1/t^2$. Write Δf as a fraction with the denominator $t^2(t+\Delta t)^2$. Divide the numerator by Δt to find $\Delta f/\Delta t$. Set $\Delta t=0$.

13 Suppose $f(t)=7t$ to $t=1$. Afterwards $f(t)=7+9(t-1)$.

(a) Find df/dt at $t=\frac{1}{2}$ and $t=\frac{3}{2}$.

(b) Why doesn't $f(t)$ have a derivative at $t=1$?

14 Find the derivative of the derivative (the *second derivative*) of $y=3x^2$. What is the third derivative?

15 Find numbers A and B so that the straight line $y=x$ fits smoothly with the curve $Y=A+Bx+x^2$ at $x=1$. Smoothly means that $y=Y$ and $dy/dx=dY/dx$ at $x=1$.

16 Find numbers A and B so that the horizontal line $y=4$ fits smoothly with the curve $y=A+Bx+x^2$ at the point $x=2$.

17 *True* (with reason) *or false* (with example):

(a) If $f(t)<0$ then $df/dt<0$.

(b) The derivative of $(f(t))^2$ is $2\,df/dt$.

(c) The derivative of $2f(t)$ is $2\,df/dt$.

(d) The derivative is the limit of Δf divided by the limit of Δt.

18 For $f(x)=1/x$ the *centered difference* $f(x+h)-f(x-h)$ is $1/(x+h)-1/(x-h)$. Subtract by using the common denominator $(x+h)(x-h)$. Then divide by $2h$ and set $h=0$. Why divide by $2h$ to obtain the correct derivative?

19 Suppose $y=mx+b$ for negative x and $y=Mx+B$ for $x\geqslant 0$. The graphs meet if ________. The two slopes are ________. The slope at $x=0$ is ________ (what is possible?).

20 The slope of $y=1/x$ at $x=1/4$ is $y'=-1/x^2=-16$. At $h=1/12$, which of these ratios is closest to -16?

$$\frac{y(x+h)-y(x)}{h} \qquad \frac{y(x)-y(x-h)}{h} \qquad \frac{y(x+h)-y(x-h)}{2h}$$

21 Find the average slope of $y=x^2$ between $x=x_1$ and $x=x_2$. What does this average approach as x_2 approaches x_1?

22 Redraw Figure 2.1 when $f(t)=3-2t$ for $t\leqslant 2$ and $f(t)=-1$ for $t\geqslant 2$. Include df/dt.

23 Redraw Figure 2.3 for the function $y(x) = 1 - (1/x)$. Include dy/dx.

24 The limit of $0/\Delta t$ as $\Delta t \to 0$ is not 0/0. Explain.

25 Guess the limits by an informal working rule. Set $\Delta t = 0.1$ and -0.1 and imagine Δt becoming smaller:

(a) $\dfrac{1+\Delta t}{2+\Delta t}$ (b) $\dfrac{|\Delta t|}{\Delta t}$

(c) $\dfrac{\Delta t + (\Delta t)^2}{\Delta t - (\Delta t)^2}$ (d) $\dfrac{t+\Delta t}{t-\Delta t}$

***26** Suppose $f(x)/x \to 7$ as $x \to 0$. Deduce that $f(0) = 0$ and $f'(0) = 7$. Give an example other than $f(x) = 7x$.

27 What is $\lim_{x\to 0} \dfrac{f(3+x)-f(3)}{x}$ if it exists? What if $x \to 1$?

Problems 28–31 use the square rule: $d(u^2)/dx = 2u(du/dx)$.

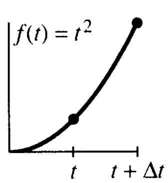
28 Take $u = x$ and find the derivative of x^2 (a new way).

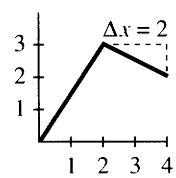
29 Take $u = x^4$ and find the derivative of x^8 (using $du/dx = 4x^3$).

30 If $u = 1$ then $u^2 = 1$. Then $d1/dx$ is 2 times $d1/dx$. How is this possible?

31 Take $u = \sqrt{x}$. The derivative of $u^2 = x$ is $1 = 2u(du/dx)$. So what is du/dx, the derivative of $\sqrt{x}$?

32 The left figure shows $f(t) = t^2$. Indicate distances $f(t + \Delta t)$ and Δt and Δf. Draw lines that have slope $\Delta f/\Delta t$ and $f'(t)$.





33 The right figure shows $f(x)$ and Δx. Find $\Delta f/\Delta x$ and $f'(2)$.

34 Draw $f(x)$ and Δx so that $\Delta f/\Delta x = 0$ but $f'(x) \neq 0$.

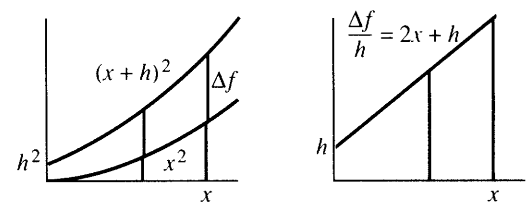
35 If $f = u^2$ then $df/dx = 2u\, du/dx$. If $g = f^2$ then $dg/dx = 2f\, df/dx$. Together those give $g = u^4$ and $dg/dx =$ ________.

36 ***True or false***, assuming $f(0) = 0$:

(a) If $f(x) \leqslant x$ for all x, then $df/dx \leqslant 1$.

(b) If $df/dx \leqslant 1$ for all x, then $f(x) \leqslant x$.

37 The graphs show Δf and $\Delta f/h$ for $f(x) = x^2$. Why is $2x + h$ the equation for $\Delta f/h$? If h is cut in half, draw in the new graphs.



38 Draw the corresponding graphs for $f(x) = \frac{1}{2}x$.

39 Draw $1/x$ and $1/(x+h)$ and $\Delta f/h$—either by hand with $h = \frac{1}{2}$ or by computer to show $h \to 0$.

40 For $y = e^x$, show on computer graphs that $dy/dx = y$.

41 *Explain the derivative in your own words.*

2.2 Powers and Polynomials

This section has two main goals. One is to find the derivatives of $f(x) = x^3$ and x^4 and x^5 (and more generally $f(x) = x^n$). The *power* or *exponent n* is at first a positive integer. Later we allow x^π and $x^{2.2}$ and every x^n.

The other goal is different. While computing these derivatives, we look ahead to their applications. In using calculus, we meet *equations with derivatives in them*—"***differential equations***." It is too early to solve those equations. But it is not too early to see the purpose of what we are doing. Our examples come from economics and biology.

With $n = 2$, the derivative of x^2 is $2x$. With $n = -1$, the slope of x^{-1} is $-1x^{-2}$. Those are two pieces in a beautiful pattern, which it will be a pleasure to discover. We begin with x^3 and its derivative $3x^2$, before jumping to x^n.

EXAMPLE 1 If $f(x) = x^3$ then $\Delta f = (x+h)^3 - x^3 = (x^3 + 3x^2h + 3xh^2 + h^3) - x^3$.

Step 1: Cancel x^3. **Step 2:** Divide by h. **Step 3:** h goes to zero.

$$\frac{\Delta f}{h} = 3x^2 + 3xh + h^2 \quad \text{approaches} \quad \frac{df}{dx} = 3x^2.$$

That is straightforward, and you see the crucial step. The power $(x+h)^3$ yields four separate terms $x^3 + 3x^2h + 3xh^2 + h^3$. (Notice 1, 3, 3, 1.) After x^3 is subtracted, we can divide by h. At the limit $(h = 0)$ we have $3x^2$.

For $f(x) = x^n$ the plan is the same. A step of size h leads to $f(x+h) = (x+h)^n$. One reason for algebra is to calculate powers like $(x+h)^n$, and if you have forgotten the binomial formula we can recapture its main point. Start with $n = 4$:

$$(x+h)(x+h)(x+h)(x+h) = x^4 + \quad ??? \quad + h^4. \tag{1}$$

Multiplying the four x's gives x^4. Multiplying the four h's gives h^4. These are the easy terms, but not the crucial ones. The subtraction $(x+h)^4 - x^4$ will remove x^4, and the limiting step $h \to 0$ will wipe out h^4 (even after division by h). ***The products that matter are those with exactly one*** h. In Example 1 with $(x+h)^3$, this key term was $3x^2h$. Division by h left $3x^2$.

With only one h, there are n places it can come from. Equation (1) has four h's in parentheses, and four ways to produce x^3h. Therefore the key term is $4x^3h$. (Division by h leaves $4x^3$.) In general there are n parentheses and n ways to produce $x^{n-1}h$, so the ***binomial formula*** contains $nx^{n-1}h$:

$$(x+h)^n = x^n + \underline{nx^{n-1}h} + \cdots + h^n. \tag{2}$$

2B For $n = 1, 2, 3, 4, \ldots$, the derivative of x^n is nx^{n-1}.

Subtract x^n from (2). Divide by h. The key term is nx^{n-1}. The rest disappears as $h \to 0$:

$$\frac{\Delta f}{\Delta x} = \frac{(x+h)^n - x^n}{h} = \frac{nx^{n-1}h + \cdots + h^n}{h} \quad \text{so} \quad \frac{df}{dx} = nx^{n-1}.$$

The terms replaced by the dots involve h^2 and h^3 and higher powers. After dividing by h, they still have at least one factor h. All those terms vanish as h approaches zero.

EXAMPLE 2 $(x+h)^4 = x^4 + \underline{4x^3h} + 6x^2h^2 + 4xh^3 + h^4$. This is $n = 4$ in detail.

Subtract x^4, divide by h, let $h \to 0$. The derivative is $4x^3$. The coefficients 1, 4, 6, 4, 1 are in Pascal's triangle below. For $(x+h)^5$ the next row is 1, 5, 10, ?.

Remark The missing terms in the binomial formula (replaced by the dots) contain all the products $x^{n-j}h^j$. An x or an h comes from each parenthesis. The binomial coefficient "*n choose j*" is ***the number of ways to choose*** *j* h's ***out of*** *n* ***parentheses***. It involves *n factorial*, which is $n(n-1)\cdots(1)$. Thus $5! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120$.

These are numbers that gamblers know and love:

$$\text{“}n \text{ choose } j\text{”} = \binom{n}{j} = \frac{n!}{j!(n-j)!}$$

1	***Pascal's***	
1 1	***triangle***	
1 2 1		
1 3 3 1	$n=3$	
1 4 6 4 1	$n=4$	

In the last row, the coefficient of x^3h is $4!/1!3! = 4\cdot3\cdot2\cdot1/1\cdot3\cdot2\cdot1 = 4$. For the x^2h^2 term, with $j=2$, there are $4\cdot3\cdot2\cdot1/2\cdot1\cdot2\cdot1 = 6$ ways to choose two h's. Notice that $1+4+6+4+1$ equals 16, which is 2^4. Each row of Pascal's triangle adds to a power of 2.

Choosing 6 numbers out of 49 in a lottery, the odds are $49\cdot48\cdot47\cdot46\cdot45\cdot44/6!$ to 1. That number is $N =$ "49 choose 6" $= 13{,}983{,}816$. It is the coefficient of $x^{43}h^6$ in $(x+h)^{49}$. If λ times N tickets are bought, the expected number of winners is λ. The chance of no winner is $e^{-\lambda}$. The chance of *one* winner is $\lambda e^{-\lambda}$. See Section 8.4.

Florida's lottery in September 1990 (these rules) had six winners out of 109,163,978 tickets.

DERIVATIVES OF POLYNOMIALS

Now we have an infinite list of functions and their derivatives:

$$x \;\; x^2 \;\; x^3 \;\; x^4 \;\; x^5 \;\cdots \qquad 1 \;\; 2x \;\; 3x^2 \;\; 4x^3 \;\; 5x^4 \;\cdots$$

The derivative of x^n ***is*** n ***times the next lower power*** x^{n-1}. ***That rule extends beyond these integers*** 1, 2, 3, 4, 5 ***to all powers***:

$f = 1/x$	has	$f' = -1/x^2$:	Example 3 of Section 2.1	$(n=-1)$
$f = 1/x^2$	has	$f' = -2/x^3$:	Example 6 of Section 2.1	$(n=-2)$
$f = \sqrt{x}$	has	$f' = \frac{1}{2}x^{-1/2}$:	true but not yet checked	$(n=\frac{1}{2})$

Remember that x^{-2} means $1/x^2$ and $x^{-1/2}$ means $1/\sqrt{x}$. Negative powers lead to *decreasing* functions, approaching zero as x gets large. Their slopes have minus signs.

Question What are the derivatives of x^{10} and $x^{2.2}$ and $x^{-1/2}$?
Answer $10x^9$ and $2.2x^{1.2}$ and $-\frac{1}{2}x^{-3/2}$. Maybe $(x+h)^{2.2}$ is a little unusual. Pascal's triangle can't deal with this fractional power, but the formula stays firm: ***After*** $x^{2.2}$ ***comes*** $2.2x^{1.2}h$. The complete binomial formula is in Section 10.5.

That list is a good start, but plenty of functions are left. What comes next is really simple. A tremendous number of new functions are "linear combinations" like

$$f(x) = 6x^3 \quad \text{or} \quad 6x^3 + \tfrac{1}{2}x^2 \quad \text{or} \quad 6x^3 - \tfrac{1}{2}x^2\,.$$

What are their derivatives? The answers are known for x^3 and x^2, and we want to multiply by 6 or divide by 2 or add or subtract. *Do the same to the derivatives*:

$$f'(x) = 18x^2 \quad \text{or} \quad 18x^2 + x \quad \text{or} \quad 18x^2 - x.$$

> **2C** The derivative of c times $f(x)$ is c times $f'(x)$.
>
> **2D** The derivative of $f(x) + g(x)$ is $f'(x) + g'(x)$.

The number c can be any constant. We can add (or subtract) any functions. The rules allow any combination of f and g: ***The derivative of*** $9f(x) - 7g(x)$ ***is*** $9f'(x) - 7g'(x)$.

The reasoning is direct. When $f(x)$ is multiplied by c, so is $f(x+h)$. The difference Δf is also multiplied by c. All averages $\Delta f/h$ contain c, so their limit is cf'. The only incomplete step is the last one (the limit). ***We still have to say what** "limit" **means.***

Rule 2D is similar. Adding $f+g$ means adding $\Delta f+\Delta g$. Now divide by h. In the limit as $h \to 0$ we reach $f'+g'$—because a limit of sums is a sum of limits. Any example is easy and so is the proof—it is the definition of limit that needs care (Section 2.6).

You can now find the derivative of every polynomial. A "polynomial" is a combination of $1, x, x^2, \ldots, x^n$—for example $9+2x-x^5$. That particular polynomial has slope $2-5x^4$. Note that the derivative of 9 is zero! A constant just raises or lowers the graph, without changing its slope. It alters the mileage before starting the car.

The disappearance of constants is one of the nice things in differential calculus. The reappearance of those constants is one of the headaches in integral calculus. When you find v from f, the starting mileage doesn't matter. The constant in f has no effect on v. (*Δf is measured by a trip meter; Δt comes from a stopwatch.*) To find distance from velocity, you need to know the mileage at the start.

A LOOK AT DIFFERENTIAL EQUATIONS (FIND y FROM dy/dx)

We know that $y=x^3$ has the derivative $dy/dx=3x^2$. Starting with the function, we found its slope. Now reverse that process. ***Start with the slope and find the function.*** This is what science does all the time—and it seems only reasonable to say so.

Begin with $dy/dx=3x^2$. The slope is given, the function y is not given.

Question Can you go backward to reach $y=x^3$?
Answer Almost but not quite. You are only entitled to say that $y=x^3+C$. The constant C is the starting value of y (when $x=0$). Then the ***differential equation*** $dy/dx=3x^2$ is solved.

Every time you find a derivative, you can go backward to solve a differential equation. The function $y=x^2+x$ has the slope $dy/dx=2x+1$. In reverse, the slope $2x+1$ produces x^2+x—and all the other functions x^2+x+C, shifted up and down. After going from distance f to velocity v, we return to $f+C$. But there is a lot more to differential equations. Here are two crucial points:

1. We reach dy/dx by way of $\Delta y/\Delta x$, but we have no system to go backward. With $dy/dx=(\sin x)/x$ we are lost. What function has this derivative?
2. Many equations have the same solution $y=x^3$. Economics has $dy/dx=3y/x$. Geometry has $dy/dx=3y^{2/3}$. These equations involve y as well as dy/dx. Function and slope are mixed together! This is typical of differential equations.

To summarize: Chapters 2–4 compute and use derivatives. Chapter 5 goes in reverse. Integral calculus discovers the function from its slope. Given dy/dx we find $y(x)$. Then Chapter 6 solves the differential equation $dy/dt=y$, function mixed with slope. Calculus moves from *derivatives* to *integrals* to *differential equations*.

This discussion of the purpose of calculus should mention a specific example. Differential equations are applied to an epidemic (like AIDS). In most epidemics the number of cases grows exponentially. The peak is quickly reached by e^t, and the epidemic dies down. Amazingly, exponential growth is not happening with AIDS—the best fit to the data through 1988 is a ***cubic polynomial*** (*Los Alamos Science*, 1989):

The number of cases fits a cubic within 2%: $y=174.6(t-1981.2)^3+340$.

This is dramatically different from other epidemics. Instead of $dy/dt = y$ we have $dy/dt = 3y/t$. Before this book is printed, we may know what has been preventing e^t (fortunately). Eventually the curve will turn away from a cubic—I hope that mathematical models will lead to knowledge that saves lives.

Added in proof: In 1989 the curve for the U.S. dropped from t^3 to t^2.

MARGINAL COST AND ELASTICITY IN ECONOMICS

First point about economics: The ***marginal*** cost and ***marginal*** income are crucially important. The average cost of making automobiles may be \$10,000. But it is the \$8000 cost of the *next car* that decides whether Ford makes it. "*The average describes the past, the marginal predicts the future.*" For bank deposits or work hours or wheat, which come in smaller units, the amounts are continuous variables. Then the word "marginal" says one thing: ***Take the derivative***.†

The average pay over all the hours we ever worked may be low. We wouldn't work another hour for that! This average is rising, but the pay for each additional hour rises faster—possibly it jumps. When \$10/hour increases to \$15/hour after a 40-hour week, a 50-hour week pays \$550. The average income is \$11/hour. The marginal income is \$15/hour—the overtime rate.

Concentrate next on cost. Let $y(x)$ be the cost of producing x tons of steel. The cost of $x + \Delta x$ tons is $y(x + \Delta x)$. The extra cost is the difference Δy. Divide by Δx, the number of extra tons. The ratio $\Delta y/\Delta x$ is ***the average cost per extra ton***. When Δx is an ounce instead of a ton, we are near the marginal cost dy/dx.

Example: When the cost is x^2, the average cost is $x^2/x = x$. The marginal cost is $2x$. Figure 2.4 has increasing slope—an example of "diminishing returns to scale."

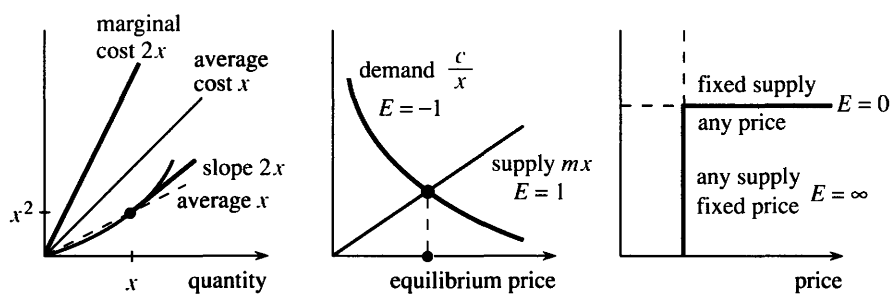


Fig. 2.4 Marginal exceeds average. Constant elasticity $E = \pm 1$. Perfectly elastic to perfectly inelastic (Γ curve).

This raises another point about economics. The units are arbitrary. In yen per kilogram the numbers look different. The way to correct for arbitrary units is to work with *percentage change* or ***relative change***. An increase of Δx tons is a relative increase of $\Delta x/x$. A cost increase Δy is a relative increase of $\Delta y/y$. Those are *dimensionless*, the same in tons/tons or dollars/dollars or yen/yen.

A third example is *the demand y at price x*. Now dy/dx is negative. But again the units are arbitrary. The demand is in liters or gallons, the price is in dollars or pesos.

†These paragraphs show how calculus applies to economics. You do *not* have to be an economist to understand them. Certainly the author is not, probably the instructor is not, possibly the student is not. We can all use dy/dx.

Relative changes are better. When the price goes up by 10%, the demand may drop by 5%. If that ratio stays the same for small increases, *the elasticity of demand is* $\frac{1}{2}$.

Actually this number should be $-\frac{1}{2}$. The price rose, the demand dropped. In our definition, the elasticity *will* be $-\frac{1}{2}$. In conversation between economists the minus sign is left out (I hope not forgotten).

DEFINITION The elasticity of the demand function $y(x)$ is

$$E(x) = \lim_{\Delta x \to 0} \frac{\Delta y/y}{\Delta x/x} = \frac{dy/dx}{y/x}. \tag{3}$$

Elasticity is "marginal" divided by "average." $E(x)$ is also relative change in y divided by relative change in x. Sometimes $E(x)$ is the same at all prices—this important case is discussed below.

EXAMPLE 1 Suppose the demand is $y = c/x$ when the price is x. The derivative $dy/dx = -c/x^2$ comes from calculus. The division $y/x = c/x^2$ is only algebra. *The ratio is* $E = -1$:

For the demand $y = c/x$, ***the elasticity is*** $(-c/x^2)/(c/x^2) = -1$.

All demand curves are compared with this one. The demand is ***inelastic*** when $|E| < 1$. It is ***elastic*** when $|E| > 1$. The demand $20/\sqrt{x}$ is inelastic ($E = -\frac{1}{2}$), while x^{-3} is elastic ($E = -3$). ***The power*** $y = cx^n$, ***whose derivative we know, is the function with constant elasticity*** n:

$$\text{if } \ y = cx^n \ \text{ then } \ dy/dx = cnx^{n-1} \ \text{ and } \ E = cnx^{n-1}/(cx^n/x) = n.$$

It is because $y = cx^n$ sets the standard that we could come so early to economics.

In the special case when $y = c/x$, consumers spend the same at all prices. Price x times quantity y remains constant at $xy = c$.

EXAMPLE 2 The supply curve has $E > 0$—supply increases with price. Now the baseline case is $y = cx$. The slope is c and the average is $y/x = c$. *The elasticity is* $E = c/c = 1$.

Compare $E = 1$ with $E = 0$ and $E = \infty$. A constant supply is "perfectly inelastic." The power n is zero and the slope is zero: $y = c$. No more is available when the harvest is over. Whatever the price, the farmer cannot suddenly grow more wheat. Lack of elasticity makes farm economics difficult.

The other extreme $E = \infty$ is "perfectly elastic." The supply is unlimited at a fixed price x. Once this seemed true of water and timber. In reality the steep curve $x =$ constant is leveling off to a flat curve $y =$ constant. Fixed price is changing to fixed supply, $E = \infty$ is becoming $E = 0$, and the supply of water follows a "gamma curve" shaped like Γ.

EXAMPLE 3 Demand is an increasing function of *income*—more income, more demand. The *income elasticity* is $E(I) = (dy/dI)/(y/I)$. A luxury has $E > 1$ (elastic). Doubling your income more than doubles the demand for caviar. A necessity has $E < 1$ (inelastic). The demand for bread does not double. Please recognize how the central ideas of calculus provide a language for the central ideas of economics.

Important note on supply = demand This is the basic equation of microeconomics. Where the supply curve meets the demand curve, the economy finds the equilibrium price. *Supply = demand assumes perfect competition.* With many suppliers, no one can raise the price. If someone tries, the customers go elsewhere.

The opposite case is a ***monopoly***—no competition. Instead of many small producers of wheat, there is one producer of electricity. An airport is a monopolist (and maybe the National Football League). If the price is raised, some demand remains.

Price fixing occurs when several producers act like a monopoly—which antitrust laws try to prevent. The price is not set by supply = demand. The calculus problem is different—***to maximize profit***. Section 3.2 locates the maximum where the marginal profit (the slope!) is zero.

Question on income elasticity From an income of \$10,000 you save \$500. The income elasticity of savings is $E = 2$. Out of the next dollar what fraction do you save?

Answer The savings is $y = cx^2$ because $E = 2$. The number c must give $500 = c(10{,}000)^2$, so c is $5 \cdot 10^{-6}$. Then the slope dy/dx is $2cx = 10 \cdot 10^{-6} \cdot 10^4 = \frac{1}{10}$. This is the marginal savings, ten cents on the dollar. ***Average savings is*** 5%, ***marginal savings is*** 10%, ***and*** $E = 2$.

2.2 EXERCISES

Read-through questions

The derivative of $f = x^4$ is $f' =$ __a__. That comes from expanding $(x+h)^4$ into the five terms __b__. Subtracting x^4 and dividing by h leaves the four terms __c__. This is $\Delta f/h$, and its limit is __d__.

The derivative of $f = x^n$ is $f' =$ __e__. Now $(x+h)^n$ comes from the __f__ theorem. The terms to look for are $x^{n-1}h$, containing only one __g__. There are __h__ of those terms, so $(x+h)^n = x^n +$ __i__ $+ \cdots$. After subtracting __j__ and dividing by h, the limit of $\Delta f/h$ is __k__. The coefficient of $x^{n-j}h^j$, not needed here, is "n choose j" = __l__, where $n!$ means __m__.

The derivative of x^{-2} is __n__. The derivative of $x^{1/2}$ is __o__. The derivative of $3x + (1/x)$ is __p__, which uses the following rules: The derivative of $3f(x)$ is __q__ and the derivative of $f(x) + g(x)$ is __r__. Integral calculus recovers __s__ from dy/dx. If $dy/dx = x^4$ then $y(x) =$ __t__.

1 Starting with $f = x^6$, write down f' and then f''. (This is "f double prime," the derivative of f'.) After ________ derivatives of x^6 you reach a constant. What constant?

2 Find a function that has x^6 as its derivative.

Find the derivatives of the functions in 3–10. Even if n is negative or a fraction, the derivative of x^n is nx^{n-1}.

3 $x^2 + 7x + 5$

4 $1 + (7/x) + (5/x^2)$

5 $1 + x + x^2 + x^3 + x^4$

6 $(x^2 + 1)^2$

7 $x^n + x^{-n}$

8 $x^n/n!$

9 $1 + x + \frac{1}{2}x^2 + \frac{1}{6}x^3 + \frac{1}{24}x^4$

10 $\frac{2}{3}x^{3/2} + \frac{2}{5}x^{5/2}$

11 Name two functions with $df/dx = 1/x^2$.

12 ***Find the mistake***: x^2 is $x + x + \cdots + x$ (with x terms). Its derivative is $1 + 1 + \cdots + 1$ (also x terms). So the derivative of x^2 seems to be x.

13 What are the derivatives of $3x^{1/3}$ and $-3x^{-1/3}$ and $(3x^{1/3})^{-1}$?

14 The slope of $x + (1/x)$ is zero when $x =$ ________. What does the graph do at that point?

15 Draw a graph of $y = x^3 - x$. Where is the slope zero?

16 If df/dx is negative, is $f(x)$ always negative? Is $f(x)$ negative for large x? If you think otherwise, give examples.

17 A rock thrown upward with velocity 16 ft/sec reaches height $f = 16t - 16t^2$ at time t.

(a) Find its average speed $\Delta f/\Delta t$ from $t = 0$ to $t = \frac{1}{2}$.

(b) Find its average speed $\Delta f/\Delta t$ from $t = \frac{1}{2}$ to $t = 1$.

(c) What is df/dt at $t = \frac{1}{2}$?

18 When f is in feet and t is in seconds, what are the units of f' and its derivative f''? In $f = 16t - 16t^2$, the first 16 is ft/sec but the second 16 is ________.

19 Graph $y = x^3 + x^2 - x$ from $x = -2$ to $x = 2$ and estimate where it is decreasing. Check the transition points by solving $dy/dx = 0$.

20 At a point where $dy/dx = 0$, what is special about the graph of $y(x)$? Test case: $y = x^2$.

21 Find the slope of $y = \sqrt{x}$ by algebra (then $h \to 0$):

$$\frac{\Delta y}{h} = \frac{\sqrt{x+h} - \sqrt{x}}{h} = \frac{\sqrt{x+h} - \sqrt{x}}{h}\,\frac{\sqrt{x+h} + \sqrt{x}}{\sqrt{x+h} + \sqrt{x}}.$$

22 Imitate Problem 21 to find the slope of $y = 1/\sqrt{x}$.

23 Complete Pascal's triangle for $n = 5$ and $n = 6$. Why do the numbers across each row add to 2^n?

24 Complete $(x + h)^5 = x^5 +$ ________. What are the binomial coefficients $\binom{5}{1}$ and $\binom{5}{2}$ and $\binom{5}{3}$?

25 Compute $(x + h)^3 - (x - h)^3$, divide by $2h$, and set $h = 0$. *Why divide by* $2h$ *to find this slope*?

26 Solve the differential equation $y'' = x$ to find $y(x)$.

27 For $f(x) = x^2 + x^3$, write out $f(x + \Delta x)$ and $\Delta f/\Delta x$. What is the limit at $\Delta x = 0$ and what rule about sums is confirmed?

28 The derivative of $(u(x))^2$ is ________ from Section 2.1. Test this rule on $u = x^n$.

29 What are the derivatives of $x^7 + 1$ and $(x + 1)^7$? Shift the graph of x^7.

30 If df/dx is $v(x)$, what functions have these derivatives?

(a) $4v(x)$ (b) $v(x) + 1$

(c) $v(x + 1)$ (d) $v(x) + v'(x)$.

31 What function $f(x)$ has fourth derivative equal to 1?

32 What function $f(x)$ has nth derivative equal to 1?

33 Suppose $df/dx = 1 + x + x^2 + x^3$. Find $f(x)$.

34 Suppose $df/dx = x^{-2} - x^{-3}$. Find $f(x)$.

35 $f(x)$ *can be its own derivative.* In the infinite polynomial $f = 1 + x + \frac{1}{2}x^2 + \frac{1}{6}x^3 +$ ________, what numbers multiply x^4 and x^5 if df/dx equals f?

36 Write down a differential equation $dy/dx =$ ________ that is solved by $y = x^2$. Make the right side involve y (not just $2x$).

37 *True or false*: (a) The derivative of x^π is πx^π.

(b) The derivative of ax^n/bx^n is a/b.

(c) If $df/dx = x^4$ and $dg/dx = x^4$ then $f(x) = g(x)$.

(d) $(f(x) - f(a))/(x - a)$ approaches $f'(a)$ as $x \to a$.

(e) The slope of $y = (x - 1)^3$ is $y' = 3(x - 1)^2$.

Problems 38–44 are about calculus in economics.

38 When the cost is $y = y_0 + cx$, find $E(x) = (dy/dx)/(y/x)$. It approaches ________ for large x.

39 From an income of $x = \$10{,}000$ you spend $y = \$1200$ on your car. If $E = \frac{1}{2}$, what fraction of your next dollar will be spent on the car? Compare dy/dx (marginal) with y/x (average).

40 Name a product whose price elasticity is

(a) high (b) low (c) negative (?)

41 The demand $y = c/x$ has $dy/dx = -y/x$. Show that $\Delta y/\Delta x$ is *not* $-y/x$. (Use numbers or algebra.) Finite steps miss the special feature of infinitesimal steps.

42 The demand $y = x^n$ has $E =$ ________. The revenue xy (price times demand) has elasticity $E =$ ________.

43 $y = 2x + 3$ grows with marginal cost 2 from the fixed cost 3. Draw the graph of $E(x)$.

44 From an income I we save $S(I)$. The *marginal* propensity to save is ________. Elasticity is not needed because S and I have the same ________. Applied to the whole economy this is (microeconomics) (macroeconomics).

45 2^t is doubled when t increases by ________. t^3 is doubled when t increases to ________t. The doubling time for AIDS is proportional to t.

46 Biology also leads to $dy/y = n\,dx/x$, for the relative growth of the head (dy/y) and the body (dx/x). Is $n > 1$ or $n < 1$ for a child?

47 What functions have $df/dx = x^9$ and $df/dx = x^n$? Why does $n = -1$ give trouble?

48 The slope of $y = x^3$ comes from this identity:

$$\frac{(x + h)^3 - x^3}{h} = (x + h)^2 + (x + h)x + x^2.$$

(a) Check the algebra. Find dy/dx as $h \to 0$.

(b) Write a similar identity for $y = x^4$.

49 (Computer graphing) Find all the points where $y = x^4 + 2x^3 - 7x^2 + 3 = 0$ and where $dy/dx = 0$.

50 The graphs of $y_1(x) = x^4 + x^3$ and $y_2(x) = 7x - 5$ touch at the point where $y_3(x) =$ ________ $= 0$. Plot $y_3(x)$ to see what is special. What does the graph of $y(x)$ do at a point where $y = y' = 0$?

51 In the Massachusetts lottery you choose 6 numbers out of 36. What is your chance to win?

52 In what circumstances would it pay to buy a lottery ticket for every possible combination, so one of the tickets would win?

2.3 The Slope and the Tangent Line

Chapter 1 started with straight line graphs. The velocity was constant (at least piecewise). The distance function was linear. Now we are facing polynomials like $x^3 - 2$ or $x^4 - x^2 + 3$, with other functions to come soon. Their graphs are definitely curved. Most functions are not close to linear—except if you focus all your attention near a single point. That is what we will do.

Over a very short range a curve looks straight. Look through a microscope, or zoom in with a computer, and there is no doubt. The graph of distance versus time becomes nearly linear. Its slope is the velocity at that moment. We want to find the line that the graph stays closest to—the "***tangent line***"—before it curves away.

The tangent line is easy to describe. We are at a particular point on the graph of $y = f(x)$. At that point x equals a and y equals $f(a)$ and the slope equals $f'(a)$. ***The tangent line goes through that point*** $x = a$, $y = f(a)$ ***with that slope*** $m = f'(a)$. Figure 2.5 shows the line more clearly than any equation, but we have to turn the geometry into algebra. We need the equation of the line.

EXAMPLE 1 Suppose $y = x^4 - x^2 + 3$. At the point $x = a = 1$, the height is $y = f(a) = 3$. The slope is $dy/dx = 4x^3 - 2x$. At $x = 1$ the slope is $4 - 2 = 2$. That is $f'(a)$:

The numbers $x = 1$, $y = 3$, $dy/dx = 2$ ***determine the tangent line.***

The equation of the tangent line is $y - 3 = 2(x - 1)$, *and this section explains why.*

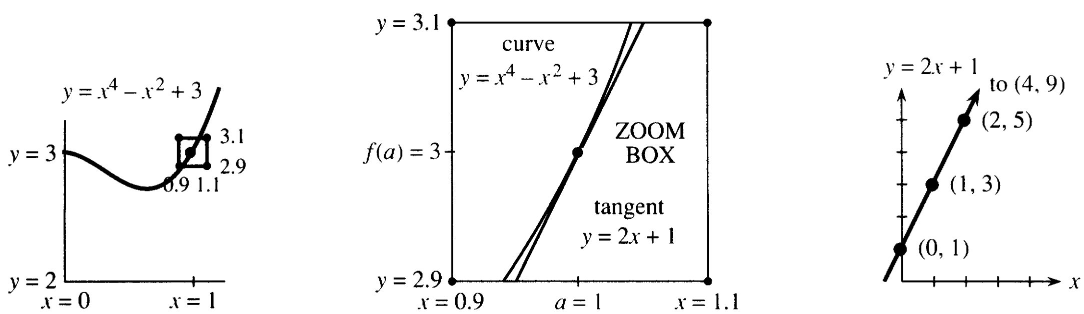


Fig. 2.5 The tangent line has the same slope 2 as the curve (especially after zoom).

THE EQUATION OF A LINE

A straight line is determined by two conditions. We know the line if we know two of its points. (We still have to write down the equation.) Also, if we know ***one point and the slope***, the line is set. That is the situation for the tangent line, which has a known slope at a known point:

1. The equation of a line has the form $y = mx + b$
2. The number m is the slope of the line, because $dy/dx = m$
3. The number b adjusts the line to go through the required point.

I will take those one at a time—first $y = mx + b$, then m, then b.

1. The graph of $y = mx + b$ is not curved. How do we know? For the specific example $y = 2x + 1$, take two points whose coordinates x, y satisfy the equation:

$$x = 0,\ y = 1 \quad \text{and} \quad x = 4,\ y = 9 \quad \text{both satisfy} \quad y = 2x + 1.$$

Those points (0, 1) and (4, 9) lie on the graph. *The point halfway between has* $x = 2$ *and* $y = 5$. That point also satisfies $y = 2x + 1$. ***The halfway point is on the graph.*** If we subdivide again, the midpoint between (0, 1) and (2, 5) is (1, 3). This also has $y = 2x + 1$. The graph contains all halfway points and must be straight.

2. What is the correct slope m for the tangent line? In our example it is $m = f'(a) = 2$. ***The curve and its tangent line have the same slope at the crucial point***: $dy/dx = 2$.

Allow me to say in another way why the line $y = mx + b$ has slope m. At $x = 0$ its height is $y = b$. At $x = 1$ its height is $y = m + b$. The graph has gone *one unit across* (0 to 1) *and m units up* (b to $m + b$). The whole idea is

$$\text{slope} = \frac{\text{distance up}}{\text{distance across}} = \frac{m}{1}. \tag{1}$$

Each unit across means m units up, to $2m + b$ or $3m + b$. A straight line keeps a constant slope, whereas the slope of $y = x^4 - x^2 + 3$ equals 2 only at $x = 1$.

3. Finally we decide on b. The tangent line $y = 2x + b$ must go through $x = 1$, $y = 3$. Therefore $b = 1$. With letters instead of numbers, $y = mx + b$ leads to $f(a) = ma + b$. So we know b:

> **2E** The equation of the tangent line has $b = f(a) - ma$:
>
> $$y = mx + f(a) - ma \qquad \text{or} \qquad y - f(a) = m(x - a). \tag{2}$$

That last form is the best. You see immediately what happens at $x = a$. The factor $x - a$ is zero. Therefore $y = f(a)$ as required. This is the ***point-slope form*** of the equation, and we use it constantly:

$$y - 3 = 2(x - 1) \qquad \text{or} \qquad \frac{y-3}{x-1} = \frac{\text{distance up}}{\text{distance across}} = \text{slope } 2.$$

EXAMPLE 2 The curve $y = x^3 - 2$ goes through $y = 6$ when $x = 2$. At that point $dy/dx = 3x^2 = 12$. The point-slope equation of the tangent line uses 2 and 6 and 12:

$$y - 6 = 12(x - 2), \quad \text{which is also} \quad y = 12x - 18.$$

There is another important line. It is *perpendicular* to the tangent line and *perpendicular* to the curve. This is the ***normal line*** in Figure 2.6. Its new feature is its slope. When the tangent line has slope m, the normal line has slope $-1/m$. (Rule: Slopes of

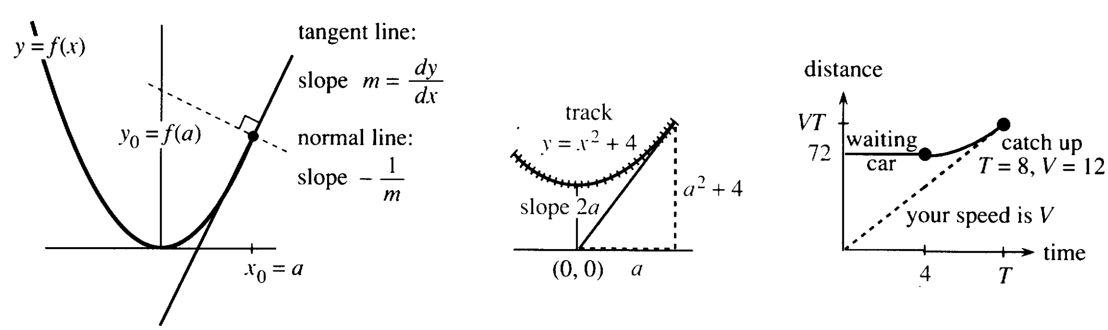


Fig. 2.6 Tangent line $y - y_0 = m(x - x_0)$. Normal line $y - y_0 = -\frac{1}{m}(x - x_0)$. Leaving a roller-coaster and catching up to a car.

perpendicular lines multiply to give -1.) Example 2 has $m=12$, so the normal line has slope $-1/12$:

$$\textit{tangent line}: y-6=12(x-2) \qquad \textit{normal line}: y-6=-\tfrac{1}{12}(x-2).$$

Light rays travel in the normal direction. So do brush fires—they move perpendicular to the fire line. Use the point-slope form! The tangent is $y=12x-18$, the normal is not $y=-\frac{1}{12}x-18$.

EXAMPLE 3 You are on a roller-coaster whose track follows $y=x^2+4$. You see a friend at (0, 0) and want to get there quickly. Where do you step off?

Solution Your path will be the tangent line (at high speed). The problem is ***to choose*** $x=a$ ***so the tangent line passes through*** $x=0$, $y=0$. When you step off at $x=a$,

the height is $y=a^2+4$ and the slope is $2a$

the equation of the tangent line is $y-(a^2+4)=2a(x-a)$

this line goes through (0, 0) if $-(a^2+4)=-2a^2$ or $a=\pm 2$.

The same problem is solved by spacecraft controllers and baseball pitchers. Releasing a ball at the right time to hit a target 60 feet away is an amazing display of calculus. Quarterbacks with a moving target should read Chapter 4 on related rates.

Here is a better example than a roller-coaster. Stopping at a red light wastes gas. It is smarter to slow down early, and then accelerate. When a car is waiting in front of you, the timing needs calculus:

EXAMPLE 4 How much must you slow down when a red light is 72 meters away? In 4 seconds it will be green. The waiting car will accelerate at 3 meters/sec^2. You cannot pass the car.

Strategy Slow down immediately to the speed V at which you will just catch that car. (If you wait and brake later, your speed will have to go below V.) At the catch-up time T, the cars have the same speed and same distance. ***Two conditions***, so the distance functions in Figure 2.6d are tangent.

Solution At time T, the other car's speed is $3(T-4)$. That shows the delay of 4 seconds. Speeds are equal when $3(T-4)=V$ or $T=\frac{1}{3}V+4$. Now require equal distances. Your distance is V times T. The other car's distance is $72+\frac{1}{2}at^2$:

$$72+\tfrac{1}{2}\cdot 3(T-4)^2=VT \quad \text{becomes} \quad 72+\tfrac{1}{2}\cdot\tfrac{1}{3}V^2=V(\tfrac{1}{3}V+4).$$

The solution is $V=12$ meters/second. This is 43 km/hr or 27 miles per hour.

Without the other car, you only slow down to $V=72/4=18$ meters/second. As the light turns green, you go through at 65 km/hr or 40 miles per hour. Try it.

THE SECANT LINE CONNECTING TWO POINTS ON A CURVE

Instead of the tangent line through one point, consider the ***secant line through two points***. For the tangent line the points came together. Now spread them apart. The point-slope form of a linear equation is replaced by the ***two-point form***.

The equation of the curve is still $y=f(x)$. The first point remains at $x=a$, $y=f(a)$. The other point is at $x=c$, $y=f(c)$. The secant line goes between them, and we want its equation. This time we don't start with the slope—but m is easy to find.

EXAMPLE 5 The curve $y = x^3 - 2$ goes through $x = 2$, $y = 6$. It also goes through $x = 3$, $y = 25$. The slope between those points is

$$m = \frac{\text{change in } y}{\text{change in } x} = \frac{25 - 6}{3 - 2} = 19.$$

The point-slope form (at the first point) is $y - 6 = 19(x - 2)$. This line automatically goes through the second point (3, 25). Check: $25 - 6$ equals $19(3 - 2)$. The secant has the right slope 19 to reach the second point. It is the *average slope* $\Delta y/\Delta x$.

A look ahead The second point is going to approach the first point. The secant slope $\Delta y/\Delta x$ will approach the tangent slope dy/dx. We discover the derivative (in the limit). That is the main point now—but not forever.

Soon you will be fast at derivatives. The exact dy/dx will be much easier than $\Delta y/\Delta x$. The situation is turned around as soon as you know that x^9 has slope $9x^8$. Near $x = 1$, the distance *up* is about 9 times the distance *across*. To find $\Delta y = 1.001^9 - 1^9$, just multiply $\Delta x = .001$ by 9. The quick approximation is .009, the calculator gives $\Delta y = .009036$. It is easier to follow the tangent line than the curve.

Come back to the secant line, and change numbers to letters. What line connects $x = a$, $y = f(a)$ to $x = c$, $y = f(c)$? A mathematician puts formulas ahead of numbers, and reasoning ahead of formulas, and ideas ahead of reasoning:

(1) The slope is $m = \dfrac{\text{distance up}}{\text{distance across}} = \dfrac{f(c) - f(a)}{c - a}$

(2) The height is $y = f(a)$ at $x = a$

(3) The height is $y = f(c)$ at $x = c$ (automatic with correct slope).

2F The ***two-point form*** uses the slope between the points:

$$\textbf{\textit{secant line}}: \quad y - f(a) = \left(\frac{f(c) - f(a)}{c - a}\right)(x - a). \qquad (3)$$

At $x = a$ the right side is zero. So $y = f(a)$ on the left side. At $x = c$ the right side has two factors $c - a$. They cancel to leave $y = f(c)$. With equation (2) for the tangent line and equation (3) for the secant line, we are ready for the moment of truth.

THE SECANT LINE APPROACHES THE TANGENT LINE

What comes now is pretty basic. It matches what we did with velocities:

$$\text{average velocity} = \frac{\Delta \text{ distance}}{\Delta \text{ time}} = \frac{f(t + \Delta t) - f(t)}{\Delta t}.$$

The limit is df/dt. We now do exactly the same thing with slopes. ***The secant line turns into the tangent line as c approaches*** a:

$$\text{slope of secant line:} \quad \frac{\Delta f}{\Delta x} = \frac{f(c) - f(a)}{c - a}$$

$$\text{slope of tangent line:} \quad \frac{df}{dx} = \text{limit of } \frac{\Delta f}{\Delta x}.$$

There stands the fundamental idea of differential calculus! You have to imagine more secant lines than I can draw in Figure 2.7, as c comes close to a. Everybody recognizes $c-a$ as Δx. Do you recognize $f(c)-f(a)$ as $f(x+\Delta x)-f(x)$? It is Δf, the change in height. All lines go through $x=a$, $y=f(a)$. ***Their limit is the tangent line.***

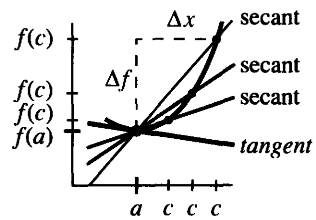


Fig. 2.7 Secants approach tangent as their slopes $\Delta f/\Delta x$ approach df/dx.

secant $\quad y-f(a)=\dfrac{f(c)-f(a)}{c-a}(x-a)$

tangent $\quad y-f(a)=f'(a)(x-a)$

Intuitively, the limit is pretty clear. The two points come together, and the tangent line touches the curve at *one* point. (It could touch again at faraway points.) Mathematically this limit can be tricky—it takes us from algebra to calculus. Algebra stays away from 0/0, but calculus gets as close as it can.

The new limit for df/dx looks different, but it is the same as before:

$$f'(a)=\lim_{c\to a}\frac{f(c)-f(a)}{c-a}. \qquad (4)$$

EXAMPLE 6 Find the secant lines and tangent line for $y=f(x)=\sin x$ at $x=0$.

The starting point is $x=0$, $y=\sin 0$. This is the origin (0, 0). The ratio of distance up to distance across is $(\sin c)/c$:

$$\textit{secant equation } y=\frac{\sin c}{c}x \qquad \textit{tangent equation } y=1x.$$

As c approaches zero, the secant line becomes the tangent line. The limit of $(\sin c)/c$ is not 0/0, which is meaningless, but 1, which is dy/dx.

EXAMPLE 7 The gold you own will be worth $\sqrt{t}$ million dollars in t years. When does the rate of increase drop to 10% of the current value, so you should sell the gold and buy a bond? At $t=25$, how far does that put you ahead of $\sqrt{t}=5$?

Solution The rate of increase is the derivative of $\sqrt{t}$, which is $1/2\sqrt{t}$. That is 10% of the current value $\sqrt{t}$ when $1/2\sqrt{t}=\sqrt{t}/10$. Therefore $2t=10$ or $t=5$. At that time you sell the gold, leave the curve, and go onto the tangent line:

$$y-\sqrt{5}=\frac{\sqrt{5}}{10}(t-5) \quad \text{becomes} \quad y-\sqrt{5}=2\sqrt{5} \quad \text{at} \quad t=25.$$

With straight interest on the bond, not compounded, you have reached $y=3\sqrt{5}=6.7$ million dollars. The gold is worth a measly five million.

2.3 EXERCISES

Read-through questions

A straight line is determined by ___a___ points, or one point and the ___b___. The slope of the tangent line equals the slope of the ___c___. The point-slope form of the tangent equation is $y-f(a)=$ ___d___.

The tangent line to $y=x^3+x$ at $x=1$ has slope ___e___. Its equation is __f__. It crosses the y axis at __g__ and the x axis at __h__. The normal line at this point (1, 2) has slope __i__. Its equation is $y-2=$ __j__. The secant line from (1, 2) to (2, __k__) has slope __l__. Its equation is $y-2=$ __m__.

The point $(c, f(c))$ is on the line $y-f(a)=m(x-a)$ provided $m=$ __n__. As c approaches a, the slope m approaches __o__. The secant line approaches the __p__ line.

1 (a) Find the slope of $y=12/x$.
(b) Find the equation of the tangent line at (2, 6).
(c) Find the equation of the normal line at (2, 6).
(d) Find the equation of the secant line to (4, 3).

2 For $y=x^2+x$ find equations for
(a) the tangent line and normal line at (1, 2);
(b) the secant line to $x=1+h$, $y=(1+h)^2+(1+h)$.

3 A line goes through (1, −1) and (4, 8). Write its equation in point-slope form. Then write it as $y=mx+b$.

4 The tangent line to $y=x^3+6x$ at the origin is $y=$ ________. Does it cross the curve again?

5 The tangent line to $y=x^3-3x^2+x$ at the origin is $y=$ ________. It is also the secant line to the point ________.

6 Find the tangent line to $x=y^2$ at $x=4$, $y=2$.

7 For $y=x^2$ the secant line from (a, a^2) to (c, c^2) has the equation ________. Do the division by $c-a$ to find the tangent line as c approaches a.

8 Construct a function that has the same slope at $x=1$ and $x=2$. Then find two points where $y=x^4-2x^2$ has the same tangent line (draw the graph).

9 Find a curve that is tangent to $y=2x-3$ at $x=5$. Find the normal line to that curve at (5, 7).

10 For $y=1/x$ the secant line from $(a, 1/a)$ to $(c, 1/c)$ has the equation ________. Simplify its slope and find the limit as c approaches a.

11 What are the equations of the tangent line and normal line to $y=\sin x$ at $x=\pi/2$?

12 If c and a both approach an in-between value $x=b$, then the secant slope $(f(c)-f(a))/(c-a)$ approaches ________.

13 At $x=a$ on the graph of $y=1/x$, compute
(a) the equation of the tangent line
(b) the points where that line crosses the axes.
The triangle between the tangent line and the axes always has area ________.

14 Suppose $g(x)=f(x)+7$. The tangent lines to f and g at $x=4$ are ________. *True or false*: The distance between those lines is 7.

15 Choose c so that $y=4x$ is tangent to $y=x^2+c$. Match heights as well as slopes.

16 Choose c so that $y=5x-7$ is tangent to $y=x^2+cx$.

17 For $y=x^3+4x^2-3x+1$, find all points where the tangent is horizontal.

18 $y=4x$ can't be tangent to $y=cx^2$. Try to match heights and slopes, or draw the curves.

19 Determine c so that the straight line joining (0, 3) and (5, −2) is tangent to the curve $y=c/(x+1)$.

20 Choose b, c, d so that the two parabolas $y=x^2+bx+c$ and $y=dx-x^2$ are tangent to each other at $x=1$, $y=0$.

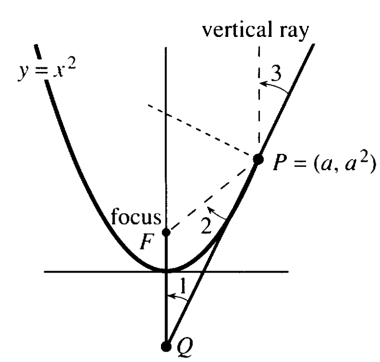
21 The graph of $f(x)=x^3$ goes through (1, 1).
(a) Another point is $x=c=1+h$, $y=f(c)=$ ________.
(b) The change in f is $\Delta f=$ ________.
(c) The slope of the secant is $m=$ ________.
(d) As h goes to zero, m approaches ________.

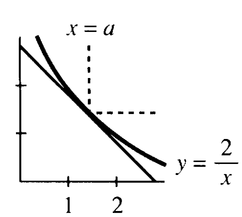
22 Construct a function $y=f(x)$ whose tangent line at $x=1$ is the same as the secant that meets the curve again at $x=3$.

23 Draw two curves bending away from each other. Mark the points P and Q where the curves are closest. At those points, the tangent lines are ________ and the normal lines are ________.

***24** If the parabolas $y=x^2+1$ and $y=x-x^2$ come closest at (a, a^2+1) and $(c, c-c^2)$, set up two equations for a and c.

25 A light ray comes down the line $x=a$. It hits the parabolic reflector $y=x^2$ at $P=(a, a^2)$.
(a) Find the tangent line at P. Locate the point Q where that line crosses the y axis.
(b) Check that P and Q are the same distance from the focus at $F=(0, \frac{1}{4})$.
(c) Show from (b) that the figure has equal angles.
(d) What law of physics makes every ray reflect off the parabola to the focus at F?





26 In a bad reflector $y=2/x$, a ray down one special line $x=a$ is reflected horizontally. What is a?

27 For the parabola $4py = x^2$, where is the slope equal to 1? At that point a vertical ray will reflect horizontally. So the focus is at (0, ________).

28 Why are these statements wrong? Make them right.

(a) If $y = 2x$ is the tangent line at (1, 2) then $y = -\frac{1}{2}x$ is the normal line.

(b) As c approaches a, the secant slope $(f(c) - f(a))/(c - a)$ approaches $(f(a) - f(a))/(a - a)$.

(c) The line through (2, 3) with slope 4 is $y - 2 = 4(x - 3)$.

29 A ball goes around a circle: $x = \cos t$, $y = \sin t$. At $t = 3\pi/4$ the ball flies off on the tangent line. Find the equation of that line and the point where the ball hits the ground ($y = 0$).

30 If the tangent line to $y = f(x)$ at $x = a$ is the same as the tangent line to $y = g(x)$ at $x = b$, find two equations that must be satisfied by a and b.

31 Draw a circle of radius 1 resting in the parabola $y = x^2$. At the touching point (a, a^2), the equation of the normal line is ________. That line has $x = 0$ when $y =$ ________. The distance to (a, a^2) equals the radius 1 when $a =$ ________. This locates the touching point.

32 Follow Problem 31 for the flatter parabola $y = \frac{1}{2}x^2$ and explain where the circle rests.

33 You are applying for a \$1000 scholarship and your time is worth \$10 a hour. If the chance of success is $1 - (1/x)$ from x hours of writing, when should you stop?

34 Suppose $|f(c) - f(a)| \leqslant |c - a|$ for every pair of points a and c. Prove that $|df/dx| \leqslant 1$.

35 From which point $x = a$ does the tangent line to $y = 1/x^2$ hit the x axis at $x = 3$?

36 If $u(x)/v(x) = 7$ find $u'(x)/v'(x)$. Also find $(u(x)/v(x))'$.

37 Find $f(c) = 1.001^{10}$ in two ways—by calculator and by $f(c) - f(a) \approx f'(a)(c - a)$. Choose $a = 1$ and $f(x) = x^{10}$.

38 At a distance Δx from $x = 1$, how far is the curve $y = 1/x$ above its tangent line?

39 At a distance Δx from $x = 2$, how far is the curve $y = x^3$ above its tangent line?

40 Based on Problem 38 or 39, the distance between curve and tangent line grows like what power $(\Delta x)^p$?

41 The tangent line to $f(x) = x^2 - 1$ at $x_0 = 2$ crosses the x axis at $x_1 =$ ________. The tangent line at x_1 crosses the x axis at $x_2 =$ ________. Draw the curve and the two lines, which are the beginning of *Newton's method* to solve $f(x) = 0$.

42 (Puzzle) The equation $y = mx + b$ requires *two* numbers, the point-slope form $y - f(a) = f'(a)(x - a)$ requires *three*, and the two-point form requires *four*: $a, f(a), c, f(c)$. How can this be?

43 Find the time T at the tangent point in Example 4, when you catch the car in front.

44 If the waiting car only accelerates at 2 meters/sec^2, what speed V must you slow down to?

45 A thief 40 meters away runs toward you at 8 meters per second. What is the smallest acceleration so that $v = at$ keeps you in front?

46 With 8 meters to go in a relay race, you slow down badly ($f = -8 + 6t - \frac{1}{2}t^2$). How fast should the next runner start (choose v in $f = vt$) so you can just pass the baton?

2.4 The Derivative of the Sine and Cosine

This section does two things. One is to compute the derivatives of $\sin x$ and $\cos x$. The other is to explain why these functions are so important. They describe ***oscillation***, which will be expressed in words and equations. You will see a "***differential equation.***" It involves the derivative of an unknown function $y(x)$.

The differential equation will say that the *second* derivative—***the derivative of the derivative***—is equal and opposite to y. In symbols this is $y'' = -y$. Distance in one direction leads to acceleration in the other direction. That makes y and y' and y'' all oscillate. The solutions to $y'' = -y$ are $\sin x$ and $\cos x$ and all their combinations.

We begin with the slope. The derivative of $y = \sin x$ is $y' = \cos x$. There is no reason for that to be a mystery, but I still find it beautiful. Chapter 1 followed a ball around a circle; the shadow went up and down. Its height was $\sin t$ and its velocity was $\cos t$.

We now find that derivative by *the standard method of limits*, when $y(x) = \sin x$:

$$\frac{dy}{dx} = \text{limit of } \frac{\Delta y}{\Delta x} = \lim_{h \to 0} \frac{\sin(x+h) - \sin x}{h}. \tag{1}$$

The sine is harder to work with than x^2 or x^3. Where we had $(x+h)^2$ or $(x+h)^3$, we now have $\sin(x+h)$. This calls for one of the basic "addition formulas" from trigonometry, reviewed in Section 1.5:

$$\sin(x+h) = \sin x \cos h + \cos x \sin h \tag{2}$$

$$\cos(x+h) = \cos x \cos h - \sin x \sin h. \tag{3}$$

Equation (2) puts $\Delta y = \sin(x+h) - \sin x$ in a new form:

$$\frac{\Delta y}{\Delta x} = \frac{\sin x \cos h + \cos x \sin h - \sin x}{h} = \sin x \left(\frac{\cos h - 1}{h}\right) + \cos x \left(\frac{\sin h}{h}\right). \tag{4}$$

The ratio splits into two simpler pieces on the right. Algebra and trigonometry got us this far, and now comes the calculus problem. ***What happens as*** $h \to 0$? It is no longer easy to divide by h. (I will not even mention the unspeakable crime of writing $(\sin h)/h = \sin$.) There are two critically important limits—the first is zero and the second is one:

$$\lim_{h \to 0} \frac{\cos h - 1}{h} = 0 \quad \text{and} \quad \lim_{h \to 0} \frac{\sin h}{h} = 1. \tag{5}$$

The careful reader will object that limits have not been defined! You may further object to computing these limits separately, before combining them into equation (4). Nevertheless—following the principle of *ideas now, rigor later*—I would like to proceed. It is entirely true that the limit of (4) comes from the two limits in (5):

$$\frac{dy}{dx} = (\sin x)(\text{first limit}) + (\cos x)(\text{second limit}) = 0 + \cos x. \tag{6}$$

The secant slope $\Delta y/\Delta x$ has approached the tangent slope dy/dx.

2G The derivative of $y = \sin x$ is $dy/dx = \cos x$.

We cannot pass over the crucial step—the two limits in (5). They contain the real ideas. ***Both ratios become*** $0/0$ ***if we just substitute*** $h = 0$. Remember that the cosine of a zero angle is 1, and the sine of a zero angle is 0. Figure 2.8a shows a small angle h (as near to zero as we could reasonably draw). The edge of length $\sin h$ is close to zero, and the edge of length $\cos h$ is near 1. Figure 2.8b shows how the *ratio* of $\sin h$ to h (both headed for zero) gives the slope of the sine curve at the start.

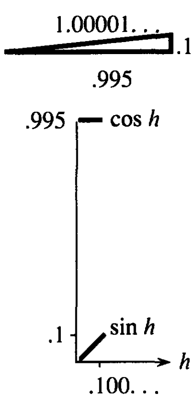


Fig. 2.8

When two functions approach zero, their ratio might do anything. We might have

$$\frac{h^2}{h} \to 0 \quad \text{or} \quad \frac{h}{h} \to 1 \quad \text{or} \quad \frac{\sqrt{h}}{h} \to \infty.$$

No clue comes from $0/0$. What matters is *whether the top or bottom goes to zero more quickly*. Roughly speaking, we want to show that $(\cos h - 1)/h$ is like h^2/h and $(\sin h)/h$ is like h/h.

Time out The graph of $\sin x$ is in Figure 2.9 (in black). The graph of $\sin(x + \Delta x)$ sits just beside it (in red). The height difference is Δf when the shift distance is Δx.

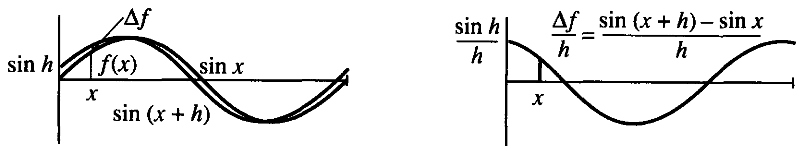


Fig. 2.9 $\sin(x+h)$ with $h = 10° = \pi/18$ radians. $\Delta f/\Delta x$ is close to $\cos x$.

Now divide by that small number Δx (or h). The second figure shows $\Delta f/\Delta x$. It is close to $\cos x$. (Look how it starts—it is not quite $\cos x$.) Mathematics will prove that the limit is $\cos x$ exactly, when $\Delta x \to 0$. Curiously, the reasoning concentrates on only one point ($x = 0$). The slope at that point is $\cos 0 = 1$.

We now prove this: $\sin \Delta x$ divided by Δx goes to 1. The sine curve starts with slope 1. By the addition formula for $\sin(x+h)$, this answer at one point will lead to the slope $\cos x$ at all points.

Question Why does the graph of $f(x + \Delta x)$ shift *left* from $f(x)$ when $\Delta x > 0$?
Answer When $x = 0$, the shifted graph is already showing $f(\Delta x)$. In Figure 2.9a, the red graph is shifted *left* from the black graph. The red graph shows $\sin h$ when the black graph shows $\sin 0$.

THE LIMIT OF $(\sin h)/h$ IS 1

There are several ways to find this limit. The direct approach is to let a computer draw a graph. Figure 2.10a is very convincing. ***The function*** $(\sin h)/h$ ***approaches*** 1 ***at the key point*** $h = 0$. So does $(\tan h)/h$. In practice, the only danger is that you might get a message like "undefined function" and no graph. (The machine may refuse to divide by zero at $h = 0$. Probably you can get around that.) Because of the importance of this limit, I want to give a mathematical proof that it equals 1.

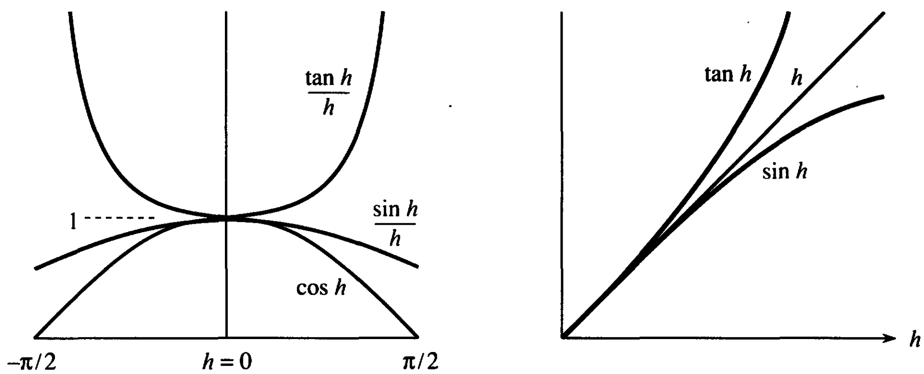


Fig. 2.10 $(\sin h)/h$ squeezed between $\cos x$ and 1; $(\tan h)/h$ decreases to 1.

Figure 2.10b indicates, but still only graphically, that $\sin h$ *stays below* h. (The first graph shows that too; $(\sin h)/h$ is below 1.) We also see that $\tan h$ *stays above* h. Remember that the tangent is the ratio of sine to cosine. Dividing by the cosine is enough to push the tangent above h. The crucial inequalities (to be proved when h is small and positive) are

$$\sin h < h \quad \text{and} \quad \tan h > h. \tag{7}$$

Since $\tan h = (\sin h)/(\cos h)$, those are the same as

$$\frac{\sin h}{h} < 1 \quad \text{and} \quad \frac{\sin h}{h} > \cos h. \tag{8}$$

What happens as h goes to zero? ***The ratio*** $(\sin h)/h$ ***is squeezed between*** $\cos h$ ***and*** 1. But $\cos h$ is approaching 1! The squeeze as $h \to 0$ leaves only one possibility for $(\sin h)/h$, which is caught in between: *The ratio* $(\sin h)/h$ *approaches* 1.

Figure 2.10 shows that "squeeze play." ***If two functions approach the same limit, so does any function caught in between.*** This is proved at the end of Section 2.6.

For negative values of h, which are absolutely allowed, the result is the same. To the left of zero, h reverses sign and $\sin h$ reverses sign. The ratio $(\sin h)/h$ is unchanged. (The sine is an odd function: $\sin(-h) = -\sin h$.) The ratio is an *even* function, symmetric around zero and approaching 1 from both sides.

The proof depends on $\sin h < h < \tan h$, which is displayed by the graph but not explained. We go back to right triangles.

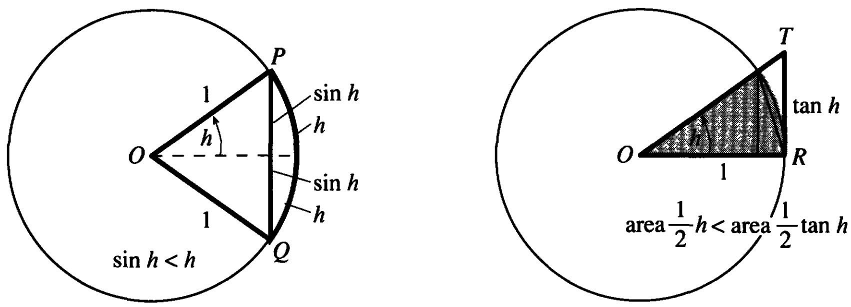


Fig. 2.11 Line shorter than arc: $2 \sin h < 2h$. Areas give $h < \tan h$.

Figure 2.11a shows why $\sin h < h$. The straight line PQ has length $2 \sin h$. The circular arc must be longer, because the shortest distance between two points is a straight line.† The arc PQ has length $2h$. (Important: *When the radius is* 1, *the arc length equals the angle.* The full circumference is 2π and the full angle is also 2π.) ***The straight distance*** $2 \sin h$ ***is less than the circular distance*** $2h$, ***so*** $\sin h < h$.

Figure 2.11b shows why $h < \tan h$. This time we look at ***areas***. The triangular area is $\frac{1}{2}$(base)(height) $= \frac{1}{2}(1)(\tan h)$. Inside that triangle is the shaded sector of the circle. Its area is $h/2\pi$ times the area of the whole circle (because the angle is that fraction of the whole angle). The circle has area $\pi r^2 = \pi$, so multiplication by $h/2\pi$ gives $\frac{1}{2}h$ for the area of the sector. Comparing with the triangle around it, $\frac{1}{2} \tan h > \frac{1}{2}h$.

The inequalities $\sin h < h < \tan h$ are now proved. The squeeze in equation (8) produces $(\sin h)/h \to 1$. Q.E.D. Problem 13 shows how to prove $\sin h < h$ from areas.

Note All angles x and h are being measured in radians. *In degrees,* $\cos x$ *is not the derivative of* $\sin x$. A degree is much less than a radian, and dy/dx is reduced by the factor $2\pi/360$.

THE LIMIT OF $(\cos h - 1)/h$ IS 0

This second limit is different. We will show that $1 - \cos h$ shrinks to zero *more quickly* than h. Cosines are connected to sines by $(\sin h)^2 + (\cos h)^2 = 1$. We start from the

†If we try to prove that, we will be here all night. Accept it as true.

known fact $\sin h < h$ and work it into a form involving cosines:

$$(1-\cos h)(1+\cos h) = 1-(\cos h)^2 = (\sin h)^2 < h^2. \tag{9}$$

Note that everything is positive. Divide through by h and also by $1+\cos h$:

$$0 < \frac{1-\cos h}{h} < \frac{h}{1+\cos h}. \tag{10}$$

Our ratio is caught in the middle. *The right side goes to zero because* $h \to 0$. This is another "*squeeze*"—there is no escape. Our ratio goes to zero.

For $\cos h - 1$ or for negative h, the signs change but minus zero is still zero. This confirms equation (6). The slope of $\sin x$ is $\cos x$.

Remark Equation (10) also shows that $1-\cos h$ is approximately $\frac{1}{2}h^2$. The 2 comes from $1+\cos h$. This is a basic purpose of calculus—to find simple approximations like $\frac{1}{2}h^2$. A "tangent parabola" $1-\frac{1}{2}h^2$ is close to the top of the cosine curve.

THE DERIVATIVE OF THE COSINE

This will be easy. The quick way to differentiate $\cos x$ is to shift the sine curve by $\pi/2$. That yields the cosine curve (solid line in Figure 2.12b). The derivative also shifts by $\pi/2$ (dotted line). ***The derivative of*** $\cos x$ ***is*** $-\sin x$.

Notice how the dotted line (the slope) goes below zero when the solid line turns downward. The slope equals zero when the solid line is level. ***Increasing functions have positive slopes. Decreasing functions have negative slopes.*** That is important, and we return to it.

There is more information in dy/dx than "function rising" or "function falling." The slope tells *how quickly* the function goes up or down. It gives the *rate of change*. The slope of $y = \cos x$ can be computed in the normal way, as the limit of $\Delta y/\Delta x$:

$$\frac{\Delta y}{\Delta x} = \frac{\cos(x+h)-\cos x}{h} = \cos x\left(\frac{\cos h - 1}{h}\right) - \sin x\left(\frac{\sin h}{h}\right)$$

$$\frac{dy}{dx} = (\cos x)(0) - (\sin x)(1) = -\sin x. \tag{11}$$

The first line came from formula (3) for $\cos(x+h)$. The second line took limits, reaching 0 and 1 as before. This confirms the graphical proof that the slope of $\cos x$ is $-\sin x$.

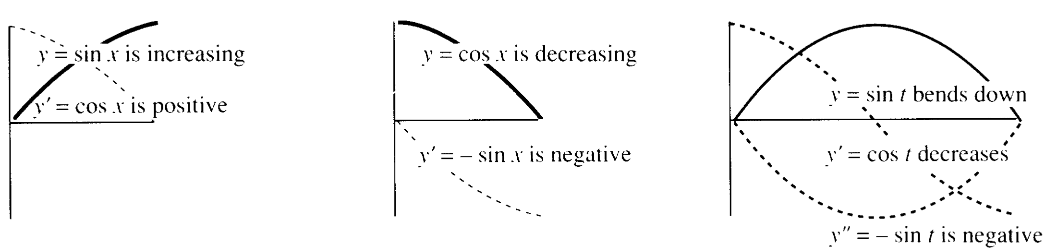


Fig. 2.12 $y(x)$ increases where y' is positive. $y(x)$ bends up where y'' is positive.

THE SECOND DERIVATIVES OF THE SINE AND COSINE

We now introduce ***the derivative of the derivative***. That is the ***second derivative*** of the original function. It tells how fast the slope is changing, not how fast y itself is changing. The second derivative is the "rate of change of the velocity." A straight line has constant slope (constant velocity), so its second derivative is zero:

$$f(t) = 5t \quad \text{has} \quad df/dt = 5 \quad \text{and} \quad d^2f/dt^2 = 0.$$

The parabola $y = x^2$ has slope $2x$ (linear) which has slope 2 (constant). Similarly

$$f(t) = \tfrac{1}{2}at^2 \quad \text{has} \quad df/dt = at \quad \text{and} \quad d^2f/dt^2 = a.$$

There stands the notation d^2f/dt^2 (or d^2y/dx^2) for the second derivative. A short form is f'' or y''. (This is pronounced f *double prime* or y *double prime*). Example: The second derivative of $y = x^3$ is $y'' = 6x$.

In the distance-velocity problem, f'' is *acceleration*. It tells how fast v is changing, while v tells how fast f is changing. Where df/dt was distance/time, the second derivative is distance/(time)^2. The acceleration due to gravity is about 32 ft/sec^2 or 9.8 m/sec^2, which means that v increases by 32 ft/sec in one second. It does not mean that the distance increases by 32 feet!

The graph of $y = \sin t$ increases at the start. Its derivative $\cos t$ is positive. However the second derivative is $-\sin t$. ***The curve is bending down while going up***. The arch is "***concave down***" because $y'' = -\sin t$ is negative.

At $t = \pi$ the curve reaches zero and goes negative. The second derivative becomes positive. *Now the curve bends upward.* The lower arch is "***concave up***."

$y'' > 0$ means that y' increases so y bends upward (concave up)

$y'' < 0$ means that y' decreases so y bends down (concave down).

Chapter 3 studies these things properly—here we get an advance look for $\sin t$.

The remarkable fact about the sine and cosine is that $y'' = -y$. That is unusual and special: *acceleration* = − *distance*. The greater the distance, the greater the force pulling back:

$$y = \sin t \quad \text{has} \quad dy/dt = +\cos t \quad \text{and} \quad d^2y/dt^2 = -\sin t = -y.$$

$$y = \cos t \quad \text{has} \quad dy/dt = -\sin t \quad \text{and} \quad d^2y/dt^2 = -\cos t = -y.$$

Question Does $d^2y/dt^2 < 0$ mean that the distance $y(t)$ is decreasing?
Answer No. Absolutely not! It means that dy/dt is decreasing, not necessarily y. At the start of the sine curve, y is still increasing but $y'' < 0$.

Sines and cosines give *simple harmonic motion*—up and down, forward and back, out and in, tension and compression. Stretch a spring, and the restoring force pulls it back. Push a swing up, and gravity brings it down. These motions are controlled by a ***differential equation***:

$$\frac{d^2y}{dt^2} = -y. \tag{12}$$

All solutions are combinations of the sine and cosine: $y = A \sin t + B \cos t$.

This is not a course on differential equations. But you have to see the purpose of calculus. It models events by equations. It models oscillation by equation (12). Your heart fills and empties. Balls bounce. Current alternates. The economy goes up and down:

$$\text{high prices} \rightarrow \text{high production} \rightarrow \text{low prices} \rightarrow \cdots$$

We can't live without oscillations (or differential equations).

2.4 EXERCISES

Read-through questions

The derivative of $y = \sin x$ is $y' =$ __a__. The second derivative (the __b__ of the derivative) is $y'' =$ __c__. The fourth derivative is $y'''' =$ __d__. Thus $y = \sin x$ satisfies the differential equations $y'' =$ __e__ and $y'''' =$ __f__. So does $y = \cos x$, whose second derivative is __g__.

All these derivatives come from one basic limit: $(\sin h)/h$ approaches __h__. The sine of .01 radians is very close to __i__. So is the __j__ of .01. The cosine of .01 is not .99, because $1 - \cos h$ is much __k__ than h. The ratio $(1 - \cos h)/h^2$ approaches __l__. Therefore $\cos h$ is close to $1 - \frac{1}{2}h^2$ and $\cos .01 \approx$ __m__. We can replace h by x.

The differential equation $y'' = -y$ leads to __n__. When y is positive, y'' is __o__. Therefore y' is __p__. Eventually y goes below zero and y'' becomes __q__. Then y' is __r__. Examples of oscillation in real life are __s__ and __t__.

1 Which of these ratios approach 1 as $h \to 0$?

(a) $\dfrac{h}{\sin h}$ (b) $\dfrac{\sin^2 h}{h^2}$ (c) $\dfrac{\sin h}{\sin 2h}$ (d) $\dfrac{\sin(-h)}{h}$

2 (Calculator) Find $(\sin h)/h$ at $h = 0.5$ and 0.1 and .01. Where does $(\sin h)/h$ go above .99?

3 Find the limits as $h \to 0$ of

(a) $\dfrac{\sin^2 h}{h}$ (b) $\dfrac{\sin 5h}{5h}$ (c) $\dfrac{\sin 5h}{h}$ (d) $\dfrac{\sin h}{5h}$

4 Where does $\tan h = 1.01h$? Where does $\tan h = h$?

5 $y = \sin x$ has period 2π, which means that $\sin x =$ ________. The limit of $(\sin(2\pi + h) - \sin 2\pi)/h$ is 1 because ________. This gives dy/dx at $x =$ ________.

6 Draw $\cos(x + \Delta x)$ next to $\cos x$. Mark the height difference Δy. Then draw $\Delta y/\Delta x$ as in Figure 2.9.

7 The key to trigonometry is $\cos^2\theta = 1 - \sin^2\theta$. Set $\sin\theta \approx \theta$ to find $\cos^2\theta \approx 1 - \theta^2$. *The square root is* $\cos\theta \approx 1 - \frac{1}{2}\theta^2$. Reason: Squaring gives $\cos^2\theta \approx$ ________ and the correction term ________ is very small near $\theta = 0$.

8 (Calculator) Compare $\cos\theta$ with $1 - \frac{1}{2}\theta^2$ for

(a) $\theta = 0.1$ (b) $\theta = 0.5$ (c) $\theta = 30°$ (d) $\theta = 3°$.

9 Trigonometry gives $\cos\theta = 1 - 2\sin^2\frac{1}{2}\theta$. The approximation $\sin\frac{1}{2}\theta \approx$ ________ leads directly to $\cos\theta \approx 1 - \frac{1}{2}\theta^2$.

10 Find the limits as $h \to 0$:

(a) $\dfrac{1 - \cos h}{h^2}$ (b) $\dfrac{1 - \cos^2 h}{h^2}$

(c) $\dfrac{1 - \cos^2 h}{\sin^2 h}$ (d) $\dfrac{1 - \cos 2h}{h}$

11 Find by calculator or calculus:

(a) $\displaystyle\lim_{h\to 0} \frac{\sin 3h}{\sin 2h}$ (b) $\displaystyle\lim_{h\to 0} \frac{1 - \cos 2h}{1 - \cos h}$.

12 Compute the slope at $x = 0$ directly from limits:

(a) $y = \tan x$ (b) $y = \sin(-x)$

13 The unmarked points in Figure 2.11 are P and S. Find the height PS and the area of triangle OPR. Prove by areas that $\sin h < h$.

14 The slopes of $\cos x$ and $1 - \frac{1}{2}x^2$ are $-\sin x$ and ________. The slopes of $\sin x$ and ________ are $\cos x$ and $1 - \frac{1}{2}x^2$.

15 Chapter 10 gives an infinite series for $\sin x$:

$$\sin x = \frac{x}{1} - \frac{x^3}{3\cdot 2\cdot 1} + \frac{x^5}{5\cdot 4\cdot 3\cdot 2\cdot 1} - \cdots.$$

From the derivative find the series for $\cos x$. Then take *its* derivative to get back to $-\sin x$.

16 A *centered difference* for $f(x) = \sin x$ is

$$\frac{f(x+h) - f(x-h)}{2h} = \frac{\sin(x+h) - \sin(x-h)}{2h} = ?$$

Use the addition formula (2). Then let $h \to 0$.

17 Repeat Problem 16 to find the slope of $\cos x$. Use formula (3) to simplify $\cos(x + h) - \cos(x - h)$.

18 Find the tangent line to $y = \sin x$ at

(a) $x = 0$ (b) $x = \pi$ (c) $x = \pi/4$

19 Where does $y = \sin x + \cos x$ have zero slope?

20 Find the derivative of $\sin(x + 1)$ in two ways:

(a) Expand to $\sin x \cos 1 + \cos x \sin 1$. Compute dy/dx.

(b) Divide $\Delta y = \sin(x + 1 + \Delta x) - \sin(x + 1)$ by Δx. Write X instead of $x + 1$. Let Δx go to zero.

21 Show that $(\tan h)/h$ is squeezed between 1 and $1/\cos h$. As $h \to 0$ the limit is ________.

22 For $y = \sin 2x$, the ratio $\Delta y/h$ is

$$\frac{\sin 2(x+h) - \sin 2x}{h} = \frac{\sin 2x(\cos 2h - 1) + \cos 2x \sin 2h}{h}.$$

Explain why the limit dy/dx is $2\cos 2x$.

23 Draw the graph of $y = \sin\frac{1}{2}x$. State its slope at $x = 0$, $\pi/2$, π, and 2π. Does $\frac{1}{2}\sin x$ have the same slopes?

24 Draw the graph of $y = \sin x + \sqrt{3}\cos x$. Its maximum value is $y =$ ________ at $x =$ ________. The slope at that point is ________.

25 By combining $\sin x$ and $\cos x$, find a combination that starts at $x = 0$ from $y = 2$ with slope 1. This combination also solves $y'' =$ ________.

26 ***True or false***, with reason:

(a) The derivative of $\sin^2 x$ is $\cos^2 x$

(b) The derivative of $\cos(-x)$ is $\sin x$

(c) A positive function has a negative second derivative.

(d) If y' is increasing then y'' is positive.

27 Find solutions to $dy/dx = \sin 3x$ and $dy/dx = \cos 3x$.

28 If $y = \sin 5x$ then $y' = 5\cos 5x$ and $y'' = -25 \sin 5x$. So this function satisfies the differential equation $y'' =$ ________.

29 If h is measured in degrees, find $\lim_{h\to 0} (\sin h)/h$. You could set your calculator in degree mode.

30 Write down a ratio that approaches dy/dx at $x = \pi$. For $y = \sin x$ and $\Delta x = .01$ compute that ratio.

31 By the square rule, the derivative of $(u(x))^2$ is $2u\, du/dx$. Take the derivative of each term in $\sin^2 x + \cos^2 x = 1$.

32 Give an example of oscillation that does not come from physics. Is it simple harmonic motion (one frequency only)?

33 Explain the second derivative in your own words.

2.5 The Product and Quotient and Power Rules

What are the derivatives of $x + \sin x$ and $x \sin x$ and $1/\sin x$ and $x/\sin x$ and $\sin^n x$? Those are made up from the familiar pieces x and $\sin x$, but we need new rules. Fortunately they are rules that apply to every function, so they can be established once and for all. If we know the separate derivatives of two functions u and v, then the derivatives of $u + v$ and uv and $1/v$ and u/v and u^n are immediately available.

This is a straightforward section, with those five rules to learn. It is also an important section, containing most of the working tools of differential calculus. But I am afraid that five rules and thirteen examples (which we need—the eyes glaze over with formulas alone) make a long list. At least the easiest rule comes first. ***When we add functions, we add their derivatives.***

Sum Rule

The derivative of the sum $u(x) + v(x)$ is $\dfrac{d}{dx}(u+v) = \dfrac{du}{dx} + \dfrac{dv}{dx}$. (1)

EXAMPLE 1 The derivative of $x + \sin x$ is $1 + \cos x$. That is tremendously simple, but it is fundamental. The interpretation for distances may be more confusing (and more interesting) than the rule itself:

> Suppose a train moves with velocity 1. The distance at time t is t. On the train a professor paces back and forth (in simple harmonic motion). His distance from his seat is $\sin t$. Then the total distance from his starting point is $t + \sin t$, and his velocity (train speed plus walking speed) is $1 + \cos t$.

If you add distances, you add velocities. Actually that example is ridiculous, because the professor's maximum speed equals the train speed (= 1). He is running like mad, not pacing. Occasionally he is standing still with respect to the ground.

The sum rule is a special case of a bigger rule called "***linearity***." It applies when we add or subtract functions and multiply them by constants—as in $3x - 4\sin x$. By linearity the derivative is $3 - 4\cos x$. The rule works for all functions $u(x)$ and $v(x)$. A *linear combination* is $y(x) = au(x) + bv(x)$, where a and b are any real numbers. Then $\Delta y/\Delta x$ is

$$\frac{au(x+\Delta x) + bv(x+\Delta x) - au(x) - bv(x)}{\Delta x} = a\frac{u(x+\Delta x) - u(x)}{\Delta x} + b\frac{v(x+\Delta x) - v(x)}{\Delta x}.$$

The limit on the left is dy/dx. The limit on the right is $a\,du/dx + b\,dv/dx$. We are allowed to take limits separately and add. The result is what we hope for:

Rule of Linearity

$$\text{The derivative of } au(x)+bv(x) \text{ is } \frac{d}{dx}(au+bv) = a\frac{du}{dx} + b\frac{dv}{dx}. \qquad (2)$$

The ***product rule*** comes next. It can't be so simple—products are not linear. The sum rule is what you would have done anyway, but products give something new. ***The derivative of u times v is not du/dx times dv/dx.*** Example: The derivative of x^5 is $5x^4$. Don't multiply the derivatives of x^3 and x^2. ($3x^2$ times $2x$ is not $5x^4$.) *For a product of two functions, the derivative has two terms.*

Product Rule (the key to this section)

$$\text{The derivative of } u(x)v(x) \text{ is } \frac{d}{dx}(uv) = u\frac{dv}{dx} + v\frac{du}{dx}. \qquad (3)$$

EXAMPLE 2 $u = x^3$ times $v = x^2$ is $uv = x^5$. The product rule leads to $5x^4$:

$$x^3\frac{dv}{dx} + x^2\frac{du}{dx} = x^3(2x) + x^2(3x^2) = 2x^4 + 3x^4 = 5x^4.$$

EXAMPLE 3 In the slope of $x \sin x$, I don't write $dx/dx = 1$ but it's there:

$$\frac{d}{dx}(x \sin x) = x \cos x + \sin x.$$

EXAMPLE 4 If $u = \sin x$ and $v = \sin x$ then $uv = \sin^2 x$. We get two equal terms:

$$\sin x \frac{d}{dx}(\sin x) + \sin x \frac{d}{dx}(\sin x) = 2 \sin x \cos x.$$

This confirms the "square rule" $2u\,du/dx$, when u is the same as v. Similarly the slope of $\cos^2 x$ is $-2 \cos x \sin x$ (minus sign from the slope of the cosine).

Question Those answers for $\sin^2 x$ and $\cos^2 x$ have opposite signs, so the derivative of $\sin^2 x + \cos^2 x$ is zero (sum rule). How do you see that more quickly?

EXAMPLE 5 The derivative of uvw is $uvw' + uv'w + u'vw$—one derivative at a time. The derivative of xxx is $xx + xx + xx$.

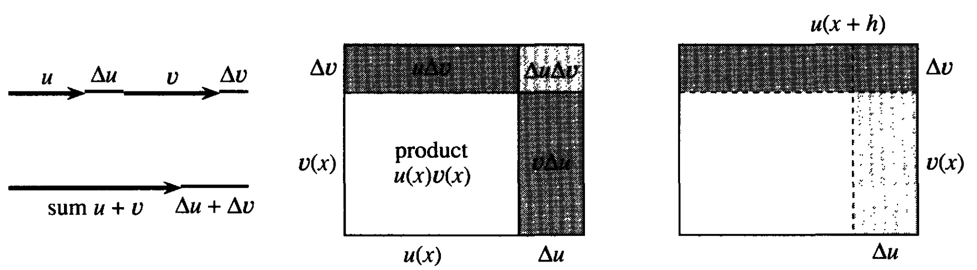


Fig. 2.13 Change in length $= \Delta u + \Delta v$. Change in area $= u\,\Delta v + v\,\Delta u + \Delta u\,\Delta v$.

After those examples we prove the product rule. Figure 2.13 explains it best. The area of the big rectangle is uv. ***The important changes in area are the two strips*** $u\,\Delta v$ ***and*** $v\,\Delta u$. The corner area $\Delta u\,\Delta v$ is much smaller. When we divide by Δx, the strips give $u\,\Delta v/\Delta x$ and $v\,\Delta u/\Delta x$. The corner gives $\Delta u\,\Delta v/\Delta x$, which approaches zero.

Notice how the sum rule is in one dimension and the product rule is in two dimensions. The rule for uvw would be in three dimensions.

The extra area comes from the whole top strip plus the side strip. By algebra,

$$u(x+h)v(x+h)-u(x)v(x)=u(x+h)[v(x+h)-v(x)]+v(x)[u(x+h)-u(x)]. \quad (4)$$

This increase is $u(x+h)\Delta v+v(x)\Delta u$—top plus side. *Now divide by* h (or Δx) *and let* $h\to 0$. The left side of equation (4) becomes the derivative of $u(x)v(x)$. The right side becomes $u(x)$ times dv/dx—*we can multiply the two limits*—plus $v(x)$ times du/dx. That proves the product rule—definitely useful.

We could go immediately to the quotient rule for $u(x)/v(x)$. But start with $u=1$. The derivative of $1/x$ is $-1/x^2$ (known). What is the derivative of $1/v(x)$?

Reciprocal Rule

$$\text{The derivative of } \frac{1}{v(x)} \text{ is } \frac{-dv/dx}{v^2}. \quad (5)$$

The proof starts with $(v)(1/v)=1$. The derivative of 1 is 0. Apply the product rule:

$$v\frac{d}{dx}\left(\frac{1}{v}\right)+\frac{1}{v}\frac{dv}{dx}=0 \quad \text{so that} \quad \frac{d}{dx}\left(\frac{1}{v}\right)=\frac{-dv/dx}{v^2}. \quad (6)$$

It is worth checking the units—in the reciprocal rule and others. A test of dimensions is automatic in science and engineering, and a good idea in mathematics. The test ignores constants and plus or minus signs, but it prevents bad errors. If v is in dollars and x is in hours, dv/dx is in *dollars per hour*. Then dimensions agree:

$$\frac{d}{dx}\left(\frac{1}{v}\right)\approx\frac{(1/\text{dollars})}{\text{hour}} \quad \text{and also} \quad \frac{-dv/dx}{v^2}\approx\frac{\text{dollars/hour}}{(\text{dollars})^2}.$$

From this test, the derivative of $1/v$ cannot be $1/(dv/dx)$. A similar test shows that Einstein's formula $e=mc^2$ is dimensionally possible. The theory of relativity might be correct! Both sides have the dimension of (mass)(distance)^2/(time)^2, when mass is converted to energy.†

EXAMPLE 6 The derivatives of x^{-1}, x^{-2}, x^{-n} are $-1x^{-2}, -2x^{-3}, -nx^{-n-1}$.

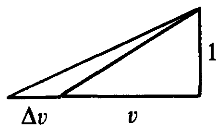
Those come from the reciprocal rule with $v=x$ and x^2 and any x^n:

$$\frac{d}{dx}(x^{-n})=\frac{d}{dx}\left(\frac{1}{x^n}\right)=-\frac{nx^{n-1}}{(x^n)^2}=-nx^{-n-1}.$$

The beautiful thing is that this answer $-nx^{-n-1}$ fits into the same pattern as x^n. ***Multiply by the exponent and reduce it by one.***

For negative and positive exponents the derivative of x^n ***is*** nx^{n-1}. (7)

†But only Einstein knew that the constant is 1.



Reciprocal $\quad \dfrac{1}{v+\Delta v} - \dfrac{1}{v} = \dfrac{-\Delta v}{v(v+\Delta v)}$

Quotient $\quad \dfrac{u+\Delta u}{v+\Delta v} - \dfrac{u}{v} = \dfrac{v\,\Delta u - u\,\Delta v}{v(v+\Delta v)}$

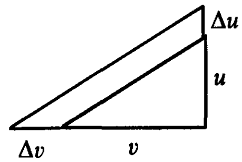


Fig. 2.14 Reciprocal rule from $(-\Delta v)/v^2$. Quotient rule from $(v\,\Delta u - u\,\Delta v)/v^2$.

EXAMPLE 7 The derivatives of $\dfrac{1}{\cos x}$ and $\dfrac{1}{\sin x}$ are $\dfrac{+\sin x}{\cos^2 x}$ and $\dfrac{-\cos x}{\sin^2 x}$.

Those come directly from the reciprocal rule. In trigonometry, $1/\cos x$ is the ***secant*** of the angle x, and $1/\sin x$ is the ***cosecant*** of x. Now we have their derivatives:

$$\frac{d}{dx}(\sec x) = \frac{\sin x}{\cos^2 x} = \frac{1}{\cos x}\frac{\sin x}{\cos x} = \sec x \tan x. \tag{8}$$

$$\frac{d}{dx}(\csc x) = -\frac{\cos x}{\sin^2 x} = -\frac{1}{\sin x}\frac{\cos x}{\sin x} = -\csc x \cot x. \tag{9}$$

Those formulas are often seen in calculus. If you have a good memory they are worth storing. Like most mathematicians, I have to check them every time before using them (maybe once a year). It is really the rules that are basic, not the formulas.

The next rule applies to the quotient $u(x)/v(x)$. That is u times $1/v$. Combining the product rule and reciprocal rule gives something new and important:

Quotient Rule

The derivative of $\dfrac{u(x)}{v(x)}$ is $\dfrac{1}{v}\dfrac{du}{dx} - u\dfrac{dv/dx}{v^2} = \dfrac{v\,du/dx - u\,dv/dx}{v^2}$.

You *must* memorize that last formula. The v^2 is familiar. The rest is new, but not very new. If $v = 1$ the result is du/dx (of course). For $u = 1$ we have the reciprocal rule. Figure 2.14b shows the difference $(u+\Delta u)/(v+\Delta v) - (u/v)$. The denominator $v(v+\Delta v)$ is responsible for v^2.

EXAMPLE 8 (only practice) If $u/v = x^5/x^3$ (which is x^2) the quotient rule gives $2x$:

$$\frac{d}{dx}\left(\frac{x^5}{x^3}\right) = \frac{x^3(5x^4) - x^5(3x^2)}{x^6} = \frac{5x^7 - 3x^7}{x^6} = 2x.$$

EXAMPLE 9 (important) For $u = \sin x$ and $v = \cos x$, the quotient is $\sin x/\cos x = \tan x$. ***The derivative of*** $\tan x$ ***is*** $\sec^2 x$. Use the quotient rule and $\cos^2 x + \sin^2 x = 1$:

$$\frac{d}{dx}\left(\frac{\sin x}{\cos x}\right) = \frac{\cos x(\cos x) - \sin x(-\sin x)}{\cos^2 x} = \frac{1}{\cos^2 x} = \sec^2 x. \tag{11}$$

Again to memorize: $(\tan x)' = \sec^2 x$. At $x = 0$, this slope is 1. The graphs of $\sin x$ and x and $\tan x$ all start with this slope (then they separate). At $x = \pi/2$ the sine curve is flat ($\cos x = 0$) and the tangent curve is vertical ($\sec^2 x = \infty$).

The slope generally blows up faster than the function. We divide by $\cos x$, once for the tangent and twice for its slope. The slope of $1/x$ is $-1/x^2$. The slope is more sensitive than the function, because of the square in the denominator.

EXAMPLE 10 $\qquad \dfrac{d}{dx}\left(\dfrac{\sin x}{x}\right) = \dfrac{x\cos x - \sin x}{x^2}$.

That one I hesitate to touch at $x = 0$. Formally it becomes $0/0$. In reality it is more like $0^3/0^2$, and the true derivative is zero. Figure 2.10 showed graphically that $(\sin x)/x$ is flat at the center point. The function is *even* (symmetric across the y axis) so its derivative can only be zero.

This section is full of rules, and I hope you will allow one more. It goes beyond x^n to $(u(x))^n$. A power of x changes to a power of $u(x)$—as in $(\sin x)^6$ or $(\tan x)^7$ or $(x^2+1)^8$. The derivative contains nu^{n-1} (copying nx^{n-1}), but ***there is an extra factor*** du/dx. Watch that factor in $6(\sin x)^5 \cos x$ and $7(\tan x)^6 \sec^2 x$ and $8(x^2+1)^7(2x)$:

Power Rule

$$\text{The derivative of } \left[u(x)\right]^n \text{ is } n\left[u(x)\right]^{n-1}\frac{du}{dx}. \qquad (12)$$

For $n = 1$ this reduces to $du/dx = du/dx$. For $n = 2$ we get the square rule $2u\, du/dx$. Next comes u^3. The best approach is to use ***mathematical induction***, which goes from each n to the next power $n+1$ by the product rule:

$$\frac{d}{dx}(u^{n+1}) = \frac{d}{dx}(u^n u) = u^n\frac{du}{dx} + u\left(nu^{n-1}\frac{du}{dx}\right) = (n+1)u^n\frac{du}{dx}.$$

That is exactly equation (12) for the power $n+1$. We get all positive powers this way, going up from $n = 1$—then the negative powers come from the reciprocal rule.

Figure 2.15 shows the power rule for $n = 1, 2, 3$. The cube makes the point best. The three thin slabs are u by u by Δu. ***The change in volume is essentially*** $3u^2\Delta u$. From multiplying out $(u+\Delta u)^3$, the exact change in volume is $3u^2\Delta u + 3u(\Delta u)^2 + (\Delta u)^3$—which also accounts for three narrow boxes and a midget cube in the corner. This is the binomial formula in a picture.

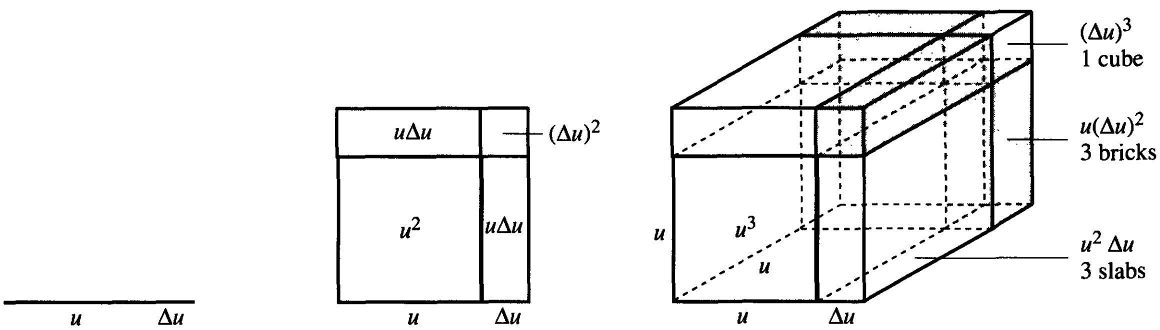


Fig. 2.15 Length change $= \Delta u$; area change $\approx 2u\,\Delta u$; volume change $\approx 3u^2\,\Delta u$.

EXAMPLE 11 $\dfrac{d}{dx}(\sin x)^n = n(\sin x)^{n-1}\cos x$. The extra factor $\cos x$ is du/dx.

Our last step finally escapes from a very undesirable restriction—that n must be a whole number. We want to allow fractional powers $n = p/q$, and keep the same formula. *The derivative of* x^n *is still* nx^{n-1}.

To deal with square roots I can write $(\sqrt{x})^2 = x$. Its derivative is $2\sqrt{x}(\sqrt{x})' = 1$. Therefore $(\sqrt{x})'$ is $1/2\sqrt{x}$, which fits the formula when $n = \frac{1}{2}$. Now try $n = p/q$:

Fractional powers Write $u = x^{p/q}$ as $u^q = x^p$. Take derivatives, assuming they exist:

$$qu^{q-1}\frac{du}{dx} = px^{p-1} \qquad \text{(power rule on both sides)}$$

$$\frac{du}{dx} = \frac{px^{-1}}{qu^{-1}} \qquad \text{(cancel } x^p \text{ with } u^q\text{)}$$

$$\frac{du}{dx} = nx^{n-1} \qquad \text{(replace } p/q \text{ by } n \text{ and } u \text{ by } x^n\text{)}$$

EXAMPLE 12 The slope of $x^{1/3}$ is $\frac{1}{3}x^{-2/3}$. The slope is infinite at $x = 0$ and zero at $x = \infty$. But the curve in Figure 2.16 keeps climbing. It doesn't stay below an "asymptote."

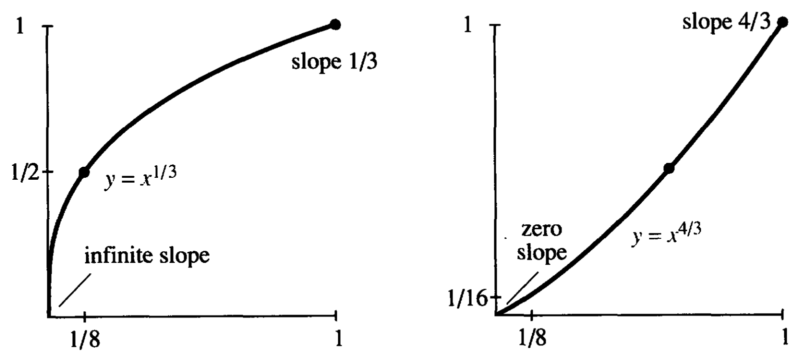


Fig. 2.16 Infinite slope of x^n versus zero slope: the difference between $0 < n < 1$ and $n > 1$.

EXAMPLE 13 The slope of $x^{4/3}$ is $\frac{4}{3}x^{1/3}$. The slope is zero at $x = 0$ and infinite at $x = \infty$. The graph climbs faster than a line and slower than a parabola ($\frac{4}{3}$ is between 1 and 2). Its slope follows the cube root curve (times $\frac{4}{3}$).

WE STOP NOW! I am sorry there were so many rules. A computer can memorize them all, but it doesn't know what they mean and you do. Together with the chain rule that dominates Chapter 4, they achieve virtually all the derivatives ever computed by mankind. We list them in one place for convenience.

Rule of Linearity	$(au + bv)' = au' + bv'$
Product Rule	$(uv)' = uv' + vu'$
Reciprocal Rule	$(1/v)' = -v'/v^2$
Quotient Rule	$(u/v)' = (vu' - uv')/v^2$
Power Rule	$(u^n)' = nu^{n-1}u'$

The power rule applies when n is *negative*, or a *fraction*, or ***any real number***. The derivative of x^π is $\pi x^{\pi-1}$, according to Chapter 6. The derivative of $(\sin x)^\pi$ is _______.
And the derivatives of all six trigonometric functions are now established:

$$(\sin x)' = \cos x \qquad (\tan x)' = \sec^2 x \qquad (\sec x)' = \sec x \tan x$$

$$(\cos x)' = -\sin x \qquad (\cot x)' = -\csc^2 x \qquad (\csc x)' = -\csc x \cot x.$$

2.5 EXERCISES

Read-through questions

The derivatives of $\sin x \cos x$ and $1/\cos x$ and $\sin x/\cos x$ and $\tan^3 x$ come from the __a__ rule, __b__ rule, __c__ rule, and __d__ rule. The product of $\sin x$ times $\cos x$ has $(uv)' = uv' +$ __e__ $=$ __f__. The derivative of $1/v$ is __g__, so the slope of $\sec x$ is __h__. The derivative of u/v is __i__, so the slope of $\tan x$ is __j__. The derivative of $\tan^3 x$ is __k__. The slope of x^n is __l__ and the slope of $(u(x))^n$ is __m__. With $n = -1$ the derivative of $(\cos x)^{-1}$ is __n__, which agrees with the rule for $\sec x$.

Even simpler is the rule of __o__, which applies to $au(x) + bv(x)$. The derivative is __p__. The slope of $3 \sin x + 4 \cos x$ is __q__. The derivative of $(3 \sin x + 4 \cos x)^2$ is __r__. The derivative of __s__ is $4 \sin^3 x \cos x$.

Find the derivatives of the functions in 1–26.

1 $(x+1)(x-1)$

2 $(x^2+1)(x^2-1)$

3 $\dfrac{1}{1+x} + \dfrac{1}{1+\sin x}$

4 $\dfrac{1}{1+x^2} + \dfrac{1}{1-\sin x}$

5 $(x-1)(x-2)(x-3)$

6 $(x-1)^2(x-2)^2$

7 $x^2 \cos x + 2x \sin x$

8 $x^{1/2}(x + \sin x)$

9 $\dfrac{x^3+1}{x+1} + \dfrac{\cos x}{\sin x}$

10 $\dfrac{x^2+1}{x^2-1} + \dfrac{\sin x}{\cos x}$

11 $x^{1/2} \sin^2 x + (\sin x)^{1/2}$

12 $x^{3/2} \sin^3 x + (\sin x)^{3/2}$

13 $x^4 \cos x + x \cos^4 x$

14 $\sqrt{x}(\sqrt{x}+1)(\sqrt{x}+2)$

15 $\frac{1}{2}x^2 \sin x - x \cos x + \sin x$

16 $(x-6)^{10} + \sin^{10} x$

17 $\sec^2 x - \tan^2 x$

18 $\csc^2 x - \cot^2 x$

19 $\dfrac{4}{(x-5)^{2/3}} + \dfrac{4}{(5-x)^{2/3}}$

20 $\dfrac{\sin x - \cos x}{\sin x + \cos x}$

21 $(\sin x \cos x)^3 + \sin 2x$

22 $x \cos x \csc x$

23 $u(x)v(x)w(x)z(x)$

24 $[u(x)]^2[v(x)]^2$

25 $\dfrac{1}{\tan x} - \dfrac{1}{\cot x}$

26 $x \sin x + \cos x$

27 A growing box has length t, width $1/(1+t)$, and height $\cos t$.

(a) What is the rate of change of the volume?

(b) What is the rate of change of the surface area?

28 With two applications of the product rule show that the derivative of uvw is $uvw' + uv'w + u'vw$. When a box with sides u, v, w grows by $\Delta u, \Delta v, \Delta w$, three slabs are added with volume $uv\,\Delta w$ and ________ and ________.

29 Find the velocity if the distance is $f(t) =$

$$5t^2 \text{ for } t \leqslant 10, \qquad 500 + 100\sqrt{t-10} \text{ for } t \geqslant 10.$$

30 A cylinder has radius $r = \dfrac{t^{3/2}}{1+t^{3/2}}$ and height $h = \dfrac{t}{1+t}$.

(a) What is the rate of change of its volume?

(b) What is the rate of change of its surface area (including top and base)?

31 The height of a model rocket is $f(t) = t^3/(1+t)$.

(a) What is the velocity $v(t)$?

(b) What is the acceleration dv/dt?

32 Apply the product rule to $u(x)u^2(x)$ to find the power rule for $u^3(x)$.

33 Find the *second* derivative of the product $u(x)v(x)$. Find the *third* derivative. Test your formulas on $u = v = x$.

34 Find functions $y(x)$ whose derivatives are

(a) x^3 (b) $1/x^3$ (c) $(1-x)^{3/2}$ (d) $\cos^2 x \sin x$.

35 Find the distances $f(t)$, starting from $f(0) = 0$, to match these velocities:

(a) $v(t) = \cos t \sin t$ (b) $v(t) = \tan t \sec^2 t$

(c) $v(t) = \sqrt{1+t}$

36 Apply the quotient rule to $(u(x))^3/(u(x))^2$ and $-v'/v^2$. The latter gives the second derivative of ________.

37 Draw a figure like 2.13 to explain the *square rule*.

38 Give an example where $u(x)/v(x)$ is increasing but $du/dx = dv/dx = 1$.

39 ***True or false***, with a good reason:

(a) The derivative of x^{2n} is $2nx^{2n-1}$.

(b) By linearity the derivative of $a(x)u(x) + b(x)v(x)$ is $a(x)\,du/dx + b(x)\,dv/dx$.

(c) The derivative of $|x|^3$ is $3|x|^2$.

(d) $\tan^2 x$ and $\sec^2 x$ have the same derivative.

(e) $(uv)' = u'v'$ is true when $u(x) = 1$.

40 The cost of u shares of stock at v dollars per share is uv dollars. Check dimensions of $d(uv)/dt$ and $u\,dv/dt$ and $v\,du/dt$.

41 If $u(x)/v(x)$ is a ratio of polynomials of degree n, what are the degrees for its derivative?

42 For $y = 5x + 3$, is $(dy/dx)^2$ the same as d^2y/dx^2?

43 If you change from $f(t) = t \cos t$ to its tangent line at $t = \pi/2$, find the two-part function df/dt.

44 Explain in your own words why the derivative of $u(x)v(x)$ has two terms.

45 A plane starts its descent from height $y = h$ at $x = -L$ to land at $(0, 0)$. Choose a, b, c, d so its landing path $y = ax^3 + bx^2 + cx + d$ is **smooth**. With $dx/dt = V =$ constant, find dy/dt and d^2y/dt^2 at $x = 0$ and $x = -L$. (To keep d^2y/dt^2 small, a coast-to-coast plane starts down $L > 100$ miles from the airport.)

2.6 Limits

You have seen enough limits to be ready for a definition. It is true that we have survived this far without one, and we could continue. But this seems a reasonable time to define limits more carefully. The goal is to achieve rigor without rigor mortis.

First you should know that limits of $\Delta y/\Delta x$ are by no means the only limits in mathematics. Here are five completely different examples. They involve $n \to \infty$, not $\Delta x \to 0$:

1. $a_n = (n-3)/(n+3)$ (for large n, ignore the 3's and find $a_n \to 1$)
2. $a_n = \frac{1}{2}a_{n-1} + 4$ (start with any a_1 and always $a_n \to 8$)
3. $a_n =$ probability of living to year n (unfortunately $a_n \to 0$)
4. $a_n =$ fraction of zeros among the first n digits of π ($a_n \to \frac{1}{10}$?)
5. $a_1 = .4$, $a_2 = .49$, $a_3 = .493$, ... *No matter what the remaining decimals are, the a's converge to a limit.* Possibly $a_n \to .493000 \ldots$, but not likely.

The problem is to say what the limit symbol $\to$ really means.

A good starting point is to ask about convergence to *zero*. When does a sequence of positive numbers approach zero? What does it mean to write $a_n \to 0$? The numbers $a_1, a_2, a_3, \ldots$, must become "small," but that is too vague. We will propose four definitions of ***convergence to zero***, and I hope the right one will be clear.

1. *All the numbers a_n are below* 10^{-10}. That may be enough for practical purposes, but it certainly doesn't make the a_n approach zero.

2. *The sequence is getting closer to zero*—each a_{n+1} is smaller than the preceding a_n. This test is met by 1.1, 1.01, 1.001, ... which converges to 1 instead of 0.

3. *For any small number you think of, at least one of the a_n's is smaller.* That pushes something toward zero, but not necessarily the whole sequence. The condition would be satisfied by $1, \frac{1}{2}, 1, \frac{1}{3}, 1, \frac{1}{4}, \ldots$, which does not approach zero.

4. *For any small number you think of, the a_n's eventually go below that number and* ***stay below***. This is the correct definition.

I want to repeat that. To test for convergence to zero, start with a small number—say 10^{-10}. The a_n's must go *below that number*. They may come back up and go below again—the first million terms make absolutely no difference. Neither do the next billion, but eventually all terms must go below 10^{-10}. After waiting longer (possibly a lot longer), all terms drop below 10^{-20}. The tail end of the sequence decides everything.

Question 1 Does the sequence $10^{-3}, 10^{-2}, 10^{-6}, 10^{-5}, 10^{-9}, 10^{-8}, \ldots$ approach 0?
Answer Yes. These up and down numbers eventually stay below any ε.

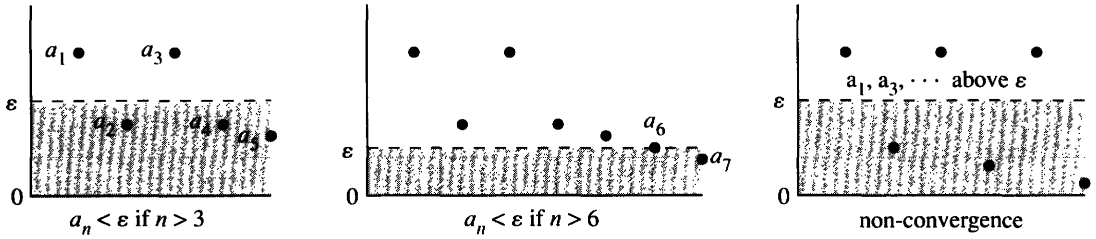


Fig. 2.17 Convergence means: Only a finite number of a's are outside any strip around L.

Question 2 Does $10^{-4}, 10^{-6}, 10^{-4}, 10^{-8}, 10^{-4}, 10^{-10}, \ldots$ approach zero?
Answer No. This sequence goes below 10^{-4} but does not stay below.

There is a recognized symbol for "an arbitrarily small positive number." By worldwide agreement, it is the Greek letter ε (***epsilon***). Convergence to zero means that ***the sequence eventually goes below*** ε ***and stays there***. The smaller the ε, the tougher the test and the longer we wait. Think of ε as the tolerance, and keep reducing it.

To emphasize that ε comes from outside, Socrates can choose it. Whatever ε he proposes, the a's must eventually be smaller. *After some a_N, all the a's are below the tolerance* ε. Here is the exact statement:

for any ε there is an N such that $a_n < \varepsilon$ if $n > N$.

Once you see that idea, the rest is easy. Figure 2.17 has $N = 3$ and then $N = 6$.

EXAMPLE 1 The sequence $\frac{1}{2}, \frac{4}{4}, \frac{9}{8}, \ldots$ starts upward but goes to zero. Notice that $1, 4, 9, \ldots, 100, \ldots$ are squares, and $2, 4, 8, \ldots, 1024, \ldots$ are powers of 2. Eventually 2^n grows faster than n^2, as in $a_{10} = 100/1024$. The ratio goes below any ε.

EXAMPLE 2 $1, 0, \frac{1}{2}, 0, \frac{1}{3}, 0, \ldots$ approaches zero. These a's do not decrease steadily (the mathematical word for steadily is "monotonically") but still their limit is zero. The choice $\varepsilon = 1/1000$ produces the right response: *Beyond a_{2001} all terms are below* $1/1000$. So $N = 2001$ for that ε.

The sequence $1, \frac{1}{2}, \frac{1}{2}, \frac{1}{3}, \frac{1}{3}, \frac{1}{3}, \ldots$ is much slower—but it also converges to zero.

Next we allow the numbers a_n to be *negative* as well as positive. They can converge upward toward zero, or they can come in from both sides. The test still requires the a_n to go inside any strip near zero (and stay there). But now the strip starts at $-\varepsilon$.

The distance from zero is the absolute value $|a_n|$. Therefore $a_n \to 0$ means $|a_n| \to 0$. The previous test can be applied to $|a_n|$:

for any ε there is an N such that $|a_n| < \varepsilon$ if $n > N$.

EXAMPLE 3 $1, -\frac{1}{2}, \frac{1}{3}, -\frac{1}{4}, \ldots$ converges to zero because $1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \ldots$ converges to zero.

It is a short step to limits other than zero. ***The limit is L if the numbers $a_n - L$ converge to zero.*** Our final test applies to the absolute value $|a_n - L|$:

for any ε there is an N such that $|a_n - L| < \varepsilon$ if $n > N$.

This is the definition of convergence! Only a finite number of a's are outside any strip around L (Figure 2.18). We write $a_n \to L$ or $\lim a_n = L$ or $\lim_{n \to \infty} a_n = L$.

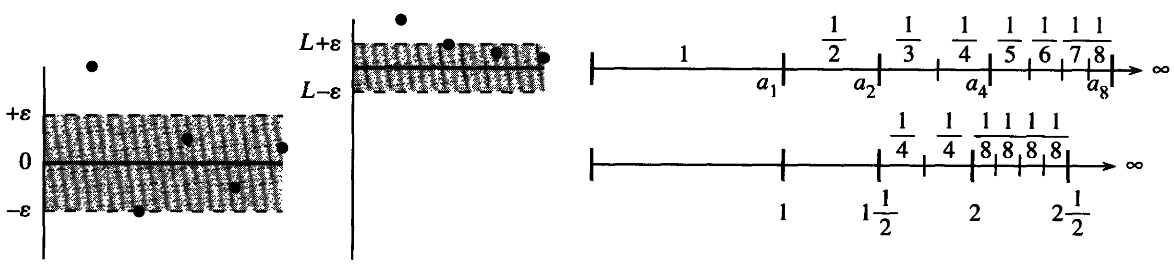


Fig. 2.18 $a_n \to 0$ in Example 3; $a_n \to 1$ in Example 4; $a_n \to \infty$ in Example 5 (but $a_{n+1} - a_n \to 0$).

EXAMPLE 4 The numbers $\frac{3}{2}, \frac{5}{4}, \frac{7}{6}, \ldots$ converge to $L = 1$. After subtracting 1 the differences $\frac{1}{2}, \frac{1}{4}, \frac{1}{6}, \ldots$ converge to zero. Those difference are $|a_n - L|$.

EXAMPLE 5 ***The sequence*** $1, 1+\frac{1}{2}, 1+\frac{1}{2}+\frac{1}{3}, 1+\frac{1}{2}+\frac{1}{3}+\frac{1}{4}, \ldots$ ***fails to converge.***

The distance between terms is getting smaller. But those numbers $a_1, a_2, a_3, a_4, \ldots$ go past any proposed limit L. The second term is $1\frac{1}{2}$. The fourth term adds on $\frac{1}{3}+\frac{1}{4}$, so a_4 goes past 2. The eighth term has four new fractions $\frac{1}{5}+\frac{1}{6}+\frac{1}{7}+\frac{1}{8}$, totaling more than $\frac{1}{8}+\frac{1}{8}+\frac{1}{8}+\frac{1}{8}=\frac{1}{2}$. Therefore a_8 exceeds $2\frac{1}{2}$. Eight more terms will add more than 8 times $\frac{1}{16}$, so a_{16} is beyond 3. The lines in Figure 2.18c are infinitely long, not stopping at any L.

In the language of Chapter 10, the harmonic series $1+\frac{1}{2}+\frac{1}{3}+\cdots$ *does not converge*. The sum is infinite, because the "partial sums" a_n go beyond every limit L (a_{5000} is past $L = 9$). We will come back to infinite series, but this example makes a subtle point: The steps between the a_n can go to zero while still $a_n \to \infty$.

Thus the condition $a_{n+1} - a_n \to 0$ is ***not sufficient*** for convergence. However this condition is ***necessary***. If we do have convergence, then $a_{n+1} - a_n \to 0$. That is a good exercise in the logic of convergence, emphasizing the difference between "sufficient" and "necessary." We discuss this logic below, after proving that [statement A] implies [statement B]:

$$\textbf{\textit{If}}\ [a_n \text{ converges to } L]\ \textbf{\textit{then}}\ [a_{n+1} - a_n \text{ converges to zero}]. \tag{1}$$

Proof Because the a_n converge, there is a number N beyond which $|a_n - L| < \varepsilon$ and also $|a_{n+1} - L| < \varepsilon$. Since $a_{n+1} - a_n$ is the sum of $a_{n+1} - L$ and $L - a_n$, its absolute value cannot exceed $\varepsilon + \varepsilon = 2\varepsilon$. Therefore $a_{n+1} - a_n$ approaches zero.

Objection by Socrates: We only got below 2ε and he asked for ε. *Our reply*: If he particularly wants $|a_{n+1} - a_n| < 1/10$, we start with $\varepsilon = 1/20$. Then $2\varepsilon = 1/10$. But this juggling is not necessary. To stay below 2ε is just as convincing as to stay below ε.

THE LOGIC OF "IF" AND "ONLY IF"

The following page is inserted to help with the language of mathematics. In ordinary language we might say "I will come if you call." Or we might say "I will come only if you call." That is different! A mathematician might even say "I will come *if and only if* you call." Our goal is to think through the logic, because it is important and not so familiar.†

Statement A above implies statement B. Statement A is $a_n \to L$; statement B is $a_{n+1} - a_n \to 0$. Mathematics has at least five ways of writing down $A \Rightarrow B$, and I though you might like to see them together. It seems excessive to have so many expressions for the same idea, but authors get desperate for a little variety. Here are the five ways that come to mind:

$A \Rightarrow B$

A implies B

if A ***then*** B

A is a ***sufficient*** condition for B

B is true ***if*** A is true

†Logical thinking is much more important than ε and δ.

EXAMPLES ***If*** [positive numbers are decreasing] ***then*** [they converge to a limit].
If [sequences a_n and b_n converge] ***then*** [the sequence $a_n + b_n$ converges].
If [$f(x)$ is the integral of $v(x)$] ***then*** [$v(x)$ is the derivative of $f(x)$].

Those are all true, but not proved. A is the hypothesis, B is the conclusion.

Now we go in the other direction. (It is called the "converse," not the inverse.) *We exchange A and B.* Of course stating the converse does not make it true! B might imply A, or it might not. In the first two examples the converse was false—the a_n can converge without decreasing, and $a_n + b_n$ can converge when the separate sequences do not. The converse of the third statement is true—and there are five more ways to state it:

$$A \Leftarrow B$$

A is implied by B

if B ***then*** A

A is a ***necessary*** condition for B

B is true ***only if*** A is true

Those words "necessary" and "sufficient" are not always easy to master. The same is true of the deceptively short phrase "if and only if." The two statements $A \Rightarrow B$ and $A \Leftarrow B$ are completely different and *they both require proof*. That means two separate proofs. But they can be stated together for convenience (when both are true):

$$A \Leftrightarrow B$$

A implies B and B implies A

A is ***equivalent*** to B

A is a ***necessary and sufficient*** condition for B

A is true ***if and only if*** B is true

EXAMPLES $[a_n \to L] \Leftrightarrow [2a_n \to 2L] \Leftrightarrow [a_n + 1 \to L + 1] \Leftrightarrow [a_n - L \to 0]$.

RULES FOR LIMITS

Calculus needs a *definition of limits*, to define dy/dx. That derivative contains two limits: $\Delta x \to 0$ and $\Delta y/\Delta x \to dy/dx$. Calculus also needs *rules for limits*, to prove the sum rule and product rule for derivatives. We started on the definition, and now we start on the rules.

Given *two convergent sequences*, $a_n \to L$ and $b_n \to M$, other sequences also converge:

Addition: $a_n + b_n \to L + M$ Subtraction: $a_n - b_n \to L - M$

Multiplication: $a_n b_n \to LM$ Division: $a_n/b_n \to L/M$ (provided $M \neq 0$)

We check the multiplication rule, which uses a convenient identity:

$$a_n b_n - LM = (a_n - L)(b_n - M) + M(a_n - L) + L(b_n - M). \tag{2}$$

Suppose $|a_n - L| < \varepsilon$ beyond some point N, and $|b_n - M| < \varepsilon$ beyond some other point N'. Then beyond the larger of N and N', the right side of (2) is small. It is less than $\varepsilon \cdot \varepsilon + M\varepsilon + L\varepsilon$. This proves that (2) gives $a_n b_n \to LM$.

An important special case is $ca_n \to cL$. (The sequence of b's is $c, c, c, c, \ldots$.) Thus a constant can be brought "outside" the limit, to give $\lim ca_n = c \lim a_n$.

THE LIMIT OF $f(x)$ AS $x \to a$

The final step is to replace sequences by functions. Instead of $a_1, a_2, \ldots$ there is a continuum of values $f(x)$. The limit is taken as x approaches a specified point a (instead of $n \to \infty$). Example: As x approaches $a = 0$, the function $f(x) = 4 - x^2$ approaches $L = 4$. As x approaches $a = 2$, the function $5x$ approaches $L = 10$. Those statements are fairly obvious, but we have to say what they mean. Somehow it must be this:

***if** x **is close to** a **then** $f(x)$ **is close to** L.*

If $x - a$ is small, then $f(x) - L$ should be small. As before, the word *small* does not say everything. We really mean "arbitrarily small," or "below any ε." The difference $f(x) - L$ must become *as small as anyone wants*, when x gets near a. In that case $\lim_{x \to a} f(x) = L$. Or we write $f(x) \to L$ as $x \to a$.

The statement is awkward because it involves *two limits*. The limit $x \to a$ is forcing $f(x) \to L$. (Previously $n \to \infty$ forced $a_n \to L$.) But it is wrong to expect the same ε in both limits. We do not and cannot require that $|x - a| < \varepsilon$ produces $|f(x) - L| < \varepsilon$. ***It may be necessary to push** x **extremely close to** a* (closer than ε). We must guarantee that if x is close enough to a, then $|f(x) - L| < \varepsilon$.

We have come to the "***epsilon-delta definition***" of limits. First, Socrates chooses ε. He has to be shown that $f(x)$ is within ε of L, for every x near a. Then somebody else (maybe Plato) replies with a number δ. That gives the meaning of "near a." Plato's goal is to get $f(x)$ within ε of L, by keeping x within δ of a:

$$\textit{\textbf{if}}\quad 0 < |x - a| < \delta \quad \textit{\textbf{then}} \quad |f(x) - L| < \varepsilon. \tag{3}$$

The input tolerance is δ (delta), the output tolerance is ε. When Plato can find a δ for every ε, Socrates concedes that the limit is L.

EXAMPLE Prove that $\lim_{x \to 2} 5x = 10$. In this case $a = 2$ and $L = 10$.

Socrates asks for $|5x - 10| < \varepsilon$. Plato responds by requiring $|x - 2| < \delta$. What δ should he choose? In this case $|5x - 10|$ is exactly 5 times $|x - 2|$. So Plato picks δ below $\varepsilon/5$ (a smaller δ is always OK). Whenever $|x - 2| < \varepsilon/5$, multiplication by 5 shows that $|5x - 10| < \varepsilon$.

Remark* 1** In Figure 2.19, Socrates chooses the height of the box. It extends above and below L, by the small number ε. Second, Plato chooses the width. He must make the box narrow enough for the graph to go ***out the sides. Then $|f(x) - L| < \varepsilon$.

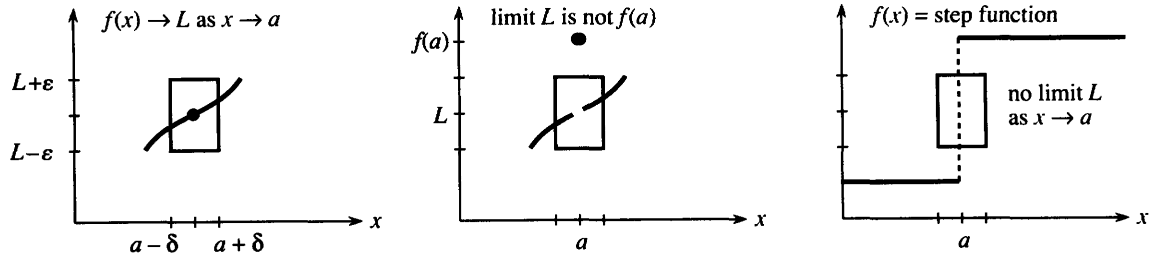


Fig. 2.19 S chooses height 2ε, then P chooses width 2δ. Graph must go out the sides.

When $f(x)$ has a jump, the box can't hold it. A step function has no limit as x approaches the jump, because the graph goes through the top or bottom—no matter how thin the box.

Remark* 2** The second figure has $f(x) \to L$, because in taking limits ***we ignore the final point $x = a$. The value $f(a)$ can be anything, with no effect on L. The first figure has more: $f(a)$ equals L. Then a special name applies—f is ***continuous***. The left figure shows a continuous function, the other figures do not.

We soon come back to continuous functions.

***Remark* 3** In the example with $f = 5x$ and $\delta = \varepsilon/5$, the number 5 was the *slope*. That choice barely kept the graph in the box—it goes out the corners. A little narrower, say $\delta = \varepsilon/10$, and the graph goes safely out the sides. *A reasonable choice is to divide* ε *by* $2|f'(a)|$. (We double the slope for safety.) I want to say why this δ works—even if the ε–δ test is seldom used in practice.

The ratio of $f(x) - L$ to $x - a$ is distance up over distance across. This is $\Delta f/\Delta x$, close to the slope $f'(a)$. When the distance across is δ, the distance up or down is near $\delta|f'(a)|$. That equals $\varepsilon/2$ for our "reasonable choice" of δ—so we are safely below ε. This choice solves most exercises. But Example 7 shows that a limit might exist even when the slope is infinite.

EXAMPLE 7 $\lim_{x \to 1^+} \sqrt{x - 1} = 0$ (***a one-sided limit***).

Notice the plus sign in the symbol $x \to 1^+$. The number x approaches $a = 1$ *only from above*. An ordinary limit $x \to 1$ requires us to accept x on both sides of 1 (the exact value $x = 1$ is not considered). Since negative numbers are not allowed by the square root, we have a *one-sided limit*. It is $L = 0$.

Suppose ε is 1/10. Then the response could be $\delta = 1/100$. A number below 1/100 has a square root below 1/10. In this case the box must be made extremely narrow, δ much smaller than ε, because the square root starts with infinite slope.

Those examples show the point of the ε–δ definition. (Given ε, look for δ. This came from Cauchy in France, not Socrates in Greece.) We also see its bad feature: The test is not convenient. Mathematicians do not go around proposing ε's and replying with δ's. We may live a strange life, but not that strange.

It is easier to establish once and for all that $5x$ approaches its obvious limit $5a$. The same is true for other familiar functions: $x^n \to a^n$ and $\sin x \to \sin a$ and $(1 - x)^{-1} \to (1 - a)^{-1}$—except at $a = 1$. ***The correct limit L comes by substituting*** $x = a$ ***into the function***. This is exactly the property of a "***continuous function***." Before the section on continuous functions, we prove the Squeeze Theorem using ε and δ.

2H ***Squeeze Theorem*** Suppose $f(x) \le g(x) \le h(x)$ for x near a. If $f(x) \to L$ and $h(x) \to L$ as $x \to a$, then the limit of $g(x)$ is also L.

Proof $g(x)$ is squeezed between $f(x)$ and $h(x)$. After subtracting L, $g(x) - L$ is between $f(x) - L$ and $h(x) - L$. Therefore

$$|g(x) - L| < \varepsilon \quad \text{if} \quad |f(x) - L| < \varepsilon \quad \text{and} \quad |h(x) - L| < \varepsilon.$$

For any ε, the last two inequalities hold in some region $0 < |x - a| < \delta$. So the first one also holds. This proves that $g(x) \to L$. Values at $x = a$ are not involved—until we get to continuous functions.

2.6 EXERCISES

Read-through questions

The limit of $a_n = (\sin n)/n$ is __a__. The limit of $a_n = n^4/2^n$ is __b__. The limit of $a_n = (-1)^n$ is __c__. The meaning of $a_n \to 0$ is: Only __d__ of the numbers $|a_n|$ can be __e__. The meaning of $a_n \to L$ is: For every __f__ there is an __g__ such that __h__ if $n >$ __i__. The sequence $1, 1+\frac{1}{2}, 1+\frac{1}{2}+\frac{1}{3}, \ldots$ is not __j__ because eventually those sums go past __k__.

The limit of $f(x) = \sin x$ as $x \to a$ is __l__. The limit of $f(x) = x/|x|$ as $x \to -2$ is __m__, but the limit as $x \to 0$ does not __n__. This function only has __o__-sided limits. The meaning of $\lim_{x \to a} f(x) = L$ is: For every ε there is a δ such that $|f(x) - L| < \varepsilon$ whenever __p__.

Two rules for limits, when $a_n \to L$ and $b_n \to M$, are $a_n + b_n \to$ __q__ and $a_n b_n \to$ __r__. The corresponding rules for functions, when $f(x) \to L$ and $g(x) \to M$ as $x \to a$, are __s__ and __t__. In all limits, $|a_n - L|$ or $|f(x) - L|$ must eventually go below and __u__ any positive __v__.

$A \Rightarrow B$ means that A is a __w__ condition for B. Then B is true __x__ A is true. $A \Leftrightarrow B$ means that A is a __y__ condition for B. Then B is true __z__ A is true.

1 What is a_4 and what is the limit L? After which N is $|a_n - L| < \frac{1}{10}$? (Calculator allowed)

(a) $-1, +\frac{1}{2}, -\frac{1}{3}, \ldots$
(b) $\frac{1}{2}, \frac{1}{2}+\frac{1}{4}, \frac{1}{2}+\frac{1}{4}+\frac{1}{6}, \ldots$
(c) $\frac{1}{2}, \frac{2}{4}, \frac{3}{8}, \ldots a_n = n/2^n$
(d) $1.1, 1.11, 1.111, \ldots$
(e) $a_n = \sqrt[n]{n}$
(f) $a_n = \sqrt{n^2+n} - n$
(g) $1+1, (1+\frac{1}{2})^2, (1+\frac{1}{3})^3, \ldots$

2 Show by example that these statements are false:

(a) If $a_n \to L$ and $b_n \to L$ then $a_n/b_n \to 1$
(b) $a_n \to L$ if and only if $a_n^2 \to L^2$
(c) If $a_n < 0$ and $a_n \to L$ then $L < 0$
(d) If infinitely many a_n's are inside every strip around zero then $a_n \to 0$.

3 Which of these statements are equivalent to $B \Rightarrow A$?

(a) If A is true so is B
(b) A is true if and only if B is true
(c) B is a sufficient condition for A
(d) A is a necessary condition for B.

4 Decide whether $A \Rightarrow B$ or $B \Rightarrow A$ or neither or both:

(a) $A = [a_n \to 1]$ $\quad B = [-a_n \to -1]$
(b) $A = [a_n \to 0]$ $\quad B = [a_n - a_{n-1} \to 0]$
(c) $A = [a_n \leqslant n]$ $\quad B = [a_n = n]$
(d) $A = [a_n \to 0]$ $\quad B = [\sin a_n \to 0]$
(e) $A = [a_n \to 0]$ $\quad B = [1/a_n$ fails to converge$]$
(f) $A = [a_n < n]$ $\quad B = [a_n/n$ converges$]$

***5** If the sequence $a_1, a_2, a_3, \ldots$ approaches zero, prove that we can put those numbers in any order and the new sequence still approaches zero.

***6** Suppose $f(x) \to L$ and $g(x) \to M$ as $x \to a$. Prove from the definitions that $f(x) + g(x) \to L + M$ as $x \to a$.

Find the limits 7–24 if they exist. An ε–δ test is not required.

7 $\displaystyle\lim_{t \to 2} \frac{t+3}{t^2-2}$ $\qquad$ **8** $\displaystyle\lim_{t \to 2} \frac{t^2+3}{t-2}$

9 $\displaystyle\lim_{x \to 0} \frac{f(x+h)-f(x)}{h}$ (careful) $\qquad$ **10** $\displaystyle\lim_{h \to 0} \frac{f(1+h)-f(1)}{h}$

11 $\displaystyle\lim_{h \to 0} \frac{\sin^2 h \cos^2 h}{h^2}$ $\qquad$ **12** $\displaystyle\lim_{x \to 0} \frac{2x \tan x}{\sin x}$

13 $\displaystyle\lim_{x \to 0^+} \frac{|x|}{x}$ (one-sided) $\qquad$ **14** $\displaystyle\lim_{x \to 0^-} \frac{|x|}{x}$ (one-sided)

15 $\displaystyle\lim_{x \to 1} \frac{\sin x}{x}$ $\qquad$ **16** $\displaystyle\lim_{c \to a} \frac{f(c)-f(a)}{c-a}$

17 $\displaystyle\lim_{x \to 5} \frac{x^2+25}{x-5}$ $\qquad$ **18** $\displaystyle\lim_{x \to 5} \frac{x^2-25}{x-5}$

19 $\displaystyle\lim_{x \to 0} \frac{\sqrt{1+x}-1}{x}$ (test $x = .01$) $\qquad$ **20** $\displaystyle\lim_{x \to 2} \frac{\sqrt{4-x}}{\sqrt{6+x}}$

21 $\displaystyle\lim_{x \to a} [f(x)-f(a)]$(?) $\qquad$ **22** $\displaystyle\lim_{x \to \pi/2} (\sec x - \tan x)$

23 $\displaystyle\lim_{x \to 0} \frac{\sin x}{\sin x/2}$ $\qquad$ **24** $\displaystyle\lim_{x \to 1} \frac{\sin(x-1)}{x^2-1}$

25 Choose δ so that $|f(x)| < \frac{1}{100}$ if $0 < x < \delta$.

$f(x) = 10x \qquad f(x) = \sqrt{x} \qquad f(x) = \sin 2x \qquad f(x) = x \sin x$

26 Which does the definition of a limit require?

(1) $|f(x) - L| < \varepsilon \;\Rightarrow\; 0 < |x - a| < \delta$
(2) $|f(x) - L| < \varepsilon \;\Leftarrow\; 0 < |x - a| < \delta$
(3) $|f(x) - L| < \varepsilon \;\Leftrightarrow\; 0 < |x - a| < \delta$

27 The definition of "$f(x) \to L$ as $x \to \infty$" is this: For any ε there is an X such that ________ $< \varepsilon$ if $x > X$. Give an example in which $f(x) \to 4$ as $x \to \infty$.

28 Give a correct definition of "$f(x) \to 0$ as $x \to -\infty$."

29 The limit of $f(x) = (\sin x)/x$ as $x \to \infty$ is ________. For $\varepsilon = .01$ find a point X beyond which $|f(x)| < \varepsilon$.

30 The limit of $f(x) = 2x/(1+x)$ as $x \to \infty$ is $L = 2$. For $\varepsilon = .01$ find a point X beyond which $|f(x) - 2| < \varepsilon$.

31 The limit of $f(x) = \sin x$ as $x \to \infty$ does not exist. Explain why not.

32 (Calculator) Estimate the limit of $\left(1+\frac{1}{x}\right)^x$ as $x \to \infty$.

33 For the polynomial $f(x) = 2x - 5x^2 + 7x^3$ find

(a) $\lim_{x\to 1} f(x)$ (b) $\lim_{x\to\infty} f(x)$

(c) $\lim_{x\to\infty} \frac{f(x)}{x^3}$ (d) $\lim_{x\to-\infty} \frac{f(x)}{x^3}$

34 For $f(x) = 6x^3 + 1000x$ find

(a) $\lim_{x\to\infty} \frac{f(x)}{x}$ (b) $\lim_{x\to\infty} \frac{f(x)}{x^2}$

(c) $\lim_{x\to\infty} \frac{f(x)}{x^4}$ (d) $\lim_{x\to\infty} \frac{f(x)}{x^3+1}$

Important rule As $x \to \infty$ the ratio of polynomials $f(x)/g(x)$ has the same limit as the ratio of their *leading terms.* $f(x) = x^3 - x + 2$ has leading term x^3 and $g(x) = 5x^6 + x + 1$ has leading term $5x^6$. Therefore $f(x)/g(x)$ behaves like $x^3/5x^6 \to 0$, $g(x)/f(x)$ behaves like $5x^6/x^3 \to \infty$, $(f(x))^2/g(x)$ behaves like $x^6/5x^6 \to 1/5$.

35 Find the limit as $x \to \infty$ if it exists:

$$\frac{3x^2+2x+1}{3+2x+x^2} \qquad \frac{x^4}{x^3+x^2} \qquad \frac{x^2+1000}{x^3-1000} \qquad x \sin \frac{1}{x}.$$

36 If a particular δ achieves $|f(x) - L| < \varepsilon$, why is it OK to choose a smaller δ?

37 The sum of $1 + r + r^2 + \cdots + r^{n-1}$ is $a_n = (1 - r^n)/(1 - r)$. What is the limit of a_n as $n \to \infty$? For which r does the limit exist?

38 If $a_n \to L$ prove that there is a number N with this property: If $n > N$ and $m > N$ then $|a_n - a_m| < 2\varepsilon$. This is Cauchy's test for convergence.

39 No matter what decimals come later, $a_1 = .4$, $a_2 = .49$, $a_3 = .493, \ldots$ approaches a limit L. How do we know (when we can't know L)? *Cauchy's test* is passed: the a's get closer to each other.

(a) From a_4 onwards we have $|a_n - a_m| <$ ________.

(b) After which a_N is $|a_m - a_n| < 10^{-7}$?

40 Choose decimals in Problem 39 so the limit is $L = .494$. Choose decimals so that your professor can't find L.

41 If every decimal in $.abcde\cdots$ is picked at random from 0, 1, ..., 9, what is the "average" limit L?

42 If every decimal is 0 or 1 (at random), what is the average limit L?

43 Suppose $a_n = \frac{1}{2}a_{n-1} + 4$ and start from $a_1 = 10$. Find a_2 and a_3 and a connection between $a_n - 8$ and $a_{n-1} - 8$. Deduce that $a_n \to 8$.

44 "For every δ there is an ε such that $|f(x)| < \varepsilon$ if $|x| < \delta$." That test is twisted around. Find ε when $f(x) = \cos x$, which does *not* converge to zero.

45 Prove the Squeeze Theorem for sequences, using ε: If $a_n \to L$ and $c_n \to L$ and $a_n \leqslant b_n \leqslant c_n$ for $n > N$, then $b_n \to L$.

46 Explain in 110 words the difference between "we will get there if you hurry" and "we will get there only if you hurry" and "we will get there if and only if you hurry."

2.7 Continuous Functions

This will be a brief section. It was originally included with limits, but the combination was too long. We are still concerned with the limit of $f(x)$ as $x \to a$, but a new number is involved. That number is $f(a)$, *the value of f at $x = a$.* For a "limit," x approached a but never reached it—so $f(a)$ was ignored. For a "continuous function," this final number $f(a)$ must be right.

May I summarize the usual (good) situation as x approaches a?

1. The number $f(a)$ exists (f is defined at a)
2. The limit of $f(x)$ exists (it was called L)
3. The limit L equals $f(a)$ ($f(a)$ is the right value)

In such a case, $f(x)$ is ***continuous*** at $x = a$. These requirements are often written in a single line: $f(x) \to f(a)$ as $x \to a$. By way of contrast, start with four functions that are *not* continuous at $x = 0$.

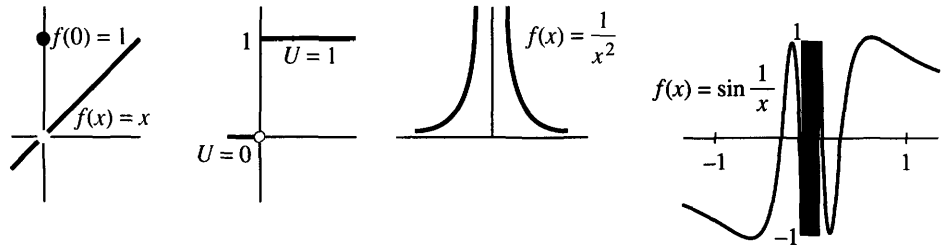


Fig. 2.20 Four types of discontinuity (others are possible) at $x = 0$.

In Figure 2.20, the first function would be continuous if it had $f(0) = 0$. But it has $f(0) = 1$. After changing $f(0)$ to the right value, the problem is gone. The discontinuity is *removable*. Examples 2, 3, 4 are more important and more serious. There is no "correct" value for $f(0)$:

2. $f(x) =$ step function (jump from 0 to 1 at $x = 0$)
3. $f(x) = 1/x^2$ (infinite limit as $x \to 0$)
4. $f(x) = \sin(1/x)$ (infinite oscillation as $x \to 0$).

The graphs show how the limit fails to exist. The step function has a ***jump discontinuity***. It has ***one-sided limits***, from the left and right. It does not have an ordinary (two-sided) limit. The limit from the left ($x \to 0^-$) is 0. The limit from the right ($x \to 0^+$) is 1. Another step function is $x/|x|$, which jumps from -1 to 1.

In the graph of $1/x^2$, the only reasonable limit is $L = +\infty$. I cannot go on record as saying that this limit exists. Officially, it doesn't—but we often write it anyway: $1/x^2 \to \infty$ as $x \to 0$. This means that $1/x^2$ goes (and stays) above every L as $x \to 0$.

In the same unofficial way we write one-sided limits for $f(x) = 1/x$:

$$\text{From the left, } \lim_{x \to 0^-} \frac{1}{x} = -\infty. \quad \text{From the right, } \lim_{x \to 0^+} \frac{1}{x} = +\infty. \tag{1}$$

Remark $1/x$ has a "***pole***" at $x = 0$. So has $1/x^2$ (a double pole). The function $1/(x^2 - x)$ has poles at $x = 0$ and $x = 1$. In each case the denominator goes to zero and the function goes to $+\infty$ or $-\infty$. Similarly $1/\sin x$ has a pole at every multiple of π (where $\sin x$ is zero). Except for $1/x^2$ these poles are "*simple*"—the functions are completely smooth at $x = 0$ when we multiply them by x:

$$(x)\left(\frac{1}{x}\right) = 1 \quad \text{and} \quad (x)\left(\frac{1}{x^2 - x}\right) = \frac{1}{x - 1} \quad \text{and} \quad (x)\left(\frac{1}{\sin x}\right) \quad \text{are continuous at } x = 0.$$

$1/x^2$ has a double pole, since it needs multiplication by x^2 (not just x). A ratio of polynomials $P(x)/Q(x)$ has poles where $Q = 0$, provided any common factors like $(x + 1)/(x + 1)$ are removed first.

Jumps and poles are the most basic discontinuities, but others can occur. The fourth graph shows that $\sin(1/x)$ has no limit as $x \to 0$. This function does not blow up; the sine never exceeds 1. At $x = \frac{1}{3}$ and $\frac{1}{4}$ and $\frac{1}{1000}$ it equals sin 3 and sin 4 and sin 1000. Those numbers are positive and negative and (?). As x gets small and $1/x$ gets large, the sine oscillates faster and faster. Its graph won't stay in a small box of height ε, no matter how narrow the box.

CONTINUOUS FUNCTIONS

DEFINITION f is "***continuous at*** $x = a$" if $f(a)$ is defined and $f(x) \to f(a)$ as $x \to a$. If f is continuous at every point where it is defined, it is a ***continuous function***.

Objection The definition makes $f(x) = 1/x$ a continuous function! It is not defined at $x = 0$, so its continuity can't fail. The logic requires us to accept this, but we don't have to like it. Certainly there is no $f(0)$ that would make $1/x$ continuous at $x = 0$.

It is amazing but true that the definition of "continuous function" is still debated (*Mathematics Teacher*, May 1989). You see the reason—we speak about a discontinuity of $1/x$, and at the same time call it a continuous function. The definition misses the difference between $1/x$ and $(\sin x)/x$. *The function $f(x) = (\sin x)/x$ can be made continuous at all x.* Just set $f(0) = 1$.

We call a function "***continuable***" if its definition can be extended *to all* x in a way that makes it continuous. Thus $(\sin x)/x$ and $\sqrt{x}$ are continuable. The functions $1/x$ and $\tan x$ are not continuable. This suggestion may not end the debate, but I hope it is helpful.

EXAMPLE 1 $\sin x$ and $\cos x$ and all polynomials $P(x)$ are continuous functions.

EXAMPLE 2 The absolute value $|x|$ is continuous. Its slope jumps (not continuable).

EXAMPLE 3 Any rational function $P(x)/Q(x)$ is continuous except where $Q = 0$.

EXAMPLE 4 The function that jumps between 1 at fractions and 0 at non-fractions is ***discontinuous everywhere***. There is a fraction between every pair of non-fractions and vice versa. (Somehow there are many more non-fractions.)

EXAMPLE 5 The function 0^{x^2} is zero for every x, except that 0^0 is not defined. So define it as zero and this function is continuous. But see the next paragraph where 0^0 has to be 1.

We could fill the book with proofs of continuity, but usually the situation is clear. "A function is continuous if you can draw its graph without lifting up your pen." At a jump, or an infinite limit, or an infinite oscillation, there is no way across the discontinuity except to start again on the other side. The function x^n is continuous for $n > 0$. It is not continuable for $n < 0$. The function x^0 equals 1 for every x, except that 0^0 is not defined. This time continuity requires $0^0 = 1$.

The interesting examples are the close ones—we have seen two of them:

EXAMPLE 6 $\dfrac{\sin x}{x}$ and $\dfrac{1 - \cos x}{x}$ are both continuable at $x = 0$.

Those were crucial for the slope of $\sin x$. The first approaches 1 and the second approaches 0. Strictly speaking we must give these functions the correct values (1 and 0) at the limiting point $x = 0$—which of course we do.

It is important to know what happens when the denominators change to x^2.

EXAMPLE 7 $\dfrac{\sin x}{x^2}$ blows up but $\dfrac{1 - \cos x}{x^2}$ has the limit $\dfrac{1}{2}$ at $x = 0$.

Since $(\sin x)/x$ approaches 1, dividing by another x gives a function like $1/x$. There is a simple pole. It is an example of 0/0, in which the zero from x^2 is reached more quickly than the zero from $\sin x$. The "***race to zero***" produces almost all interesting problems about limits.

For $1-\cos x$ and x^2 the race is almost even. Their ratio is 1 to 2:

$$\frac{1-\cos x}{x^2}=\frac{1-\cos^2 x}{x^2(1+\cos x)}=\frac{\sin^2 x}{x^2}\cdot\frac{1}{1+\cos x}\to\frac{1}{1+1}\text{ as } x\to 0.$$

This answer $\frac{1}{2}$ will be found again (more easily) by "l'Hôpital's rule." Here I emphasize not the answer but the problem. A central question of differential calculus is *to know how fast the limit is approached.* ***The speed of approach is exactly the information in the derivative.***

These three examples are all continuous at $x=0$. The race is controlled by the slope—because $f(x)-f(0)$ is nearly $f'(0)$ times x:

derivative of $\sin x$ is 1 $\leftrightarrow$ $\sin x$ decreases like x

derivative of $\sin^2 x$ is 0 $\leftrightarrow$ $\sin^2 x$ decreases faster than x

derivative of $x^{1/3}$ is ∞ $\leftrightarrow$ $x^{1/3}$ decreases more slowly than x.

DIFFERENTIABLE FUNCTIONS

The absolute value $|x|$ is continuous at $x=0$ but has no derivative. The same is true for $x^{1/3}$. ***Asking for a derivative is more than asking for continuity.*** The reason is fundamental, and carries us back to the key definitions:

Continuous at x: $f(x+\Delta x)-f(x)\to 0$ as $\Delta x\to 0$

Derivative at x: $\dfrac{f(x+\Delta x)-f(x)}{\Delta x}\to f'(x)$ as $\Delta x\to 0$.

In the first case, Δf goes to zero (maybe slowly). In the second case, Δf goes to zero *as fast as* Δx (because $\Delta f/\Delta x$ has a limit). That requirement is stronger:

> **2I** At a point where $f(x)$ has a derivative, the function must be continuous. But $f(x)$ can be continuous with no derivative.

Proof The limit of $\Delta f=(\Delta x)(\Delta f/\Delta x)$ is $(0)(df/dx)=0$. So $f(x+\Delta x)-f(x)\to 0$.

The continuous function $x^{1/3}$ has no derivative at $x=0$, because $\frac{1}{3}x^{-2/3}$ blows up. The absolute value $|x|$ has no derivative because its slope jumps. The remarkable function $\frac{1}{2}\cos 3x+\frac{1}{4}\cos 9x+\cdots$ is continuous at *all points* and has a derivative at *no points*. You can draw its graph without lifting your pen (but not easily—it turns at every point). To most people, it belongs with space-filling curves and unmeasurable areas—in a box of curiosities. Fractals used to go into the same box! They are beautiful shapes, with boundaries that have no tangents. The theory of fractals is very alive, for good mathematical reasons, and we touch on it in Section 3.7.

I hope you have a clear idea of these basic definitions of calculus:

1 ***Limit*** $(n\to\infty$ or $x\to a)$ **2** ***Continuity*** (at $x=a$) **3** ***Derivative*** (at $x=a$).

Those go back to ε and δ, but it is seldom necessary to follow them so far. In the same way that economics describes many transactions, or history describes many events, a function comes from many values $f(x)$. A few points may be special, like market crashes or wars or discontinuities. At other points df/dx is the best guide to the function.

This chapter ends with two essential facts about ***a continuous function on a closed interval.*** The interval is $a \leqslant x \leqslant b$, written simply as $[a, b]$.† At the endpoints a and b we require $f(x)$ to approach $f(a)$ and $f(b)$.

Extreme Value Property A continuous function on the finite interval $[a, b]$ has a maximum value M and a minimum value m. There are points $x_{\max}$ and $x_{\min}$ in $[a, b]$ where it reaches those values:

$$f(x_{\max}) = M \geqslant f(x) \geqslant f(x_{\min}) = m \quad \text{for all } x \text{ in } [a, b].$$

Intermediate Value Property If the number F is between $f(a)$ and $f(b)$, there is a point c between a and b where $f(c) = F$. Thus if F is between the minimum m and the maximum M, there is a point c between $x_{\min}$ and $x_{\max}$ where $f(c) = F$.

Examples show why we require closed intervals and continuous functions. For $0 < x \leqslant 1$ the function $f(x) = x$ never reaches its minimum (zero). If we close the interval by defining $f(0) = 3$ (discontinuous) the minimum is still not reached. Because of the jump, the intermediate value $F = 2$ is also not reached. The idea of continuity was inescapable, after Cauchy defined the idea of a limit.

2.7 EXERCISES

Read-through questions

Continuity requires the __a__ of $f(x)$ to exist as $x \to a$ and to agree with __b__. The reason that $x/|x|$ is not continuous at $x = 0$ is __c__. This function does have __d__ limits. The reason that $1/\cos x$ is discontinuous at __e__ is __f__. The reason that $\cos(1/x)$ is discontinuous at $x = 0$ is __g__. The function $f(x) =$ __h__ has a simple pole at $x = 3$, where f^2 has a __i__ pole.

The power x^n is continuous at all x provided n is __j__. It has no derivative at $x = 0$ when n is __k__. $f(x) = \sin(-x)/x$ approaches __l__ as $x \to 0$, so this is a __m__ function provided we define $f(0) =$ __n__. A "continuous function" must be continuous at all __o__. A "continuable function" can be extended to every point x so that __p__.

If f has a derivative at $x = a$ then f is necessarily __q__ at $x = a$. The derivative controls the speed at which $f(x)$ approaches __r__. On a closed interval $[a, b]$, a continuous f has the __s__ value property and the __t__ value property. It reaches its __u__ M and its __v__ m, and it takes on every value __w__.

In Problems 1–20, find the numbers c that make $f(x)$ into (A) a continuous function and (B) a differentiable function. In one case $f(x) \to f(a)$ at every point, in the other case $\Delta f/\Delta x$ has a limit at every point.

1 $f(x) = \begin{cases} \sin x & x < 1 \\ c & x \geqslant 1 \end{cases}$

2 $f(x) = \begin{cases} \cos^3 x & x \neq \pi \\ c & x = \pi \end{cases}$

3 $f(x) = \begin{cases} cx & x < 0 \\ 2cx & x \geqslant 0 \end{cases}$

4 $f(x) = \begin{cases} cx & x < 1 \\ 2cx & x \geqslant 1 \end{cases}$

5 $f(x) = \begin{cases} c + x & x < 0 \\ c^2 + x^2 & x \geqslant 0 \end{cases}$

6 $f(x) = \begin{cases} x^3 & x \neq c \\ -8 & x = c \end{cases}$

7 $f(x) = \begin{cases} 2x & x < c \\ x + 1 & x \geqslant c \end{cases}$

8 $f(x) = \begin{cases} x^c & x \neq 0 \\ 0 & x = 0 \end{cases}$

9 $f(x) = \begin{cases} (\sin x)/x^2 & x \neq 0 \\ c & x = 0 \end{cases}$

10 $f(x) = \begin{cases} x + c & x \leqslant c \\ 1 & x > c \end{cases}$

11 $f(x) = \begin{cases} c & x \neq 4 \\ 1/x^3 & x = 4 \end{cases}$

12 $f(x) = \begin{cases} c & x \leqslant 0 \\ \sec x & x \geqslant 0 \end{cases}$

13 $f(x) = \begin{cases} \dfrac{x^2 + c}{x - 1} & x \neq 1 \\ 2 & x = 1 \end{cases}$

14 $f(x) = \begin{cases} \dfrac{x^2 - 1}{x - c} & x \neq c \\ 2c & x = c \end{cases}$

15 $f(x) = \begin{cases} (\tan x)/x & x \neq 0 \\ c & x = 0 \end{cases}$

16 $f(x) = \begin{cases} x^2 & x \leqslant c \\ 2x & x > c \end{cases}$

17 $f(x) = \begin{cases} (c + \cos x)/x & x \neq 0 \\ 0 & x = 0 \end{cases}$

18 $f(x) = |x + c|$

†The interval $[a, b]$ is ***closed*** (endpoints included). The interval (a, b) is ***open*** (a and b left out). The infinite interval $[0, \infty)$ contains all $x \geqslant 0$.

19 $f(x)=\begin{cases}(\sin x - x)/x^c & x \neq 0 \\ 0 & x=0\end{cases}$ **20** $f(x)=|x^2+c^2|$

Construct your own $f(x)$ with these discontinuities at $x=1$.

21 Removable discontinuity

22 Infinite oscillation

23 Limit for $x \to 1^+$, no limit for $x \to 1^-$

24 A double pole

25 $\lim_{x\to 1^-} f(x) = 4 + \lim_{x \to 1^+} f(x)$

26 $\lim_{x\to 1} f(x)=\infty$ but $\lim_{x\to 1}(x-1)f(x)=0$

27 $\lim_{x\to 1}(x-1)f(x)=5$

28 The statement "$3x \to 7$ as $x \to 1$" is false. Choose an ε for which no δ can be found. The statement "$3x \to 3$ as $x \to 1$" is true. For $\varepsilon = \frac{1}{2}$ choose a suitable δ.

29 How many derivatives f', f'', ... are continuable functions?

(a) $f=x^{3/2}$ (b) $f=x^{3/2}\sin x$ (c) $f=(\sin x)^{5/2}$

30 Find one-sided limits at points where there is no two-sided limit. Give a 3-part formula for function (c).

(a) $\dfrac{|x|}{7x}$ (b) $\sin|x|$ (c) $\dfrac{d}{dx}|x^2-1|$

31 Let $f(1)=1$ and $f(-1)=1$ and $f(x)=(x^2-x)/(x^2-1)$ otherwise. Decide whether f is continuous at

(a) $x=1$ (b) $x=0$ (c) $x=-1$.

***32** Let $f(x)=x^2\sin 1/x$ for $x\neq 0$ and $f(0)=0$. If the limits exist, find

(a) $\lim_{x\to 0} f(x)$ (b) df/dx at $x=0$ (c) $\lim_{x\to 0} f'(x)$.

33 If $f(0)=0$ and $f'(0)=3$, rank these functions from smallest to largest as x decreases to zero:

$$f(x),\ \ x,\ \ xf(x),\ \ f(x)+2x,\ \ 2(f(x)-x),\ \ (f(x))^2.$$

34 Create a discontinuous function $f(x)$ for which $f^2(x)$ is continuous.

35 *True or false*, with an example to illustrate:

(a) If $f(x)$ is continuous at all x, it has a maximum value M.

(b) If $f(x) \leqslant 7$ for all x, then f reaches its maximum.

(c) If $f(1)=1$ and $f(2)=-2$, then somewhere $f(x)=0$.

(d) If $f(1)=1$ and $f(2)=-2$ and f is continuous on $[1, 2]$, then somewhere on that interval $f(x)=0$.

36 The functions $\cos x$ and $2x$ are continuous. Show from the ________ property that $\cos x = 2x$ at some point between 0 and 1.

37 Show by example that these statements are false:

(a) If a function reaches its maximum and minimum then the function is continuous.

(b) If $f(x)$ reaches its maximum and minimum and all values between $f(0)$ and $f(1)$, it is continuous at $x=0$.

(c) (mostly for instructors) If $f(x)$ has the intermediate value property between all points a and b, it must be continuous.

38 Explain with words and a graph why $f(x)=x\sin(1/x)$ is continuous but has no derivative at $x=0$. Set $f(0)=0$.

39 Which of these functions are *continuable*, and why?

$$f_1(x)=\begin{cases}\sin x & x<0\\ \cos x & x>1\end{cases} \qquad f_2(x)=\begin{cases}\sin 1/x & x<0\\ \cos 1/x & x>1\end{cases}$$

$$f_3(x)=\frac{x}{\sin x} \text{ when } \sin x \neq 0 \qquad f_4(x)=x^0+0^{x^2}$$

40 Explain the difference between a continuous function and a continuable function. Are continuous functions always continuable?

***41** $f(x)$ is any continuous function with $f(0)=f(1)$.

(a) Draw a typical $f(x)$. Mark where $f(x)=f(x+\frac{1}{2})$.

(b) Explain why $g(x)=f(x+\frac{1}{2})-f(x)$ has $g(\frac{1}{2})=-g(0)$.

(c) Deduce from (b) that (a) is always possible: There *must* be a point where $g(x)=0$ and $f(x)=f(x+\frac{1}{2})$.

42 Create an $f(x)$ that is continuous only at $x=0$.

43 If $f(x)$ is continuous and $0 \leqslant f(x) \leqslant 1$ for all x, then there is a point where $f(x^*)=x^*$. Explain with a graph and prove with the intermediate value theorem.

44 In the ε–δ definition of a limit, change $0<|x-a|<\delta$ to $|x-a|<\delta$. Why is $f(x)$ now *continuous* at $x=a$?

45 A function has a ________ at $x=0$ if and only if $(f(x)-f(0))/x$ is ________ at $x=0$.

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:

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

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

$$df = v\,dt \quad \textit{is exactly true.}$$

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

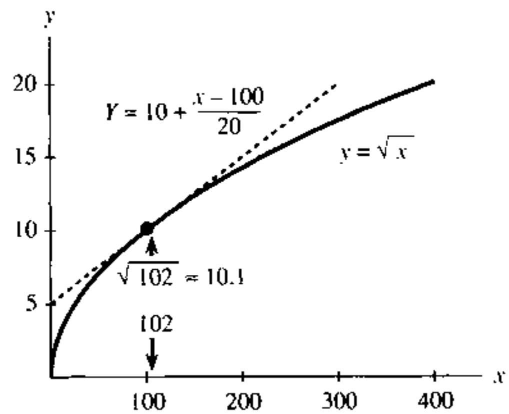
In Section 2.3 we found the tangent line to $y = f(x)$. At $x = a$, the slope of the curve and the slope of the line are $f'(a)$. For points on the line, start at $y = f(a)$. Add the slope times the "increment" $x - a$:

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

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 $x=0$:

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

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:

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

DIFFERENTIALS

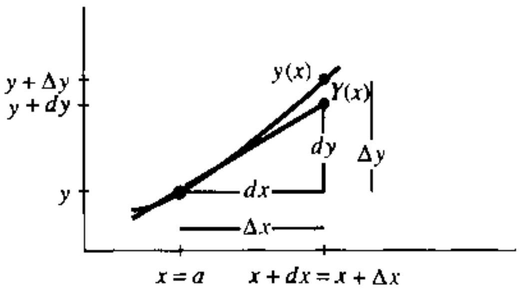
There is one more notation for this linear approximation. It has to be presented, because it is often used. The notation is suggestive and confusing at the same time—

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 $f'(x)dx$.



Δy = 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 $m_0 +$ ________ 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 $\sqrt{1 + x} \approx$ ________.

30 In Problem 29 the numbers you see are less than 1.05, 1.025, ... The second derivative of $\sqrt{1 + x}$ is ________ 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 ________ 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 $x = 1$.

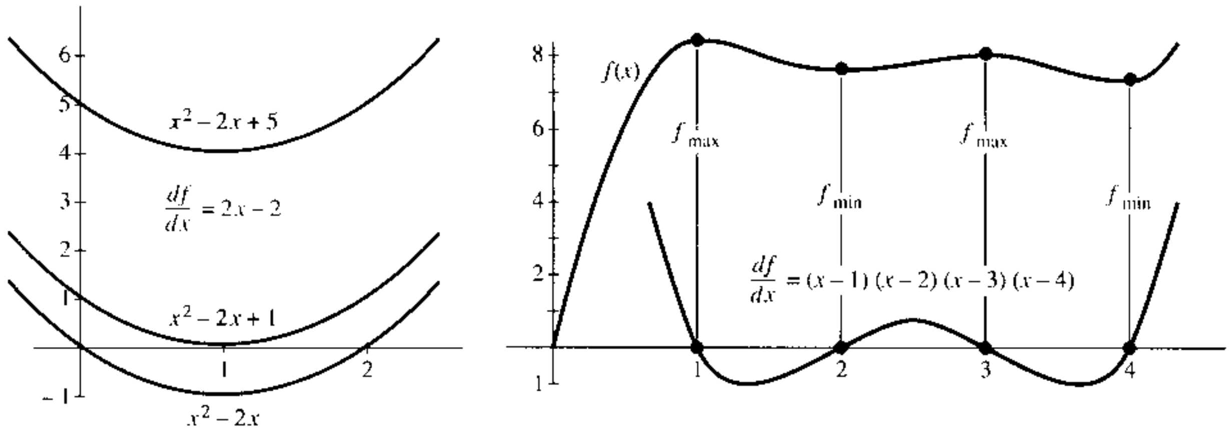


Fig. 3.3 Slopes are $- +$. Slope is $+ - + - +$ so f is up-down-up-down-up.

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.

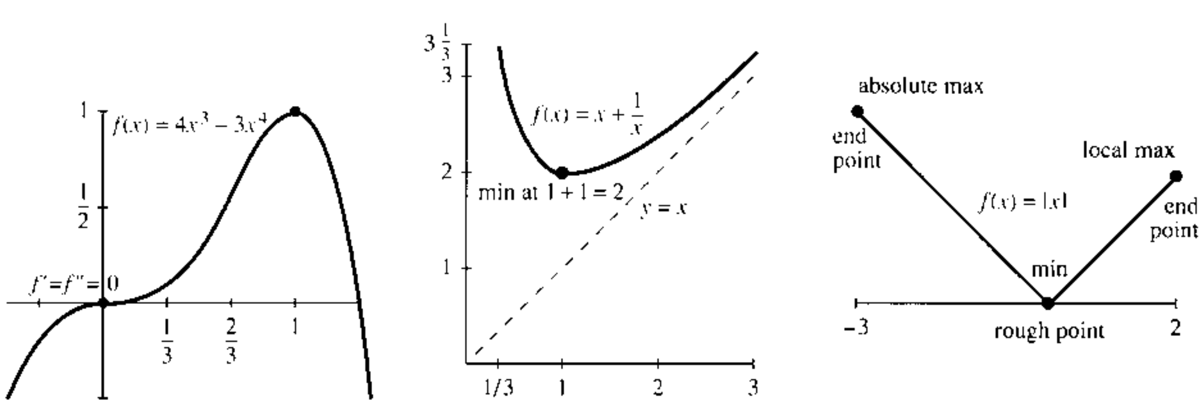


Fig. 3.4 The graphs of $4x^3 - 3x^4$ and $x + x^{-1}$. Check rough points and endpoints.

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

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 $2x$.

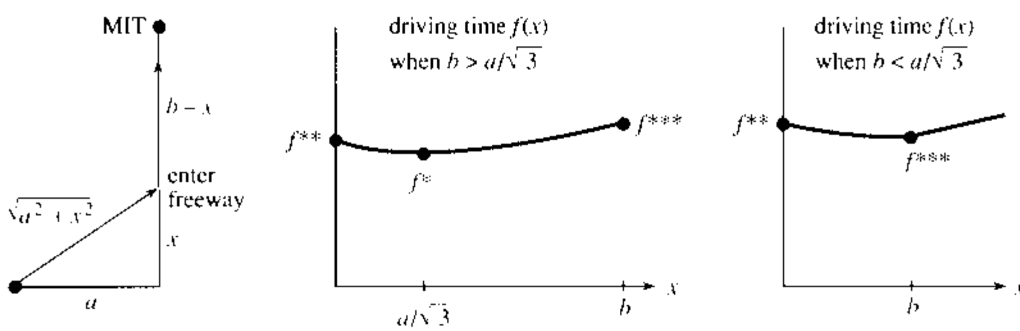


Fig. 3.5 Join the freeway at x — minimize the driving time $f(x)$.

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 $x \leq b$, 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 $\sin x = \frac{1}{2}$. 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.

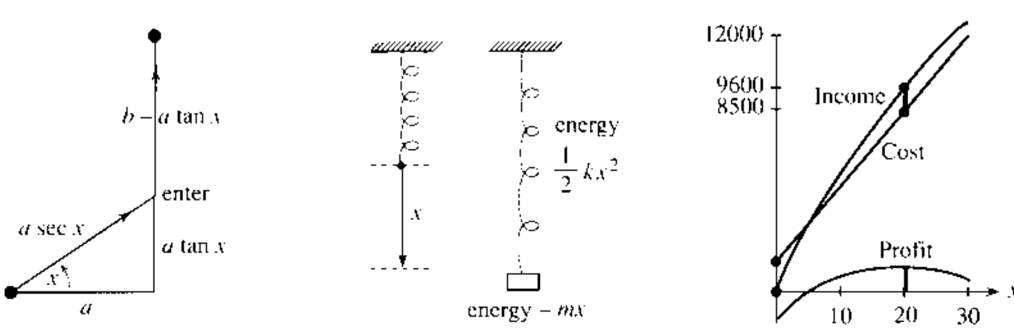


Fig. 3.6 (a) Driving at angle x. (b) Energies of spring and mass. (c) Profit = income − cost.

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 $x = 20$

$$\textit{profit} = \textit{income} - \textit{cost} = 9600 - 8500 = 1100.$$

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.

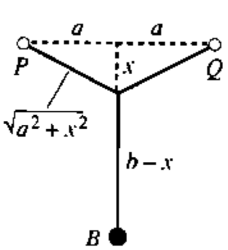
28 On the longest and shortest days, in June and December, why does the length of day change the least?

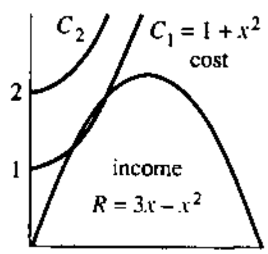
29 Find the shortest **Y** connecting P, Q, and B in the figure. Originally B was a birdfeeder. The length of **Y** is $L(x) = (b-x) + 2\sqrt{a^2 + x^2}$.

(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 $C(x) = c + dx + ex^2$.

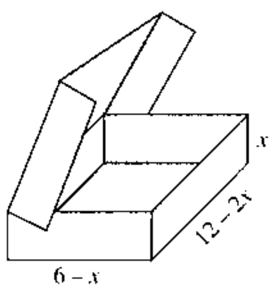
31 The profit is $\Pi(x) =$ ________. The average profit per pizza is $=$ ________. The marginal profit per additional pizza is $d\Pi/dx =$ ________. We should maximize the (profit) (average profit) (marginal profit).

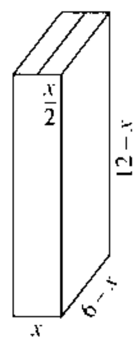
32 We receive $R(x) = ax + bx^2$ when the price per pizza is $p(x) =$ ________. In reverse: When the price is p we sell $x =$ ________ pizzas (a function of p). We expect $b < 0$ because ________.

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″ × 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.

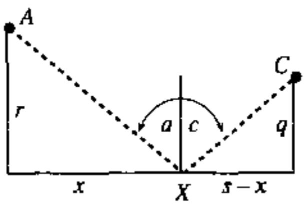
51 Find the closed cylinder of volume $V = \pi r^2 h = 16\pi$ that has the least surface area.

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 $(1-2x^2)/\sqrt{1-x^2}$.

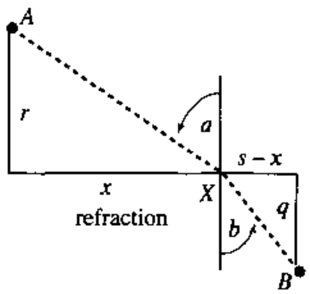
In 53–56, x and y are nonnegative numbers with $x + y = 10$. Maximize and minimize:

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.

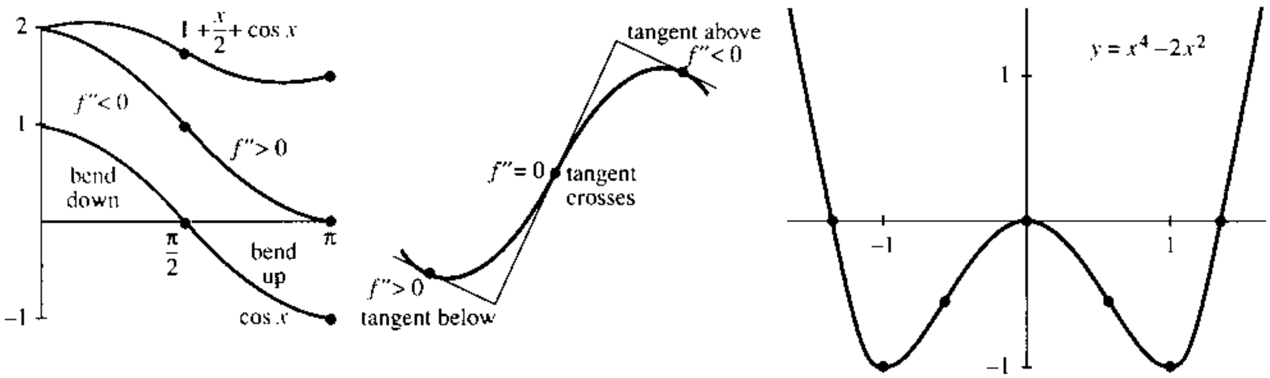


Fig. 3.7 Increasing slope = concave up ($f'' > 0$). Concave down is $f'' < 0$. Inflection point $f'' = 0$.

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***":

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

†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:

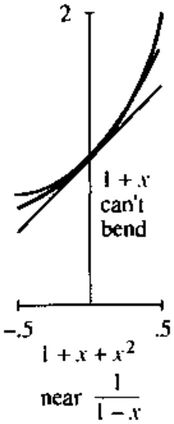


Fig. 3.8

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 $\frac{1}{16}$:

linear approximation $\left(\text{error} \approx \frac{3}{8}x^2\right)$:	.0194	.0053	.0014
quadratic approximation $\left(\text{error} \approx \frac{-5}{16}x^3\right)$:	−.00401	−.00055	−.00007

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 $f(\Delta x) \approx f(0) + f'(0)\Delta x + \frac{1}{2}f''(0)(\Delta x)^2 +$ ________. 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 ________ 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 ________. Then the tangent line is ________ 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.

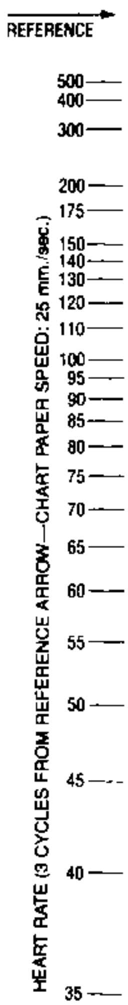
47 Repeat Problem 46 on a printout showing $y(x) = x^3 - 4x^2 + x + 2$ and dy/dx and d^2y/dx^2 on the same graph.

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.

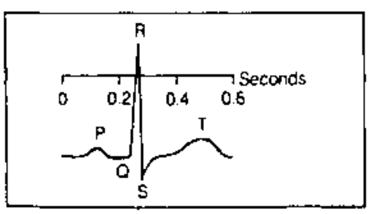
This section combines two approaches. One is to study actual machine-produced graphs (especially electrocardiograms). The other is to understand the mathematics of graphs—slope, concavity, asymptotes, shifts, and scaling. We introduce the ***centering transform*** and ***zoom transform***. These two 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.

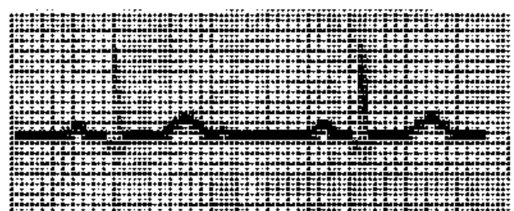
READING AN ELECTROCARDIOGRAM (ECG or EKG)



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.

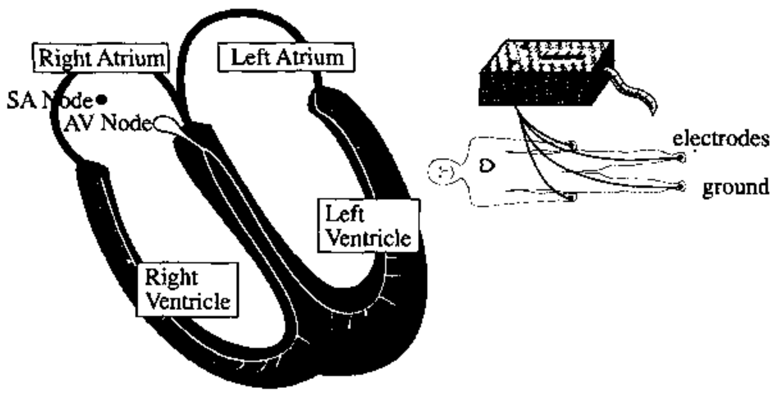


Fig. 3.9 Happy person with a heart and a normal electrocardiogram.

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 tbe 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:

I*schemia*: Reduced blood supply, upside-down T wave in the chest leads.
I*njury*: An elevated segment between S and T means a recent attack.
I*nfarction*: 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 $\quad f(x) = \dfrac{x^2}{1 - x^2} \qquad f'(x) = \dfrac{2x}{(1 - x^2)^2} \qquad f''(x) = \dfrac{2 + 6x^2}{(1 - x^2)^3}$

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

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 $f'' = 0$.

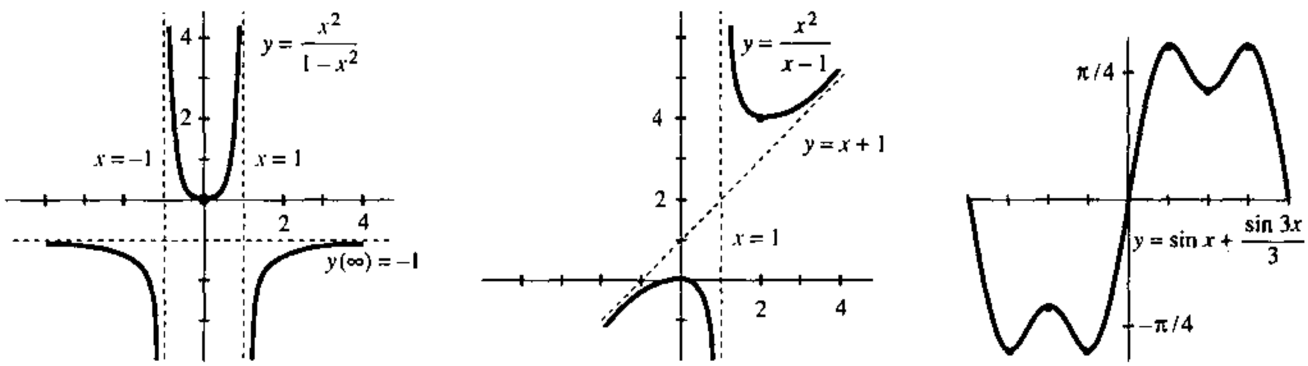


Fig. 3.11 The graphs of $x^2/(1-x^2)$ and $x^2/(x-1)$ and $\sin x + \frac{1}{3}\sin 3x$.

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

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*:

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

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

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

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:

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

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.

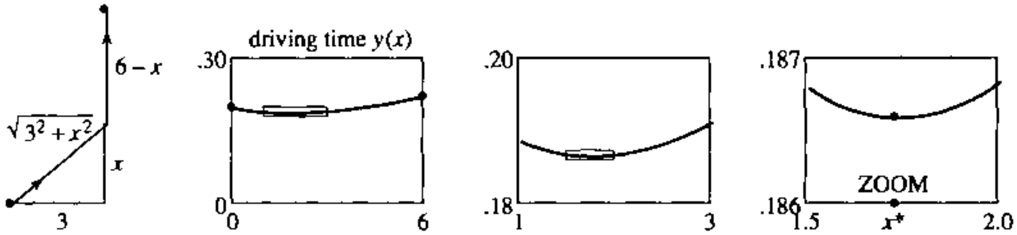


Fig. 3.12 Enter at x. The graph of driving time $y(x)$. Zoom boxes locate x^*.

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.

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.

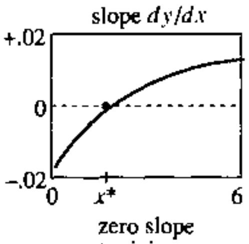


Fig. 3.13

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 $x^* = \sqrt{3} = 1.73205\ldots$

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 tbe 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 $\mathbf{y}/c=2(\mathbf{x}/c)^2$. The number 2 disappears if $c=d=2$. With the right centering and the right zoom, every parabola that opens upward is $\mathbf{y}=\mathbf{x}^2$.

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 $\mathbf{y}=dF(\mathbf{x}/c)$. 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**:

$$y=f(x) \quad \text{becomes} \quad Y=f(X+a)-b \quad \text{and then} \quad \mathbf{y}=d\left[f\left(\frac{\mathbf{x}}{c}+a\right)-b\right].$$

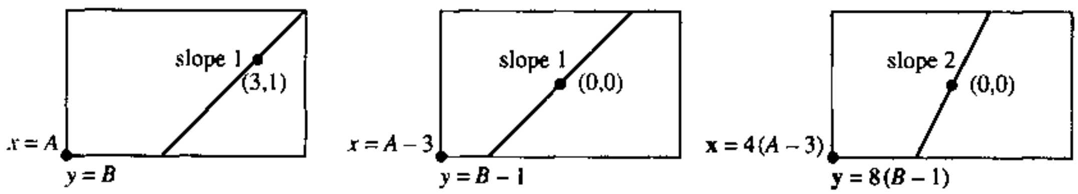


Fig. 3.14 Change of coordinates by centering and zoom. Calculators still show (x, y).

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 $-1 \leqslant c(\mathbf{x}-a) \leqslant 1$ 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 $\mathbf{y}=\mathbf{x}^2$). 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 $\mathbf{y}=\cos \mathbf{t}$.

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 $\mathbf{x}=cX$ and __v__. The __w__ transform changes $Y=F(X)$ to $\mathbf{y}=$ __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 $\mathbf{xy}$ 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.

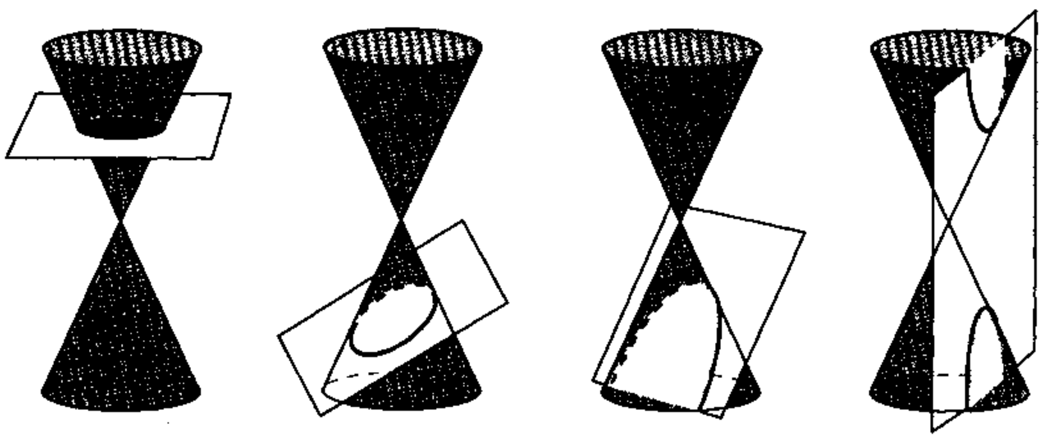


Fig. 3.15 The cutting plane gets steeper: circle to ellipse to parabola to hyperbola.

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:

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

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

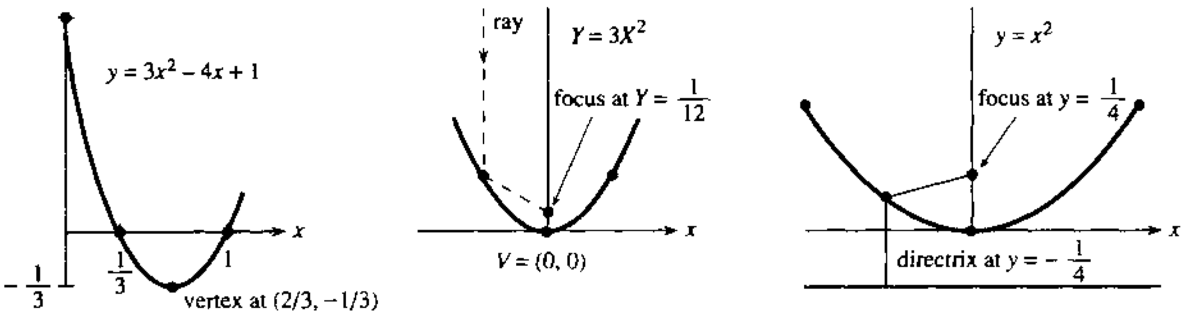


Fig. 3.16 Parabola with minimum at V. Rays reflect to focus. Centered in (b), rescaled in (c).

A parabola has another important point—the ***focus***. Its distance from the vertex is called p. The special parabola $\mathbf{y} = \mathbf{x}^2$ has $p = 1/4$, and other parabolas $Y = aX^2$ have $p = 1/4a$. You magnify by a factor a to get $\mathbf{y} = \mathbf{x}^2$. 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 tbat with tbe distance to the focus at $(0, \frac{1}{4})$— this is the square root below. Out comes the special parabola $y = x^2$:

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

The exercises give practice with all the steps we have taken—center the parabola to $Y = aX^2$, rescale it to $\mathbf{y} = \mathbf{x}^2$, 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":

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

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.

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

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 $x^2 \leqslant a^2$, 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):

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

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:

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

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.

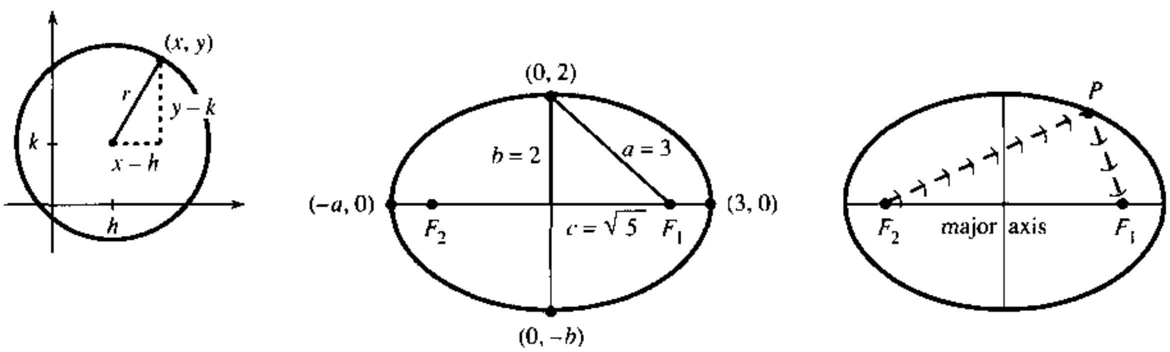


Fig. 3.17 Uncentered circle. Centered ellipse $x^2/3^2 + y^2/2^2 = 1$. *The distance from center to far right is also* $a = 3$. All rays from F_2 reflect to F_1.

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:

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

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:

$$\text{from } F_1 \text{ and } F_2 \text{ to } (x, y): \quad \sqrt{(x-c)^2+y^2} + \sqrt{(x+c)^2+y^2} = 2a. \qquad (5)$$

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.

HYPERBOLAS $\dfrac{y^2}{a^2} - \dfrac{x^2}{b^2} = 1$

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 −):

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

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***:

$$\text{if } \frac{x^2}{b^2} = 1000 \quad \text{then} \quad \frac{y^2}{a^2} = 1001. \quad \text{So } \frac{y}{a} \text{ is close to } \frac{x}{b} \text{ or } -\frac{x}{b}.$$

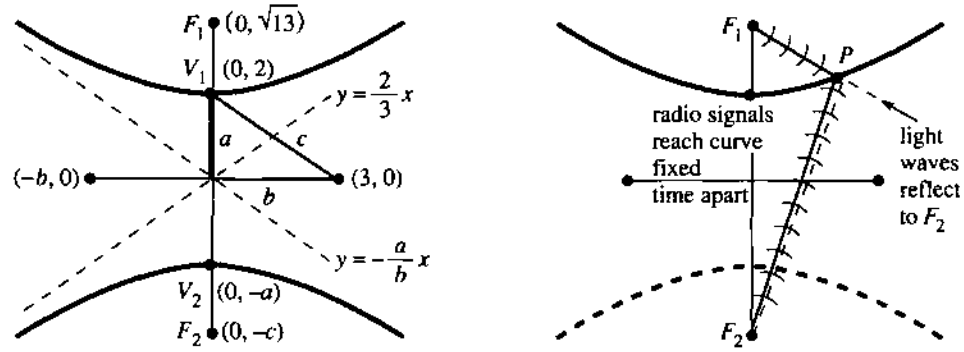


Fig. 3.18 The hyperbola $\frac{1}{4}y^2 - \frac{1}{9}x^2 = 1$ has $a = 2$, $b = 3$, $c = \sqrt{4+9}$. The distances to F_1 and F_2 differ by $2a = 4$.

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':

$$\begin{aligned} x &= x' \cos\alpha - y' \sin\alpha \\ y &= x' \sin\alpha + y' \cos\alpha \end{aligned} \quad \text{and} \quad \begin{aligned} x' &= x\cos\alpha + y\sin\alpha \\ y' &= -y\sin\alpha + x\cos\alpha. \end{aligned} \tag{7}$$

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:

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

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 $\mathbf{y} = \mathbf{x}^2$ or $\mathbf{x}^2 + \mathbf{y}^2 = 1$ or $\mathbf{y}^2 - \mathbf{x}^2 = 1$.

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

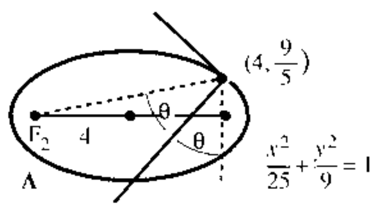
25 Decide between circle-ellipse-parabola-hyperbola, based on the XY equation with $X = x - 1$ and $Y = y + 3$.

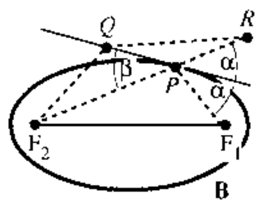
(a) $x^2 - 2x + y^2 + 6y = 6$
(b) $x^2 - 2x - y^2 - 6y = 6$
(c) $x^2 - 2x + 2y^2 + 12y = 6$
(d) $x^2 - 2x - y = 6$.

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 $s = -1/$(slope of tangent) = ________. The slope of the line from F_2 is S = ________. 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 ________ with center at ________ and radius ________. 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 ________ with foci at ________.

39 $y^2 - 4y$ is part of $(y - 2)^2 =$ ________ and $2x^2 + 12x$ is part of $2(x + 3)^2 =$ ________. Therefore $y^2 - 4y - 2x^2 - 12x = 0$ gives the hyperbola $(y - 2)^2 - 2(x + 3)^2 =$ ________. Its center is ________ and it opens to the ________.

40 Following Problem 39 turn $y^2 + 2y = x^2 + 10x$ into $Y^2 = X^2 + C$ with X, Y, and C equal to ________.

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 ________ 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 ________ and no real roots if ________. 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 $x_{12} = .75, \quad x_{13} = .73, \quad x_{14} = .74, \quad \ldots, \quad x_{29} = .7391, \quad x_{30} = .7391.$

The goal is to explain why the x's approach $x^* = .739085 \ldots$. 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 $x_0, x_1, x_2, \ldots$.

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

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

Those numbers $0, 4, 6, 7, 7\frac{1}{2}, \ldots$ 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 $8, 8, 8, \ldots$. When we start at $x_0 = 12$, the sequence goes back toward 8:

$$x_1 = \tfrac{1}{2} \cdot 12 + 4 = 10, \qquad x_2 = \tfrac{1}{2} \cdot 10 + 4 = 9, \qquad x_3 = \tfrac{1}{2} \cdot 9 + 4 = 8.5, \qquad \ldots$$

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 $.7391\ldots$ is special because it equals cos $.7391\ldots$. ***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 $\frac{1}{4}, \frac{1}{16}, \frac{1}{256}, \ldots$ 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 $4, 16, 256, \ldots$ 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 $F'\Delta x$:***

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

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 $y = x$.

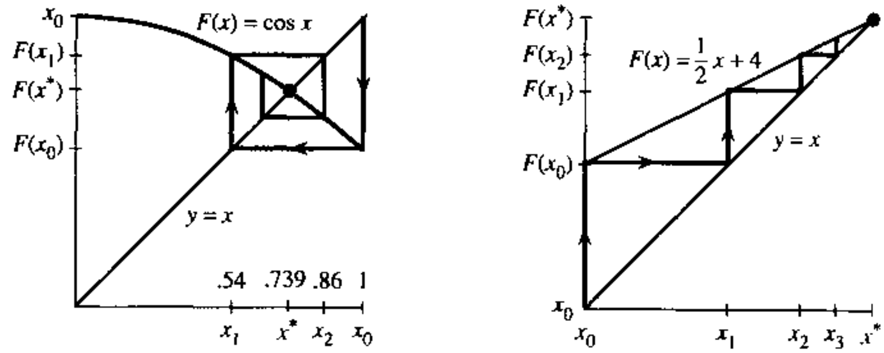


Fig. 3.19 Cobwebs go from (x_0, x_0) to (x_0, x_1) to (x_1, x_1)—line to curve to line.

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 `LINE: PLOT X,Y` followed by `EXE`. Now move the cursor vertically to $y = F(x)$ and press `EXE`. Then move horizontally to $y = x$ and press `EXE`. Continue. Each step draws a line.

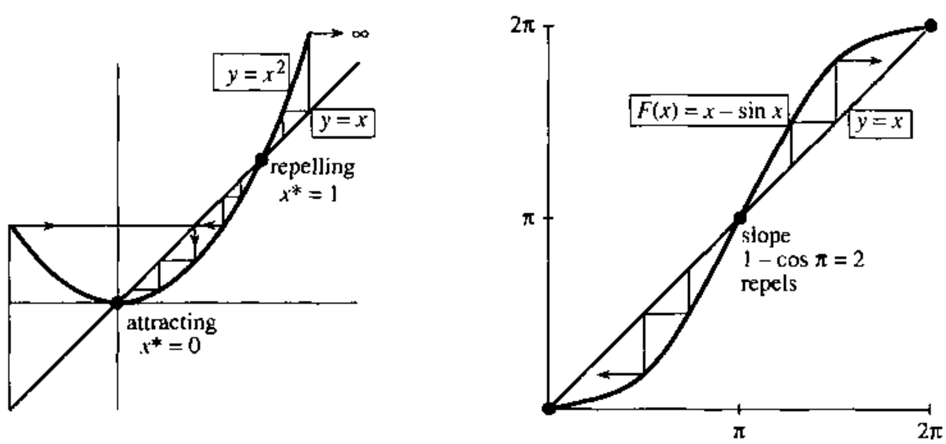


Fig. 3.20 Converging and diverging cobwebs: $F(x) = x^2$ and $F(x) = x - \sin x$.

For the TI-81 (and also the Casio) a short program produces a cobweb. Store $F(x)$ in the Y= function slot Y1. 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:

$$x_n \textit{ converges to } x^* \textit{ if and only if } -1 < 1 - ca < 1. \tag{5}$$

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 $x^* = x^* - cf(x^*)$:

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

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^*:

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

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:

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

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.

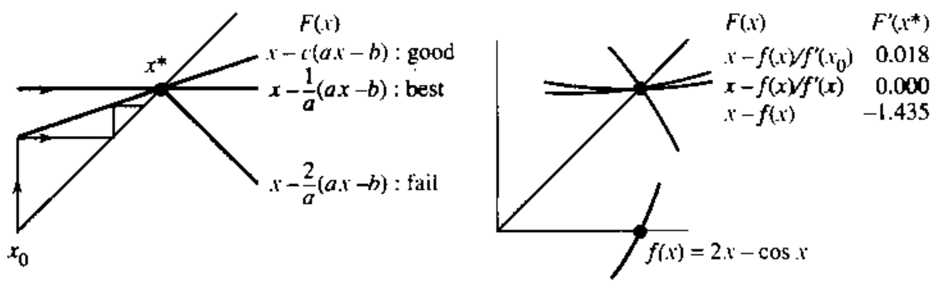


Fig. 3.21 The error multiplier is $m = 1 - cf'(x^*)$. Newton has $c = 1/f'(x_n)$ and $m \to 0$.

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 $c_n = 1/f'(x_n)$:

$x_0 = .50$	$c = 1$	$c = 1/f'(x_0)$	$c_n = 1/f'(x_n)$
$x_1 =$	.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 $(\text{error})_{n+1} \leq M(\text{error})_n^2$. 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 $x^* = .450\ldots$.

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 superattracting 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 $x_1, x_2, \ldots$ 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 $x_1, x_2, \ldots$ 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 $0, -1, 0, -1, \ldots$.

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}:

$$\textit{Newton's method} \qquad x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}. \tag{3}$$

Usually this iteration $x_{n+1} = F(x_n)$ converges quickly to x^*.

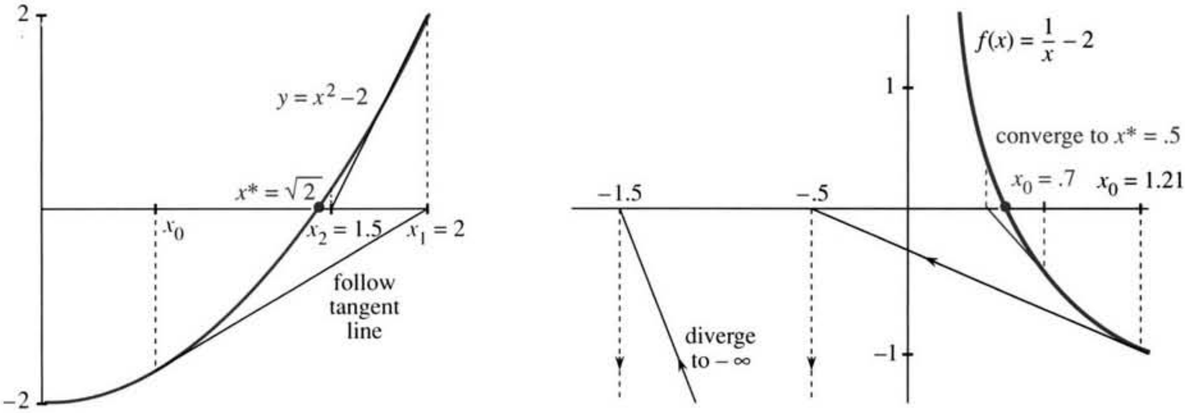


Fig. 3.22 Newton's method along tangent lines from x_0 to x_1 to x_2.

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

$$x_{n+1} = x_n - \frac{x_n^2 - b}{2x_n} = x_n - \frac{1}{2}x_n + \frac{b}{2x_n}. \tag{4}$$

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

$$F(x) = \frac{1}{2}\left(x + \frac{b}{x}\right) \quad \text{and} \quad F'(x) = \frac{1}{2}\left(1 - \frac{b}{x^2}\right). \tag{5}$$

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:

$$x_{n+1} = 2x_n - 2x_n^2 \quad \text{becomes} \quad x_{n+1} - \tfrac{1}{2} = -2(x_n - \tfrac{1}{2})^2. \tag{8}$$

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* $f(x) = 0$.

The examples $x^2 = b$ and $1/x = a$ show fast convergence or failure. In Chapter 13, and in reality, Newton's method solves much harder equations. Here I am going to choose a third example that came from pure curiosity about what might happen. The results are absolutely amazing. The equation is $x^2 = -1$.

EXAMPLE 3 ***What happens to Newton's method if you ask it to solve $f(x) = x^2 + 1 = 0$?***

The only solutions are the imaginary numbers $x^* = i$ and $x^* = -i$. There is no real square root of -1. Newton's method might as well give up. But it has no way to know that! The tangent line still crosses the axis at a new point x_{n+1}, even if the curve $y = x^2 + 1$ never crosses. Equation (5) still gives the iteration for $b = -1$:

$$x_{n+1} = \frac{1}{2}\left(x_n - \frac{1}{x_n}\right) = F(x_n). \tag{9}$$

The x's cannot approach i or $-i$ (nothing is imaginary). So what do they do?

The starting guess $x_0 = 1$ is interesting. It is followed by $x_1 = 0$. Then x_2 divides by zero and blows up. I expected other sequences to go to infinity. But the experiments showed something different (and mystifying). When x_n is large, x_{n+1} is less than half as large. After $x_n = 10$ comes $x_{n+1} = \frac{1}{2}(10 - \frac{1}{10}) = 4.95$. After much indecision and a long wait, a number near zero eventually appears. Then the next guess divides by that small number and goes far out again. This reminded me of "chaos."

It is tempting to retreat to ordinary examples, where Newton's method is a big success. By trying exercises from the book or equations of your own, you will see that the fast convergence to $\sqrt{4}$ is very typical. The function can be much more complicated than $x^2 - 4$ (in practice it certainly is). The iteration for $2x = \cos x$ was in the previous section, and the error was squared at every step. If Newton's method starts close to x^*, its convergence is overwhelming. That has to be the main point of this section: ***Follow the tangent line.***

Instead of those good functions, may I stay with this strange example $x^2 + 1 = 0$? It is not so predictable, and maybe not so important, but somehow it is more interesting. There is no real solution x^*, and Newton's method $x_{n+1} = \frac{1}{2}(x_n - 1/x_n)$ bounces around. We will now discover x_n.

A FORMULA FOR x_n

The key is an exercise from trigonometry books. Most of those problems just give practice with sines and cosines, but this one exactly fits $\frac{1}{2}(x_n - 1/x_n)$:

$$\frac{1}{2}\left(\frac{\cos\theta}{\sin\theta} - \frac{\sin\theta}{\cos\theta}\right) = \frac{\cos 2\theta}{\sin 2\theta} \quad \text{or} \quad \frac{1}{2}\left(\cot\theta - \frac{1}{\cot\theta}\right) = \cot 2\theta$$

In the left equation, the common denominator is $2\sin\theta\cos\theta$ (which is $\sin 2\theta$). The numerator is $\cos^2\theta - \sin^2\theta$ (which is $\cos 2\theta$). Replace cosine/sine by cotangent, and the identity says this:

If $x_0 = \cot\theta$ then $x_1 = \cot 2\theta$. Then $x_2 = \cot 4\theta$. Then $x_n = \cot 2^n\theta$.

This is the formula. *Our points are on the cotangent curve.* Figure 3.23 starts from $x_0 = 2 = \cot\theta$, and every iteration doubles the angle.

Example A The sequence $x_0 = 1, x_1 = 0, x_2 = \infty$ matches the cotangents of $\pi/4$, $\pi/2$, and π. This sequence blows up because x_2 has a division by $x_1 = 0$.

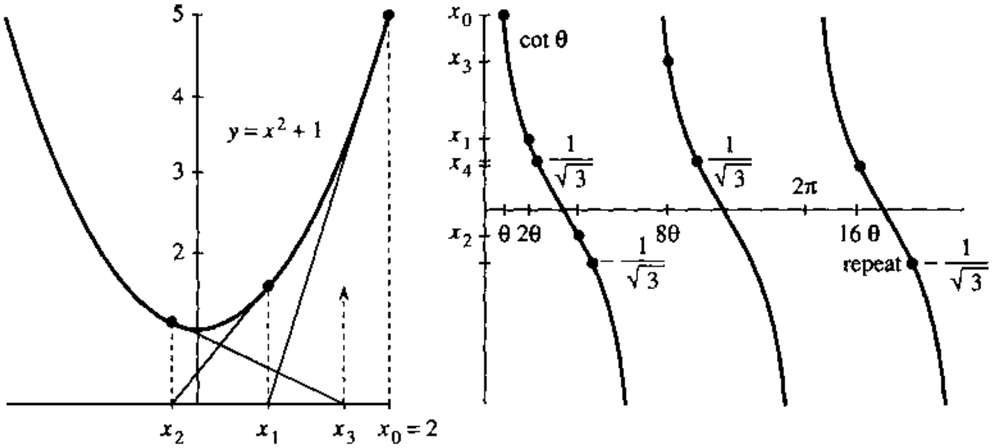


Fig. 3.23 Newton's method for $x^2 + 1 = 0$. Iteration gives $x_n = \cot 2^n\theta$.

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 $4\theta, 8\theta, \ldots$ 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