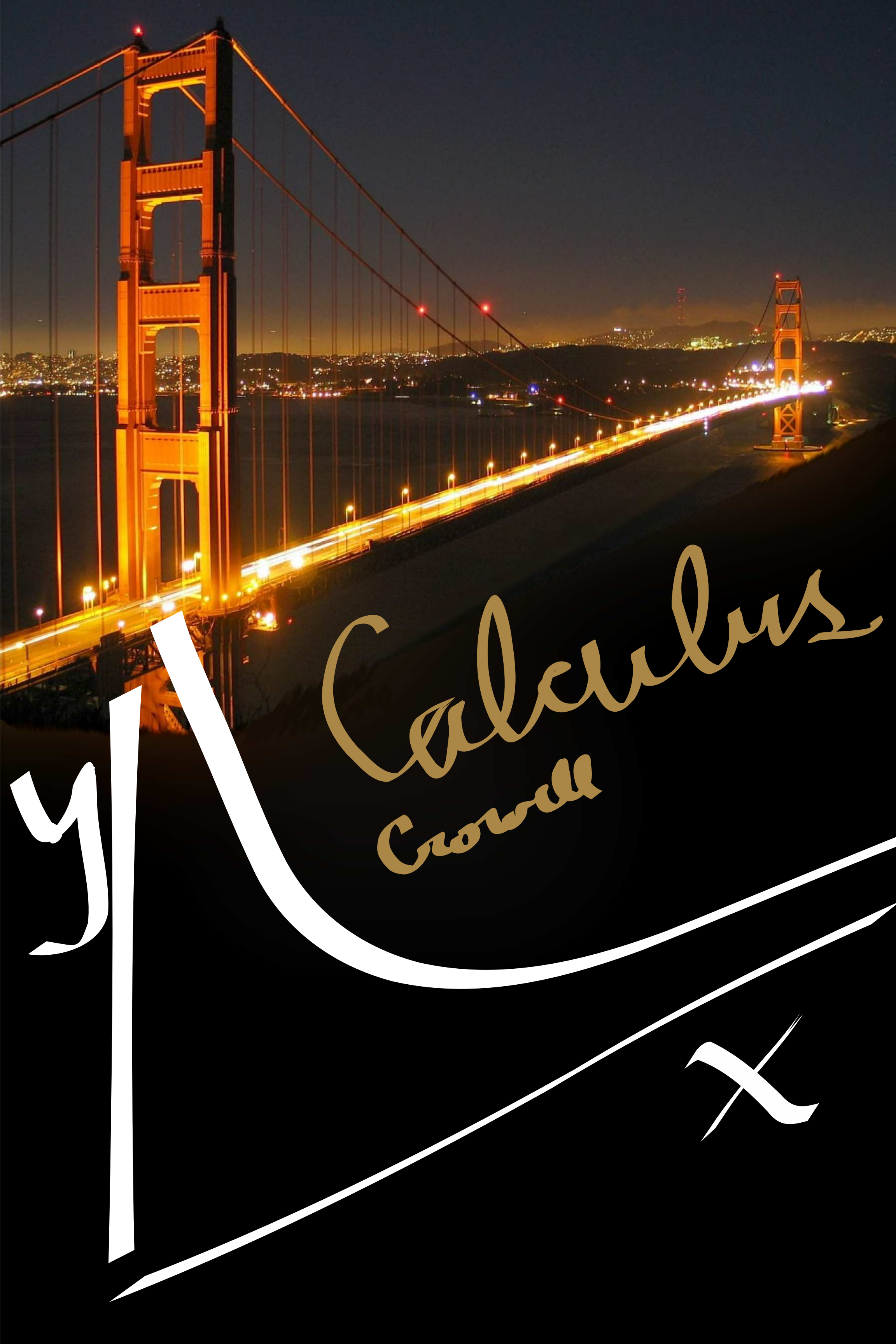

Calculus
Crowell
y
x

Calculus

Benjamin Crowell

Light and Matter
Fullerton, California
www.lightandmatter.com

rev. April 15, 2009

1 Rates of Change

2 To infinity — and beyond!

3 Integration

4 Techniques

5 Improper integrals

6 Sequences and Series

7 Complex number techniques

Preface

Calculus isn't a hard subject.

Algebra is hard. I still remember my encounter with algebra. It was my first taste of abstraction in mathematics, and it gave me quite a few black eyes and bloody noses.

Geometry is hard. For most people, geometry is the first time they have to do proofs using formal, axiomatic reasoning.

I teach physics for a living. Physics is hard. There's a reason that people believed Aristotle's bogus version of physics for centuries: it's because the real laws of physics are counterintuitive.

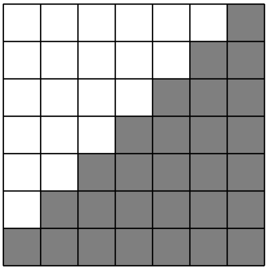
Calculus, on the other hand, is a very straightforward subject that rewards intuition, and can be easily visualized. Silvanus Thompson, author of one of the most popular calculus texts ever written, opined that "considering how many fools can calculate, it is surprising that it should be thought either a difficult or a tedious task for any other fool to master the same tricks."

Since I don't teach calculus, I can't require anyone to read this book. For that reason, I've written it so that you can go through it and get to the dessert course without having to eat too many Brussels sprouts and Lima beans along the way. The development of any mathematical subject involves a large number of boring details that have little to do with the main thrust of the topic. These details I've relegated to a chapter in the back of the book, and the reader who has an interest in mathematics as a career — or who enjoys a nice heavy pot roast before moving on to dessert — will want to read those details when the main text suggests the possibility of a detour.

1 Rates of Change

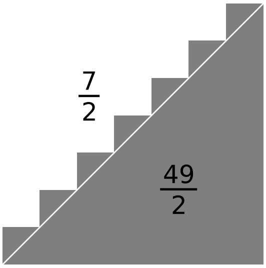
1.1 Change in discrete steps

Toward the end of the eighteenth century, a German elementary school teacher decided to keep his pupils busy by assigning them a long, boring arithmetic problem. To oversimplify a little bit (which is what textbook authors always do when they tell you about history), I'll say that the assignment was to add up all the numbers from one to a hundred. The children set to work on their slates, and the teacher lit his pipe, confident of a long break. But almost immediately, a boy named Carl Friedrich Gauss brought up his answer: 5,050.



a / Adding the numbers from 1 to 7.

Figure a suggests one way of solving this type of problem. The filled-in columns of the graph represent the numbers from 1 to 7, and adding them up means finding the area of the shaded region. Roughly half the square is shaded in, so if we want only an approximate solution, we can simply calculate $7^2/2 = 24.5$.



b / A trick for finding the sum.

But, as suggested in figure b, it's not much more work to get an exact result. There are seven sawteeth sticking out out above the diagonal, with a total area of 7/2, so the total shaded area is $(7^2 + 7)/2 = 28$. In general, the sum of the first n numbers will be $(n^2 + n)/2$, which explains Gauss's result: $(100^2 + 100)/2 = 5,050$.

Two sides of the same coin

Problems like this come up frequently. Imagine that each household in a certain small town sends a total of one ton of garbage to the dump every year. Over time, the garbage accumulates in the dump, taking up more and more space.

c / Carl Friedrich Gauss (1777-1855), a long time after graduating from elementary school.

Let's label the years as $n = 1$, 2, 3,..., and let the function[1] $x(n)$ represent the amount of garbage that has accumulated by the end of year n. If the population is constant, say 13 households, then garbage accumulates at a constant rate, and we have $x(n) = 13n$.

But maybe the town's population is growing. If the population starts out as 1 household in year 1, and then grows to 2 in year 2, and so on, then we have the same kind of problem that the young Gauss solved. After 100 years, the accumulated amount of garbage will be 5,050 tons. The pile of refuse grows more and more every year; the rate of change of x is not constant. Tabulating the examples we've done so far, we have this:

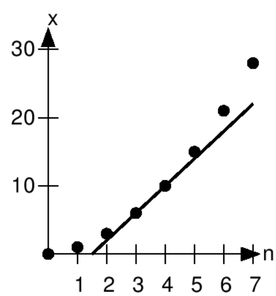
[1] Recall that when x is a function, the notation $x(n)$ means the output of the function when the input is n. It doesn't represent multiplication of a number x by a number n.

rate of change	*accumulated result*
13	$13n$
n	$(n^2+n)/2$

The rate of change of the function x can be notated as $\dot{x}$. Given the function $\dot{x}$, we can always determine the function x for any value of n by doing a running sum.

Likewise, if we know x, we can determine $\dot{x}$ by subtraction. In the example where $x = 13n$, we can find $\dot{x} = x(n) - x(n-1) = 13n - 13(n-1) = 13$. Or if we knew that the accumulated amount of garbage was given by $(n^2+n)/2$, we could calculate the town's population like this:

$$\frac{n^2+n}{2} - \frac{(n-1)^2+(n-1)}{2}$$
$$= \frac{n^2+n-\left(n^2-2n+1+n-1\right)}{2}$$
$$= n$$



d / $\dot{x}$ is the slope of x.

The graphical interpretation of

this is shown in figure d: on a graph of $x = (n^2 + n)/2$, the slope of the line connecting two successive points is the value of the function $\dot{x}$.

In other words, the functions x and $\dot{x}$ are like different sides of the same coin. If you know one, you can find the other — with two caveats.

First, we've been assuming implicitly that the function x starts out at $x(0) = 0$. That might not be true in general. For instance, if we're adding water to a reservoir over a certain period of time, the reservoir probably didn't start out completely empty. Thus, if we know $\dot{x}$, we can't find out everything about x without some further information: the starting value of x. If someone tells you $\dot{x} = 13$, you can't conclude $x = 13n$, but only $x = 13n + c$, where c is some constant. There's no such ambiguity if you're going the opposite way, from x to $\dot{x}$. Even if $x(0) \neq 0$, we still have $\dot{x} = 13n + c - [13(n - 1) + c] = 13$.

Second, it may be difficult, or even impossible, to find a *formula* for the answer when we want to determine the running sum x given a formula for the rate of change $\dot{x}$. Gauss had a flash of insight that led him to the result $(n^2 + n)/2$, but in general we might only be able to use a computer spreadsheet to calculate a number for the running sum, rather than an equation that would be valid for all values of n.

Some guesses

Even though we lack Gauss's genius, we can recognize certain patterns. One pattern is that if $\dot{x}$ is a function that gets bigger and bigger, it seems like x will be a function that grows even faster than $\dot{x}$. In the example of $\dot{x} = n$ and $x = (n^2 + n)/2$, consider what happens for a large value of n, like 100. At this value of n, $\dot{x} = 100$, which is pretty big, but even without pawing around for a calculator, we know that x is going to turn out really *really* big. Since n is large, n^2 is quite a bit bigger than n, so roughly speaking, we can approximate $x \approx n^2/2 = 5,000$. 100 may be a big number, but 5,000 is a lot bigger. Continuing in this way, for $n = 1000$ we have $\dot{x} = 1000$, but $x \approx 500,000$ — now x has far outstripped $\dot{x}$. This can be a fun game to play with a calculator: look at which functions grow the fastest. For instance, your calculator might have an x^2 button, an e^x button, and a button for $x!$ (the factorial function, defined as $x! = 1 \cdot 2 \cdot \ldots \cdot x$, e.g., $4! = 1 \cdot 2 \cdot 3 \cdot 4 = 24$). You'll find that 50^2 is pretty big, but e^{50} is incomparably greater, and 50! is so big that it causes an error.

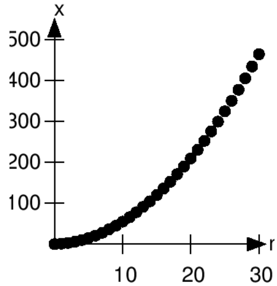
All the x and $\dot{x}$ functions we've seen so far have been polynomials. If x is a polynomial, then of course we can find a polynomial for $\dot{x}$ as well, because if x is a polynomial, then $x(n)-x(n-1)$ will be one too. It also looks like every polynomial we could choose for $\dot{x}$ might also correspond to an x that's a polynomial. And not only that, but it looks as though there's a pattern in the power of n. Suppose x is a polynomial, and the highest power of n it contains is a certain number — the "order" of the polynomial. Then $\dot{x}$ is a polynomial of that order minus one. Again, it's fairly easy to prove this going one way, passing from x to $\dot{x}$, but more difficult to prove the opposite relationship: that if $\dot{x}$ is a polynomial of a certain order, then x must be a polynomial with an order that's greater by one.

We'd imagine, then, that the running sum of $\dot{x} = n^2$ would be a polynomial of order 3. If we calculate $x(100) = 1^2 + 2^2 + \ldots + 100^2$ on a computer spreadsheet, we get 338,350, which looks suspiciously close to $1,000,000/3$. It looks like $x(n) = n^3/3 + \ldots$, where the dots represent terms involving lower powers of n such as n^2.

e / Isaac Newton (1643-1727)

1.2 Continuous change

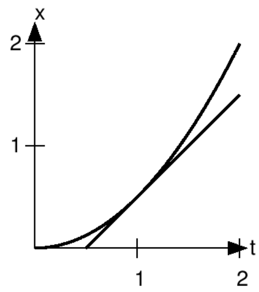
Did you notice that I sneaked something past you in the example of water filling up a reservoir? The x and $\dot{x}$ functions I've been using as examples have all been functions defined on the integers, so they represent change that happens in discrete steps, but the flow of water into a reservoir is smooth and continuous. Or is it? Water is made out of molecules, after all. It's just that water molecules are so small that we don't notice them as individuals. Figure f shows a graph that is discrete, but almost appears continuous because the scale has been chosen so that the points blend together visually.



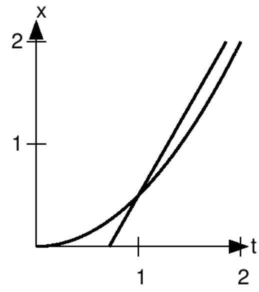
f / On this scale, the graph of $(n^2+n)/2$ appears almost continuous.

The physicist Isaac Newton started thinking along these lines in the

1660's, and figured out ways of analyzing x and $\dot{x}$ functions that were truly continuous. The notation $\dot{x}$ is due to him (and he only used it for continuous functions). Because he was dealing with the continuous *flow* of change, he called his new set of mathematical techniques the method of *fluxions*, but nowadays it's known as the calculus.



g / The function $x(t) = t^2/2$, and its tangent line at the point $(1, 1/2)$.



h / This line isn't a tangent line: it crosses the graph.

Newton was a physicist, and he needed to invent the calculus as part of his study of how objects move. If an object is moving in one dimension, we can specify its position with a variable x, and x will then be a function of time, t. The rate of change of its position, $\dot{x}$, is its speed, or velocity. Earlier experiments by Galileo had established that when a ball rolled down a slope, its position was proportional to t^2, so Newton inferred that a graph like figure g would be typical for any object moving under the influence of a constant force. (It could be $7t^2$, or $t^2/42$, or anything else proportional to t^2, depending on the force acting on the object and the object's mass.)

Because the functions are continuous, not discrete, we can no longer define the relationship between x and $\dot{x}$ by saying x is a running sum of $\dot{x}$'s, or that $\dot{x}$ is the difference between two successive x's. But we already found a geometrical relationship between the two functions in the discrete case, and that can serve as our definition for the continuous case: x is the area under the graph of $\dot{x}$, or, if you like, $\dot{x}$ is the slope of the tangent line on the graph of x. For now we'll concentrate on the slope idea.

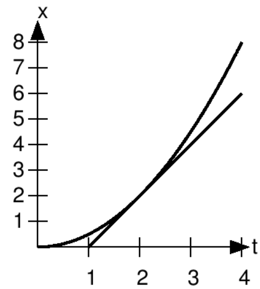
The tangent line is defined as the line that passes through the graph at a certain point, but, unlike the one in figure h, doesn't cut across the graph.[2] By measuring with a ruler on figure g, we find that the slope is very close to 1, so evi-

[2]For a more formal definition, see page 121.

dently $\dot{x}(1) = 1$. To prove this, we construct the function representing the line: $\ell(t) = t - 1/2$. We want to prove that this line doesn't cross the graph of $x(t) = t^2/2$. The difference between the two functions, $x - \ell$, is the polynomial $t^2/2 - t + 1/2$, and this polynomial will be zero for any value of t where the line touches or crosses the curve. We can use the quadratic formula to find these points, and the result is that there is only one of them, which is $t = 1$. Since $x - \ell$ is positive for at least some points to the left and right of $t = 1$, and it only equals zero at $t = 1$, it must never be negative, which means that the line always lies below the curve, never crossing it.

A derivative

That proves that $\dot{x}(1) = 1$, but it was a lot of work, and we don't want to do that much work to evaluate $\dot{x}$ at every value of t. There's a way to avoid all that, and find a formula for $\dot{x}$. Compare figures g and i. They're both graphs of the same function, and they both look the same. What's different? The only difference is the scales: in figure i, the t axis has been shrunk by a factor of 2, and the x axis by a factor of 4. The graph looks the same, because doubling t quadruples $t^2/2$. The tangent line here is the tangent line at $t = 2$, not $t = 1$, and although it looks like the same line as the one in figure g, it isn't, because the scales are different. The line in figure g had a slope of rise/run $= 1/1 = 1$, but this one's slope is $4/2 = 2$. That means $\dot{x}(2) = 2$. In general, this scaling argument shows that $\dot{x}(t) = t$ for any t.



i / The function $t^2/2$ again. How is this different from figure g?

This is called *differentiating*: finding a formula for the function $\dot{x}$, given a formula for the function x. The term comes from the idea that for a discrete function, the slope is the difference between two successive values of the function. The function $\dot{x}$ is referred to as the *derivative* of the function x, and the art of differentiating is differential calculus. The opposite process, computing a formula for x when given $\dot{x}$, is called integrating, and makes up the field of integral calculus; this terminology is based on the idea that computing a running sum is like putting together (integrating) many little pieces.

Note the similarity between this result for continuous functions,

$$x = t^2/2 \qquad \dot{x} = t \qquad ,$$

and our earlier result for discrete ones,

$$x = (n^2 + n)/2 \qquad \dot{x} = n \qquad .$$

The similarity is no coincidence. A continuous function is just a smoothed-out version of a discrete one. For instance, the continuous version of the staircase function shown in figure b on page 9 would simply be a triangle without the saw teeth sticking out; the area of those ugly sawteeth is what's represented by the $n/2$ term in the discrete result $x = (n^2 + n)/2$, which is the only thing that makes it different from the continuous result $x = t^2/2$.

Properties of the derivative

It follows immediately from the definition of the derivative that multiplying a function by a constant multiplies its derivative by the same constant, so for example since we know that the derivative of $t^2/2$ is t, we can immediately tell that the derivative of t^2 is $2t$, and the derivative of $t^2/17$ is $2t/17$.

Also, if we add two functions, their derivatives add. To give a good example of this, we need to have another function that we can differentiate, one that isn't just some multiple of t^2. An easy one is t: the derivative of t is 1, since the graph of $x = t$ is a line with a slope of 1, and the tangent line lies right on top of the original line.

The derivative of a constant is zero, since a constant function's graph is a horizontal line, with a slope of zero. We now know enough to differentiate a second-order polynomial.

Example 1

The derivative of $5t^2 + 2t$ is the derivative of $5t^2$ plus the derivative of $2t$, since derivatives add. The derivative of $5t^2$ is 5 times the derivative of t^2, and the derivative of $2t$ is 2 times the derivative of t, so putting everything together, we find that the derivative of $5t^2 + 2t$ is $(5)(2t) + (2)(1) = 10t + 2$.

Example 2

▷ An insect pest from the United States is inadvertently released in a village in rural China. The pests spread outward at a rate of s kilometers per year, forming a widening circle of contagion. Find the number of square kilometers per year that become newly infested. Check that the units of the result make sense. Interpret the result.

▷ Let t be the time, in years, since the pest was introduced. The radius of the circle is $r = st$, and its area is $a = \pi r^2 = \pi(st)^2$. To make this look like a polynomial, we have to rewrite this as $a = (\pi s^2)t^2$. The derivative is

$$\dot{a} = (\pi s^2)(2t)$$
$$\dot{a} = (2\pi s^2)t$$

The units of s are km/year, so squaring it gives $\text{km}^2/\text{year}^2$. The 2 and the

π are unitless, and multiplying by t gives units of km^2/year, which is what we expect for $\dot{a}$, since it represents the number of square kilometers per year that become infested.

Interpreting the result, we notice a couple of things. First, the rate of infestation isn't constant; it's proportional to t, so people might not pay so much attention at first, but later on the effort required to combat the problem will grow more and more quickly. Second, we notice that the result is proportional to s^2. This suggests that anything that could be done to reduce s would be very helpful. For instance, a measure that cut s in half would reduce $\dot{a}$ by a factor of four.

Higher-order polynomials

So far, we have the following results for polynomials up to order 2:

function	*derivative*
1	0
t	1
t^2	$2t$

Interpreting 1 as t^0, we detect what seems to be a general rule, which is that the derivative of t^k is kt^{k-1}. The proof is straightforward but not very illuminating if carried out with the methods developed in this chapter, so I've relegated it to page 122. It can be proved much more easily using the methods of chapter 2.

Example 3

▷ If $x = 2t^7 - 4t + 1$, find $\dot{x}$.

▷ This is similar to example 1, the only difference being that we can now handle higher powers of t. The derivative of t^7 is $7t^6$, so we have

$$\begin{aligned} \dot{x} &= (2)(7t^6) + (-4)(1) + 0 \\ &= 14t^6 + -4 \end{aligned}$$

The second derivative

I described how Galileo and Newton found that an object subject to an external force, starting from rest, would have a velocity $\dot{x}$ that was proportional to t, and a position x that varied like t^2. The proportionality constant for the velocity is called the acceleration, a, so that $\dot{x} = at$ and $x = at^2/2$. For example, a sports car accelerating from a stop sign would have a large acceleration, and its velocity at at a given time would therefore be a large number. The acceleration can be thought of as the derivative of the derivative of x, written $\ddot{x}$, with two dots. In our example, $\ddot{x} = a$. In general, the acceleration doesn't need to be constant. For example, the sports car will eventually have to stop accelerating, perhaps because the backward force of air friction becomes as great as the force pushing it forward. The total force acting on the car would then be zero, and the car would continue in motion at a constant speed.

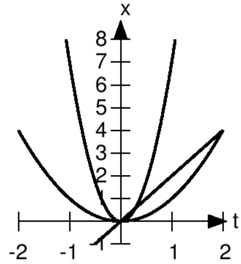
Example 4

Suppose the pilot of a blimp has just

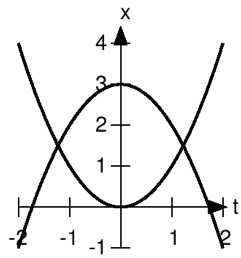
turned on the motor that runs its propeller, and the propeller is spinning up. The resulting force on the blimp is therefore increasing steadily, and let's say that this causes the blimp to have an acceleration $\ddot{x} = 3t$, which increases steadily with time. We want to find the blimp's velocity and position as functions of time.

For the velocity, we need a polynomial whose derivative is $3t$. We know that the derivative of t^2 is $2t$, so we need to use a function that's bigger by a factor of $3/2$: $\dot{x} = (3/2)t^2$. In fact, we could add any constant to this, and make it $\dot{x} = (3/2)t^2 + 14$, for example, where the 14 would represent the blimp's initial velocity. But since the blimp has been sitting dead in the air until the motor started working, we can assume the initial velocity was zero. Remember, any time you're working backwards like this to find a function whose derivative is some other function (integrating, in other words), there is the possibility of adding on a constant like this.

Finally, for the position, we need something whose derivative is $(3/2)t^2$. The derivative of t^3 would be $3t^2$, so we need something half as big as this: $x = t^3/2$.



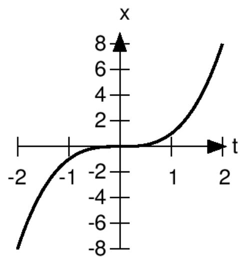
j / The functions $2t$, t^2 and $7t^2$.



k / The functions t^2 and $3 - t^2$.

The second derivative can be interpreted as a measure of the curvature of the graph, as shown in figure j. The graph of the function $x = 2t$ is a line, with no curvature. Its first derivative is 2, and its second derivative is zero. The function t^2 has a second derivative of 2, and the more tightly curved function $7t^2$ has a bigger second derivative, 14.

Positive and negative signs of the second derivative indicate concavity. In figure k, the function t^2 is like a cup with its mouth pointing up. We say that it's "concave up," and this corresponds to its positive second derivative. The function $3-t^2$, with a second derivative less than zero, is concave down. Another way of saying it is that if you're driving along a road shaped like t^2, going in the direction of increasing t, then your steering wheel is turned to the left, whereas on a road shaped like $3 - t^2$ it's turned to the right.



l / The functions t^3 has an inflection point at $t = 0$.

Figure l shows a third possibility. The function t^3 has a derivative $3t^2$, which equals zero at $t = 0$. This called a point of inflection. The concavity of the graph is down on the left, up on the right. The inflection point is where it switches from one concavity to the other. In the alternative description in terms of the steering wheel, the inflection point is where your steering wheel is crossing from left to right.

1.3 Applications

Maxima and minima

When a function goes up and then smoothly turns around and comes back down again, it has zero slope at the top. A place where $\dot{x} = 0$, then, could represent a place where x was at a maximum. On the other hand, it could be concave up, in which case we'd have a minimum.

Example 5

▷ Fred receives a mysterious e-mail tip telling him that his investment in a certain stock will have a value given by $x = -2t^4 + (6.4577 \times 10^{10})t$, where $t \geq 2005$ is the year. Should he sell at some point? If so, when?

▷ If the value reaches a maximum at some time, then the derivative should be zero then. Taking the derivative and setting it equal to zero, we have

$$0 = -8t^3 + 6.4577 \times 10^{10}$$
$$t = \left(\frac{6.4577 \times 10^{10}}{8}\right)^{1/3}$$
$$t = \pm 2006.0 \qquad .$$

Obviously the solution at $t = -2006.0$ is bogus, since the stock market didn't exist four thousand years ago, and the tip only claimed the function would be valid for $t \geq 2005$.

Should Fred sell on New Year's eve of 2006?

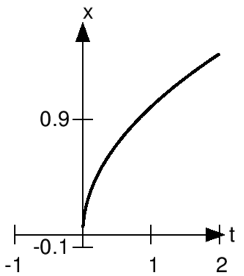
But this could be a maximum, a minimum, or an inflection point. Fred definitely does *not* want to sell at $t = 2006$ if it's a minimum! To check which of the three possibilities hold, Fred takes the second derivative:

$$\ddot{x} = -24t^2 \qquad .$$

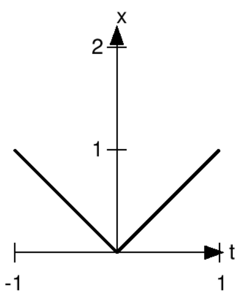
Plugging in $t = 2006.0$, we find that the second derivative is negative at that time, so it is indeed a maximum.

Implicit in this whole discussion was the assumption that the maximum or minimum where the function was smooth. There are some other possibilities.

In figure m, the function's minimum occurs at an end-point of its domain.



m / The function $x = \sqrt{t}$ has a minimum at $t = 0$, which is not a place where $\dot{x} = 0$. This point is the edge of the function's domain.



n / The function $x = |t|$ has a minimum at $t = 0$, which is not a place where $\dot{x} = 0$. This is a point where the function isn't differentiable.

Another possibility is that the function can have a minimum or maximum at some point where its derivative isn't well defined. Figure n shows such a situation. There is a kink in the function at $t = 0$, so a wide variety of lines could be placed through the graph there, all with different slopes and all staying on one side of the graph. There is no uniquely defined tangent line, so the derivative is undefined.

Example 6

▷ Rancher Rick has a length of cyclone fence L with which to enclose a rectangular pasture. Show that he can enclose the greatest possible area by forming a square with sides of length $L/4$.

▷ If the width and length of the rectangle are t and u, and Rick is going to use up all his fencing material, then the perimeter of the rectangle, $2t + 2u$, equals L, so for a given width, t, the length is $u = L/2 - t$. The area is $a = tu = t(L/2 - t)$. The function only means anything realistic for $0 \le t \le L/2$, since for values of t outside this region either the width or the height of the rectangle would be negative. The function $a(t)$ could therefore have a maximum either at a place where $\dot{a} = 0$, or at the endpoints of the function's domain. We can eliminate the latter possibility, because the area is zero at the endpoints.

To evaluate the derivative, we first need to reexpress a as a polynomial:

$$a = -t^2 + \frac{L}{2}t \quad .$$

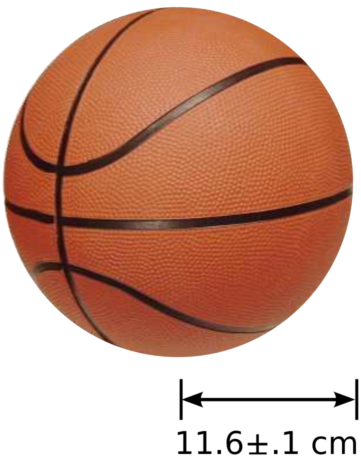
The derivative is

$$\dot{a} = -2t + \frac{L}{2} \quad .$$

Setting this equal to zero, we find $t = L/4$, as claimed. This is a maximum, not a minimum or an inflection point, because the second derivative is the constant $\ddot{a} = -2$, which is negative for all t, including $t = L/4$.

Propagation of errors

The Women's National Basketball Association says that balls used in its games should have a radius of 11.6 cm, with an allowable range of error of plus or minus 0.1 cm (one millimeter). How accurately can we determine the ball's volume?



o / How accurately can we determine the ball's volume?

The equation for the volume of a sphere gives $V = (4/3)\pi r^3 = 6538\ \text{cm}^3$ (about six and a half liters). We have a function $V(r)$, and we want to know how much of an effect will be produced on the function's output V if its input r is changed by a certain small amount. Since the amount by which r can be changed is small compared to r, it's reasonable to take the tangent line as an approximation to the actual graph. The slope of the tangent line is the derivative of V, which is $4\pi r^2$. (This is the ball's surface area.) Setting (slope) = (rise)/(run) and solving for the rise, which represents the change in V, we find that it could be off by as much as $(4\pi r^2)(0.1\ \text{cm}) = 170\ \text{cm}^3$. The volume of the ball can therefore be expressed as $6500 \pm 170\ \text{cm}^3$, where the original figure of 6538 has been rounded off to the nearest hundred in order to avoid creating the impression that the 3 and the 8 actually mean anything — they clearly don't, since the possible error is out in the hundreds' place.

This calculation is an example of a very common situation that occurs in the sciences, and even in everyday life, in which we base a calculation on a number that has some range of uncertainty in it, causing a corresponding range of uncertainty in the final result. This is called propagation of errors. The idea is that the derivative expresses how sensitive the function's output is to its input.

The example of the basketball could also have been handled without calculus, simply by recalculating the volume using a radius that was raised from 11.6 to 11.7 cm, and finding the difference between the two volumes. Understanding it in terms of calculus, however, gives us a different way of getting at the same ideas, and often allows us to understand more deeply what's going on. For example, we noticed in passing that the derivative of the volume was simply the surface area of the ball, which provides a nice geometric visualization. We can imagine inflating the ball so that its radius is increased by a millimeter. The amount of added volume equals the surface area of the ball multiplied by one millimeter, just as the amount of volume added to the world's oceans by global warming equals the oceans' surface area multiplied by the added depth.

For an example of an insight that we would have missed if we hadn't applied calculus, consider how much error is incurred in the measurement of the width of a book if the ruler is placed on the book at a slightly incorrect angle, so that it doesn't form an angle of exactly 90 degrees with spine. The measurement has its minimum (and correct) value if the ruler is placed at exactly 90 degrees. Since the function has a minimum at this angle, its derivative is zero. That means that we expect essentially no error in the measurement if the ruler's angle is just a tiny bit off. This gives us the insight that it's not worth fiddling excessively over the angle in this measurement. Other sources of error will be more important. For example, is the book a uniform rectangle? Are we using the worn end of the ruler as its zero, rather than letting the ruler hang over both sides of the book and subtracting the two measurements?

Problems

1 Graph the function t^2 in the neighborhood of $t = 3$, draw a tangent line, and use its slope to verify that the derivative equals $2t$ at this point. ▷ Solution, p. 146

2 Graph the function $\sin e^t$ in the neighborhood of $t = 0$, draw a tangent line, and use its slope to estimate the derivative. Answer: 0.5403023058. (You will of course not get an answer this precise using this technique.)
▷ Solution, p. 146

3 Differentiate the following functions with respect to t: 1, 7, t, $7t$, t^2, $7t^2$, t^3, $7t^3$.
▷ Solution, p. 147

4 Differentiate $3t^7 - 4t^2 + 6$ with respect to t. ▷ Solution, p. 147

5 Differentiate $at^2 + bt + c$ with respect to t.
▷ Solution, p. 147 [Thompson, 1919]

6 Find two different functions whose derivatives are the constant 3, and give a geometrical interpretation. ▷ Solution, p. 147

7 Find a function x whose derivative is $\dot{x} = t^7$. In other words, integrate the given function. ▷ Solution, p. 148

8 Find a function x whose derivative is $\dot{x} = 3t^7$. In other words, integrate the given function. ▷ Solution, p. 148

9 Find a function x whose derivative is $\dot{x} = 3t^7 - 4t^2 + 6$. In other words, integrate the given function. ▷ Solution, p. 148

10 Let t be the time that has elapsed since the Big Bang. In that time, light, traveling at speed c, has been able to travel a maximum distance ct. The portion of the universe that we can observe is therefore a sphere of radius ct, with volume $v = (4/3)\pi r^3 = (4/3)\pi(ct)^3$. Compute the rate $\dot{v}$ at which the observable universe is expanding, and check that your answer has the right units, as in example 2 on page 15. ▷ Solution, p. 148

11 Kinetic energy is a measure of an object's quantity of motion; when you buy gasoline, the energy you're paying for will be converted into the car's kinetic energy (actually only some of it, since the engine isn't perfectly efficient). The kinetic energy of an object with mass m and velocity v is given by $K = (1/2)mv^2$. For a car accelerating at a steady rate, with $v = at$, find the rate $\dot{K}$ at which the engine is required to put out kinetic energy. $\dot{K}$, with units of energy over time, is known as the *power*. Check that your answer has the right units, as in example 2 on page 15. ▷ Solution, p. 148

12 A metal square expands and contracts with temperature, the lengths of its sides varying according to the equation $\ell = (1 + \alpha T)\ell_o$. Find the rate of change of its surface area with respect to temperature. That is, find $\dot{\ell}$, where the variable with respect to which you're differentiating is the temperature, T. Check that your answer has the right units, as in example 2 on page 15.

▷ Solution, p. 149

13 Find the second derivative of $2t^3 - t$. ▷ Solution, p. 149

14 Locate any points of inflection of the function $t^3 + t^2$. Verify by graphing that the concavity of the function reverses itself at this point. ▷ Solution, p. 149

15 Let's see if the rule that the derivative of t^k is kt^{k-1} also works for $k < 0$. Use a graph to test one particular case, choosing one particular negative value of k, and one particular value of t. If it works, what does that tell you about the rule? If it doesn't work?

▷ Solution, p. 149

16 Two atoms will interact via electrical forces between their protons and electrons. To put them at a distance r from one another (measured from nucleus to nucleus), a certain amount of energy E is required, and the minimum energy occurs when the atoms are in equilibrium, forming a molecule. Often a fairly good approximation to the energy is the Lennard-Jones expression

$$E(r) = k\left[\left(\frac{a}{r}\right)^{12} - 2\left(\frac{a}{r}\right)^{6}\right] \quad ,$$

where k and a are constants. Note that, as proved in chapter 2, the rule that the derivative of t^k is kt^{k-1} also works for $k < 0$. Show that there is an equilibrium at $r = a$. Verify (either by graphing or by testing the second derivative) that this is a minimum, not a maximum or a point of inflection.

▷ Solution, p. 151

17 Prove that the total number of maxima and minima possessed by a third-order polynomial is at most two. ▷ Solution, p. 152

18 Euclid proved that the volume of a pyramid equals $(1/3)bh$, where b is the area of its base, and h its height. A pyramidal tent without tent-poles is erected by blowing air into it under pressure. The area of the base is easy to measure accurately, because the base is nailed down, but the height fluctuates somewhat and is hard to measure accurately. If the amount of uncertainty in the measured height is plus or minus e_h, find the amount of possible error e_V in the volume. ▷ Solution, p. 152

19 A hobbyist is going to measure the height to which her model rocket rises at the peak of its trajectory. She plans to take a digital photo from far away and then do trigonometry to determine the height, given the baseline from the launchpad to the camera and the angular height of the rocket as determined from analysis of the photo. Comment on the error incurred by the inability to snap the photo at exactly the right moment.

▷ Solution, p. 152

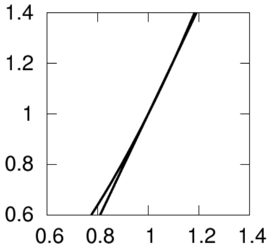
2 To infinity — and beyond!

a / Gottfried Leibniz (1646-1716)

Little kids readily pick up the idea of infinity. "When I grow up, I'm gonna have a million Barbies." "Oh yeah? Well, I'm gonna have a billion." "Well, I'm gonna have infinity Barbies." "So what? I'll have two infinity of them." Adults laugh, convinced that infinity, ∞, is the biggest number, so 2∞ can't be any bigger. This is the idea behind a joke in the movie Toy Story. Buzz Lightyear's slogan is "To infinity — and beyond!" We assume there *isn't* any beyond. Infinity is supposed to be the biggest there is, so by definition there can't be anything bigger, right?

2.1 Infinitesimals

Actually mathematicians have invented several many different logical systems for working with infinity, and in most of them infinity does come in different sizes and flavors. Newton, as well as the German mathematician Leibniz who invented calculus independently,[1] had a strong intuitive idea that calculus was really about numbers that were infinitely small: infinitesimals, the opposite of infinities. For instance, consider the number $1.1^2 = 1.21$. That 2 in the first decimal place is the same 2 that appears in the expression $2t$ for the derivative of t^2.



b / A close-up view of the function $x = t^2$, showing the line that connects the points $(1, 1)$ and $(1.1, 1.21)$.

[1] There is some dispute over this point. Newton and his supporters claimed that Leibniz plagiarized Newton's ideas, and merely invented a new notation for them.

Figure b shows the idea visually. The line connecting the points $(1, 1)$ and $(1.1, 1.21)$ is almost indistinguishable from the tangent line on this scale. Its slope is $(1.21 - 1)/(1.1 - 1) = 2.1$, which is very close to the tangent line's slope of 2. It was a good approximation because the points were close together, separated by only 0.1 on the t axis.

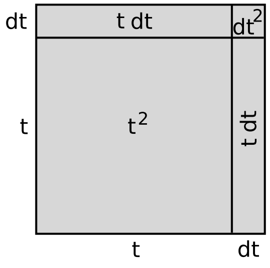
If we needed a better approximation, we could try calculating $1.01^2 = 1.0201$. The slope of the line connecting the points $(1, 1)$ and $(1.01, 1.0201)$ is 2.01, which is even closer to the slope of the tangent line.

Another method of visualizing the idea is that we can interpret $x = t^2$ as the area of a square with sides of length t, as suggested in figure c. We increase t by an infinitesimally small number $\mathrm{d}t$. The d is Leibniz's notation for a very small difference, and $\mathrm{d}t$ is to be read is a single symbol, "dee-tee," not as a number d multiplied by a number t. The idea is that $\mathrm{d}t$ is smaller than any ordinary number you could imagine, but it's not zero. The area of the square is increased by $\mathrm{d}x = 2t\mathrm{d}t + \mathrm{d}t^2$, which is analogous to the finite numbers 0.21 and 0.0201 we calculated earlier. Where before we divided by a finite change in t such as 0.1 or 0.01, now we divide by $\mathrm{d}t$, producing

$$\begin{aligned}\frac{\mathrm{d}x}{\mathrm{d}t} &= \frac{2t\,\mathrm{d}t + \mathrm{d}t^2}{\mathrm{d}t} \\ &= 2t + \mathrm{d}t\end{aligned}$$

for the derivative. On a graph like figure b, $\mathrm{d}x/\mathrm{d}t$ is the slope of the tangent line: the change in x divided by the changed in t.

But adding an infinitesimal number $\mathrm{d}t$ onto $2t$ doesn't really change it by any amount that's even theoretically measurable in the real world, so the answer is really $2t$. Evaluating it at $t = 1$ gives the exact result, 2, that the earlier approximate results, 2.1 and 2.01, were getting closer and closer to.



c / A geometrical interpretation of the derivative of t^2.

Example 7

To show the power of infinitesimals and the Leibniz notation, let's prove that the derivative of t^3 is $3t^2$:

$$\begin{aligned}\frac{\mathrm{d}x}{\mathrm{d}t} &= \frac{(t+\mathrm{d}t)^3 - t^3}{\mathrm{d}t} \\ &= \frac{3t^2\,\mathrm{d}t + 3t\,\mathrm{d}t^2 + \mathrm{d}t^3}{\mathrm{d}t} \\ &= 3t^2 + \ldots \qquad ,\end{aligned}$$

where the dots indicate infinitesimal terms that we can neglect.

This result required significant sweat and ingenuity when proved on page 122 by the methods of chapter 1, and not only that but the old method would have required a completely different method of proof for a function that wasn't a polynomial, whereas the new one can be applied more generally, as we'll see presently in examples 8-10.

It's easy to get the mistaken impression that infinitesimals exist in some remote fairyland where we can never touch them. This may be true in the same artsy-fartsy sense that we can never truly understand $\sqrt{2}$, because its decimal expansion goes on forever, and we therefore can never compute it exactly. But in practical work, that doesn't stop us from working with $\sqrt{2}$. We just approximate it as, e.g., 1.41. Infinitesimals are no more or less mysterious than irrational numbers, and in particular we can represent them concretely on a computer. If you go to `lightandmatter.com/calc/inf`, you'll find a web-based calculator called Inf, which can handle infinite and infinitesimal numbers. It has a built-in symbol, `d`, which represents an infinitesimally small number such as the dx's and dt's we've been handling symbolically.

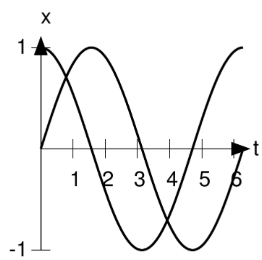
Let's use Inf to verify that the derivative of t^3, evaluated at $t = 1$, is equal to 3, as found by plugging in to the result of example 7. The : symbol is the prompt that shows you Inf is ready to accept your typed input.

```
: ((1+d)^3-1)/d
  3+3d+d^2
```

As claimed, the result is 3, or close enough to 3 that the infinitesimal error doesn't matter in real life. It might look like Inf did this example by using algebra to simplify the expression, but in fact Inf doesn't know anything about algebra. One way to see this is to use Inf to compare `d` with various real numbers:

```
: d<1
  true
: d<0.01
  true
: d<10^-10
  true
: d<0
  false
```

If `d` were just a variable being treated according to the axioms of algebra, there would be no way to tell how it compared with other numbers without having some special information. Inf doesn't know algebra, but it does know that `d` is a positive number that is less than any positive *real* number that can be represented using decimals or scientific notation.



d / Graphs of $\sin t$, and its derivative $\cos t$.

Example 8

The derivative of $x = \sin t$, with t in units of radians, is

$$\frac{dx}{dt} = \frac{\sin(t + dt) - \sin t}{dt} \quad ,$$

and with the trig identity $\sin(\alpha + \beta) = \sin\alpha\cos\beta + \cos\alpha\sin\beta$, this becomes

$$= \frac{\sin t\ \cos dt + \cos t\ \sin dt - \sin t}{dt} \quad .$$

Applying the small-angle approximations $\sin u \approx u$ and $\cos u \approx 1$, we have

$$\begin{aligned} \frac{dx}{dt} &= \frac{\cos t\, dt}{dt} + \ldots \\ &= \cos t + \ldots \quad , \end{aligned}$$

where "..." represents the error caused by the small-angle approximations.

This is essentially all there is to the computation of the derivative, except for the remaining technical point that we haven't proved that the small-angle approximations are good enough. In example 7 on page 26, when we calculated the derivative of t^3, the resulting expression for the quotient dx/dt came out in a form in which we could inspect the "..." terms and verify before discarding them that they were infinitesimal. The issue is less trivial in the present example. This point is addressed more rigorously on page 123.

Figure d shows the graphs of the function and its derivative. Note how the two graphs correspond. At $t = 0$, the slope of $\sin t$ is at its largest, and is positive; this is where the derivative, $\cos t$, attains its maximum positive value of 1. At $t = \pi/2$, $\sin t$ has reached a maximum, and has a slope of zero; $\cos t$ is zero here. At $t = \pi$, in the middle of the graph, $\sin t$ has its maximum negative slope, and $\cos t$ is at its most negative extreme of -1.

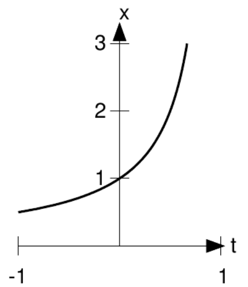
Physically, $\sin t$ could represent the position of a pendulum as it moved back and forth from left to right, and $\cos t$ would then be the pendulum's velocity.

Example 9

What about the derivative of the cosine? The cosine and the sine are really the same function, shifted to the left or right by $\pi/2$. If the derivative of the sine is the same as itself, but shifted to the left by $\pi/2$, then the derivative of the cosine must be a cosine shifted to the left by $\pi/2$:

$$\begin{aligned} \frac{d\ \cos t}{dt} &= \cos(t + \pi/2) \\ &= -\sin t \quad . \end{aligned}$$

The next example will require a little trickery. By the end of this chapter you'll learn general techniques for cranking out any derivative cookbook-style, without having to come up with any tricks.



e / The function $x = 1/(1-t)$.

Example 10

▷ Find the derivative of $1/(1-t)$, evaluated at $t=0$.

▷ The graph shows what the function looks like. It blows up to infinity at $t=1$, but it's well behaved at $t=0$, where it has a positive slope.

For insight, let's calculate some points on the curve. The point at which we're differentiating is $(0,1)$. If we put in a small, positive value of t, we can observe how much the result increases relative to 1, and this will give us an approximation to the derivative. For example, we find that at $t=0.001$, the function has the value 1.001001001001, and so the derivative is approximately $(1.001-1)/(.001-0)$, or about 1. We can therefore conjecture that the derivative is exactly 1, but that's not the same as proving it.

But let's take another look at that number 1.001001001001. It's clearly a repeating decimal. In other words, it appears that

$$\frac{1}{1-1/1000} = 1+\frac{1}{1000}+\left(\frac{1}{1000}\right)^2+\ldots$$

and we can easily verify this by multiplying both sides of the equation by $1-1/1000$ and collecting like powers. This is a special case of the geometric series

$$\frac{1}{1-t} = 1+t+t^2+\ldots \quad ,$$

which can be derived[2] by doing synthetic division (the equivalent of long division for polynomials), or simply verified, after forming the conjecture based on the numerical example above, by multiplying both sides by $1-t$.

As we'll see in section 2.2, and have been implicitly assuming so far, infinitesimals obey all the same elementary laws of algebra as the real numbers, so the above derivation also holds for an infinitesimal value of t. We can verify the result using Inf:

```
: 1/(1-d)
  1+d+d^2+d^3+d^4
```

Notice, however, that the series is truncated after the first five terms. This is similar to the truncation that happens when you ask your calculator to find $\sqrt{2}$ as a decimal.

The result for the derivative is

$$\begin{aligned}\frac{dx}{dt} &= \frac{\left(1+dt+dt^2+\ldots\right)-1}{1+dt-1}\\ &= 1+\ldots \quad .\end{aligned}$$

[2]As a technical aside, it's not necessary for our present purposes to go into the issue of how to make the most general possible definition of what is meant by a sum like this one which has an infinite number of terms; the only fact we'll need here is that the error in *finite* sum obtained by leaving out the "..." has only higher powers of t.

f / Bishop George Berkeley (1685-1753)

2.2 Safe use of infinitesimals

The idea of infinitesimally small numbers has always irked purists. One prominent critic of the calculus was Newton's contemporary George Berkeley, the Bishop of Cloyne. Although some of his complaints are clearly wrong (he denied the possibility of the second derivative), there was clearly something to his criticism of the infinitesimals. He wrote sarcastically, "They are neither finite quantities, nor quantities infinitely small, nor yet nothing. May we not call them ghosts of departed quantities?"

Infinitesimals seemed scary, because if you mishandled them, you could prove absurd things. For example, let $\mathrm{d}u$ be an infinitesimal. Then $2\mathrm{d}u$ is also infinitesimal. Therefore both $1/\mathrm{d}u$ and $1/(2\mathrm{d}u)$ equal infinity, so $1/\mathrm{d}u = 1/(2\mathrm{d}u)$. Multiplying by $\mathrm{d}u$ on both sides, we have a proof that $1 = 1/2$.

In the eighteenth century, the use of infinitesimals became like adultery: commonly practiced, but shameful to admit to in polite circles. Those who used them learned certain rules of thumb for handling them correctly. For instance, they would identify the flaw in my proof of $1 = 1/2$ as my assumption that there was only one size of infinity, when actually $1/\mathrm{d}u$ should be interpreted as an infinity twice as big as $1/(2\mathrm{d}u)$. The use of the symbol ∞ played into this trap, because the use of a single symbol for infinity implied that infinities only came in one size. However, the practitioners of infinitesimals had trouble articulating a clear set of principles for their proper use, and couldn't prove that a self-consistent system could be built around them.

By the twentieth century, when I learned calculus, a clear consensus had formed that infinite and infinitesimal numbers weren't numbers at all. A notation like $\mathrm{d}x/\mathrm{d}t$, my calculus teacher told me, wasn't really one number divided by another, it was merely a symbol for something called a limit,

$$\lim_{\Delta t \to 0} \frac{\Delta x}{\Delta t} \qquad ,$$

where Δx and Δt represented finite changes. I'll give a formal definition (actually two different formal definitions) of the term "limit" in section 2.9, but intuitively the concept is that is that we can get as good an approximation to the derivative as we like, provided that we make Δt small enough.

That satisfied me until we got to a certain topic (implicit differentiation) in which we were encouraged to break the dx away from the dt, leaving them on opposite sides of the equation. I buttonholed my teacher after class and asked why he was now doing what he'd told me you couldn't really do, and his response was that dx and dt weren't really numbers, but most of the time you could get away with treating them as if they were, and you would get the right answer in the end. *Most of the time!?* That bothered me. How was I supposed to know when it *wasn't* "most of the time?"

g / Abraham Robinson (1918-1974)

But unknown to me and my teacher, mathematician Abraham Robinson had already shown in the 1960's that it was possible to construct a self-consistent number system that included infinite and infinitesimal numbers. He called it the hyperreal number system, and it included the real numbers as a subset.[3]

Moreover, the rules for what you can and can't do with the hyperreals turn out to be extremely simple. Take any true statement about the real numbers. Suppose it's possible to translate it into a statement about the hyperreals in the most obvious way, simply by replacing the word "real" with the word "hyperreal." Then the translated statement is also true. This is known as the *transfer principle.*

Let's look back at my bogus proof of $1 = 1/2$ in light of this simple principle. The final step of the proof, for example, is perfectly valid: multiplying both sides of the equation by the same thing. The following statement about the real numbers is true:

> For any real numbers a, b, and c, if $a = b$, then $ac = bc$.

This can be translated in an obvious way into a statement about the hyperreals:

> For any hyperreal numbers a, b, and c, if $a = b$, then $ac = bc$.

[3]The main text of this book treats infinitesimals with the minimum fuss necessary in order to avoid the common goofs. More detailed discussions are often relegated to the back of the book, as in example 8 on page 28. The reader who wants to learn even more about the hyperreal system should consult the list of further reading on page 171.

However, what about the statement that both $1/\mathrm{d}u$ and $1/(2\mathrm{d}u)$ equal infinity, so they're equal to each other? This isn't the translation of a statement that's true about the reals, so there's no reason to believe it's true when applied to the hyperreals — and in fact it's false.

What the transfer principle tells us is that the real numbers as we normally think of them are not unique in obeying the ordinary rules of algebra. There are completely different systems of numbers, such as the hyperreals, that also obey them.

How, then, are the hyperreals even different from the reals, if everything that's true of one is true of the other? But recall that the transfer principle doesn't guarantee that every statement about the reals is also true of the hyperreals. It only works if the statement about the reals can be translated into a statement about the hyperreals in the most simple, straightforward way imaginable, simply by replacing the word "real" with the word "hyperreal." Here's an example of a true statement about the reals that can't be translated in this way:

> For any real number a, there is an integer n that is greater than a.

This one can't be translated so simplemindedly, because it refers to a subset of the reals called the integers. It might be possible to translate it somehow, but it would require some insight into the correct way to translate that word "integer." The transfer principle doesn't apply to this statement, which indeed is false for the hyperreals, because the hyperreals contain infinite numbers that are greater than all the integers. In fact, the contradiction of this statement can be taken as a definition of what makes the hyperreals special, and different from the reals: we assume that there is at least one hyperreal number, H, which is greater than all the integers.

As an analogy from everyday life, consider the following statements about the student body of the high school I attended:

1. Every student at my high school had two eyes and a face.
2. Every student at my high school who was on the football team was a jerk.

Let's try to translate these into statements about the population of California in general. The student body of my high school is like the set of real numbers, and the present-day population of California is like the hyperreals. Statement 1 can be translated mindlessly into a statement that every Californian has two eyes and a face; we simply substitute "every Californian" for "every student at my high school." But statement 2 isn't so easy, because it refers to the subset of students who were on the football team, and it's not obvious what the corresponding subset of Californians would be. Would it include everybody who played high school, college, or pro football? Maybe it shouldn't include the pros, because they belong to an organization covering a region bigger than California. Statement 2 is the kind of statement that the transfer principle doesn't apply to.[4]

Example 11

As a nontrivial example of how to apply the transfer principle, let's consider how to handle expressions like the one that occurred when we wanted to differentiate t^2 using infinitesimals:

$$\frac{\mathrm{d}\left(t^2\right)}{\mathrm{d}t} = 2t + \mathrm{d}t \qquad .$$

I argued earlier than $2t+\mathrm{d}t$ is so close to $2t$ that for all practical purposes, the answer is really $2t$. But is it really valid in general to say that $2t+\mathrm{d}t$ is the same hyperreal number as $2t$? No. We can apply the transfer principle to the following statement about the reals:

> For any real numbers a and b, with $b \neq 0$, $a+b \neq a$.

Since $\mathrm{d}t$ isn't zero, $2t+\mathrm{d}t \neq 2t$.

More generally, example 11 leads us to visualize every number as being surrounded by a "halo" of numbers that don't equal it, but differ from it by only an infinitesimal amount. Just as a magnifying glass would allow you to see the fleas on a dog, you would need an infinitely strong microscope to see this halo. This is similar to the idea that every integer is surrounded by a bunch of fractions that would round off to that integer. We can define the *standard part* of a finite hyperreal number, which means the unique real number that differs from it infinitesimally. For instance, the standard part of $2t+\mathrm{d}t$, notated $\mathrm{st}(2t+\mathrm{d}t)$, equals $2t$. The derivative of a function should actually be defined as the standard part of $\mathrm{d}x/\mathrm{d}t$, but we often write $\mathrm{d}x/\mathrm{d}t$ to mean the derivative, and don't worry about the distinction.

One of the things Bishop Berkeley disliked about infinitesimals was the idea that they existed in a kind of hierarchy, with $\mathrm{d}t^2$ being not just infinitesimally small, but infinitesimally small compared to the infinitesimal $\mathrm{d}t$. If $\mathrm{d}t$ is the flea on a dog, then $\mathrm{d}t^2$ is a submicroscopic flea that lives on the flea, as in Swift's doggerel: "Big fleas have little fleas/ On their backs to ride 'em,/ and little fleas have lesser fleas,/And so, ad infinitum." Berkeley's criticism was off the mark here: there is such a hierarchy. Our basic assumption about the hyperreals was that they contain at least one infinite number, H, which is bigger than all the integers. If this is true, then

[4] For a slightly more precise and formal statement of the transfer principle, see page 125.

$1/H$ must be less than $1/2$, less than $1/100$, less then $1/1,000,000$ — less than $1/n$ for any integer n. Therefore the hyperreals are guaranteed to include infinitesimals as well, and so we have at least three levels to the hierarchy: infinities comparable to H, finite numbers, and infinitesimals comparable to $1/H$. If you can swallow that, then it's not too much of a leap to add more rungs to the ladder, like extra-small infinitesimals that are comparable to $1/H^2$. If this seems a little crazy, it may comfort you to think of statements about the hyperreals as descriptions of limiting processes involving real numbers. For instance, in the sequence of numbers $1.1^2 = 1.21$, $1.01^2 = 1.0201$, $1.001^2 = 1.002001$, ..., it's clear that the number represented by the digit 1 in the final decimal place is getting smaller faster than the contribution due to the digit 2 in the middle.

One subtle issue here, which I avoided mentioning in the differentiation of the sine function on page 28, is whether the transfer principle is sufficient to let us define all the functions that appear as familiar keys on a calculator: x^2, $\sqrt{x}$, $\sin x$, $\cos x$, e^x, and so on. After all, these functions were originally defined as rules that would take a real number as an input and give a real number as an output. It's not trivially obvious that their definitions can naturally be extended to take a hyperreal number as an input and give back a hyperreal as an output. Essentially the answer is that we can apply the transfer principle to them just as we would to statements about simple arithmetic, but I've discussed this a little more on page 132.

2.3 The product rule

When I first learned calculus, it seemed to me that if the derivative of $3t$ was 3, and the derivative of $7t$ was 7, then the derivative of t multiplied by t ought to be just plain old t, not $2t$. The reason there's a factor of 2 in the correct answer is that t^2 has two reasons to grow as t gets bigger: it grows because the first factor of t is increasing, but also because the second one is. In general, it's possible to find the derivative of the product of two functions any time we know the derivatives of the individual functions.

The product rule
If x and y are both functions of t, then the derivative of their product is

$$\frac{\mathrm{d}(xy)}{\mathrm{d}t} = \frac{\mathrm{d}x}{\mathrm{d}t} \cdot y + x \cdot \frac{\mathrm{d}y}{\mathrm{d}t} \qquad .$$

The proof is easy. Changing t by an infinitesimal amount $\mathrm{d}t$ changes the product xy by an amount

$$\begin{aligned} &(x + \mathrm{d}x)(y + \mathrm{d}y) - xy \\ &= y\mathrm{d}x + x\mathrm{d}y + \mathrm{d}x\mathrm{d}y \qquad , \end{aligned}$$

and dividing by dt makes this into

$$\frac{\mathrm{d}x}{\mathrm{d}t} \cdot y + x \cdot \frac{\mathrm{d}y}{\mathrm{d}t} + \frac{\mathrm{d}x\mathrm{d}y}{\mathrm{d}t} \qquad ,$$

whose standard part is the result to be proved.

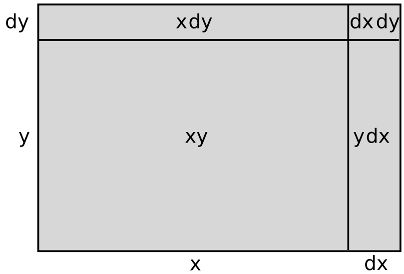
Example 12

▷ Find the derivative of the function $t \sin t$.

▷

$$\begin{aligned} \frac{\mathrm{d}(t \sin t)}{\mathrm{d}t} &= t \cdot \frac{\mathrm{d}(\sin t)}{\mathrm{d}t} + \frac{\mathrm{d}t}{\mathrm{d}t} \cdot \sin t \\ &= t \cos t + \sin t \end{aligned}$$

Figure h gives the geometrical interpretation of the product rule. Imagine that the king, in his castle at the southwest corner of his rectangular kingdom, sends out a line of infantry to expand his territory to the north, and a line of cavalry to take over more land to the east. In a time interval $\mathrm{d}t$, the cavalry, which moves faster, covers a distance $\mathrm{d}x$ greater than that covered by the infantry, $\mathrm{d}y$. However, the strip of territory conquered by the cavalry, $y\mathrm{d}x$, isn't as great as it could have been, because in our example y isn't as big as x.



h / A geometrical interpretation of the product rule.

A helpful feature of the Leibniz notation is that one can easily use it to check whether the units of an answer make sense. If we measure distances in meters and time in seconds, then xy has units of square meters (area), and so does the change in the area, $\mathrm{d}(xy)$. Dividing by $\mathrm{d}t$ gives the number of square meters per second being conquered. On the right-hand side of the product rule, $\mathrm{d}x/\mathrm{d}t$ has units of meters per second (velocity), and multiplying it by y makes the units square meters per second, which is consistent with the left-hand side. The units of the second term on the right likewise check out. Some beginners might be tempted to guess that the product rule would be $\mathrm{d}(xy)/\mathrm{d}t = (\mathrm{d}x/\mathrm{d}t)(\mathrm{d}y/\mathrm{d}t)$, but the Leibniz notation instantly reveals that this can't be the case, because then the units on the left, m^2/s, wouldn't match the ones on the right, $\mathrm{m}^2/\mathrm{s}^2$.

Because this unit-checking feature is so helpful, there is a special way of writing a second derivative in the Leibniz notation. What Newton called $\ddot{x}$, Leibniz wrote as

$$\frac{\mathrm{d}^2 x}{\mathrm{d}t^2} \quad .$$

Although the different placement of the 2's on top and bottom seems strange and inconsistent to many beginners, it actually works out nicely. If x is a distance, measured in meters, and t is a time, in units of seconds, then the second derivative is supposed to have units of acceleration, in units of meters per second per second, also written $(\mathrm{m/s})/\mathrm{s}$, or $\mathrm{m/s}^2$. (The acceleration of falling objects on Earth is 9.8 $\mathrm{m/s}^2$ in these units.) The Leibniz notation is meant to suggest exactly this: the top of the fraction looks like it has units of meters, because we're not squaring x, while the bottom of the fraction looks like it has units of seconds, because it looks like we're squaring $\mathrm{d}t$. Therefore the units come out right. It's important to realize, however, that the symbol d isn't a number (not a real one, and not a hyperreal one, either), so we can't really square it; the notation is not to be taken as a literal statement about infinitesimals.

Example 13

A tricky use of the product rule is to find the derivative of $\sqrt{t}$. Since $\sqrt{t}$ can be written as $t^{1/2}$, we might suspect that the rule $\mathrm{d}(t^k)/\mathrm{d}t = kt^{k-1}$ would work, giving a derivative $\frac{1}{2}t^{-1/2} = 1/(2\sqrt{t})$. However, the methods used to prove that rule in chapter 1 only work if k is an integer, so the best we could do would be to confirm our conjecture approximately by graphing.

Using the product rule, we can write $f(t) = \mathrm{d}\sqrt{t}/\mathrm{d}t$ for our unknown derivative, and back into the result using the product rule:

$$\begin{aligned}\frac{\mathrm{d}t}{\mathrm{d}t} &= \frac{\mathrm{d}(\sqrt{t}\sqrt{t})}{\mathrm{d}t} \\ &= f(t)\sqrt{t} + \sqrt{t}f(t) \\ &= 2f(t)\sqrt{t}\end{aligned}$$

But $\mathrm{d}t/\mathrm{d}t = 1$, so $f(t) = 1/(2\sqrt{t})$ as claimed.

The trick used in example 13 can also be used to prove that the power rule $\mathrm{d}(x^n)/\mathrm{d}x = nx^{n-1}$ applies to cases where n is an integer less than 0, but I'll instead prove this on page 41 by a technique that doesn't depend on a trick, and also applies to values of n that aren't integers.

2.4 The chain rule

Figure i shows three clowns on seesaws. If the leftmost clown moves down by a distance $\mathrm{d}x$, the middle one will come up by $\mathrm{d}y$, but this will also cause the one on the right to move down by $\mathrm{d}z$. If we want to predict how much the rightmost clown will move in response to a certain amount of motion by the leftmost one, we have

$$\frac{\mathrm{d}z}{\mathrm{d}x} = \frac{\mathrm{d}z}{\mathrm{d}y} \cdot \frac{\mathrm{d}y}{\mathrm{d}x} \qquad .$$

This relation, called the chain rule, allows us to calculate a derivative of a function defined by one function inside another. The proof, given on page 133, is essentially just the application of the transfer principle. (As is often the case, the proof using the hyperreals is much simpler than the one using real numbers and limits.)

Example 14

▷ Find the derivative of the function $z(x) = \sin(x^2)$.

▷ Let $y(x) = x^2$, so that $z(x) = \sin(y(x))$. Then

$$\begin{aligned}\frac{dz}{dx} &= \frac{dz}{dy} \cdot \frac{dy}{dx} \\ &= \cos(y) \cdot 2x \\ &= 2x\cos(x^2)\end{aligned}$$

The way people usually say it is that the chain rule tells you to take the derivative of the outside function, the sine in this case, and then multiply by the derivative of "the inside stuff," which here is the square. Once you get used to doing it, you don't need to invent a third, intermediate variable, as we did here with y.

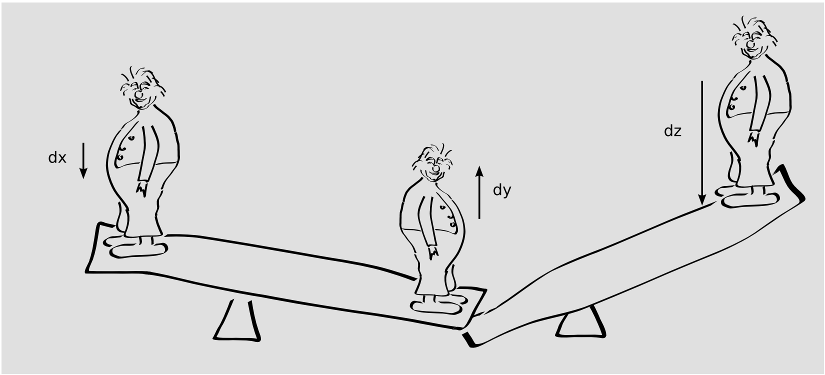
Example 15

Let's express the chain rule without the use of the Leibniz notation. Let the function f be defined by $f(x) = g(h(x))$. Then the derivative of f is given by $f'(x) = g'(h(x)) \cdot h'(x)$.

2.5 Exponentials and logarithms

The exponential

The exponential function e^x, where $e = 2.71828\ldots$ is the base of natural logarithms, comes constantly up in applications as diverse as credit-card interest, the growth of animal populations, and electric circuits. For its derivative



i / Three clowns on seesaws demonstrate the chain rule.

we have

$$\begin{aligned}\frac{\mathrm{d}e^x}{\mathrm{d}x} &= \frac{e^{x+\mathrm{d}x} - e^x}{\mathrm{d}x} \\ &= \frac{e^x e^{\mathrm{d}x} - e^x}{\mathrm{d}x} \\ &= e^x\,\frac{e^{\mathrm{d}x} - 1}{\mathrm{d}x}\end{aligned}$$

The second factor, $\left(e^{\mathrm{d}x} - 1\right)/\mathrm{d}x$, doesn't have x in it, so it must just be a constant. Therefore we know that the derivative of e^x is simply e^x, multiplied by some unknown constant,

$$\frac{\mathrm{d}e^x}{\mathrm{d}x} = c\,e^x.$$

A rough check by graphing at, say $x = 0$, shows that the slope is close to 1, so c is close to 1. Numerical calculation also shows that, for example, $(e^{0.001} - 1)/0.001 = 1.00050016670838$ is very close to 1. But how do we know it's exactly one when $\mathrm{d}x$ is really infinitesimal? We can use Inf:

```
: [exp(d)-1]/d
  1+0.5d+...
```

(The ... indicates where I've snipped some higher-order terms out of the output.) It seems clear that c is equal to 1 except for negligible terms involving higher powers of $\mathrm{d}x$. A rigorous proof is given on page 133.

Example 16

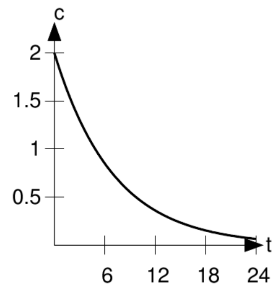
▷ The concentration of a foreign substance in the bloodstream generally falls off exponentially with time as $c = c_o e^{-t/a}$, where c_o is the initial concentration, and a is a constant. For caffeine in adults, a is typically about 7 hours. An example is shown in figure j. Differentiate the concentration with respect to time, and interpret the result. Check that the units of the result make sense.

▷ Using the chain rule,

$$\begin{aligned}\frac{dc}{dt} &= c_o e^{-t/a} \cdot \left(-\frac{1}{a}\right) \\ &= -\frac{c_o}{a} e^{-t/a}\end{aligned}$$

This can be interpreted as the rate at which caffeine is being removed from the blood and put into the person's urine. It's negative because the concentration is decreasing. According to the original expression for x, a substance with a large a will take a long time to reduce its concentration, since t/a won't be very big unless we have large t on top to compensate for the large a on the bottom. In other words, larger values of a represent substances that the body has a harder time getting rid of efficiently. The derivative has a on the bottom, and the interpretation of this is that for a drug that is hard to eliminate, the rate at which it is removed from the blood is low.

It makes sense that a has units of time, because the exponential function has to have a unitless argument, so the units of t/a have to cancel out. The units of the result come from the factor of c_o/a, and it makes sense that the units are concentration divided by time, because the result represents the rate at which the concentration is changing.



j / Example 16. A typical graph of the concentration of caffeine in the blood, in units of milligrams per liter, as a function of time, in hours.

Example 17

▷ Find the derivative of the function $y = 10^x$.

▷ In general, one of the tricks to doing calculus is to rewrite functions in forms that you know how to handle. This one can be rewritten as a base-e exponent:

$$\begin{aligned}y &= 10^x \\ \ln y &= \ln\left(10^x\right) \\ \ln y &= x \ln 10 \\ y &= e^{x \ln 10}\end{aligned}$$

Applying the chain rule, we have the derivative of the exponential, which is just the same exponential, multiplied by the derivative of the inside stuff:

$$\frac{dy}{dx} = e^{x \ln 10} \cdot \ln 10 \qquad .$$

In other words, the "c" referred to in the discussion of the derivative of e^x becomes $c = \ln 10$ in the case of the base-10 exponential.

The logarithm

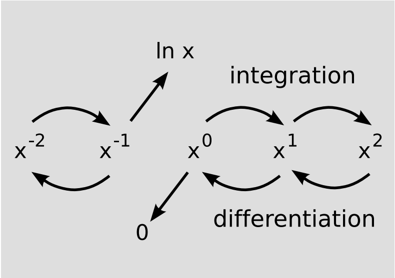
The natural logarithm is the function that undoes the exponential. In a situation like this, we have

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{1}{\mathrm{d}x/\mathrm{d}y} \qquad ,$$

where on the left we're thinking of y as a function of x, and on the right we consider x to be a function of y. Applying this to the natural logarithm,

$$\begin{aligned} y &= \ln x \\ x &= e^y \\ \frac{\mathrm{d}x}{\mathrm{d}y} &= e^y \\ \frac{\mathrm{d}y}{\mathrm{d}x} &= \frac{1}{e^y} \\ &= \frac{1}{x} \\ \frac{\mathrm{d}\ln x}{\mathrm{d}x} &= \frac{1}{x} \qquad . \end{aligned}$$

This is noteworthy because it shows that there must be an exception to the rule that the derivative of x^n is nx^{n-1}, and the integral of x^{n-1} is x^n/n. (On page 37 I remarked that this rule could be proved using the product rule for negative integer values of k, but that I would give a simpler, less tricky, and more general proof later. The proof is example 18 below.) The integral of x^{-1} is not $x^0/0$, which wouldn't make sense anyway because it involves division by zero.[5] Likewise the derivative of $x^0 = 1$ is $0x^{-1}$, which is zero. Figure k shows the idea. The functions x^n form a kind of ladder, with differentiation taking us down one rung, and integration taking us up. However, there are two special



k / Differentiation and integration of functions of the form x^n. Constants out in front of the functions are not shown, so keep in mind that, for example, the derivative of x^2 isn't x, it's $2x$.

[5] Speaking casually, one can say that division by zero gives infinity. This is often a good way to think when trying to connect mathematics to reality. However, it doesn't really work that way according to our rigorous treatment of the hyperreals. Consider this statement: "For a nonzero real number a, there is no real number b such that $a = 0b$." This means that we can't divide a by 0 and get b. Applying the transfer principle to this statement, we see that the same is true for the hyperreals: division by zero is undefined. However, we can divide a finite number by an infinitesimal, and get an infinite result, which is almost the same thing.

cases where differentiation takes us off the ladder entirely.

Example 18

▷ Prove $\mathrm{d}(x^n)/\mathrm{d}x = nx^{n-1}$ for any real value of n, not just an integer.

▷

$$\begin{aligned} y &= x^n \\ &= e^{n \ln x} \end{aligned}$$

By the chain rule,

$$\begin{aligned} \frac{\mathrm{d}y}{\mathrm{d}x} &= e^{n \ln x} \cdot \frac{n}{x} \\ &= x^n \cdot \frac{n}{x} \\ &= nx^{n-1} \end{aligned}.$$

(For $n = 0$, the result is zero.)

When I started the discussion of the derivative of the logarithm, I wrote $y = \ln x$ right off the bat. That meant I was implicitly assuming x was positive. More generally, the derivative of $\ln |x|$ equals $1/x$, regardless of the sign (see problem 26 on page 61).

2.6 Quotients

So far we've been successful with a divide-and-conquer approach to differentiation: the product rule and the chain rule offer methods of breaking a function down into simpler parts, and finding the derivative of the whole thing based on knowledge of the derivatives of the parts. We know how to find the derivatives of sums, differences, and products, so the obvious next step is to look for a way of handling division. This is straightforward, since we know that the derivative of the function function $1/u = u^{-1}$ is $-u^{-2}$. Let u and v be functions of x. Then by the product rule,

$$\frac{\mathrm{d}(v/u)}{\mathrm{d}x} = \frac{\mathrm{d}v}{\mathrm{d}x} \cdot \frac{1}{u} + v \cdot \frac{\mathrm{d}(1/u)}{\mathrm{d}x}$$

and by the chain rule,

$$\frac{\mathrm{d}(v/u)}{\mathrm{d}x} = \frac{\mathrm{d}v}{\mathrm{d}x} \cdot \frac{1}{u} - v \cdot \frac{1}{u^2} \frac{\mathrm{d}u}{\mathrm{d}x}$$

This is so easy to rederive on demand that I suggest not memorizing it.

By the way, notice how the notation becomes a little awkward when we want to write a derivative like $\mathrm{d}(v/u)/\mathrm{d}x$. When we're differentiating a complicated function, it can be uncomfortable trying to cram the expression into the top of the $\mathrm{d}\ldots/\mathrm{d}\ldots$ fraction. Therefore it would be more common to write such an expression like this:

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{v}{u}\right)$$

This could be considered an abuse of notation, making d look like a number being divided by another number $\mathrm{d}x$, when actually d is meaningless on its own. On the other hand, we can consider the symbol $\mathrm{d}/\mathrm{d}x$ to represent the operation of differentiation with respect to x; such an interpretation will seem more natural to those who have been inculcated with the taboo against considering infinitesimals as numbers in the first place.

Using the new notation, the quotient rule becomes

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{v}{u}\right)=\frac{1}{u}\cdot\frac{\mathrm{d}v}{\mathrm{d}x}-\frac{v}{u^2}\cdot\frac{\mathrm{d}u}{\mathrm{d}x}\qquad .$$

The interpretation of the minus sign is that if u increases, v/u decreases.

Example 19

▷ Differentiate $y = x/(1+3x)$, and check that the result makes sense.

▷ We identify v with x and u with $1+x$. The result is

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{v}{u}\right)&=\frac{1}{u}\cdot\frac{\mathrm{d}v}{\mathrm{d}x}-\frac{v}{u^2}\cdot\frac{\mathrm{d}u}{\mathrm{d}x}\\&=\frac{1}{1+3x}-\frac{3x}{(1+3x)^2}\end{aligned}$$

One way to check that the result makes sense it to consider extreme values of x. For very large values of x, the 1 on the bottom of $x/(1+3x)$ becomes negligible compared to the $3x$, and the function y approaches $x/3x = 1/3$ as a limit. Therefore we expect that the derivative $\mathrm{d}y/\mathrm{d}x$ should approach zero, since the derivative of a constant is zero. It works: plugging in bigger and bigger numbers for x in the expression for the derivative does give smaller and smaller results. (In the second term, the denominator gets bigger faster than the numerator, because it has a square in it.)

Another way to check the result is to verify that the units work out. Suppose arbitrarily that x has units of gallons. (If the 3 on the bottom is unitless, then the 1 would have to represent 1 gallon, since you can't add things that have different units.) The function y is defined by an expression with units of gallons divided by gallons, so y is unitless. Therefore the derivative $\mathrm{d}y/\mathrm{d}x$ should have units of inverse gallons. Both terms in the expression for the derivative do have those units, so the units of the answer check out.

2.7 Differentiation on a computer

In this chapter you've learned a set of rules for evaluating derivatives: derivatives of products, quotients, functions inside other functions, etc. Because these rules exist, it's always possible to find a formula for a function's derivative, given the formula for the original function. Not only that, but there is no real creativity required, so a computer can be programmed to do all the drudgery. For example, you can download a free, open-source program called Yacas from `yacas.sourceforge.net` and install it on a Windows or Linux machine. There is even a version you can run in a web browser without installing any special software: `http://yacas.sourceforge.net/yacasconsole.html` .
A typical session with Yacas looks like this:

Example 20

```
D(x) x^2
   2*x
D(x) Exp(x^2)
   2*x*Exp(x^2)
D(x) Sin(Cos(Sin(x)))
   -Cos(x)*Sin(Sin(x))
      *Cos(Cos(Sin(x)))
```

Upright type represents your input, and italicized type is the program's output.

First I asked it to differentiate x^2 with respect to x, and it told me the result was $2x$. Then I did the derivative of e^{x^2}, which I also could have done fairly easily by hand. (If you're trying this out on a computer as you real along, make sure to capitalize functions like Exp, Sin, and Cos.) Finally I tried an example where I didn't know the answer off the top of my head, and that would have been a little tedious to calculate by hand.

Unfortunately things are a little less rosy in the world of integrals. There are a few rules that can help you do integrals, e.g., that the integral of a sum equals the sum of the integrals, but the rules don't cover all the possible cases. Using Yacas to evaluate the integrals of the same functions, here's what happens.[6]

Example 21

```
Integrate(x) x^2
   x^3/3
Integrate(x) Exp(x^2)
   Integrate(x)Exp(x^2)
Integrate(x)
      Sin(Cos(Sin(x)))
   Integrate(x)
        Sin(Cos(Sin(x)))
```

[6]If you're trying these on your own computer, note that the long input line for the function sin cos sin x shouldn't be broken up into two lines as shown in the listing.

The first one works fine, and I can easily verify that the answer is correct, by taking the derivative of $x^3/3$, which is x^2. (The answer could have been $x^3/3+7$, or $x^3/3+c$, where c was any constant, but Yacas doesn't bother to tell us that.) The second and third ones don't work, however; Yacas just spits back the input at us without making any progress on it. And it may not be because Yacas isn't smart enough to figure out these integrals. The function e^{x^2} can't be integrated at all in terms of a formula containing ordinary operations and functions such as addition, multiplication, exponentiation, trig functions, exponentials, and so on.

That's not to say that a program like this is useless. For example, here's an integral that I wouldn't have known how to do, but that Yacas handles easily:

Example 22

```
Integrate(x) Sin(Ln(x))
   (x*Sin(Ln(x)))/2
        -(x*Cos(Ln(x)))/2
```

This one is easy to check by differentiating, but I could have been marooned on a desert island for a decade before I could have figured it out in the first place. There are various rules, then, for integration, but they don't cover all possible cases as the rules for differentiation do, and sometimes it isn't obvious which rule to apply. Yacas's ability to integrate sin ln x shows that it had a rule in its bag of tricks that

I don't know, or didn't remember, or didn't realize applied to this integral.

Back in the 17th century, when Newton and Leibniz invented calculus, there were no computers, so it was a big deal to be able to find a simple formula for your result. Nowadays, however, it may not be such a big deal. Suppose I want to find the derivative of $\sin\cos\sin x$, evaluated at $x = 1$. I can do something like this on a calculator:

Example 23

```
 sin cos sin 1 =
    0.61813407
 sin cos sin 1.0001 =
    0.61810240
(0.61810240-0.61813407)
       /.0001 =
    -0.3167
```

I have the right answer, with plenty of precision for most realistic applications, although I might have never guessed that the mysterious number -0.3167 was actually $-(\cos 1)(\sin\sin 1)(\cos\cos\sin 1)$. This could get a little tedious if I wanted to graph the function, for instance, but then I could just use a computer spreadsheet, or write a little computer program. In this chapter, I'm going to show you how to do derivatives and integrals using simple computer programs, using Yacas. The following little Yacas program does the same thing as the set of calculator operations shown above:

Example 24

```
1  f(x):=Sin(Cos(Sin(x)))
2  x:=1
3  dx:=.0001
4  N( (f(x+dx)-f(x))/dx )
     -0.3166671628
```

(I've omitted all of Yacas's output except for the final result.) Line 1 defines the function we want to differentiate. Lines 2 and 3 give values to the variables x and dx. Line 4 computes the derivative; the `N( )` surrounding the whole thing is our way of telling Yacas that we want an approximate numerical result, rather than an exact symbolic one.

An interesting thing to try now is to make dx smaller and smaller, and see if we get better and better accuracy in our approximation to the derivative.

Example 25

```
5  g(x,dx):=
        N( (f(x+dx)-f(x))/dx )
6  g(x,.1)
     -0.3022356406
7  g(x,.0001)
     -0.3166671628
8  g(x,.0000001)
     -0.3160458019
9  g(x,.00000000000000001)
     0
```

Line 5 defines the derivative function. It needs to know both x and dx. Line 6 computes the derivative using $\mathrm{d}x = 0.1$, which we expect to be a lousy approximation, since $\mathrm{d}x$ is really supposed to be infinitesimal, and 0.1 isn't even that small. Line 7 does it with the same value of dx we used earlier. The two results agree exactly in the first decimal place, and approximately in the second, so we can be pretty sure that the derivative is -0.32 to two figures of precision. Line 8 ups the ante, and produces a result that looks accurate to at least 3 decimal places. Line 9 attempts to produce fantastic precision by using an extremely small value of dx. Oops — the result isn't better, it's worse! What's happened here is that Yacas computed $f(x)$ and $f(x + \mathrm{d}x)$, but they were the same to within the precision it was using, so $f(x+\mathrm{d}x) - f(x)$ rounded off to zero.[7]

Example 25 demonstrates the concept of how a derivative can be defined in terms of a limit:

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x}$$

The idea of the limit is that we can theoretically make $\Delta y/\Delta x$ approach as close as we like to dy/dx, provided we make Δx sufficiently small. In reality, of course, we eventually run into the limits of our ability to do the computation, as in the bogus result generated on line 9 of the example.

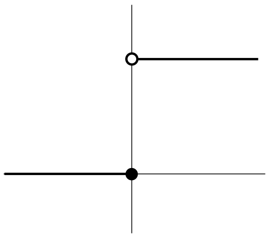
[7]Yacas can do arithmetic to any precision you like, although you may run into practical limits due to the amount of memory your computer has and the speed of its CPU. For fun, try `N(Pi,1000)`, which tells Yacas to compute π numerically to 1000 decimal places.

2.8 Continuity

Intuitively, a continuous function is one whose graph has no sudden jumps in it; the graph is all a single connected piece. Formally, a function $f(x)$ is defined to be continuous if for any real x and any infinitesimal dx, $f(x + \mathrm{d}x) - f(x)$ is infinitesimal.

Example 26

Let the function f be defined by $f(x) = 0$ for $x \le 0$, and $f(x) = 1$ for $x > 0$. Then $f(x)$ is discontinuous, since for d$x > 0$, $f(0+\mathrm{d}x) - f(0) = 1$, which isn't infinitesimal.



l / Example 26. The black dot indicates that the endpoint of the lower ray is part of the ray, while the white one shows the contrary for the ray on the top.

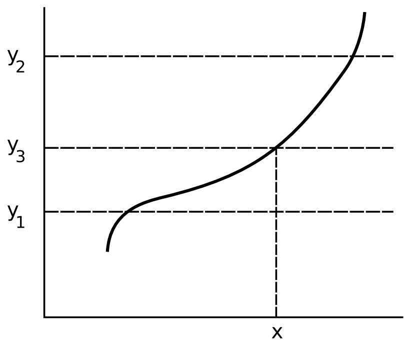
If a function is discontinuous at a given point, then it is not differentiable at that point. On the other hand, the example $y = |x|$ shows that a function can be continuous without being differentiable.

In most cases, there is no need

to invoke the definition explicitly in order to check whether a function is continuous. Most of the functions we work with are defined by putting together simpler functions as building blocks. For example, let's say we're already convinced that the functions defined by $g(x) = 3x$ and $h(x) = \sin x$ are both continuous. Then if we encounter the function $f(x) = \sin(3x)$, we can tell that it's continuous because its definition corresponds to $f(x) = h(g(x))$. The functions g and h have been set up like a bucket brigade, so that g takes the input, calculates the output, and then hands it off to h for the final step of the calculation. This method of combining functions is called *composition*. The composition of two continuous functions is also continuous. Just watch out for division. The function $f(x) = 1/x$ is continuous everywhere except at $x = 0$, so for example $1/\sin(x)$ is continuous everywhere except at multiples of π, where the sine has zeroes.

The intermediate value theorem

Another way of thinking about continuous functions is given by the *intermediate value theorem*. Intuitively, it says that if you are moving continuously along a road, and you get from point A to point B, then you must also visit every other point along the road; only by teleporting (by moving discontinuously) could you avoid doing so. More formally, the theorem states that if y is a continuous real-valued function on the real interval from a to b, and if y takes on values y_1 and y_2 at certain points within this interval, then for any y_3 between y_1 and y_2, there is some real x in the interval for which $y(x) = y_3$.



m / The intermediate value theorem states that if the function is continuous, it must pass through y_3.

The intermediate value theorem seems so intuitively appealing that if we want to set out to prove it, we may feel as though we're being asked to prove a proposition such as, "a number greater than 10 exists." If a friend wanted to bet you a six-pack that you couldn't prove this with complete mathematical rigor, you would have to get your friend to spell out very explicitly what she thought were the facts about integers that you were allowed to start with as initial assumptions. Are you allowed to assume that 1 exists? Will she grant you that if a number n exists, so does $n + 1$? The intermediate value theorem is similar. It's stated as a theorem about certain types of functions, but its truth isn't so much a matter of the properties of functions as the properties of the underlying number system. For the reader with a interest in pure mathematics, I've discussed this in more detail on page 136 and given an abbreviated proof. (Most introductory calculus texts do not prove it at all.)

Example 27

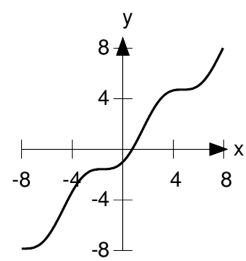
▷ Show that there is a solution to the equation $10^x + x = 1000$.

▷ We expect there to be a solution near $x = 3$, where the function $f(x) = 10^x + x = 1003$ is just a little too big. On the other hand, $f(2) = 102$ is much too small. Since f has values above and below 1000 on the interval from 2 to 3, and f is continuous, the intermediate value theorem proves that a solution exists between 2 and 3. If we wanted to find a better numerical approximation to the solution, we could do it using Newton's method, which is introduced in section 4.1.

Example 28

▷ Show that there is at least one solution to the equation $\cos x = x$, and give bounds on its location.

▷ This is a transcendental equation, and no amount of fiddling with algebra and trig identities will ever give a closed-form solution, i.e., one that can be written down with a finite number of arithmetic operations to give an exact result. However, we can easily prove that at least one solution exists, by applying the intermediate value theorem to the function $x - \cos x$. The cosine function is bounded between -1 and 1, so this function must be negative for $x < -1$ and positive for $x > 1$. By the intermediate value theorem, there must be a solution in the interval $-1 \leq x \leq 1$. The graph, n, verifies this, and shows that there is only one solution.



n / The function $x - \cos x$ constructed in example 28.

Example 29

▷ Prove that every odd-order polynomial P with real coefficients has at least one real root x, i.e., a point at which $P(x) = 0$.

▷ Example 28 might have given the impression that there was nothing to be learned from the intermediate value theorem that couldn't be determined by graphing, but this example clearly can't be solved by graphing, because we're trying to prove a general result for all polynomials.

To see that the restriction to odd orders is necessary, consider the polynomial $x^2 + 1$, which has no real roots because $x^2 > 0$ for any real number x.

To fix our minds on a concrete example for the odd case, consider the polynomial $P(x) = x^3 - x + 17$. For large values of x, the linear and constant terms will be negligible compared to the x^3 term, and since x^3 is positive for large values of x and negative for large negative ones, it follows that P is sometimes positive and sometimes negative.

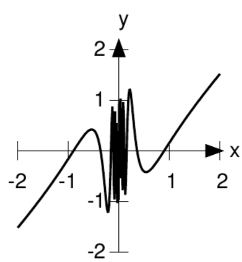
Making this argument more general and rigorous, suppose we had a polynomial of odd order n that always had the same sign for real x. Then by the transfer principle the same would hold for any hyperreal value of x. Now if x is infinite then the lower-order terms are infinitesimal compared to the x^n term, and the sign of the result is determined entirely by the x^n term, but x^n and $(-x)^n$ have opposite signs, and therefore $P(x)$ and $P(-x)$ have opposite signs. This is a contradiction, so we have disproved the assumption that P always had the same sign for real x. Since P is sometimes negative and sometimes positive, we conclude by the intermediate value theorem that it is zero somewhere.

Example 30

▷ Show that the equation $x = \sin 1/x$ has infinitely many solutions.

▷ This is another example that can't be solved by graphing; there is clearly no way to prove, just by looking at a graph like o, that it crosses the x axis *infinitely* many times. The graph does, however, help us to gain intuition for what's going on. As x gets smaller and smaller, $1/x$ blows up, and $\sin 1/x$ oscillates more and more rapidly. The function f is undefined at 0, but it's continuous everywhere else, so we can apply the intermediate value theorem to any interval that doesn't include 0.

We want to prove that for any positive u, there exists an x with $0 < x < u$ for which $f(x)$ has either desired sign. Suppose that this fails for some real u. Then by the transfer principle the nonexistence of any real x with the desired property also implies the nonexistence of any such hyperreal x. But for an infinitesimal x the sign of f is determined entirely by the sine term, since the sine term is finite and the linear term infinitesimal. Clearly $\sin 1/x$ can't have a single sign for all values of x less than u, so this is a contradiction, and the proposition succeeds for any u. It follows from the intermediate value theorem that there are infinitely many solutions to the equation.



o / The function $x - \sin 1/x$.

The extreme value theorem

In chapter 1, we saw that locating maxima and minima of functions may in general be fairly difficult, because there are so many different ways in which a function can attain an extremum: e.g., at an endpoint, at a place where its derivative is zero, or at a nondifferentiable kink. The following theorem allows us to make a very general statement about all these possible cases, assuming only continuity.

The *extreme value theorem* states that if f is a continuous real-valued function on the real-number interval defined by $a \le x \le b$, then f has maximum and minimum values on that interval, which are attained at specific points in the interval.

Let's first see why the assumptions are necessary. If we weren't combined to a finite interval, then $y = x$ would be a counterexample, because it's continuous and doesn't have any maximum or minimum value. If we didn't assume continuity, then we could have a function defined as $y = x$ for $x < 1$, and $y = 0$ for $x \ge 1$; this function never gets bigger than 1, but it never attains a value of 1 for any specific value of x.

The extreme value theorem is proved, in a somewhat more general form, on page 139.

Example 31

▷ Find the maximum value of the polynomial $P(x) = x^3 + x^2 + x + 1$ for $-5 \le x \le 5$.

▷ Polynomials are continuous, so the extreme value theorem guarantees that such a maximum exists. Suppose we try to find it by looking for a place where the derivative is zero. The derivative is $3x^2 + 2x + 1$, and setting it equal to zero gives a quadratic equation, but application of the quadratic formula shows that it has no real solutions. It appears that the function doesn't have a maximum anywhere (even outside the interval of interest) that looks like a smooth peak. Since it doesn't have kinks or discontinuities, there is only one other type of maximum it could have, which is a maximum at one of its endpoints. Plugging in the limits, we find $P(-5) = -104$ and $P(5) = 156$, so we conclude that the maximum value on this interval is 156.

2.9 Limits

Historically, the calculus of infinitesimals as created by Newton and Leibniz was reinterpreted in the nineteenth century by Cauchy, Bolzano, and Weierstrass in terms of limits. All mathematicians learned both languages, and switched back and forth between them effortlessly, like the lady I overheard in a Southern California supermarket telling her mother, "Let's get that one, *con los* nuts." Those who had been trained in infinitesimals might hear a statement using the language of limits, but translate it mentally into infinitesimals; to them, every statement about limits was really a statement about infinitesimals. To their younger colleagues, trained using limits, every statement about infinitesimals was really to be understood as shorthand for a limiting process. When Robinson laid the rigorous foundations for the hyperreal number system in the 1960's, a common objection was that it was really nothing new, because every statement about infinitesimals was really just a different way of expressing a corresponding statement about limits; of course the same could have been said about Weierstrass's work of the preceding century! In reality, all practitioners of calculus had realized all along that different approaches worked better for different problems; problem 11 on page 76 is an example of a result that is much easier to prove with infinitesimals than with limits.

The Weierstrass definition of a limit is this:

Definition of the limit
We say that ℓ is the limit of the function $f(x)$ as x approaches a, written

$$\lim_{x \to a} f(x) = \ell \qquad ,$$

if the following is true: for any real number ϵ, there exists another real number δ such that for all x in the interval $a-\delta \le x \le a+\delta$, the value of f lies within the range from $\ell-\epsilon$ to $\ell+\epsilon$.

Intuitively, the idea is that if I want you to make $f(x)$ close to ℓ, I just have to tell you how close, and you can tell me that it will be that close as long as x is within a certain distance of a.

In terms of infinitesimals, we have:

Definition of the limit
We say that ℓ is the limit of the function $f(x)$ as x approaches a, written

$$\lim_{x \to a} f(x) = \ell \qquad ,$$

if the following is true: for any infinitesimal number $\mathrm{d}x$, the value of $f(a+\mathrm{d}x)$ is finite, and the standard part of $f(a+\mathrm{d}x)$ equals ℓ.

The two definitions are equivalent.

Sometimes a limit can be evaluated simply by plugging in numbers:

Example 32

▷ Evaluate

$$\lim_{x \to 0} \frac{1}{1+x} \qquad .$$

▷ Plugging in $x = 0$, we find that the limit is 1.

In some examples, plugging in fails if we try to do it directly, but can be made to work if we massage the expression into a different form:

Example 33

▷ Evaluate

$$\lim_{x \to 0} \frac{\frac{2}{x}+7}{\frac{1}{x}+8686} \qquad .$$

▷ Plugging in $x = 0$ fails because division by zero is undefined.

Intuitively, however, we expect that the limit will be well defined, and will equal 2, because for very small values of x, the numerator is dominated by the $2/x$ term, and the denominator by the $1/x$ term, so the 7 and 8686 terms will matter less and less as x gets smaller and smaller.

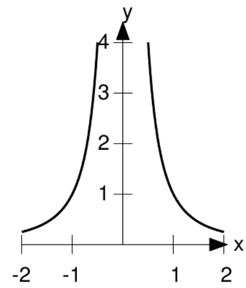
To demonstrate this more rigorously, a trick that works is to multiply both the top and the bottom by x, giving

$$\frac{2+7x}{1+8686x} \qquad ,$$

which equals 2 when we plug in $x = 0$, so we find that the limit is zero.

This example is a little subtle, because when x *equals* zero, the function is not defined, and moreover it would not be valid to multiply both the top and the bottom by x. In general, it's not valid algebra to multiply both the top and the bottom of a fraction by 0, because the result is $0/0$, which is undefined. But we *didn't* actually multiply both the top and the bottom by zero, because we never let x equal zero. Both the Weierstrass definition and the definition in terms of infinitesimals only refer to the properties of the function in a region very close to the limiting point, not at the limiting point itself.

This is an example in which the function was not well defined at a certain point, and yet the limit of the function was well defined as we approached that point. In a case like this, where there is only one point missing from the domain of the function, it is natural to extend the definition of the function by filling in the "gap tooth." Example 35 below shows that this kind of filling-in procedure is not always possible.



p / Example 34, the function $1/x^2$.

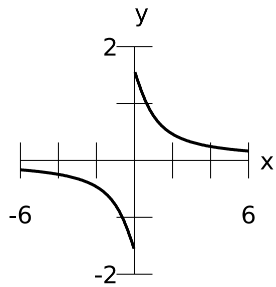
Example 34

▷ Investigate the limiting behavior of $1/x^2$ as x approaches 0, and 1.

▷ At $x = 1$, plugging in works, and we

find that the limit is 1.

At $x = 0$, plugging in doesn't work, because division by zero is undefined. Applying the definition in terms of infinitesimals to the limit as x approaches 0, we need to find out whether $1/(0 + dx)^2$ is finite for infinitesimal dx, and if so, whether it always has the same standard part. But clearly $1/(0 + dx)^2 = dx^{-2}$ is always infinite, and we conclude that this limit is undefined.



q / Example 35, the function $\tan^{-1}(1/x)$.

Example 35

▷ Investigate the limiting behavior of $f(x) = \tan^{-1}(1/x)$ as x approaches 0.

▷ Plugging in doesn't work, because division by zero is undefined.

In the definition of the limit in terms of infinitesimals, the first requirement is that $f(0 + dx)$ be finite for infinitesimal values of dx. The graph makes this look plausible, and indeed we can prove that it is true by the transfer principle. For any real x we have $-\pi/2 \le f(x) \le \pi/2$, and by the transfer principle this holds for the hyperreals as well, and therefore $f(0 + dx)$ is finite.

The second requirement is that the standard part of $f(0 + dx)$ have a uniquely defined value. The graph shows that we really have two cases to consider, one on the right side of the graph, and one on the left. Intuitively, we expect that the standard part of $f(0 + dx)$ will equal $\pi/2$ for positive dx, and $-\pi/2$ for negative, and thus the second part of the definition will not be satisfied. For a more formal proof, we can use the transfer principle. For real x with $0 < x < 1$, for example, f is always positive and greater than 1, so we conclude based on the transfer principle that $f(0 + dx) > 1$ for positive infinitesimal dx. But on similar grounds we can be sure that $f(0 + dx) < -1$ when dx is negative and infinitesimal. Thus the standard part of $f(0 + dx)$ can have different values for different infinitesimal values of dx, and we conclude that the limit is undefined.

In examples like this, we can define a kind of one-sided limit, notated like this:

$$\lim_{x\to 0^-} \tan^{-1}\frac{1}{x} = -\frac{\pi}{2}$$
$$\lim_{x\to 0^+} \tan^{-1}\frac{1}{x} = \frac{\pi}{2} \qquad ,$$

where the notations $x \to 0^-$ and $x \to 0^+$ are to be read "as x approaches zero from below," and "as x approaches zero from above."

L'Hôpital's rule

Consider the limit

$$\lim_{x\to 0} \frac{\sin x}{x} \qquad .$$

Plugging in doesn't work, because we get 0/0. Division by zero is undefined, both in the real number system and in the hyperreals. A nonzero number divided by a small number gives a big number; a nonzero number divided by a very small number gives a very big number; and a nonzero number divided by an infinitesimal number gives an infinite number. On the other hand, dividing *zero* by zero means looking for a solution to the equation $0 = 0q$, where q is the result of the division. But any q is a solution of this equation, so even speaking casually, it's not correct to say that 0/0 is infinite; it's not infinite, it's anything we like.

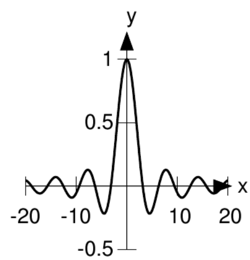
Since plugging in zero didn't work, let's try estimating the limit by plugging in a number for x that's small, but not zero. On a calculator,

$$\frac{\sin 0.00001}{0.00001} = 0.999999999983333$$

It looks like the limit is 1. We can confirm our conjecture to higher precision using Yacas's ability to do high-precision arithmetic:

```
N(Sin(10^-20)/10^-20,50)
    0.99999999999999999
    99999999999999999999
    99998333333333
```

It's looking pretty one-ish. This is the idea of the Weierstrass definition of a limit: it seems like we can get an answer as close to 1 as we like, if we're willing to make x as close to 0 as necessary. The graph helps to make this plausible.



r / The graph of $\sin x/x$.

The general idea here is that for small values of x, the small-angle approximation $\sin x \approx x$ obtains, and as x gets smaller and smaller, the approximation gets better and better, so $\sin x/x$ gets closer and closer to 1.

But we still haven't proved rigorously that the limit is exactly 1. Let's try using the definition of the limit in terms of infinitesimals.

$$\begin{aligned}\lim_{x\to 0}\frac{\sin x}{x} &= \operatorname{st}\left[\frac{\sin(0+\mathrm{d}x)}{0+\mathrm{d}x}\right]\\ &= \operatorname{st}\left[\frac{\mathrm{d}x+\ldots}{\mathrm{d}x}\right] \qquad ,\end{aligned}$$

where ... stands for terms of order $\mathrm{d}x^2$. So

$$\begin{aligned}\lim_{x\to 0}\frac{\sin x}{x} &= \operatorname{st}\left[1+\frac{\ldots}{\mathrm{d}x}\right] \qquad ,\\ &= 1 \qquad .\end{aligned}$$

We can check our work using Inf:

```
: (sin d)/d
  1+(-0.16667)d^2+...
```

(The ... is where I've snipped trailing terms from the output.)

This is a special case of a the following rule for calculating limits involving 0/0:

L'Hôpital's rule
If u and v are functions with $u(a) = 0$ and $v(a) = 0$, the derivatives $\dot{v}(a)$ and $\dot{v}(a)$ are defined, and the derivative $\dot{v}(a) \neq 0$, then

$$\lim_{x\to a} \frac{u}{v} = \frac{\dot{u}(a)}{\dot{v}(a)} \quad .$$

Proof: Since $u(a) = 0$, and the derivative $\mathrm{d}u/\mathrm{d}x$ is defined at a, $u(a+\mathrm{d}x) = \mathrm{d}u$ is infinitesimal, and likewise for v. By the definition of the limit, the limit is the standard part of

$$\frac{u}{v} = \frac{\mathrm{d}u}{\mathrm{d}v} = \frac{\mathrm{d}u/\mathrm{d}x}{\mathrm{d}v/\mathrm{d}x} \quad ,$$

where by assumption the numerator and denominator are both defined (and finite, because the derivative is defined in terms of the standard part). The standard part of a quotient like p/q equals the quotient of the standard parts, provided that both p and q are finite (which we've established), and $q \neq 0$ (which is true by assumption). But the standard part of $\mathrm{d}u/\mathrm{d}x$ is the definition of the derivative $\dot{u}$, and likewise for $\mathrm{d}v/\mathrm{d}x$, so this establishes the result.

By the way, the housetop accent on the "ô" in L'Hôpital means that in Old French it used to be spelled and pronounced "L'Hospital," but the "s" later became silent, so they stopped writing it. So yes, it is the same word as "hospital."

Example 36

▷ Evaluate

$$\lim_{x\to 0} \frac{e^x - 1}{x}$$

▷ Taking the derivatives of the top and bottom, we find $e^x/1$, which equals 1 when evaluated at $x = 0$.

Example 37

▷ Evaluate

$$\lim_{x\to 1} \frac{x-1}{x^2 - 2x + 1}$$

▷ Plugging in $x = 1$ fails, because both the top and the bottom are zero. Taking the derivatives of the top and bottom, we find $1/(2x - 2)$, which blows up to infinity when $x = 1$. To symbolize the fact that the limit is undefined, and undefine because it blows up to infinity, we write

$$\lim_{x\to 1} \frac{x-1}{x^2 - 2x + 1} = \infty$$

In the following example, we have to use L'Hôpital's rule twice before we get an answer.

Example 38

▷ Evaluate

$$\lim_{x\to\pi} \frac{1+\cos x}{(x-\pi)^2}$$

▷ Applying L'Hôpital's rule gives

$$\frac{-\sin x}{2(x-\pi)} \quad ,$$

which still produces $0/0$ when we plug in $x = \pi$. Going again, we get

$$\frac{-\cos x}{2} = \frac{1}{2} \quad .$$

Another perspective on indeterminate forms

An expression like $0/0$, called an indeterminate form, can be thought of in a different way in terms of infinitesimals. Suppose I tell you I have two infinitesimal numbers d and e in my pocket, and I ask you whether d/e is finite, infinite, or infinitesimal. You can't tell, because d and e might not be infinitesimals of the same order of magnitude. For instance, if $e = 37d$, then $d/e = 1/37$ is finite; but if $e = d^2$, then d/e is infinite; and if $d = e^2$, then d/e is infinitesimal. Acting this out with numbers that are small but not infinitesimal,

$$\frac{.001}{.037} = \frac{1}{37}$$
$$\frac{.001}{.000001} = 1000$$
$$\frac{.000001}{.001} = .001 \quad .$$

On the other hand, suppose I tell you I have an infinitesimal number d and a finite number x, and I ask you to speculate about d/x. You know for sure that it's going to be infinitesimal. Likewise, you can be sure that x/d is infinite. These aren't indeterminate forms.

We can do something similar with infinite numbers. If H and K are both infinite, then $H - K$ is indeterminate. It could be infinite, for example, if H was positive infinite and $K = H/2$. On the other hand, it could be finite if $H = K + 1$. Acting this out with big but finite numbers,

$$1000 - 500 = 500$$
$$1001 - 1000 = 1 \quad .$$

Example 39

▷ If H is a positive infinite number, is $\sqrt{H+1} - \sqrt{H-1}$ finite, infinite, infinitesimal, or indeterminate?

▷ Trying it with a finite, big number, we have

$$\sqrt{1000001} - \sqrt{999999}$$
$$= 1.00000000020373 \times 10^{-3} \quad ,$$

which is clearly a wannabe infinitesimal. We can verify the result using Inf:

```
: H=1/d
  d^-1
: sqrt(H+1)-sqrt(H-1)
  d^1/2+0.125d^5/2+...
```

For convenience, the first line of input defines an infinite number H in terms of the calculator's built-in infinitesimal

d. The result has only positive powers of d, so it's clearly infinitesimal.

More rigorously, we can rewrite the expression as $\sqrt{H}(\sqrt{1+1/H}-\sqrt{1-1/H})$. Since the derivative of the square root function $\sqrt{x}$ evaluated at $x = 1$ is 1/2, we can approximate this as

$$\begin{aligned}\sqrt{H}&\left[1+\frac{1}{2H}+\ldots-\left(1-\frac{1}{2H}+\ldots\right)\right]\\ &=\sqrt{H}\left[\frac{1}{H}+\ldots\right]\\ &=\frac{1}{\sqrt{H}}\qquad,\end{aligned}$$

which is infinitesimal.

Limits at infinity

The definition of the limit in terms of infinitesimals extends immediately to limiting processes where x gets bigger and bigger, rather than closer and closer to some finite value. For example, the function $3 + 1/x$ clearly gets closer and closer to 3 as x gets bigger and bigger. If a is an infinite number, then the definition says that evaluating this expression at $a + \mathrm{d}x$, where $\mathrm{d}x$ is infinitesimal, gives a result whose standard part is 3. It doesn't matter that a happens to be infinite, the definition still works. We also note that in this example, it doesn't matter what infinite number a is; the limit equals 3 for *any* infinite a. We can write this fact as

$$\lim_{x\to\infty}\left(3+\frac{1}{x}\right)=3\qquad,$$

where the symbol ∞ is to be interpreted as "nyeah nyeah, I don't even care what infinite number you put in here, I claim it will work out to 3 no matter what." The symbol ∞ is *not* to be interpreted as standing for any specific infinite number. That would be the type of fallacy that lay behind the bogus proof on page 30 that $1 = 1/2$, which assumed that all infinities had to be the same size.

A somewhat different example is the arctangent function. The arctangent of 1000 equals approximately 1.5698, and inputting bigger and bigger numbers gives answers that appear to get closer and closer to $\pi/2 \approx 1.5707$. But the arctangent of -1000 is approximately -1.5698, i.e., very close to $-\pi/2$. From these numerical observations, we conjecture that

$$\lim_{x\to a}\tan^{-1}x$$

equals $\pi/2$ for positive infinite a, but $-\pi/2$ for negative infinite a. It would not be correct to write

$$\lim_{x\to\infty}\tan^{-1}x=\frac{\pi}{2}\qquad\text{[wrong]}\qquad,$$

because it *does* matter what infinite number we pick. Instead we write

$$\begin{aligned}\lim_{x\to+\infty}\tan^{-1}x&=\frac{\pi}{2}\\ \lim_{x\to-\infty}\tan^{-1}x&=-\frac{\pi}{2}\qquad.\end{aligned}$$

Some expressions don't have this kind of limit at all. For example, if you take the sines of big numbers like a thousand, a million, etc., on your calculator, the results are essentially random numbers lying between -1 and 1. They don't settle down to any particular value, because the sine function oscillates back and forth forever. To prove formally that $\lim_{x\to+\infty} \sin x$ is undefined, consider that the sine function, defined on the real numbers, has the property that you can always change its result by at least 0.1 if you add either 1.5 or -1.5 to its input. For example, $\sin(.8) \approx 0.717$, and $\sin(.8-1.5) \approx -0.644$. Applying the transfer principle to this statement, we find that the same is true on the hyperreals. Therefore there cannot be any value ℓ that differs infinitesimally from $\sin a$ for all positive infinite values of a.

Often we're interested in finding the limit as x approaches infinity of an expression that is written as an indeterminate form like H/K, where both H and K are infinite.

Example 40

▷ Evaluate the limit

$$\lim_{x\to\infty} \frac{2x+7}{x+8686} \quad .$$

▷ Intuitively, if x gets large enough the constant terms will be negligible, and the top and bottom will be dominated by the $2x$ and x terms, respectively, giving an answer that approaches 2.

One way to verify this is to divide both the top and the bottom by x, giving

$$\frac{2+\frac{7}{x}}{1+\frac{8686}{x}} \quad .$$

If x is infinite, then the standard part of the top is 2, the standard part of the bottom is 1, and the standard part of the whole thing is therefore 2.

Another approach is to use L'Hôpital's rule. The derivative of the top is 2, and the derivative of the bottom is 1, so the limit is 2/1=2.

Problems

1 Carry out a calculation like the one in example 7 on page 26 to show that the derivative of t^4 equals $4t^3$. ▷ Solution, p. 153

2 Example 9 on page 28 gave a tricky argument to show that the derivative of $\cos t$ is $-\sin t$. Prove the same result using the method of example 8 instead.
▷ Solution, p. 153

3 Suppose H is a big number. Experiment on a calculator to figure out whether $\sqrt{H+1} - \sqrt{H-1}$ comes out big, normal, or tiny. Try making H bigger and bigger, and see if you observe a trend. Based on these numerical examples, form a conjecture about what happens to this expression when H is infinite. ▷ Solution, p. 153

4 Suppose $\mathrm{d}x$ is a small but finite number. Experiment on a calculator to figure out how $\sqrt{\mathrm{d}x}$ compares in size to $\mathrm{d}x$. Try making $\mathrm{d}x$ smaller and smaller, and see if you observe a trend. Based on these numerical examples, form a conjecture about what happens to this expression when $\mathrm{d}x$ is infinitesimal. ▷ Solution, p. 154

5 To which of the following statements can the transfer principle be applied? If you think it can't be applied to a certain statement, try to prove that the statement is false for the hyperreals, e.g., by giving a counterexample.

(a) For any real numbers x and y, $x + y = y + x$.
(b) The sine of any real number is between -1 and 1.
(c) For any real number x, there exists another real number y that is greater than x.
(d) For any real numbers $x \neq y$, there exists another real number z such that $x < z < y$.
(e) For any real numbers $x \neq y$, there exists a rational number z such that $x < z < y$. (A rational number is one that can be expressed as an integer divided by another integer.)
(f) For any real numbers x, y, and z, $(x + y) + z = x + (y + z)$.
(g) For any real numbers x and y, either $x < y$ or $x = y$ or $x > y$.
(h) For any real number x, $x + 1 \neq x$. ▷ Solution, p. 154

6 If we want to pump air or water through a pipe, common sense tells us that it will be easier to move a larger quantity more quickly through a fatter pipe. Quantitatively, we can define the resistance, R, which is the ratio of the pressure difference produced by the pump to the rate of flow. A fatter pipe will have a lower resistance. Two pipes can be used in parallel, for instance when you turn on the water both in the kitchen and in the bathroom, and in this situation, the two pipes let more water flow than either would have let flow by itself, which tells us that they act like a single pipe with some lower resistance. The equation for their combined resistance is $R =$

$1/(1/R_1+1/R_2)$. Analyze the case where one resistance is finite, and the other infinite, and give a physical interpretation. Likewise, discuss the case where one is finite, but the other is infinitesimal.

7 Naively, we would imagine that if a spaceship traveling at $u = 3/4$ of the speed of light was to shoot a missile in the forward direction at $v = 3/4$ of the speed of light (relative to the ship), then the missile would be traveling at $u + v = 3/2$ of the speed of light. However, Einstein's theory of relativity tells us that this is too good to be true, because nothing can go faster than light. In fact, the relativistic equation for combining velocities in this way is not $u + v$, but rather $(u + v)/(1 + uv)$. In ordinary, everyday life, we never travel at speeds anywhere near the speed of light. Show that the nonrelativistic result is recovered in the case where both u and v are infinitesimal.

8 Differentiate $(2x + 3)^{100}$ with respect to x. ▷ Solution, p. 154

9 Differentiate $(x + 1)^{100}(x + 2)^{200}$ with respect to x.

▷ Solution, p. 154

10 Differentiate the following with respect to x: e^{7x}, e^{e^x}. (In the latter expression, as in all exponentials nested inside exponentials, the evaluation proceeds from the top down, i.e., $e^{(e^x)}$, not $(e^e)^x$.)

▷ Solution, p. 155

11 Differentiate $a\sin(bx + c)$ with respect to x.

▷ Solution, p. 155

12 Find a function whose derivative with respect to x equals $a\sin(bx + c)$. That is, find an integral of the given function.

▷ Solution, p. 155

13 Use the chain rule to differentiate $((x^2)^2)^2$, and show that you get the same result you would have obtained by differentiating x^8.

[M. Livshits]

14 The range of a gun, when elevated to an angle θ, is given by

$$R = \frac{2v^2}{g}\sin\theta\,\cos\theta \quad .$$

Find the angle that will produce the maximum range.

▷ Solution, p. 155

15 The hyperbolic cosine function is defined by

$$\cosh x = \frac{e^x + e^{-x}}{2} \quad .$$

Find any minima and maxima of this function.

▷ Solution, p. 156

16 Show that the function $\sin(\sin(\sin x))$ has maxima and minima at all the same places where $\sin x$ does, and at no other places. ▷ Solution, p. 156

17 In free fall, the acceleration will not be exactly constant, due to air resistance. For example, a skydiver does not speed up indefinitely until opening her chute, but rather approaches a certain maximum velocity at which the upward force of air resistance cancels out the force of gravity. The expression for the distance dropped by of a free-falling object, with air resistance, is[8]

$$d = A \ln \left[\cosh\left(t\sqrt{\frac{g}{A}}\right)\right] \qquad ,$$

where g is the acceleration the object would have without air resistance, the function cosh has been defined in problem 15, and A is a constant that depends on the size, shape, and mass of the object, and the density of the air. (For a sphere of mass m and diameter d dropping in air, $A = 4.11m/d^2$. Cf. problem 9, p. 101.)
(a) Differentiate this expression to find the velocity. Hint: In order to simplify the writing, start by defining some other symbol to stand for the constant $\sqrt{g/A}$.
(b) Show that your answer can be reexpressed in terms of the function tanh defined by $\tanh x = (e^x - e^{-x})/(e^x + e^{-x})$.
(c) Show that your result for the velocity approaches a constant for large values of t.
(d) Check that your answers to parts b and c have units of velocity.

▷ Solution, p. 157

[8]Jan Benacka and Igor Stubna, *The Physics Teacher*, 43 (2005) 432.

18 Differentiate $\tan\theta$ with respect to θ. ▷ Solution, p. 157

19 Differentiate $\sqrt[3]{x}$ with respect to x. ▷ Solution, p. 158

20 Differentiate the following with respect to x:
(a) $y = \sqrt{x^2 + 1}$
(b) $y = \sqrt{x^2 + a^2}$
(c) $y = 1/\sqrt{a + x}$
(d) $y = a/\sqrt{a - x^2}$

▷ Solution, p. 158 [Thompson, 1919]

21 Differentiate $\ln(2t + 1)$ with respect to t. ▷ Solution, p. 158

22 If you know the derivative of $\sin x$, it's not necessary to use the product rule in order to differentiate $3\sin x$, but show that using the product rule gives the right result anyway. ▷ Solution, p. 158

23 The Γ function (capital Greek letter gamma) is a continuous mathematical function that has the property $\Gamma(n) = 1 \cdot 2 \cdot \ldots \cdot (n-1)$ for n an integer. $\Gamma(x)$ is also well defined for values of x that are not integers, e.g., $\Gamma(1/2)$ happens to be $\sqrt{\pi}$. Use computer software that is capable of evaluating the Γ function to determine numerically the derivative of $\Gamma(x)$ with respect to x, at $x = 2$. (In Yacas, the function is called Gamma.)

▷ Solution, p. 158

24 For a cylinder of fixed surface area, what proportion of length to radius will give the maximum volume?

▷ Solution, p. 158

25 This problem is a variation on problem 11 on page 22. Einstein found that the equation $K = (1/2)mv^2$ for kinetic energy was only a good approximation for speeds much less than the speed of light, c. At speeds comparable to the speed of light, the correct equation is

$$K = \frac{\frac{1}{2}mv^2}{\sqrt{1 - v^2/c^2}} \qquad .$$

(a) As in the earlier, simpler problem, find the power $\mathrm{d}K/\mathrm{d}t$ for an object accelerating at a steady rate, with $v = at$.
(b) Check that your answer has the right units.
(c) Verify that the power required becomes infinite in the limit as v approaches c, the speed of light. This means that no material object can go as fast as the speed of light. ▷ Solution, p. 159

26 Prove, as claimed on page 41, that the derivative of $\ln|x|$ equals $1/x$, for both positive and negative x. ▷ Solution, p. 160

27 Use a trick similar to the one used in example 13 to prove that the power rule $\mathrm{d}(x^k)/\mathrm{d}x = kx^{k-1}$ applies to cases where k is an integer less than 0.
▷ Solution, p. 161 ⋆

28 The plane of Euclidean geometry is today often described as the set of all coordinate pairs (x, y), where x and y are real. We could instead imagine the plane F that is defined in the same way, but with x and y taken from the set of hyperreal numbers. As a third alternative, there is the plane G in which the finite hyperreals are used. In E, Euclid's parallel postulate holds: given a line and a point not on the line, there exists exactly one line passing through the point that does not intersect the line. Does the parallel postulate hold in F? In G? Is it valid to associate only E with the plane described by Euclid's axioms? ⋆

29 (a) Prove, using the Weierstrass definition of the limit, that if $\lim_{x\to a} f(x) = F$ and $\lim_{x\to a} g(x) = G$ both exist, them $\lim_{x\to a}[f(x) + g(x)] = F + G$, i.e., that the limit of a sum is the sum of the limits. (b) Prove the same thing using the definition of the limit in terms of infinitesimals.
▷ Solution, p. 161

30 Sketch the graph of the function $e^{-1/x}$, and evaluate the following four limits:

$$\lim_{x\to 0^+} e^{-1/x}$$
$$\lim_{x\to 0^-} e^{-1/x}$$
$$\lim_{x\to +\infty} e^{-1/x}$$
$$\lim_{x\to -\infty} e^{-1/x}$$

31 Verify the following limits.

$$\lim_{s\to 1}\frac{s^3-1}{s-1}=3$$
$$\lim_{\theta\to 0}\frac{1-\cos\theta}{\theta^2}=\frac{1}{2}$$
$$\lim_{x\to\infty}\frac{5x^2-2x}{x}=\infty$$
$$\lim_{n\to\infty}\frac{n(n+1)}{(n+2)(n+3)}=1$$
$$\lim_{x\to\infty}\frac{ax^2+bx+c}{dx^2+ex+f}=\frac{a}{d}$$

[Granville, 1911]

32 Prove that the linear function $y=ax+b$ is continuous, first using the definition of continuity in terms of infinitesimals, and then using the definition in terms of limits.

33 Discuss the following statement: *The repeating decimal* $0.999\ldots$ *is infinitesimally less than one.* ▷ Solution, p. 161

34 Example 15 on page 37 expressed the chain rule without the Leibniz notation, writing a function f defined by $f(x)=g(h(x))$. Suppose that you're trying to remember the rule, and two of the possibilities that come to mind are $f'(x)=g'(h(x))$ and $f'(x)=g'(h(x))h(x)$. Show that neither of these can possibly be right, by considering the case where x has units. You may find it helpful to convert both expressions back into the Leibniz notation.

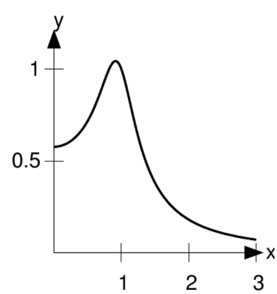
▷ Solution, p. 161

35 When you tune in a radio station using an old-fashioned rotating dial you don't have to be exactly tuned in to the right frequency in order to get the station. If you did, the tuning would be infinitely sensitive, and you'd never be able to receive any signal at all! Instead, the tuning has a certain amount of "slop" intentionally designed into it. The strength of the received signal s can be expressed in terms of the dial's setting f by a function of the form

$$s=\frac{1}{\sqrt{a(f^2-f_o^2)^2+bf^2}}\quad ,$$

where a, b, and f_o are constants. This functional form is in fact very general, and is encountered in many other physical contexts. The graph below shows the resulting bell-shaped curve. Find the frequency f at which the maximum response occurs, and show that if b is small, the maximum occurs close to, but not exactly at, f_o.

▷ Solution, p. 162



The function of problem 35, with a = 3, b = 1, and f_o = 1.

3 Integration

3.1 Definite and indefinite integrals

Because any formula can be differentiated symbolically to find another formula, the main motivation for doing derivatives numerically would be if the function to be differentiated wasn't known in symbolic form. A typical example might be a two-person network computer game, in which player A's computer needs to figure out player B's velocity based on knowledge of how her position changes over time. But in most cases, it's numerical integration that's interesting, not numerical differentiation.

As a warm-up, let's see how to do a running sum of a discrete function using Yacas. The following program computers the sum $1+2+\ldots+100$ discussed to on page 9. Now that we're writing real computer programs with Yacas, it would be a good idea to enter each program into a file before trying to run it. In fact, some of these examples won't run properly if you just start up Yacas and type them in one line at a time. If you're using Adobe Reader to read this book, you can do `Tools>Basic>Select`, select the program, copy it into a file, and then edit out the line numbers.

Example 41

```
1 n := 1;
2 sum := 0;
3 While (n<=100) [
4   sum := sum+n;
5   n := n+1;
6 ];
7 Echo(sum);
```

The semicolons are to separate one instruction from the next, and they become necessary now that we're doing real programming. Line 1 of this program defines the variable `n`, which will take on all the values from 1 to 100. Line 2 says that we haven't added anything up yet, so our running sum is zero do far. Line 3 says to keep on repeating the instructions inside the square brackets until `n` goes past 100. Line 4 updates the running sum, and line 5 updates the value of `n`. If you've never done any programming before, a statement like `n:=n+1` might seem like nonsense — how can a number equal itself plus one? But that's why we use the `:=` symbol; it says that we're redefining `n`, not stating an equation. If `n` was previously 37, then after this statement is executed, `n` will be redefined as 38. To run the program on a Linux computer, do this (assuming you saved the program in a file named `sum.yacas`):

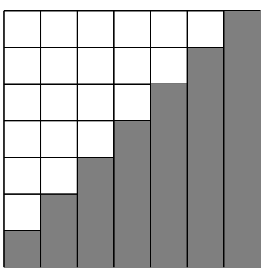
```
% yacas -pc sum.yacas
```

5050

Here the `%` symbol is the computer's prompt. The result is 5,050, as expected. One way of stating this result is

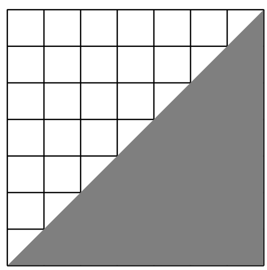
$$\sum_{n=1}^{100} n = 5050 \qquad .$$

The capital Greek letter Σ, sigma, is used because it makes the "s" sound, and that's the first sound in the word "sum." The $n = 1$ below the sigma says the sum starts at 1, and the 100 on top says it ends at 100. The n is what's known as a dummy variable: it has no meaning outside the context of the sum. Figure a shows the graphical interpretation of the sum: we're adding up the areas of a series of rectangular strips. (For clarity, the figure only shows the sum going up to 7, rather than 100.)



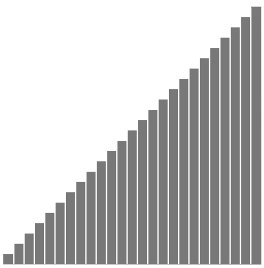
a / Graphical interpretation of the sum $1+2+\ldots+7$.

Now how about an integral? Figure b shows the graphical interpretation of what we're trying to do: find the area of the shaded triangle. This is an example we know how to do symbolically, so we can do it numerically as well, and check the answers against each other. Symbolically, the area is given by the integral. To integrate the function $\dot{x}(t) = t$, we know we need some function with a t^2 in it, since we want something whose derivative is t, and differentiation reduces the power by one. The derivative of t^2 would be $2t$ rather than t, so what we want is $x(t) = t^2/2$. Let's compute the area of the triangle that stretches along the t axis from 0 to 100: $x(100) = 100^/2 = 5000$.



b / Graphical interpretation of the integral of the function $\dot{x}(t) = t$.

Figure c shows how to accomplish the same thing numerically. We break up the area into a whole bunch of very skinny rectangles. Ideally, we'd like to make the width of each rectangle be an infinitesimal number $\mathrm{d}x$, so that we'd be adding up an infinite number of infinitesimal areas. In reality, a computer can't do that, so we divide up the interval from $t = 0$ to $t = 100$ into H rectangles, each with finite width $\mathrm{d}t = 100/H$. Instead of making H infinite, we make it the largest number we can without making the computer take too long to add up the areas of the rectangles.



c / Approximating the integral numerically.

Example 42

```
tmax := 100;
H := 1000;
dt := tmax/H;
sum := 0;
t := 0;
While (t<=tmax) [
  sum := N(sum+t*dt);
  t := N(t+dt);
];
Echo(sum);
```

In example 42, we split the interval from $t = 0$ to 100 into $H = 1000$ small intervals, each with width $\mathrm{d}t = 0.1$. The result is 5,005, which agrees with the symbolic result to three digits of precision. Changing H to 10,000 gives $5,000.5$, which is one more digit. Clearly as we make the number of rectangles greater and greater, we're converging to the correct result of 5,000.

In the Leibniz notation, the thing we've just calculated, by two different techniques, is written like this:

$$\int_0^{100} t\,\mathrm{d}t = 5,000$$

It looks a lot like the Σ notation, with the Σ replaces by a flattened-out letter "S." The t is a dummy variable. What I've been casually referring to as an integral is really two different but closely related things, known as the definite integral and the indefinite integral.

Definition of the indefinite integral
If $\dot{x}$ is a function, then a function x is an indefinite integral of $\dot{x}$ if, as implied by the notation, $\mathrm{d}x/\mathrm{d}t = \dot{x}$.

Interpretation: Doing an indefinite integral means doing the opposite of differentiation. All the possible indefinite integrals are the same function except for an additive constant.

Example 43

▷ Find the indefinite integral of the function $\dot{x}(t) = t$.

▷ Any function of the form

$$x(t) = t^2/2 + c \quad ,$$

where c is a constant, is an indefinite integral of this function, since its derivative is t.

Definition of the definite integral
If $\dot{x}$ is a function, then the definite integral of $\dot{x}$ from a to b is defined as

$$\int_a^b \dot{x}(t)\mathrm{d}t$$
$$= \lim_{H\to\infty} \sum_{i=0}^{H} \dot{x}\left(a+i\Delta t\right) \qquad ,$$

where $\Delta t = (b-a)/H$.

Interpretation: What we're calculating is the area under the graph of $\dot{x}$, from a to b. (If the graph dips below the t axis, we interpret the area between it and the axis as a negative area.) The thing inside the limit is a calculation like the one done in example 42, but generalized to $a \neq 0$. If H was infinite, then Δt would be an infinitesimal number $\mathrm{d}t$.

3.2 The fundamental theorem of calculus

The fundamental theorem of calculus
Let x be an indefinite integral of $\dot{x}$, and let $\dot{x}$ be a continuous function (one whose graph is a single connected curve). Then

$$\int_a^b \dot{x}(t)\mathrm{d}t = x(b) - x(a) \qquad .$$

Interpretation: In the simple examples we've been doing so far, we were able to choose an indefinite integral such that $x(0) = 0$. In that case, $x(t)$ is interpreted as the area from 0 to t, so in the expression $x(b) - x(a)$, we're taking the area from 0 to a, but subtracting out the area from 0 to b, which gives the area from a to b. If we choose an indefinite integral with a different c, the c's will just cancel out anyway in the difference $x(b) - x(a)$.

The fundamental theorem is proved on page 134.

Example 44

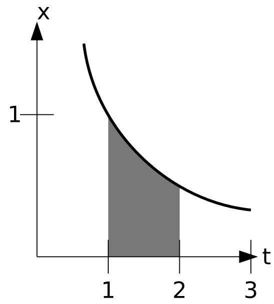
▷ Interpret the indefinite integral

$$\int_1^2 \frac{1}{t}\,\mathrm{d}t \qquad .$$

graphically; then evaluate it it both symbolically and numerically, and check that the two results are consistent.

▷ Figure d shows the graphical interpretation. The numerical calculation requires a trivial variation on the program from example 42:

```
a := 1;
b := 2;
H := 1000;
dt := (b-a)/H;
```



d / The indefinite integral $\int_1^2 (1/t)\mathrm{d}t$.

```
sum := 0;
t := a;
While (t<=b) [
  sum := N(sum+(1/t)*dt);
  t := N(t+dt);
];
Echo(sum);
```

The result is 0.693897243, and increasing H to 10,000 gives 0.6932221811, so we can be fairly confident that the result equals 0.693, to 3 decimal places.

Symbolically, the indefinite integral is $x = \ln t$. Using the fundamental theorem of calculus, the area is $\ln 2 - \ln 1 \approx 0.693147180559945$.

Judging from the graph, it looks plausible that the shaded area is about 0.7.

This is an interesting example, because the natural log blows up to negative infinity as t approaches 0, so it's not possible to add a constant onto the indefinite integral and force it to be equal to 0 at $t = 0$. Nevertheless, the fundamental theorem of calculus still works.

3.3 Properties of the integral

Let f and g be two functions of x, and let c be a constant. We already know that for derivatives,

$$\frac{\mathrm{d}}{\mathrm{d}x}(f+g) = \frac{\mathrm{d}f}{\mathrm{d}x} + \frac{\mathrm{d}g}{\mathrm{d}x}$$

and

$$\frac{\mathrm{d}}{\mathrm{d}x}(cf) = c\frac{\mathrm{d}f}{\mathrm{d}x} \qquad .$$

But since the indefinite integral is just the operation of undoing a derivative, the same kind of rules must hold true for indefinite integrals as well:

$$\int (f+g)\mathrm{d}x = \int f\mathrm{d}x + \int g\mathrm{d}x$$

and

$$\int (cf)\mathrm{d}x = c\int f\mathrm{d}x \qquad .$$

And since a definite integral can be found by plugging in the upper and lower limits of integration into the indefinite integral, the same properties must be true of definite integrals as well.

Example 45

▷ Evaluate the indefinite integral

$$\int (x + 2\sin x)\mathrm{d}x \qquad .$$

▷ Using the additive property, the integral becomes

$$\int x\mathrm{d}x + \int 2\sin x\mathrm{d}x \qquad .$$

Then the property of scaling by a constant lets us change this to

$$\int x\mathrm{d}x + 2\int \sin x\mathrm{d}x \quad .$$

We need a function whose derivative is x, which would be $x^2/2$, and one whose derivative is $\sin x$, which must be $-\cos x$, so the result is

$$\frac{1}{2}x^2 - 2\cos x + c \quad .$$

3.4 Applications

Averages

In the story of Gauss's problem of adding up the numbers from 1 to 100, one interpretation of the result, 5,050, is that the average of all the numbers from 1 to 100 is 50.5. This is the ordinary definition of an average: add up all the things you have, and divide by the number of things. (The result in this example makes sense, because half the numbers are from 1 to 50, and half are from 51 to 100, so the average is half-way between 50 and 51.)

Similarly, a definite integral can also be thought of as a kind of average. In general, if y is a function of x, then the average, or mean, value of y on the interval from $x = a$ to b can be defined as

$$\bar{y} = \frac{1}{b-a}\int_a^b y\,\mathrm{d}x \quad .$$

In the continuous case, dividing by $b - a$ accomplishes the same thing as dividing by the number of things in the discrete case.

Example 46

▷ Show that the definition of the average makes sense in the case where the function is a constant.

▷ If y is a constant, then we can take it outside of the integral, so

$$\begin{aligned}\bar{y} &= \frac{1}{b-a}y\int_a^b 1\,\mathrm{d}x\\ &= \frac{1}{b-a}y\,x\Big|_a^b\\ &= \frac{1}{b-a}y(b-a)\\ &= y\end{aligned}$$

Example 47

▷ Find the average value of the function $y = x^2$ for values of x ranging from 0 to 1.

$$\begin{aligned}\bar{y} &= \frac{1}{1-0}\int_0^1 x^2\,\mathrm{d}x\\ &= \frac{1}{3}x^3\Big|_0^1\\ &= \frac{1}{3}\end{aligned}$$

The mean value theorem
If the continuous function $y(x)$ has the average value $\bar{y}$ on the interval from $x = a$ to b, then y attains its average value at least once in that interval, i.e., there exists ξ with $a < \xi < b$ such that $y(\xi) = \bar{y}$.

The mean value theorem is proved on page 141. The special case in which $\bar{y} = 0$ is known as Rolle's theorem.

Example 48

▷ Verify the mean value theorem for $y = x^2$ on the interval from 0 to 1.

▷ The mean value is 1/3, as shown in example 47. This value is achieved at $x = \sqrt{1/3} = 1/\sqrt{3}$, which lies between 0 and 1.

Work

In physics, work is a measure of the amount of energy transferred by a force; for example, if a horse sets a wagon in motion, the horse's force on the wagon is putting some energy of motion into the wagon. When a force F acts on an object that moves in the direction of the force by an infinitesimal distance $\mathrm{d}x$, the infinitesimal work done is $\mathrm{d}W = F\mathrm{d}x$. Integrating both sides, we have $W = \int_a^b F\mathrm{d}x$, where the force may depend on x, and a and b represent the initial and final positions of the object.

Example 49

▷ A spring compressed by an amount x relative to its relaxed length provides a force $F = kx$. Find the amount of work that must be done in order to compress the spring from $x = 0$ to $x = a$. (This is the amount of energy stored in the spring, and that energy will later be released into the toy bullet.)

▷

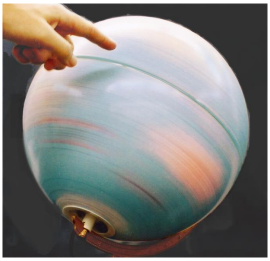
$$\begin{aligned} W &= \int_0^a F\mathrm{d}x \\ &= \int_0^a kx\mathrm{d}x \\ &= \left.\frac{1}{2}kx^2\right|_0^a \\ &= \frac{1}{2}ka^2 \end{aligned}$$

The reason W grows like a^2, not just like a, is that as the spring is compressed more, more and more effort is required in order to compress it.

Probability

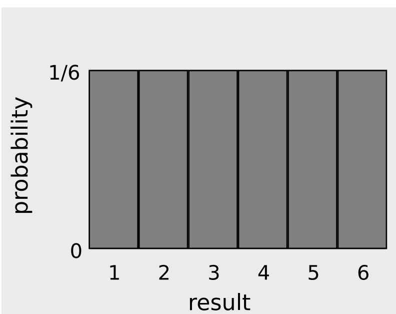
Mathematically, the probability that something will happen can be specified with a number ranging from 0 to 1, with 0 representing impossibility and 1 representing certainty. If you flip a coin, heads and tails both have probabilities of 1/2. The sum of the probabilities of all the possible outcomes has to have probability 1. This is called *normalization.*

So far we've discussed random processes having only two possible outcomes: yes or no, win or lose, on or off. More generally, a random process could have a result that is a number. Some processes yield integers, as when you roll a die and get a result from one to six, but some are not restricted to whole numbers, e.g., the height of



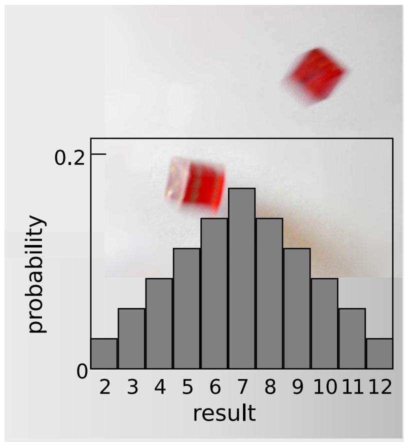
e / Normalization: the probability of picking land plus the probability of picking water adds up to 1.

a human being, or the amount of time that a uranium-238 atom will exist before undergoing radioactive decay. The key to handling these continuous random variables is the concept of the area under a curve, i.e., an integral.



f / Probability distribution for the result of rolling a single die.

Consider a throw of a die. If the die is "honest," then we expect all six values to be equally likely. Since all six probabilities must add up to 1, then probability of any particular value coming up must be $1/6$. We can summarize this in a graph, f. Areas under the curve can be interpreted as total probabilities. For instance, the area under the curve from 1 to 3 is $1/6+1/6+1/6 = 1/2$, so the probability of getting a result from 1 to 3 is $1/2$. The function shown on the graph is called the probability distribution.



g / Rolling two dice and adding them up.

Figure g shows the probabilities of various results obtained by rolling two dice and adding them together, as in the game of craps. The probabilities are not all the same. There is a small probability of getting a two, for example, because there is only one way to do it, by rolling a one and then another one. The probability of rolling a seven is high because there are six different ways to do it: 1+6, 2+5, etc.

If the number of possible outcomes is large but finite, for example the number of hairs on a dog, the graph would start to look like a smooth curve rather than a ziggurat.

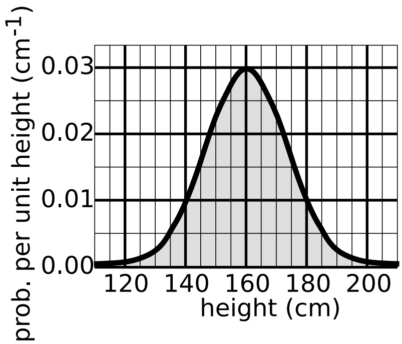
What about probability distributions for random numbers that are not integers? We can no longer make a graph with probability on the y axis, because the probability of getting a given exact number is typically zero. For instance, there is zero probability that a person will be *exactly* 200 cm tall, since there are infinitely many possible results that are close to 200 but not exactly two, for example 199.99999999687687658766. It doesn't usually make sense, therefore, to talk about the probability of a single numerical result, but it does make sense to talk about the probability of a certain range of results. For instance, the probability that a randomly chosen person will be more than 170 cm and less than 200 cm in height is a perfectly reasonable thing to discuss. We can still summarize the probability information on a graph, and we can still interpret areas under the curve as probabilities.

But the y axis can no longer be a unitless probability scale. In the example of human height, we want the x axis to have units of meters, and we want areas under the curve to be unitless probabilities. The area of a single square on the graph paper is then

$$\begin{aligned}(\text{unitless area of a square})\\ = (\text{width of square}\\ \text{with distance units})\\ \times(\text{height of square}) \quad .\end{aligned}$$

If the units are to cancel out, then the height of the square must evidently be a quantity with units of inverse centimeters. In other words, the y axis of the graph is to be interpreted as probability per unit height, not probability.

Another way of looking at it is that the y axis on the graph gives a derivative, $\mathrm{d}P/\mathrm{d}x$: the infinitesimally small probability that x will lie in the infinitesimally small range covered by $\mathrm{d}x$.



h / A probability distribution for human height.

Example 50

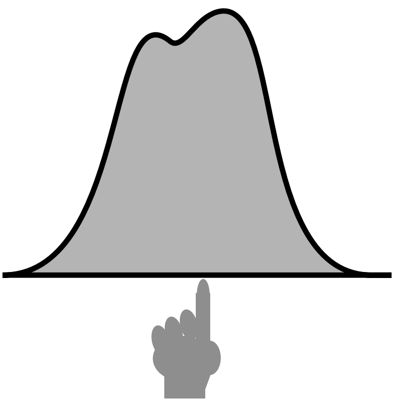
▷ A computer language will typically have a built-in subroutine that produces a fairly random number that is equally likely to take on any value in the range from 0 to 1. If you take the absolute value of the difference between two such numbers, the probability distribution is of the form $dP/dx = k(1-x)$. Find the value of the constant k that is required by normalization.

▷

$$\begin{aligned}1 &= \int_0^1 k(1-x)\,dx \\ &= kx - \frac{1}{2}kx^2\Big|_0^1 \\ &= k - k/2 \\ k &= 2\end{aligned}$$

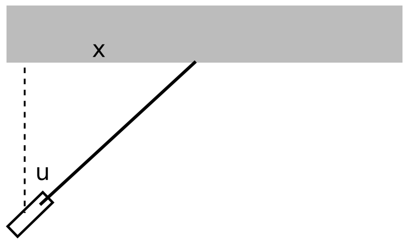
Self-Check

Compare the number of people with heights in the range of 130-135 cm to the number in the range 135-140. ▷ Answer, p. 145



i / The average can be interpreted as the balance point of the probability distribution.

When one random variable is related to another in some mathematical way, the chain rule can be used to relate their probability distributions.



j / Example 51.

Example 51

▷ A laser is placed one meter away from a wall, and spun on the ground to give it a random direction, but if the angle u shown in figure j doesn't come out in the range from 0 to $\pi/2$, the laser is spun again until an angle in the desired range is obtained. Find the probability distribution of the distance x shown in the figure. The derivative $d\tan^{-1} z/dz = 1/(1+z^2)$ will be required (see example 57, page 80).

▷ Since any angle between 0 and $\pi/2$ is equally likely, the probability distribution dP/du must be a constant, and normalization tells us that the constant must be $dP/du = 2/\pi$.

The laser is one meter from the wall, so the distance x, measured in meters, is given by $x = \tan u$. For the probability distribution of x, we have

$$\begin{aligned}\frac{dP}{dx} &= \frac{dP}{du}\cdot\frac{du}{dx}\\ &= \frac{2}{\pi}\cdot\frac{d\tan^{-1}x}{dx}\\ &= \frac{2}{\pi(1+x^2)}\end{aligned}$$

Note that the range of possible values of x theoretically extends from 0 to infinity. Problem 6 on page 90 deals with this.

If the next Martian you meet asks you, "How tall is an adult human?," you will probably reply with a statement about the average human height, such as "Oh, about 5 feet 6 inches." If you wanted to explain a little more, you could say, "But that's only an average. Most people are somewhere between 5 feet and 6 feet tall." Without bothering to draw the relevant bell curve for your new extraterrestrial acquaintance, you've summarized the relevant information by giving an average and a typical range of variation. The average of a probability distribution can be defined geometrically as the horizontal position at which it could be balanced if it was constructed out of cardboard, i. This is a different way of working with averages than the one we did earlier. Before, had a graph of y versus x, we implicitly assumed that all values of x were equally likely, and we found an average value of y. In this new method using probability distributions, the variable we're averaging is on the x axis, and the y axis tells us the relative probabilities of the various x values.

For a discrete-valued variable with n possible values, the average would be

$$\bar{x} = \sum_{i=0}^{n} x\,P(x) \qquad ,$$

and in the case of a continuous variable, this becomes an integral,

$$\bar{x} = \int_a^b x\,\frac{dP}{dx}\,dx \qquad .$$

Example 52

▷ For the situation described in example 50, find the average value of x.

▷

$$\begin{aligned}\bar{x} &= \int_0^1 x\,\frac{dP}{dx}\,dx\\ &= \int_0^1 x\cdot 2(1-x)\,dx\\ &= 2\int_0^1 (x-x^2)\,dx\\ &= 2\left(\frac{1}{2}x^2-\frac{1}{3}x^3\right)\Big|_0^1\\ &= \frac{1}{3}\end{aligned}$$

Sometimes we don't just want to know the average value of a certain variable, we also want to have some idea of the amount of variation above and below the average. The most common way of measuring this is the *standard deviation*, defined by

$$\sigma = \sqrt{\int_a^b (x-\bar{x})^2 \frac{\mathrm{d}P}{\mathrm{d}x}\,\mathrm{d}x} \qquad .$$

The idea here is that if there was no variation at all above or below the average, then the quantity $(x-\bar{x})$ would be zero whenever $\mathrm{d}P/\mathrm{d}x$ was nonzero, and the standard deviation would be zero. The reason for taking the square root of the whole thing is so that the result will have the same units as x.

Example 53

▷ For the situation described in example 50, find the standard deviation of x.

▷ The square of the standard deviation is

$$\begin{aligned}\sigma^2 &= \int_0^1 (x-\bar{x})^2 \frac{\mathrm{d}P}{\mathrm{d}x}\,\mathrm{d}x \\ &= \int_0^1 (x-1/3)^2 \cdot 2(1-x)\,\mathrm{d}x \\ &= 2\int_0^1 \left(-x^3+\frac{5}{3}x^2-\frac{7}{9}x+\frac{1}{9}\right)\mathrm{d}x \\ &= \frac{1}{18} \qquad ,\end{aligned}$$

so the standard deviation is

$$\begin{aligned}\sigma &= \frac{1}{\sqrt{18}} \\ &\approx 0.236\end{aligned}$$

Problems

1 Write a computer program similar to the one in example 44 on page 66 to evaluate the definite integral

$$\int_0^1 e^{x^2} \quad .$$

▷ Solution, p. 164

2 Evaluate the integral

$$\int_0^{2\pi} \sin x \, \mathrm{d}x \qquad ,$$

and draw a sketch to explain why your result comes out the way it does. ▷ Solution, p. 164

3 Sketch the graph that represents the definite integral

$$\int_0^2 -x^2 + 2x \qquad ,$$

and estimate the result roughly from the graph. Then evaluate the integral exactly, and check against your estimate.

▷ Solution, p. 165

4 Make a rough guess as to the average value of $\sin x$ for $0 < x < \pi$, and then find the exact result and check it against your guess.

▷ Solution, p. 166

5 Show that the mean value theorem's assumption of continuity is necessary, by exhibiting a discontinuous function for which the theorem fails. ▷ Solution, p. 166

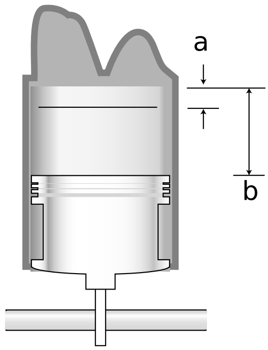
6 Show that the fundamental theorem of calculus's assumption of continuity for $\dot{x}$ is necessary, by exhibiting a discontinuous function for which the theorem fails.

▷ Solution, p. 166

7 Sketch the graphs of $y = x^2$ and $y = \sqrt{x}$ for $0 \le x \le 1$. Graphically, what relationship should exist between the integrals $\int_0^1 x^2 \, \mathrm{d}x$ and $\int_0^1 \sqrt{x} \, \mathrm{d}x$? Compute both integrals, and verify that the results are related in the expected way.

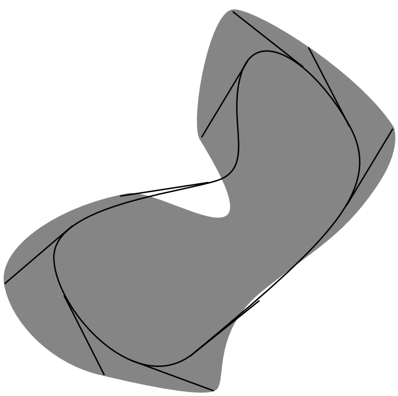
8 In a gasoline-burning car engine, the exploding air-gas mixture makes a force on the piston, and the force tapers off as the piston expands, allowing the gas to expand. (a) In the approximation $F = k/x$, where x is the position of the piston, find the work done on the piston as it travels from $x = a$ to $x = b$, and show that the result only depends on the ratio b/a. This ratio is known as the compression ratio of the engine. (b) A better approximation, which takes into account the cooling of the air-gas mixture as it expands, is $F = kx^{-1.4}$. Compute the work done in this case.

9 A certain variable x varies randomly from -1 to 1, with probability distribution $\mathrm{d}P/\mathrm{d}x = k\left(1 - x^2\right)$.
(a) Determine k from the requirement of normalization.
(b) Find the average value of x.
(c) Find its standard deviation.



Problem 8.

10 A perfectly elastic ball bounces up and down forever, always coming back up to the same height h. Find its average height. $\star$



Problem 11.

11 The figure shows a curve with a tangent line segment of length 1 that sweeps around it, forming a new curve that is usually outside the old one. Prove Holditch's theorem, which states that the new curve's area differs from the old one's by π. (This is an example of a result that is much more difficult to prove without making use of infinitesimals.) $\star$

4 Techniques

4.1 Newton's method

In the 1958 science fiction novel **Have Space Suit — Will Travel**, by Robert Heinlein, Kip is a high school student who wants to be an engineer, and his father is trying to convince him to stretch himself more if he wants to get anything out of his education:

"Why did Van Buren fail of re-election? How do you extract the cube root of eighty-seven?"

Van Buren had been a president; that was all I remembered. But I could answer the other one. "If you want a cube root, you look in a table in the back of the book."

Dad sighed. "Kip, do you think that table was brought down from on high by an archangel?"

We no longer use tables to compute roots, but how does a pocket calculator do it? A technique called Newton's method allows us to calculate the inverse of any function efficiently, including cases that aren't preprogrammed into a calculator. In the example from the novel, we know how to calculate the function $y = x^3$ fairly accurately and quickly for any given value of x, but we want to turn the equation around and find x when $y = 87$. We start with a rough mental guess: since $4^3 = 64$ is a little too small, and $5^3 = 125$ is much too big, we guess $x \approx 4.3$. Testing our guess, we have $4.3^3 = 79.5$. We want y to get bigger by 7.5, and we can use calculus to find approximately how much bigger x needs to get in order to accomplish that:

$$\begin{aligned}\Delta x &= \frac{\Delta x}{\Delta y}\Delta y \\ &\approx \frac{\mathrm{d}x}{\mathrm{d}y}\Delta y \\ &= \frac{\Delta y}{\mathrm{d}y/\mathrm{d}x} \\ &= \frac{\Delta y}{3x^2} \\ &= \frac{\Delta y}{3x^2} \\ &= 0.14\end{aligned}$$

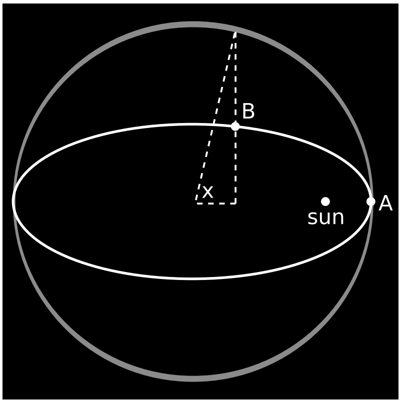
Increasing our value of x to $4.3 + 0.14 = 4.44$, we find that $4.44^3 = 87.5$ is a pretty good approximation to 87. If we need higher precision, we can go through the process again with $\Delta y = -0.5$, giving

$$\begin{aligned}\Delta x &\approx \frac{\Delta y}{3x^2} \\ &= 0.14 \\ x &= 4.43 \\ x^3 &= 86.9 \qquad .\end{aligned}$$

This second iteration gives an excellent approximation.

Example 54

▷ Figure 54 shows the astronomer Johannes Kepler's analysis of the motion



a / Example 54.

of the planets. The ellipse is the orbit of the planet around the sun. At $t = 0$, the planet is at its closest approach to the sun, A. At some later time, the planet is at point B. The angle x (measured in radians) is defined with reference to the imaginary circle encompassing the orbit. Kepler found the equation

$$2\pi\frac{t}{T} = x - e\sin x \quad ,$$

where the period, T, is the time required for the planet to complete a full orbit, and the eccentricity of the ellipse, e, is a number that measures how much it differs from a circle. The relationship is complicated because the planet speeds up as it falls inward toward the sun, and slows down again as it swings back away from it.

The planet Mercury has $e = 0.206$. Find the angle x when Mercury has completed $1/4$ of a period.

▷ We have

$$y = x - (0.206)\sin x \quad ,$$

and we want to find x when $y = 2\pi/4 = 1.57$. As a first guess, we try $x = \pi/2$ (90 degrees), since the eccentricity of Mercury's orbit is actually much smaller than the example shown in the figure, and therefore the planet's speed doesn't vary all that much as it goes around the sun. For this value of x we have $y = 1.36$, which is too small by 0.21.

$$\begin{aligned}\Delta x &\approx \frac{\Delta y}{\mathrm{d}y/\mathrm{d}x}\\ &= \frac{0.21}{1-(0.206)\cos x}\\ &= 0.21\end{aligned}$$

(The derivative $\mathrm{d}y/\mathrm{d}x$ happens to be 1 at $x = \pi/2$.) This gives a new value of x, 1.57+.21=1.78. Testing it, we have $y = 1.58$, which is correct to within rounding errors after only one iteration. (We were only supplied with a value of e accurate to three significant figures, so we can't get a result with precision better than about that level.)

4.2 Implicit differentiation

We can differentiate any function that is written as a formula, and find a result in terms of a formula. However, sometimes the original problem can't be written in any nice way as a formula. For example, suppose we want to find $\mathrm{d}y/\mathrm{d}x$ in a case where the relationship between x and y is given by the following equation:

$$y^7 + y = x^7 + x^2 \quad .$$

There is no equivalent of the quadratic formula for seventh-order polynomials, so we have no way to solve for one variable in terms of the other in order to differentiate it. However, we can still find $\mathrm{d}y/\mathrm{d}x$ in terms of x and y. Suppose we let x grow to $x+\mathrm{d}x$. Then for example the x^2 term will grow to $(x+\mathrm{d}x)^2 = x + 2\mathrm{d}x + \mathrm{d}x^2$. The squared infinitesimal is negligible, so the increase in x^2 was really just $2\mathrm{d}x$, and we've really just computer the derivative of x^2 with respect to x and multiplied it by $\mathrm{d}x$. In symbols,

$$\begin{aligned} \mathrm{d}(x^2) &= \frac{\mathrm{d}(x^2)}{\mathrm{d}x} \cdot \mathrm{d}x \\ &= 2x\,\mathrm{d}x \qquad . \end{aligned}$$

That is, the change in x^2 is $2x$ times the change in x. Doing this to both sides of the original equation, we have

$$\begin{aligned} \mathrm{d}(y^7+y) &= \mathrm{d}(x^7+x^2) \\ 7y^6\,\mathrm{d}y + 1\,\mathrm{d}y &= 7x^6\,\mathrm{d}x + 2x\,\mathrm{d}x \\ (7y^6+1)\mathrm{d}y &= (7x^6+2x)\mathrm{d}x \\ \frac{\mathrm{d}y}{\mathrm{d}x} &= \frac{7x^6+2x}{7y^6+1} \qquad . \end{aligned}$$

This still doesn't give us a formula for the derivative in terms of x alone, but it's not entirely useless. For instance, if we're given a numerical value of x, we can always use Newton's method to find y, and then evaluate the derivative.

4.3 Methods of integration

Change of variable

Sometimes and unfamiliar-looking integral can be made into a familiar one by substituting a a new variable for an old one. For example, we know how to integrate $1/x$ — the answer is $\ln x$ — but what about

$$\int \frac{\mathrm{d}x}{2x+1} \qquad ?$$

Let $u = 2x+1$. Differentiating both sides, we have $\mathrm{d}u = 2\mathrm{d}x$, or $\mathrm{d}x = \mathrm{d}u/2$, so

$$\begin{aligned} \int \frac{\mathrm{d}x}{2x+1} &= \int \frac{\mathrm{d}u/2}{u} \\ &= \frac{1}{2}\ln u + c \\ &= \frac{1}{2}\ln(2x+1) + c \qquad . \end{aligned}$$

In the case of a definite integral, we have to remember to change the limits of integration to reflect the new variable.

Example 55

▷ Evaluate $\int_3^4 \mathrm{d}x/(2x+1)$.

▷ As before, let $u = 2x+1$.

$$\begin{aligned} \int_{x=3}^{x=4} \frac{\mathrm{d}x}{2x+1} &= \int_{u=7}^{u=9} \frac{\mathrm{d}u/2}{u} \\ &= \frac{1}{2}\ln u \Big|_{u=7}^{u=9} \end{aligned}$$

Here the notation $|_{u=7}^{u=9}$ means to evaluate the function at 7 and 9, and subtract the former from the latter. The result is

$$\int_{x=3}^{x=4} \frac{dx}{2x+1} = \frac{1}{2}(\ln 9 - \ln 7) = \frac{1}{2}\ln\frac{9}{7} \quad .$$

Sometimes, as in the next example, a clever substitution is the secret to doing a seemingly impossible integral.

Example 56

▷ Evaluate

$$\int \frac{e^{\sqrt{x}}}{\sqrt{x}} dx \quad .$$

▷ The only hope for reducing this to a form we can do is to let $u = \sqrt{x}$. Then $dx = d(u^2) = 2u du$, so

$$\begin{aligned}\int \frac{e^{\sqrt{x}}}{\sqrt{x}} dx &= \int \frac{e^u}{u} \cdot 2u\,du \\ &= 2\int e^u du \\ &= 2e^u \\ &= 2e^{\sqrt{x}} \quad .\end{aligned}$$

Example 56 really isn't so tricky, since there was only one logical choice for the substitution that had any hope of working. The following is a little more dastardly.

Example 57

▷ Evaluate

$$\int \frac{dx}{1+x^2} \quad .$$

▷ The substitution that works is $x = \tan u$. First let's see what this does to the expression $1 + x^2$. The familiar identity

$$\sin^2 u + \cos^2 u = 1 \quad ,$$

when divided by $\cos^2 u$, gives

$$\tan^2 u + 1 = \sec^2 u \quad ,$$

so $1 + x^2$ becomes $\sec^2 u$. But differentiating both sides of $x = \tan u$ gives

$$\begin{aligned}dx &= d\left[\sin u(\cos u)^{-1}\right] \\ &= (d\sin u)(\cos u)^{-1} \\ &\quad + (\sin u)d\left[(\cos u)^{-1}\right] \\ &= \left(1+\tan^2 u\right) du \\ &= \sec^2 u\,du \quad ,\end{aligned}$$

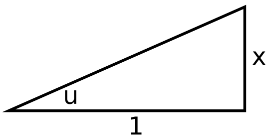
so the integral becomes

$$\begin{aligned}\int \frac{dx}{1+x^2} &= \int \frac{\sec^2 u du}{\sec^2 u} \\ &= u + c \\ &= \tan^{-1} x + c \quad .\end{aligned}$$

What mere mortal would ever have suspected that the substitution $x = \tan u$ was the one that was needed in example 57? One possible answer is to give up and do the integral on a computer:

```
Integrate(x) 1/(1+x^2)
  ArcTan(x)
```

Another possible answer is that you can usually smell the possibility of this type of substitution, involving a trig function, when the thing to be integrated contains something reminiscent of the Pythagorean theorem, as suggested by figure b. The $1 + x^2$ looks like what you'd get if you had a right triangle with legs 1 and x, and were using the Pythagorean theorem to find its hypotenuse.



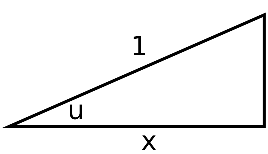
b / The substitution $x = \tan u$.

Example 58

▷ Evaluate $\int \mathrm{d}x/\sqrt{1-x^2}$.

▷ The $\sqrt{1-x^2}$ looks like what you'd get if you had a right triangle with hypotenuse 1 and a leg of length x, and were using the Pythagorean theorem to find the other leg, as in figure c. This motivates us to try the substitution $x = \cos u$, which gives $\mathrm{d}x = -\sin u\ \mathrm{d}u$ and $\sqrt{1-x^2} = \sqrt{1-\cos^2 u} = \sin u$. The result is

$$\begin{aligned}\int \frac{\mathrm{d}x}{\sqrt{1-x^2}} &= \int \frac{-\sin u\,\mathrm{d}u}{\sin u}\\ &= u + c\\ &= \cos^{-1} x \quad .\end{aligned}$$

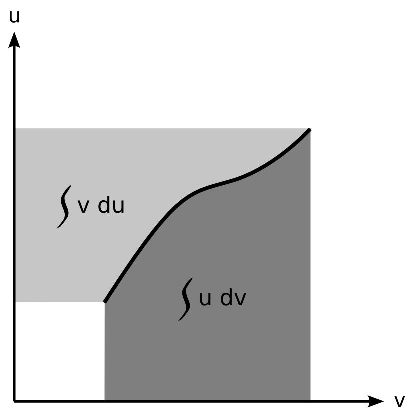


c / The substitution $x = \cos u$.

Integration by parts

Figure d shows a technique called integration by parts. If the integral $\int v\mathrm{d}u$ is easier than the integral $\int u\mathrm{d}v$, then we can calculate the easier one, and then by simple geometry determine the one we wanted. Identifying the large rectangle that surrounds both shaded areas, and the small white rectangle on the lower left, we have

$$\begin{aligned}\int u\,\mathrm{d}v =&(\text{area of large rectangle})\\ &-(\text{area of small rectangle})\\ &\int v\,\mathrm{d}u \quad .\end{aligned}$$



d / Integration by parts.

In the case of an indefinite integral, we have a similar relationship derived from the product rule:

$$\begin{aligned} \mathrm{d}(uv) &= u\,\mathrm{d}v + v\,\mathrm{d}u \\ u\,\mathrm{d}v &= \mathrm{d}(uv) - v\,\mathrm{d}u \end{aligned}$$

Integrating both sides, we have the following relation.

Integration by parts

$$\int u\,\mathrm{d}v = uv - \int v\,\mathrm{d}u \qquad .$$

Since a definite integral can always be done by evaluating an indefinite integral at its upper and lower limits, one usually uses this form. Integrals don't usually come prepackaged in a form that makes it obvious that you should use integration by parts. What the equation for integration by parts tells us is that if we can split up the integrand into two factors, one of which (the $\mathrm{d}v$) we know how to integrate, we have the option of changing the integral into a new form in which that factor becomes its integral, and the other factor becomes its derivative. If we choose the right way of splitting up the integrand into parts, the result can be a simplification.

Example 59

▷ Evaluate

$$\int x \cos x \,\mathrm{d}x$$

▷ There are two obvious possibilities for splitting up the integrand into factors,

$$u\,\mathrm{d}v = (x)(\cos x\,\mathrm{d}x)$$

or

$$u\,\mathrm{d}v = (\cos x)(x\,\mathrm{d}x) \qquad .$$

The first one is the one that lets us make progress. If $u = x$, then $\mathrm{d}u = \mathrm{d}x$, and if $\mathrm{d}v = \cos x\,\mathrm{d}x$, then integration gives $v = \sin x$.

$$\begin{aligned} \int x\cos x\,\mathrm{d}x &= \int u\,\mathrm{d}v \\ &= uv - \int v\,\mathrm{d}u \\ &= x\sin x - \int \sin x\,\mathrm{d}x \\ &= x\sin x + \cos x \end{aligned}$$

Of the two possibilities we considered for u and $\mathrm{d}v$, the reason this one helped was that differentiating x gave $\mathrm{d}x$, which was simpler, and integrating $\cos x\mathrm{d}x$ gave $\sin x$, which was no more complicated than before. The second possibility would have made things worse rather than better, because integrating $x\mathrm{d}x$ would have given $x^2/2$, which would have been more complicated rather than less.

Example 60

▷ Evaluate $\int \ln x\,\mathrm{d}x$.

▷ This one is a little tricky, because it isn't explicitly written as a product, and yet we can attack it using integration by parts. Let $u = \ln x$ and $\mathrm{d}v = \mathrm{d}x$.

$$\begin{aligned}\int \ln x\, \mathrm{d}x &= \int u\, \mathrm{d}v \\ &= uv - \int v\, \mathrm{d}u \\ &= x \ln x - \int x \frac{\mathrm{d}x}{x} \\ &= x \ln x - x\end{aligned}$$

Partial fractions

Given a function like

$$\frac{-1}{x-1} + \frac{1}{x+1} \qquad ,$$

we can rewrite it over a common denominator like this:

$$\begin{aligned}&\left(\frac{-1}{x-1}\right)\left(\frac{x+1}{x+1}\right) \\ &+ \left(\frac{1}{x+1}\right)\left(\frac{x-1}{x-1}\right) \\ &= \frac{-x-1+x-1}{(x-1)(x+1)} \\ &= \frac{-2}{x^2-1} \qquad .\end{aligned}$$

But note that the original form is easily integrated to give

$$\int \left(\frac{-1}{x-1} + \frac{1}{x+1}\right) \mathrm{d}x$$
$$= -\ln(x-1) + \ln(x+1) + c \qquad ,$$

while faced with the form $-2/(x^2-1)$, we wouldn't have known how to integrate it.

The idea of the method of partial fractions is that if we want to do an integral of the form

$$\int \frac{\mathrm{d}x}{P(x)} \qquad ,$$

where $P(x)$ is an nth order polynomial, we can always rewrite $1/P$ as

$$\frac{1}{P(x)} = \frac{A_1}{x-r_1} + \ldots \frac{A_n}{x-r_n} \qquad ,$$

where $r_1 \ldots r_n$ are the roots of the polynomial, i.e., the solutions of the equation $P(r) = 0$. If the polynomial is second-order, you can find the roots r_1 and r_2 using the quadratic formula; I'll assume for the time being that they're real. For higher-order polynomials, there is no surefire, easy way of finding the roots by hand, and you'd be smart simply to use computer software to do it. In Yacas, you can find the real roots of a polynomial like this:

```
FindRealRoots(x^4-5*x^3
    -25*x^2+65*x+84)
  {3.,7.,-4.,-1.}
```

(I assume it uses Newton's method to find them.) The constants A_i can then be determined by algebra, or by the trick of evaluating $1/P(x)$ for a value of x very close to one of the roots. In the example of the polynomial $x^4 - 5x^3 - 25x^2 + 65x + 84$, let $r_1 \ldots r_4$ be the roots in the order in which they were returned by Yacas. Then A_1 can be found by evaluating $1/P(x)$ at $x = 3.000001$:

```
P(x):=x^4-5*x^3-25*x^2
  +65*x+84
N(1/P(3.000001))
  -8928.5702094768
```

We know that for x very close to 3, the expression

$$\frac{1}{P}=\frac{A_1}{x-3}+\frac{A_2}{x-7}+\frac{A_3}{x+4}+\frac{A_4}{x+1}$$

will be dominated by the A_1 term, so

$$-8930 \approx \frac{A_1}{3.000001-3}$$
$$A_1 \approx (-8930)(10^{-6}) \qquad .$$

By the same method we can find the other four constants:

```
dx:=.000001
N(1/P(7+dx),30)*dx
  0.2840908276e-2
N(1/P(-4+dx),30)*dx
  -0.4329006192e-2
N(1/P(-1+dx),30)*dx
  0.1041666664e-1
```

(The `N( ,30)` construct is to tell Yacas to do a numerical calculation rather than an exact symbolic one, and to use 30 digits of precision, in order to avoid problems with rounding errors.) Thus,

$$\frac{1}{P}=\frac{-8.93\times10^{-3}}{x-3}+\frac{2.84\times10^{-3}}{x-7}-\frac{4.33\times10^{-3}}{x+4}+\frac{1.04\times10^{-2}}{x+1} \qquad .$$

The desired integral is

$$\int\frac{dx}{P(x)}=-8.93\times10^{-3}\ln(x-3)+2.84\times10^{-3}\ln(x-7)-4.33\times10^{-3}\ln(x+4)+1.04\times10^{-2}\ln(x+1)+c \qquad .$$

There are some possible complications: (1) The same factor may occur more than once, as in $x^3-5x^2+7x-3=(x-1)(x-1)(x-3)$. In this example, we have to look for an answer of the form $A/(x-1)+B/(x-1)^2+C/(x-3)$, the solution being $-.25/(x-1)-.5/(x-1)^2+.25/(x-3)$. (2) The roots may be complex. This is no show-stopper if you're using computer software that handles complex numbers gracefully. (You can choose a c that makes the result real.) In fact, as discussed in section 7.3, some beautiful things can happen with complex roots. But as an alternative, any polynomial with real coefficients can be factored into linear and quadratic factors with real coefficients. For each quadratic factor $Q(x)$, we then have a partial fraction of the form $(A+Bx)/Q(x)$, where A and B can be determined by algebra. In Yacas, this can be done using the `Apart` function.

Example 61

▷ Evaluate the integral

$$\int\frac{dx}{(x^4-8*x^3+8*x^2-8*x+7}$$

using the method of partial fractions.

▷ First we use Yacas to look for real roots of the polynomial:

```
FindRealRoots(x^4-8*x^3
  +8*x^2-8*x+7)
   {1.,7.}
```

Unfortunately this polynomial seems to have only two real roots; the rest are complex. We can divide out the factor $(x-1)(x-7)$, but that still leaves us with a second-order polynomial, which has no real roots. One approach would be to factor the polynomial into the form $(x-1)(x-7)(x-p)(x-q)$, where p and q are complex, as in section 7.3. Instead, let's use Yacas to expand the integrand in terms of partial fractions:

```
Apart(1/(x^4-8*x^3
  +8*x^2-8*x+7))
   ((2*x)/25+3/50)/(x^2+1)
     +1/(300*(x-7))
     +(-1)/(12*(x-1))
```

We can now rewrite the integral like this:

$$\begin{aligned}&\frac{2}{25}\int\frac{x\,\mathrm{d}x}{x^2+1}\\+&\frac{3}{50}\int\frac{\mathrm{d}x}{x^2+1}\\+&\frac{1}{300}\int\frac{\mathrm{d}x}{x-7}\\-&\frac{1}{12}\int\frac{\mathrm{d}x}{x-1}\end{aligned}$$

which we can evaluate as follows:

$$\begin{aligned}&\frac{1}{25}\ln(x^2+1)\\+&\frac{3}{50}\tan^{-1}x\\+&\frac{1}{300}\ln(x-7)\\-&\frac{1}{12}\ln(x-1)\\+&c\end{aligned}$$

In fact, Yacas should be able to do the whole integral for us from scratch, but it's best to understand how these things work under the hood, and to avoid being completely dependent on one particular piece of software. As an illustration of this gem of wisdom, I found that when I tried to make Yacas evaluate the integral in one gulp, it choked because the calculation became too complicated! Because I understood the ideas behind the procedure, I was still able to get a result through a mixture of computer calculations and working it by hand. Someone who didn't have the knowledge of the technique might have tried the integral using the software, seen it fail, and concluded, incorrectly, that the integral was one that simply couldn't be done. A computer is no substitute for understanding.

Problems

1 Graph the function $y = e^x - 7x$ and get an approximate idea of where any of its zeroes are (i.e., for what values of x we have $y(x) = 0$). Use Newton's method to find the zeroes to three significant figures of precision.

2 The relationship between x and y is given by $xy = \sin y + x^2y^2$.
(a) Use Newton's method to find the nonzero solution for y when $x = 3$. Answer: $y = 0.2231$
(b) Find $\mathrm{d}y/\mathrm{d}x$ in terms of x and y, and evaluate the derivative at the point on the curve you found in part a. Answer: $\mathrm{d}y/\mathrm{d}x = -0.0379$
Based on an example by Craig B. Watkins.

3 Suppose you want to evaluate

$$\int \frac{\mathrm{d}x}{1+\sin 2x} \qquad ,$$

and you've found

$$\int \frac{\mathrm{d}x}{1+\sin x} = -\tan\left(\frac{\pi}{4} - \frac{x}{2}\right)$$

in a table of integrals. Use a change of variable to find the answer to the original problem.

4 Evaluate

$$\int \frac{\sin x\mathrm{d}x}{1+\cos x} \qquad .$$

5 Evaluate

$$\int \frac{\sin x\mathrm{d}x}{1+\cos^2 x} \qquad .$$

6 Evaluate

$$\int xe^{-x^2}\,\mathrm{d}x \qquad .$$

7 Evaluate

$$\int xe^x\,\mathrm{d}x \qquad .$$

8 Use integration by parts to evaluate the following integrals.

$$\int \sin^{-1} x\,\mathrm{d}x$$

$$\int \cos^{-1} x\,\mathrm{d}x$$

$$\int \tan^{-1} x\,\mathrm{d}x$$

9 Evaluate

$$\int x^2 \sin x\,\mathrm{d}x \qquad .$$

Hint: Use integration by parts more than once.

10 Evaluate

$$\int \frac{\mathrm{d}x}{x^2 - x - 6} \qquad .$$

11 Evaluate

$$\int \frac{\mathrm{d}x}{x^3 + 3x^2 - 4} \qquad .$$

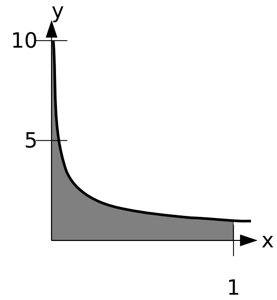
12 Evaluate

$$\int \frac{\mathrm{d}x}{x^3 - x^2 + 4x - 4} \qquad .$$

5 Improper integrals

5.1 Integrating a function that blows up

When we integrate a function that blows up to infinity at some point in the interval we're integrating, the result may be either finite or infinite.



a / The integral $\int_0^1 \mathrm{d}x/\sqrt{x}$ is finite.

Example 62

▷ Integrate the function $y = 1/\sqrt{x}$ from $x = 0$ to $x = 1$.

▷ The function blows up to infinity at one end of the region of integration, but let's just try evaluating it, and see what happens.

$$\begin{aligned}\int_0^1 x^{-1/2}\,\mathrm{d}x &= 2x^{1/2}\Big|_0^1 \\ &= 2\end{aligned}$$

The result turns out to be finite. Intuitively, the reason for this is that the spike at $x = 0$ is very skinny, and gets skinny fast as we go higher and higher up.

Example 63

▷ Integrate the function $y = 1/x^2$ from $x = 0$ to $x = 1$.

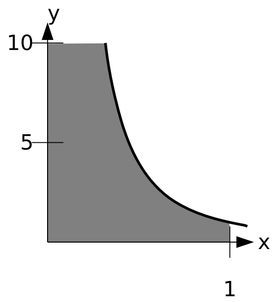
▷

$$\begin{aligned}\int_0^1 x^{-2}\,\mathrm{d}x &= -x^{-1}\Big|_0^1 \\ &= -1 + \frac{1}{0}\end{aligned}$$

Division by zero is undefined, so the result is undefined.

Another way of putting it, using the hyperreal number system, is that if we were to integrate from ϵ to 1, where ϵ was an infinitesimal number, then the result would be $-1 + 1/\epsilon$, which is infinite. The smaller we make ϵ, the bigger the infinite result we get out.

Intuitively, the reason that this integral comes out infinite is that the spike at $x = 0$ is fat, and doesn't get skinny fast enough.



b / The integral $\int_0^1 \mathrm{d}x/x^2$ is infinite.

These two examples were examples of improper integrals.

5.2 Limits of integration at infinity

Another type of improper integral is one in which one of the limits of integration is infinite. The notation

$$\int_a^\infty f(x)\,\mathrm{d}x$$

means the limit of $\int_a^H f(x)\,\mathrm{d}x$, where H is made to grow bigger and bigger. Alternatively, we can think of it as an integral in which the top end of the interval of integration is an infinite hyperreal number. A similar interpretation applies when the lower limit is $-\infty$, or when both limits are infinite.

Example 64

▷ Evaluate

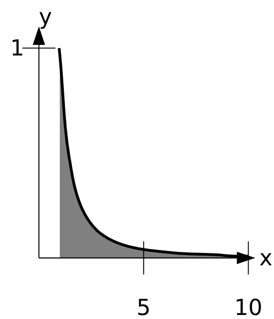
$$\int_1^\infty x^{-2}\,\mathrm{d}x$$

▷

$$\begin{aligned}\int_1^H x^{-2}\,\mathrm{d}x &= -x^{-1}\Big|_1^H \\ &= -\frac{1}{H}+1\end{aligned}$$

As H gets bigger and bigger, the result gets closer and closer to 1, so the result of the improper integral is 1.

Note that this is the same graph as in example 62, but with the x and y axes interchanged; this shows that the two different types of improper integrals really aren't so different.



c / The integral $\int_1^\infty \mathrm{d}x/x^2$ is finite.

Example 65

▷ Newton's law of gravity states that the gravitational force between two objects is given by $F = Gm_1m_2/r^2$, where G is a constant, m_1 and m_2 are the objects' masses, and r is the center-to-center distance between them. Compute the work that must be done to take an object from the earth's surface, at $r = a$, and remove it to $r = \infty$.

▷

$$\begin{aligned}
W &= \int_a^\infty \frac{Gm_1 m_2}{r^2}\,\mathrm{d}r \\
&= Gm_1 m_2 \int_a^\infty r^{-2}\,\mathrm{d}r \\
&= -Gm_1 m_2\ r^{-1}\Big|_a^\infty \\
&= \frac{Gm_1 m_2}{a}
\end{aligned}$$

The answer is inversely proportional to a. In other words, if we were able to start from higher up, less work would have to be done.

Problems

1 Integrate

$$\int_0^\infty e^{-x} \mathrm{d}x \qquad ,$$

or show that it diverges.

2 Integrate

$$\int_1^\infty \frac{\mathrm{d}x}{x} \qquad ,$$

or show that it diverges.

3 Integrate

$$\int_0^1 \frac{\mathrm{d}x}{x} \qquad ,$$

or show that it diverges.

4 Integrate

$$\int_0^\infty e^{-x} \cos x \, \mathrm{d}x$$

or show that it diverges.

5 Prove that

$$\int_0^\infty e^{-e^x} \, \mathrm{d}x$$

converges, but don't evaluate it.

6 (a) Verify that the probability distribution $\mathrm{d}P/\mathrm{d}x$ given in example 51 on page 72 is properly normalized.
(b) Find the average value of x, or show that it diverges.
(c) Find the standard deviation of x, or show that it diverges.

7 Prove

$$\int_0^\infty e^{-x} x^n \, \mathrm{d}x = n! \qquad .$$

⋆

6 Sequences and Series

6.1 Infinite sequences

Consider an infinite sequence of numbers like 1/2, 2/3, 3/4, 4/5, ... We want to define this as approaching 1, or "converging to 1." The way to do this is to make a function $f(n)$, which is only well defined for integer values of n. Then $f(1) = 1/2$, $f(2) = 2/3$, and in general $f(n) = n/(n+1)$. With just a little tinkering, our definitions of limits can be applied to this type of function (see problem 1 on page 100).

6.2 Infinite series

A related question is how to rigorously define the sum of infinitely many numbers, which is referred to as an infinite *series*. An example is the geometric series $1 + x + x^2 + x^3 + \ldots = 1/(1-x)$, which we used casually on page 29. The general concept of an infinite series goes back to ancient Greek mathematics. Various supposed paradoxes about infinite series, such as Zeno's paradox, were exhibited, influencing Euclid to sidestep the issue in his *Elements*, where in Book IX, Proposition 35 he provides only an expression $(1-x^n)/(1-x)$ for the nth partial sum of the geometric series. The case where n gets so big that x^n becomes negligible is left to the reader's imagination, as in one of those scenes in a romance novel that ends with something like "...and she surrendered..." For those with modern training, the idea is that an infinite sum like $1 + 1 + 1 + \ldots$ would clearly give an infinite result, but this is only because the terms are all staying the same size. If the terms get smaller and smaller, and get smaller fast enough, then the result can be finite. For example, consider the geometric series in the case where $x = 1/2$, for which we expect the result $1/(1-1/2) = 2$. We have

$$1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \ldots \qquad ,$$

which at the successive steps of addition equals 1, $1\frac{1}{2}$, $1\frac{3}{4}$, $1\frac{7}{8}$, $1\frac{15}{16}$, ... We're getting closer and closer to 2, cutting the distance in half at each step. Clearly we can get as close as we like to 2, if we're willing to add enough terms.

Note that we ended up wanting to talk about the partial sums of the series. This is the right way to get a rigorous definition of the convergence of series in general. In the case of the geometric series, for example, we can define a sequence of the partial sums 1, $1+x$, $1+x+x^2$, ... We can then define convergence and limits of series in terms of convergence and limits of the partial sums.

It's instructive to see what happens to the geometric series with $x = 0.1$. The geometric series becomes

$$1 + 0.1 + 0.01 + 0.001 + \ldots \qquad .$$

The partial sums are 1, 1.1, 1.11, 1.111, ... We can see vividly here that adding another term will only affect the result in a certain decimal place, without affecting any of the earlier ones. For instance, if we needed a result that was valid to three digits past the decimal place, we could stop at 1.111, being assured that we had attained a good enough approximation. If we wanted an exact result, we could also observe that multiplying the result by 9 would give $9.999\ldots$, which is the same as 10, so the result must be $10/9$, which is in agreement with $1/(1 - 1/10) = 10/9$.

One thing to watch out for with infinite series is that the axioms of the real number system only talk about finite sums, so it's easy to get wrong results by attempting to apply them to infinite ones (see problem 2 on page 100).

6.3 Tests for convergence

There are many different tests that can be used to determine whether a sequence or series converges. I'll briefly state three of the most useful, with sketches of their proofs.

Bounded and increasing sequences: A sequence that always increases, but never surpasses a certain value, converges.

This amounts to a restatement of the compactness axiom for the real numbers stated on page 137, and is therefore to be interpreted not so much as a statement about sequences but as one about the real number system. In particular, it fails if interpreted as a statement about sequences confined entirely to the rational number system, as we can see from the sequence 1, 1.4, 1.41, 1.414, ... consisting of the successive decimal approximations to $\sqrt{2}$, which does not converge to any rational-number value.

Example 66

▷ Prove that the geometric series $1 + 1/2 + 1/4 + \ldots$ converges.

▷ The sequence of partial sums is increasing, since each term is positive. Each term closes half of the remaining gap separating the previous partial sum from 2, so the sum never surpasses 2. Since the partial sums are increasing and bounded, they converge to a limit.

Once we know that a particular series converges, we can also easily infer the convergence of other series whose terms get smaller faster. For example, we can be certain that if the geometric series converges, so does the series

$$\frac{1}{1} + \frac{1}{1 \times 2} + \frac{1}{1 \times 2 \times 3} + \ldots \qquad ,$$

whose terms get smaller faster than any base raised to the power n.

Alternating series with decreasing terms: If the terms of a series alternate in sign and approach zero, then the series converges.

Sketch of a proof: The even partial sums form an increasing sequence, the odd sums a decreasing one. Neither of these sequences of partial sums can be unbounded, since the difference between partial sums n and $n+1$ would then have to be unbounded, but this difference is simply the nth term, and the terms approach zero. Since the even partial sums are increasing and bounded, they converge to a limit, and similarly for the odd ones. The two limits must be equal, since the terms approach zero.

Example 67

▷ Prove that the series $1 - 1/2 + 1/3 - 1/4 + \ldots$ converges.

▷ Its convergence follows because it is an alternating series with decreasing terms. The sum turns out to be $\ln 2$, although the convergence of the series is so slow that an extremely large number of terms is required in order to obtain a decent approximation,

The integral test: If the terms of a series a_n are positive and decreasing, and $f(x)$ is a positive and decreasing function on the real number line such that $f(n) = a_n$, then the sum of a_n from $n = 1$ to ∞ converges if and only if $\int_1^\infty f(x)\mathrm{d}x$ does.

Sketch of proof: Since the theorem is supposed to hold for both convergence and divergence, and is also an "if and only if," there are actually four cases to prove, of which we pick the representative one where the integral is known to converge and we want to prove convergence of the corresponding sum. The sum and the integral can be interpreted as the areas under two graphs: one like a smooth ramp and one like a staircase. Sliding the staircase half a unit to the left, it lies entirely underneath the ramp, and therefore the area under it is also finite.

Example 68

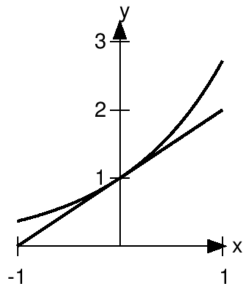
▷ Prove that the series $1+1/2+1/3+\ldots$ diverges.

▷ The integral of $1/x$ is $\ln x$, which diverges as x approaches infinity, so the series diverges as well.

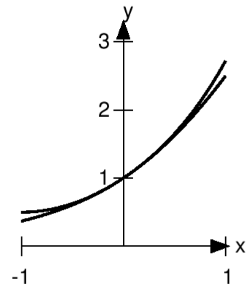
6.4 Taylor series

If you calculate $e^{0.1}$ on your calculator, you'll find that it's very close to 1.1. This is because the tangent line at $x = 0$ on the graph of e^x has a slope of 1 ($\mathrm{d}e^x/\mathrm{d}x = e^x = 1$ at $x = 0$), and the tangent line is a good approximation to the exponential curve as long as we don't get too far away from the point of tangency.

How big is the error? The



a / The function e^x, and the tangent line at $x = 0$.



b / The function e^x, and the approximation $1 + x + x^2/2$.

actual value of $e^{0.1}$ is $1.10517091807565\ldots$, which differs from 1.1 by about 0.005. If we go farther from the point of tangency, the approximation gets worse. At $x = 0.2$, the error is about 0.021, which is about four times bigger. In other words, doubling x seems to roughly quadruple the error, so the error is proportional to x^2; it seems to be about $x^2/2$. Well, if we want a handy-dandy, super-accurate estimate of e^x for small values of x, why not just account for this error. Our new and improved estimate is

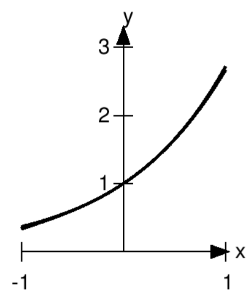
$$e^x \approx 1 + x + \frac{1}{2}x^2$$

for small values of x.

Figure b shows that the approximation is now extremely good for sufficiently small values of x. The difference is that whereas $1 + x$ matched both the y-intercept and the slope of the curve, $1+x+x^2/2$ matches the curvature as well. Recall that the second derivative is a measure of curvature. The second derivatives of the function and its approximation are

$$\frac{\mathrm{d}}{\mathrm{d}x}e^x = 1$$
$$\frac{\mathrm{d}}{\mathrm{d}x}\left(1 + x + \frac{1}{2}x^2\right) = 1$$

We can do even better. Suppose

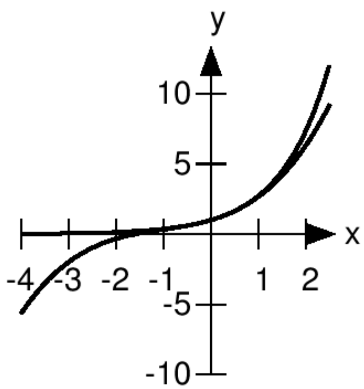


c / The function e^x, and the approximation $1 + x + x^2/2 + x^3/6$.

we want to match the third derivatives. All the derivatives of e^x, evaluated at $x = 0$, are 1, so we just need to add on a term proportional to x^3 whose third derivative is one. Taking the first derivative will bring down a factor of 3 in front, and taking and the second derivative will give a 2, so to cancel these out we need the third-order term to be (1/2)(1/3):

$$e^x \approx 1 + x + \frac{1}{2}x^2 + \frac{1}{2 \cdot 3}x^3$$

Figure c shows the result. For a significant range of x values close to zero, the approximation is now so good that we can't even see the difference between the two functions on the graph.



d / The function e^x, and the approximation $1 + x + x^2/2 + x^3/6$, on a wider scale.

On the other hand, figure d shows that the cubic approximation for somewhat larger negative and positive values of x is poor — worse, in fact, than the linear approximation, or even the constant approximation $e^x = 1$. This is to be expected, because any polynomial will blow up to either positive or negative infinity as x approaches negative infinity, whereas the function e^x is supposed to get very close to zero for large negative x. The idea here is that derivatives are *local* things: they only measure the properties of a function very close to the point at which they're evaluated, and they don't necessarily tell us anything about points far away.

It's a remarkable fact, then, that by taking enough terms in a polynomial approximation, we can always get as good an approximation to e^x as necessary — it's just that a large number of terms may be required for large values of x. In other words, the *infinite series*

$$1 + x + \frac{1}{2}x^2 + \frac{1}{2 \cdot 3}x^3 + \ldots$$

always gives exactly e^x. But what is the pattern here that would allows us to figure out, say, the fourth-order and fifth-order terms that were swept under the rug with the symbol "..."? Let's do the fifth-order term as an example. The point of adding in a fifth-order term is to make the fifth derivative of the approximation equal to the fifth derivative of e^x, which is 1. The first, second, ... derivatives of

x^5 are

$$\frac{\mathrm{d}}{\mathrm{d}x}x^5 = 5x^4$$
$$\frac{\mathrm{d}^2}{\mathrm{d}x^2}x^5 = 5 \cdot 4x^3$$
$$\frac{\mathrm{d}^3}{\mathrm{d}x^3}x^5 = 5 \cdot 4 \cdot 3x^2$$
$$\frac{\mathrm{d}^4}{\mathrm{d}x^4}x^5 = 5 \cdot 4 \cdot 3 \cdot 2x$$
$$\frac{\mathrm{d}^5}{\mathrm{d}x^5}x^5 = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1$$

The notation for a product like $1 \cdot 2 \cdot \ldots \cdot n$ is $n!$, read "n factorial." So to get a term for our polynomial whose fifth derivative is 1, we need $x^5/5!$. The result for the infinite series is

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} \qquad ,$$

where the special case of $0! = 1$ is assumed.[1] This infinite series is called the *Taylor series* for e^x, evaluated around $x = 0$, and it's true, although I haven't proved it, that this particular Taylor series always converges to e^x, no matter how far x is from zero.

In general, the Taylor series around $x = 0$ for a function y is

$$T_0(x) = \sum_{n=0}^{\infty} a_n x^n \qquad ,$$

where the condition for equality of the nth order derivative is

$$a_n = \frac{1}{n!} \frac{\mathrm{d}^n y}{\mathrm{d}x^n}\bigg|_{x=0} \qquad .$$

[1] This makes sense, because, for example, 4!=5!/5, 3!=4!/4, etc., so we should have 0!=1!/1.

Here the notation $|_{x=0}$ means that the derivative is to be evaluated at $x = 0$.

A Taylor series can be used to approximate other functions besides e^x, and when you ask your calculator to evaluate a function such as a sine or a cosine, it may actually be using a Taylor series to do it. Taylor series are also the method Inf uses to calculate most expressions involving infinitesimals. In example 10 on page 29, we saw that when Inf was asked to calculate $1/(1-d)$, where d was infinitesimal, the result was the geometric series:

```
: 1/(1-d)
  1+d+d^2+d^3+d^4
```

These are also the the first five terms of the Taylor series for the function $y = 1/(1-x)$, evaluated around $x = 0$. That is, the geometric series $1 + x + x^2 + x^3 + \ldots$ is really just one special example of a Taylor series, as demonstrated in the following example.

Example 69

▷ Find the Taylor series of $y = 1/(1-x)$ around $x = 0$.

▷ Rewriting the function as $y = (1-x)^{-1}$ and applying the chain rule, we have

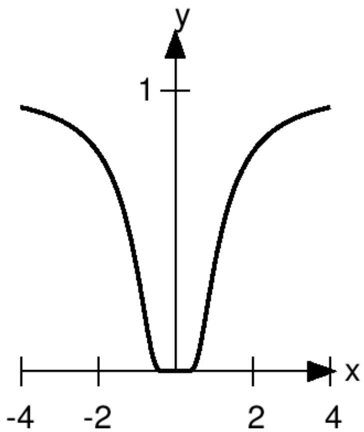
$$
\begin{aligned}
y\big|_{x=0} &= 1\\
\left.\frac{dy}{dx}\right|_{x=0} &= (1-x)^{-2}\Big|_{x=0} = 1\\
\left.\frac{d^2y}{dx^2}\right|_{x=0} &= 2(1-x)^{-3}\Big|_{x=0} = 2\\
\left.\frac{d^3y}{dx^3}\right|_{x=0} &= 2\cdot 3(1-x)^{-4}\Big|_{x=0} = 2\cdot 3\\
&\ldots
\end{aligned}
$$

The pattern is that the nth derivative is $n!$. The Taylor series therefore has $a_n = n!/n! = 1$:

$$\frac{1}{1-x} = 1 + x + x^2 + x^3 + \ldots$$

If you flip back to page 92 and compare the rate of convergence of the geometric series for $x = 0.1$ and 0.5, you'll see that the sum converged much more quickly for $x = 0.1$ than for $x = 0.5$. In general, we expect that any Taylor series will converge more quickly when x is smaller. Now consider what happens at $x = 1$. The series is now $1 + 1 + 1 + \ldots$, which gives an infinite result, and we shouldn't have expected any better behavior, since attempting to evaluate $1/(1-x)$ at $x = 1$ gives division by zero. For $x > 1$, the results become nonsense. For example, $1/(1-2) = -1$, which is finite, but the geometric series gives $1 + 2 + 4 + \ldots$, which is infinite.

In general, every function's Taylor series around $x = 0$ converges for all values of x in the range defined by $|x| < r$, where r is some number, known as the radius of convergence. Also, if the function is defined by putting together other functions that are well behaved (in the sense of converging to their own Taylor series in the relevant region), then the Taylor series will not only converge but converge to the *correct* value. For the function e^x, the radius happen to be infinite, whereas for $1/(1-x)$ it equals 1. The following example shows a worst-case scenario.



e / The function e^{-1/x^2} never converges to its Taylor series.

Example 70

The function $y = e^{-1/x^2}$, shown in figure e, never converges to its Taylor series, except at $x = 0$. This is because the Taylor series for this function, evaluated around $x = 0$ is exactly zero! At $x = 0$, we have $y = 0$, $dy/dx = 0$,

$d^2y/dx^2 = 0$, and so on for every derivative. The zero function matches the function $y(x)$ and all its derivatives to all orders, and yet is useless as an approximation to $y(x)$. The radius of convergence of the Taylor series is infinite, but it doesn't give correct results except at $x = 0$. The reason for this is that y was built by composing two functions, $w(x) = -1/x^2$ and $y(w) = e^w$. The function w is badly behaved at $x = 0$ because it blows up there. In particular, it doesn't have a well-defined Taylor series at $x = 0$.

Example 71

▷ Find the Taylor series of $y = \sin x$, evaluated around $x = 0$.

▷ The first few derivatives are

$$\begin{aligned}\frac{d}{dx}\sin x &= \cos x\\ \frac{d^2}{dx^2}\sin x &= -\sin x\\ \frac{d^3}{dx^3}\sin x &= -\cos x\\ \frac{d^4}{dx^4}\sin x &= \sin x\\ \frac{d^5}{dx^5}\sin x &= \cos x\end{aligned}$$

We can see that there will be a cycle of sin, cos, − sin, and − cos, repeating indefinitely. Evaluating these derivatives at $x = 0$, we have 0, 1, 0, −1, ... All the even-order terms of the series are zero, and all the odd-order terms are $\pm 1/n!$. The result is

$$\sin x = x - \frac{1}{3!}x^3 + \frac{1}{5!}x^5 - \ldots \quad .$$

The linear term is the familiar small-angle approximation $\sin x \approx x$.

The radius of convergence of this series turns out to be infinite. Intuitively the reason for this is that the factorials grow extremely rapidly, so that the successive terms in the series eventually start diminish quickly, even for large values of x.

A function's Taylor series doesn't have to be evaluated around $x = 0$. The Taylor series around some other center $x = c$ is given by

$$T_c(x) = \sum_{n=0}^{\infty} a_n (x-c)^n \quad ,$$

where

$$\frac{a_n}{n!} = \left.\frac{d^n y}{dx^n}\right|_{x=c} \quad .$$

To see that this is the right generalization, we can do a change of variable, defining a new function $g(x) = f(x-c)$. The radius of convergence is to be measured from the center c rather than from 0.

Example 72

▷ Find the Taylor series of $\ln x$, evaluated around $x = 1$.

▷ Evaluating a few derivatives, we get

$$\begin{aligned}\frac{d}{dx}\ln x &= x^{-1}\\ \frac{d^2}{dx^2}\ln x &= -x^{-2}\\ \frac{d^3}{dx^3}\ln x &= 2x^{-3}\\ \frac{d^4}{dx^4}\ln x &= -6x^{-4}\end{aligned}$$

Note that evaluating these at $x = 0$ wouldn't have worked, since division by zero is undefined; this is because $\ln x$ blows up to negative infinity at $x = 0$. Evaluating them at $x = 1$, we find that the *nth* derivative equals $\pm(n-1)!$, so the coefficients of the Taylor series are $\pm(n-1)!/n! = \pm 1/n$, except for the $n = 0$ term, which is zero because $\ln 1 = 0$. The resulting series is

$$\ln x = (x-1) - \frac{1}{2}(x-1)^2 + \frac{1}{3}(x-1)^3 + \ldots \quad .$$

We can predict that its radius of convergence can't be any greater than 1, because $\ln x$ blows up at 0, which is at a distance of 1 from 1.

Problems

1 Modify the Weierstrass definition of the limit to apply to infinite sequences. ▷ Solution, p. 163

2 (a) Prove that the infinite series $1-1+1-1+1-1+\ldots$ does not converge to any limit, using the generalization of the Weierstrass limit found in problem 1. (b) Criticize the following argument. The series given in part a equals zero, because addition is associative, so we can rewrite it as $(1-1)+(1-1)+(1-1)+\ldots$ ▷ Solution, p. 163

3 Use the integral test to prove the convergence of the geometric series for $0 < x < 1$.
▷ Solution, p. 163

4 Determine the convergence or divergence of the following series.
(a) $1+1/2^2+1/3^2+\ldots$
(b) $1/\ln\ln 3 - 1/\ln\ln 6 + 1/\ln\ln 9 - 1/\ln\ln 12 + \ldots$
(c)

$$\frac{2\sqrt{2}}{9801}\sum_{k=0}^{\infty}\frac{(4k)!(1103+26390k)}{(k!)^4 396^{4k}}$$

▷ Solution, p. 163

5 Find the Taylor series expansion of $\cos x$ around $x=0$. Check your work by combining the first two terms of this series with the first term of the sine function from example 71 on page 98 to verify that the trig identity $\sin^2 x + \cos^2 x = 1$ holds for terms up to order x^2.

6 In classical physics, the kinetic energy K of an object of mass m moving at velocity v is given by $K=\frac{1}{2}mv^2$. For example, if a car is to start from a stoplight and then accelerate up to v, this is the theoretical minimum amount of energy that would have to be used up by burning gasoline. (In reality, a car's engine is not 100% efficient, so the amount of gas burned is greater.)

Einstein's theory of relativity states that the correct equation is actually

$$K=\left(\frac{1}{\sqrt{1-\frac{v^2}{c^2}}}-1\right)mc^2 \qquad ,$$

where c is the speed of light. The fact that it diverges as $v \rightarrow c$ is interpreted to mean that no object can be accelerated to the speed of light.

Expand K in a Taylor series, and show that the first nonvanishing term is equal to the classical expression. This means that for velocities that are small compared to the speed of light, the classical expression is a good approximation, and Einstein's theory does not contradict any of the prior empirical evidence from which the classical expression was inferred.

7 Expand $(1+x)^{1/3}$ in a Taylor series around $x=0$. The value $x=28$ lies outside this series' radius of convergence, but we can nevertheless use it to extract the cube root of 28 by recognizing that $28^{1/3}=3(28/27)^{1/3}$. Calculate the root to four significant figures of precision, and check it in the obvious way.

8 Find the Taylor series expansion of $\log_2 x$ around $x=1$, and use it to evaluate $\log_2 1.0595$ to four significant figures of precision. Check your result by using the fact that 1.0595 is approximately the twelfth root of 2. This number is the ratio of the frequencies of two successive notes of the chromatic scale in music, e.g., C and D-flat.

9 In free fall, the acceleration will not be exactly constant, due to air resistance. For example, a skydiver does not speed up indefinitely until opening her chute, but rather approaches a certain maximum velocity at which the upward force of air resistance cancels out the force of gravity. If an object is dropped from a height h, and the time it takes to reach the ground is used to measure the acceleration of gravity, g, then the relative error in the result due to air resistance is[2]

$$\begin{aligned} E &= \frac{g-g_{vacuum}}{g} \\ &= 1-\frac{2b}{\ln^2\left(e^b+\sqrt{e^{2b}-1}\right)} \end{aligned} ,$$

where $b=h/A$, and A is a constant that depends on the size, shape, and mass of the object, and the density of the air. (For a sphere of mass m and diameter d dropping in air, $A=4.11m/d^2$. Cf. problem 17, p. 60.) Evaluate the constant and linear terms of the Taylor series for the function $E(b)$.

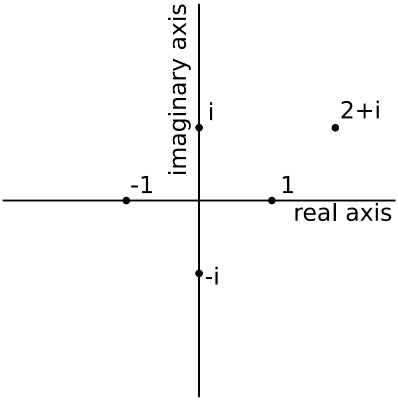
10 (a) Prove that the convergence of an infinite series is unaffected by omitting some initial terms. (b) Similarly, prove that convergence is unaffected by multiplying all the terms by some constant factor.

[2] Jan Benacka and Igor Stubna, *The Physics Teacher*, 43 (2005) 432.

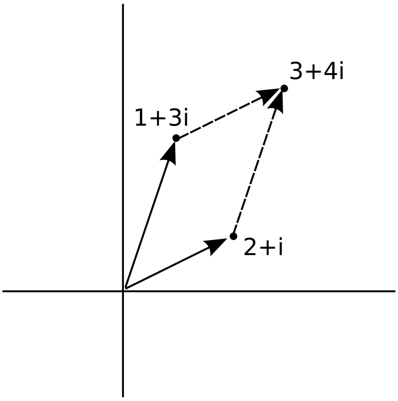
7 Complex number techniques

7.1 Review of complex numbers

For a more detailed treatment of complex numbers, see ch. 3 of James Nearing's free book at `http://www.physics.miami.edu/nearing/mathmethods/`.

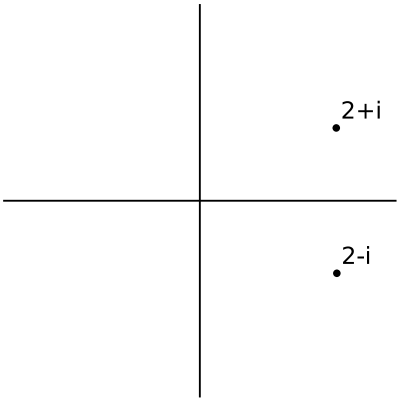


a / Visualizing complex numbers as points in a plane.



b / Addition of complex numbers is just like addition of vectors, although the real and imaginary axes don't actually represent directions in space.

We assume there is a number, i, such that $i^2 = -1$. The square roots of -1 are then i and $-i$. (In electrical engineering work, where i stands for current, j is sometimes used instead.) This gives rise to a number system, called the complex numbers, containing the real numbers as a subset. Any complex number z can be written in the form $z = a + bi$, where a and b are real, and a and b are then referred to as the real and imaginary parts of z. A number with a zero real part is called an imaginary number. The complex numbers can be visualized as a plane, figure a, with the real number line placed horizontally like the x axis of the familiar $x-y$ plane, and the imaginary numbers running along the y axis. The complex numbers are complete in a way that the real numbers aren't: every nonzero complex number has two square roots. For example, 1 is a real



c / A complex number and its conjugate.

number, so it is also a member of the complex numbers, and its square roots are -1 and 1. Likewise, -1 has square roots i and $-i$, and the number i has square roots $1/\sqrt{2}+i/\sqrt{2}$ and $-1/\sqrt{2}-i/\sqrt{2}$.

Complex numbers can be added and subtracted by adding or subtracting their real and imaginary parts, figure b. Geometrically, this is the same as vector addition.

The complex numbers $a+bi$ and $a-bi$, lying at equal distances above and below the real axis, are called complex conjugates. The results of the quadratic formula are either both real, or complex conjugates of each other. The complex conjugate of a number z is notated as $\bar{z}$ or z^*.

The complex numbers obey all the same rules of arithmetic as the reals, except that they can't be ordered along a single line. That is, it's not possible to say whether one complex number is greater than another. We can compare them in terms of their magnitudes (their distances from the origin), but two distinct complex numbers may have the same magnitude, so, for example, we can't say whether 1 is greater than i or i is greater than 1.

Example 73

▷ Prove that $1/\sqrt{2}+i/\sqrt{2}$ is a square root of i.

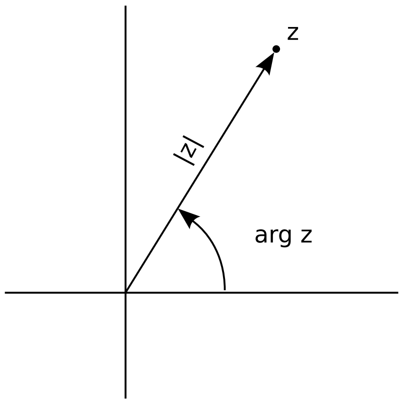
▷ Our proof can use any ordinary rules of arithmetic, except for ordering.

$$\begin{aligned}(\frac{1}{\sqrt{2}}+\frac{i}{\sqrt{2}})^2 &= \frac{1}{\sqrt{2}}\cdot\frac{1}{\sqrt{2}}+\frac{1}{\sqrt{2}}\cdot\frac{i}{\sqrt{2}}\\ &\quad+\frac{i}{\sqrt{2}}\cdot\frac{1}{\sqrt{2}}+\frac{i}{\sqrt{2}}\cdot\frac{i}{\sqrt{2}}\\ &=\frac{1}{2}(1+i+i-1)\\ &=i\end{aligned}$$

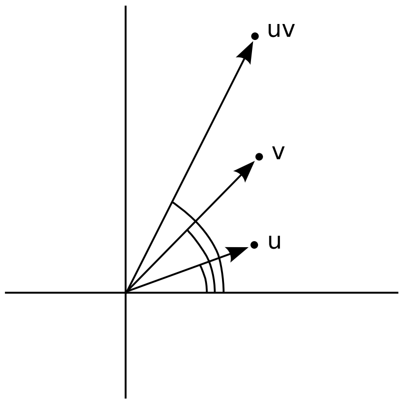
Example 73 showed one method of multiplying complex numbers. However, there is another nice interpretation of complex multiplication. We define the argument of a complex number, figure d, as its angle in the complex plane, measured counterclockwise from the positive real axis. Multiplying two complex numbers then corresponds to multiplying their magnitudes, and adding their arguments, figure e.

Self-Check

Using this interpretation of multiplication, how could you find the square



d / A complex number can be described in terms of its magnitude and argument.



e / The argument of *uv* is the sum of the arguments of *u* and *v*.

roots of a complex number? ▷ Answer, p. 145

Example 74

The magnitude $|z|$ of a complex number z obeys the identity $|z|^2 = z\bar{z}$. To prove this, we first note that $\bar{z}$ has the same magnitude as z, since flipping it to the other side of the real axis doesn't change its distance from the origin. Multiplying z by $\bar{z}$ gives a result whose magnitude is found by multiplying their magnitudes, so the magnitude of $z\bar{z}$ must therefore equal $|z|^2$. Now we just have to prove that $z\bar{z}$ is a positive real number. But if, for example, z lies counterclockwise from the real axis, then $\bar{z}$ lies clockwise from it. If z has a positive argument, then $\bar{z}$ has a negative one, or vice-versa. The sum of their arguments is therefore zero, so the result has an argument of zero, and is on the positive real axis. [1]

This whole system was built up in order to make every number have square roots. What about cube roots, fourth roots, and so on? Does it get even more weird when you want to do those as well? No. The complex number system we've already discussed is sufficient to handle all of them. The nicest way of thinking about it is in terms of roots of polynomials. In the real number system, the polynomial $x^2 - 1$ has two roots, i.e., two values of x (plus and minus one) that we can plug in to the polynomial and get zero. Because it has these two real roots, we can rewrite the polynomial as $(x-1)(x+1)$. However, the polynomial x^2+1 has no real roots. It's ugly that in the real number system, some second-

[1] I cheated a little. If z's argument is 30 degrees, then we could say $\bar{z}$'s was -30, but we could also call it 330. That's OK, because 330+30 gives 360, and an argument of 360 is the same as an argument of zero.

order polynomials have two roots, and can be factored, while others can't. In the complex number system, they all can. For instance, $x^2 + 1$ has roots i and $-i$, and can be factored as $(x - i)(x + i)$. In general, the fundamental theorem of algebra states that in the complex number system, any nth-order polynomial can be factored completely into n linear factors, and we can also say that it has n complex roots, with the understanding that some of the roots may be the same. For instance, the fourth-order polynomial $x^4 + x^2$ can be factored as $(x-i)(x+i)(x-0)(x-0)$, and we say that it has four roots, i, $-i$, 0, and 0, two of which happen to be the same. This is a sensible way to think about it, because in real life, numbers are always approximations anyway, and if we make tiny, random changes to the coefficients of this polynomial, it will have four distinct roots, of which two just happen to be very close to zero. I've given a proof of the fundamental theorem of algebra on page 142.

7.2 Euler's formula

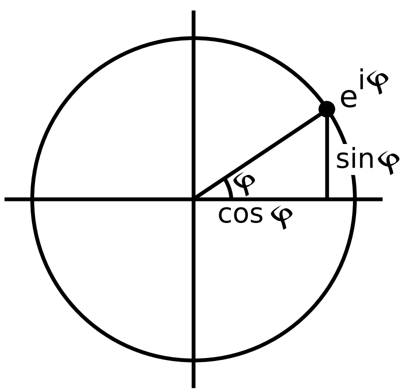
Having expanded our horizons to include the complex numbers, it's natural to want to extend functions we knew and loved from the world of real numbers so that they can also operate on complex numbers. The only really natural way to do this in general is to use Taylor series. A particularly beautiful thing happens with the functions e^x, $\sin x$, and $\cos x$:

$$\begin{aligned} e^x &= 1 + \frac{1}{2!}x^2 + \frac{1}{3!}x^3 + \ldots \\ \cos x &= 1 - \frac{1}{2!}x^2 + \frac{1}{4!}x^4 - \ldots \\ \sin x &= x - \frac{1}{3!}x^3 + \frac{1}{5!}x^5 - \ldots \end{aligned}$$

If $x = i\phi$ is an imaginary number, we have

$$e^{i\phi} = \cos\phi + i\sin\phi \qquad ,$$

a result known as Euler's formula. The geometrical interpretation in the complex plane is shown in figure f.



f / The complex number $e^{i\phi}$ lies on the unit circle.

Although the result may seem like something out of a freak show at first, applying the definition[2] of the

[2]See page 133 for an explanation of

exponential function makes it clear how natural it is:

$$e^x = \lim_{n\to\infty}\left(1+\frac{x}{n}\right)^n \qquad .$$

When $x = i\phi$ is imaginary, the quantity $(1 + i\phi/n)$ represents a number lying just above 1 in the complex plane. For large n, $(1 + i\phi/n)$ becomes very close to the unit circle, and its argument is the small angle ϕ/n. Raising this number to the nth power multiplies its argument by n, giving a number with an argument of ϕ.

g / Leonhard Euler (1707-1783)

Euler's formula is used frequently in physics and engineering.

Example 75

▷ Write the sine and cosine functions in terms of exponentials.

▷ Euler's formula for $x = -i\phi$ gives $\cos\phi - i\sin\phi$, since $\cos(-\theta) = \cos\theta$, and $\sin(-\theta) = -\sin\theta$.

$$\cos x = \frac{e^{ix}+e^{-ix}}{2}$$
$$\sin x = \frac{e^{ix}-e^{-ix}}{2i}$$

Example 76

▷ Evaluate

$$\int e^x \cos x \, dx$$

▷ This seemingly impossible integral becomes easy if we rewrite the cosine in terms of exponentials:

$$\begin{aligned}\int e^x \cos x\,dx &= \int e^x\left(\frac{e^{ix}+e^{-ix}}{2}\right)dx \\ &= \frac{1}{2}\int \left(e^{(1+i)x}+e^{(1-i)x}\right)dx \\ &= \frac{1}{2}\left(\frac{e^{(1+i)x}}{1+i}+\frac{e^{(1-i)x}}{1-i}\right)+c\end{aligned}$$

Since this result is the integral of a real-valued function, we'd like it to be real, and in fact it is, since the first and second terms are complex conjugates of one another. If we wanted to, we could use Euler's theorem to convert it back to a manifestly real result.[3]

where this definition comes from and why it makes sense.

[3]In general, the use of complex number techniques to do an integral could result in a complex number, but that complex number would be a constant, which could be subsumed within the usual constant of integration.

7.3 Partial fractions revisited

Suppose we want to evaluate the integral

$$\int \frac{\mathrm{d}x}{x^2+1}$$

by the method of partial fractions. The quadratic formula tells us that the roots are i and $-i$, setting $1/(x^2+1) = A/(x+i) + B/(x-i)$ gives $A = i/2$ and $B = -i/2$, so

$$\begin{aligned}\int \frac{\mathrm{d}x}{x^2+1} &= \frac{i}{2}\int \frac{\mathrm{d}x}{x+i} \\ &\quad - \frac{i}{2}\int \frac{\mathrm{d}x}{x-i} \\ &= \frac{i}{2}\ln(x+i) \\ &\quad - \frac{i}{2}\ln(x-i) \\ &= \frac{i}{2}\ln\frac{x+i}{x-i} \qquad .\end{aligned}$$

The attractive thing about this approach, compared with the method used on page 80, is that it doesn't require any tricks. If you came across this integral ten years from now, you could pull out your old calculus book, flip through it, and say, "Oh, here we go, there's a way to integrate one over a polynomial — partial fractions." On the other hand, it's odd that we started out trying to evaluate an integral that had nothing but real numbers, and came out with an answer that isn't even obviously a real number.

But what about that expression $(x+i)/(x-i)$? Let's give it a name, w. The numerator and denominator are complex conjugates of one another. Since they have the same magnitude, we must have $|w| = 1$, i.e., w is a complex number that lies on the unit circle, the kind of complex number that Euler's formula refers to. The numerator has an argument of $\tan^{-1}(1/x) = \pi/2 - \tan^{-1}x$, and the denominator has the same argument but with the opposite sign. Division means subtracting arguments, so $\arg w = \pi - 2\tan^{-1}x$. That means that the result can be rewritten using Euler's formula as

$$\begin{aligned}\int \frac{\mathrm{d}x}{x^2+1} &= \frac{i}{2}\ln e^{i(\pi-2\tan^{-1}x)} \\ &= \frac{i}{2}\cdot i(\pi - 2\tan^{-1}x) \\ &= \tan^{-1}x + c \qquad .\end{aligned}$$

In other words, it's the same result we found before, but found without the need for trickery.

Problems

1 Find $\arg i$, $\arg(-i)$, and $\arg 37$, where $\arg z$ denotes the argument of the complex number z.

2 Visualize the following multiplications in the complex plane using the interpretation of multiplication in terms of multiplying magnitudes and adding arguments: $(i)(i) = -1$, $(i)(-i) = 1$, $(-i)(-i) = -1$.

3 If we visualize z as a point in the complex plane, how should we visualize $-z$?

4 Find four different complex numbers z such that $z^4 = 1$.

5 Compute the following:

$$|1+i| \quad , \quad \arg(1+i) \quad ,$$
$$\left|\frac{1}{1+i}\right| \quad , \quad \arg\left(\frac{1}{1+i}\right) \quad ,$$
$$\frac{1}{1+i}$$

6 Write the function $\tan x$ in terms of complex exponentials.

7 Evaluate

$$\int \frac{\mathrm{d}x}{x^3 - x^2 + 4x - 4} \quad .$$

8 Evaluate

$$\int e^{-ax} \cos bx \, \mathrm{d}x \quad .$$

8 Iterated integrals

8.1 Integrals inside integrals

In various applications, you need to do integrals stuck inside other integrals. These are known as iterated integrals, or double integrals, triple integrals, etc. Similar concepts crop up all the time even when you're not doing calculus, so let's start by imagining such an example. Suppose you want to count how many squares there are on a chess board, and you don't know how to multiply eight times eight. You could start from the upper left, count eight squares across, then continue with the second row, and so on, until you how counted every square, giving the result of 64. In slightly more formal mathematical language, we could write the following recipe: for each row, r, from 1 to 8, consider the columns, c, from 1 to 8, and add one to the count for each one of them. Using the sigma notation, this becomes

$$\sum_{r=1}^{8}\sum_{c=1}^{8} 1 \qquad .$$

If you're familiar with computer programming, then you can think of this as a sum that could be calculated using a loop nested inside another loop. To evaluate the result (again, assuming we don't know how to multiply, so we have to use brute force), we can first evaluate the inside sum, which equals 8, giving

$$\sum_{r=1}^{8} 8 \qquad .$$

Notice how the "dummy" variable c has disappeared. Finally we do the outside sum, over r, and find the result of 64.

Now imagine doing the same thing with the pixels on a TV screen. The electron beam sweeps across the screen, painting the pixels in each row, one at a time. This is really no different than the example of the chess board, but because the pixels are so small, you normally think of the image on a TV screen as continuous rather than discrete. This is the idea of an integral in calculus. Suppose we want to find the area of a rectangle of width a and height b, and we don't know that we can just multiply to get the area ab. The brute force way to do this is to break up the rectangle into a grid of infinitesimally small squares, each having width $\mathrm{d}x$ and height $\mathrm{d}y$, and therefore the infinitesimal area $\mathrm{d}A = \mathrm{d}x\mathrm{d}y$. For convenience, we'll imagine that the rectangle's lower left corner is at the origin. Then the area is given by this integral:

$$\begin{aligned} \text{area} &= \int_{y=0}^{b} \int_{x=0}^{a} \mathrm{d}A \\ &= \int_{y=0}^{b} \int_{x=0}^{a} \mathrm{d}x\,\mathrm{d}y \end{aligned}$$

Notice how the leftmost integral sign, over y, and the rightmost differential, $\mathrm{d}y$, act like bookends, or the pieces of bread on a sandwich. Inside them, we have the integral sign that runs over x, and the differential $\mathrm{d}x$ that matches it on the right. Finally, on the innermost layer, we'd normally have the thing we're integrating, but here's it's 1, so I've omitted it. Writing the lower limits of the integrals with $x =$ and $y =$ helps to keep it straight which integral goes with with differential. The result is

$$\begin{aligned} \text{area} &= \int_{y=0}^{b} \int_{x=0}^{a} \mathrm{d}A \\ &= \int_{y=0}^{b} \int_{x=0}^{a} \mathrm{d}x\,\mathrm{d}y \\ &= \int_{y=0}^{b} \left(\int_{x=0}^{a} \mathrm{d}x\right) \mathrm{d}y \\ &= \int_{y=0}^{b} a\,\mathrm{d}y \\ &= a\int_{y=0}^{b} \mathrm{d}y \\ &= ab \qquad . \end{aligned}$$

Area of a triangle *Example 77*

▷ Find the area of a 45-45-90 right triangle having legs a.

▷ Let the triangle's hypotenuse run from the origin to the point (a, a), and let its legs run from the origin to $(0, a)$, and then to (a, a). In other words, the triangle sits on top of its hypotenuse. Then the integral can be set up the same way as the one before, but for a particular value of y, values of x only run from 0 (on the y axis) to y (on the hypotenuse). We then have

$$\begin{aligned} \text{area} &= \int_{y=0}^{a} \int_{x=0}^{y} \mathrm{d}A \\ &= \int_{y=0}^{a} \int_{x=0}^{y} \mathrm{d}x\,\mathrm{d}y \\ &= \int_{y=0}^{a} \left(\int_{x=0}^{y} \mathrm{d}x\right) \mathrm{d}y \\ &= \int_{y=0}^{a} y\,\mathrm{d}y \\ &= \frac{1}{2}a^2 \end{aligned}$$

Note that in this example, because the upper end of the x values depends on the value of y, it makes a difference which order we do the integrals in. The x integral has to be on the inside, and we have to do it first.

Volume of a cube *Example 78*

▷ Find the volume of a cube with sides of length a.

▷ This is a three-dimensional example, so we'll have integrals nested three deep, and the thing we're integrating is the volume $\mathrm{d}V = \mathrm{d}x\mathrm{d}y\mathrm{d}z$.

$$\begin{aligned}
\text{volume} &= \int_{z=0}^{a}\int_{y=0}^{a}\int_{x=0}^{a} \mathrm{d}V \\
&= \int_{z=0}^{a}\int_{y=0}^{a}\int_{x=0}^{a} \mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z \\
&= \int_{z=0}^{a}\int_{y=0}^{a} a\,\mathrm{d}y\,\mathrm{d}z \\
&= a\int_{z=0}^{a}\int_{y=0}^{a} \mathrm{d}y\,\mathrm{d}z \\
&= a\int_{z=0}^{a} a\,\mathrm{d}z \\
&= a^2\int_{z=0}^{a} \mathrm{d}z \\
&= a^3
\end{aligned}$$

Area of a circle *Example 79*

▷ Find the area of a circle.

▷ To make it easy, let's find the area of a semicircle and then double it. Let the circle's radius be r, and let it be centered on the origin and bounded below by the x axis. Then the curved edge is given by the equation $R^2 = x^2 + y^2$, or $y = \sqrt{R^2 - x^2}$. Since the y integral's limit depends on x, the x integral has to be on the outside. The area is

$$\begin{aligned}
\text{area} &= \int_{x=-R}^{r}\int_{y=0}^{\sqrt{R^2-x^2}} \mathrm{d}y\,\mathrm{d}x \\
&= \int_{x=-R}^{r} \sqrt{R^2 - x^2}\mathrm{d}x \\
&= r\int_{x=-R}^{r} \sqrt{1-(x/R)^2}\,\mathrm{d}x \qquad .
\end{aligned}$$

Substituting $u = x/R$,

$$\text{area} = R^2\int_{u=-1}^{1} \sqrt{1-u^2}\,\mathrm{d}u$$

The definite integral equals π, as you can find using a trig substitution or simply by looking it up in a table, and the result is, as expected, $\pi R^2/2$ for the area of the semicircle. Doubling it, we find the expected result of πR^2 for a full circle.

8.2 Applications

Up until now, the integrand of the innermost integral has always been 1, so we really could have done all the double integrals as single integrals. The following example is one in which you really need to do iterated integrals.

a / The famous tightrope walker Charles Blondin uses a long pole for its large moment of inertia.

Moments of inertia *Example 80*

The moment of inertia is a measure of how difficult it is to start an object rotating (or stop it). For example, tightrope walkers carry long poles because they want something with a big moment of inertia. The moment of inertia is defined by $I = \int R^2 \mathrm{d}m$, where dm is the mass of an infinitesimally small portion of the object, and R is the distance from the axis of rotation.

To start with, let's do an example that doesn't require iterated integrals. Let's calculate the moment of inertia of a thin rod of mass M and length L about a line perpendicular to the rod and passing through its center.

$$I = \int R^2 \mathrm{d}m$$

$$= \int_{-L/2}^{L/2} x^2 \, \frac{M}{L} \, \mathrm{d}x$$

[$r = |x|$, so $R^2 = x^2$]

$$= \frac{1}{12} ML^2$$

Now let's do one that requires iterated integrals: the moment of inertia of a cube of side b, for rotation about an axis that passes through its center and is parallel to four of its faces.

Let the origin be at the center of the cube, and let x be the rotation axis.

$$I = \int R^2 \mathrm{d}m$$

$$= \rho \int R^2 \mathrm{d}V$$

$$= \rho \int_{b/2}^{b/2} \int_{b/2}^{b/2} \int_{b/2}^{b/2} \left(y^2 + z^2\right) \mathrm{d}x \, \mathrm{d}y \, \mathrm{d}z$$

$$= \rho b \int_{b/2}^{b/2} \int_{b/2}^{b/2} \left(y^2 + z^2\right) \mathrm{d}y \, \mathrm{d}z$$

The fact that the last step is a trivial integral results from the symmetry of the problem. The integrand of the remaining double integral breaks down into two terms, each of which depends on only one of the variables, so we break it into two integrals,

$$I = \rho b \int_{b/2}^{b/2} \int_{b/2}^{b/2} y^2 \mathrm{d}y \, \mathrm{d}z$$
$$+ \rho b \int_{b/2}^{b/2} \int_{b/2}^{b/2} z^2 \mathrm{d}y \, \mathrm{d}z$$

which we know have identical results. We therefore only need to evaluate one of them and double the result:

$$I = 2\rho b \int_{b/2}^{b/2} \int_{b/2}^{b/2} z^2 \mathrm{d}y \, \mathrm{d}\, z$$

$$= 2\rho b^2 \int_{b/2}^{b/2} z^2 \mathrm{d}z$$

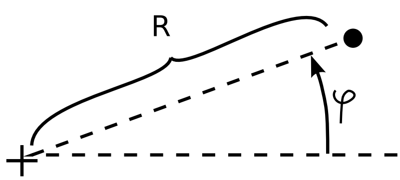
$$= \frac{1}{6} \rho b^5$$

$$= \frac{1}{6} Mb^2$$

8.3 Polar coordinates

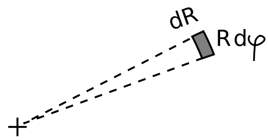
b / René Descartes (1596-1650)

Philosopher and mathematician René Descartes originated the idea of describing plane geometry using (x, y) coordinates measured from a pair of perpendicular coordinate axes. These rectangular coordinates are known as Cartesian coordinates, in his honor.



c / Polar coordinates.

As a logical extension of Descartes' idea, one can find different ways of defining coordinates on the plane, such as the polar coordinates in figure c. In polar coordinates, the differential of area, figure d can be written as $\mathrm{d}a = R\,\mathrm{d}R\,\mathrm{d}\phi$. The idea is that since $\mathrm{d}R$ and $\mathrm{d}\phi$ are infinitesimally small, the shaded area in the figure is very nearly a rectangle, measuring $\mathrm{d}R$ is one dimension and $R\,\mathrm{d}\phi$ in the other. (The latter follows from the definition of radian measure.)



d / The differential of area in polar coordinates

Example 81

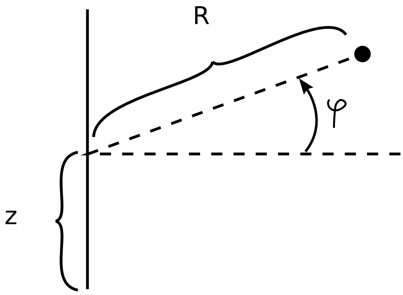
▷ A disk has mass *M* and radius *b*. Find its moment of inertia for rotation about the axis passing perpendicularly through its center.

▷

$$
\begin{aligned}
I &= \int R^2 \mathrm{d}M \\
&= \int R^2 \frac{\mathrm{d}M}{\mathrm{d}a}\,\mathrm{d}a \\
&= \int R^2 \frac{M}{\pi b^2}\,\mathrm{d}a \\
&= \frac{M}{\pi b^2} \int_{R=0}^{b} \int_{\phi=0}^{2\pi} R^2 \cdot R\,\mathrm{d}\phi\,\mathrm{d}R \\
&= \frac{M}{\pi b^2} \int_{R=0}^{b} R^3 \int_{\phi=0}^{2\pi} \mathrm{d}\phi\,\mathrm{d}R \\
&= \frac{2M}{b^2} \int_{R=0}^{b} R^3\,\mathrm{d}R \\
&= \frac{Mb^4}{2}
\end{aligned}
$$

8.4 Spherical and cylindrical coordinates

In cylindrical coordinates (R, ϕ, z), z measures distance along the axis, R measures distance from the axis, and ϕ is an angle that wraps around the axis.



e / Cylindrical coordinates.

The differential of volume in cylindrical coordinates can be written as $\mathrm{d}v = R\,\mathrm{d}R\,\mathrm{d}z\mathrm{d}\phi$. This follows from adding a third dimension, along the z axis, to the rectangle in figure d.

Example 82

▷ Show that the expression for dv has the right units.

▷ Angles are unitless, since the definition of radian measure involves a distance divided by a distance. Therefore the only factors in the expression that have units are R, dR, and dz. If these three factors are measured, say, in meters, then their product has units of cubic meters, which is correct for a volume.

Example 83

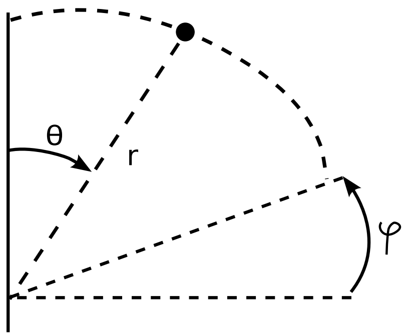
▷ Find the volume of a cone whose height is h and whose base has radius b.

▷ Let's plan on putting the z integral on the outside of the sandwich. That means we need to express the radius r_{max} of the cone in terms of z. This comes out nice and simple if we imagine the cone upside down, with its tip at the origin. Then since we have $r_{max}(z=0)=0$, and $r_{max}(h)=b$, evidently $r_{max}=zb/h$.

$$\begin{aligned}
v &= \int \mathrm{d}v \\
&= \int_{z=0}^{h}\int_{r=0}^{zb/h}\int_{\phi=0}^{2\pi} R\,\mathrm{d}\phi\,\mathrm{d}R\,\mathrm{d}z \\
&= 2\pi\int_{z=0}^{h}\int_{r=0}^{zb/h} R\,\mathrm{d}R\,\mathrm{d}z \\
&= 2\pi\int_{z=0}^{h}(zb/h)^2/2\,\mathrm{d}z \\
&= \pi(b/h)^2\int_{z=0}^{h} z^2\,\mathrm{d}z \\
&= \frac{\pi b^2 h}{3}
\end{aligned}$$

As a check, we note that the answer has units of volume. This is the classical result, known by the ancient Egyptians, that a cone has one third the volume of its enclosing cylinder.

In spherical coordinates (r, θ, ϕ), the coordinate r measures the distance from the origin, and θ and ϕ are analogous to latitude and longitude, except that θ is measured down from the pole rather than from the equator.



f / Spherical coordinates.

The differential of volume in spherical coordinates is $\mathrm{d}v = r^2 \sin\theta \,\mathrm{d}r\,\mathrm{d}\theta\,\mathrm{d}\phi$.

Example 84

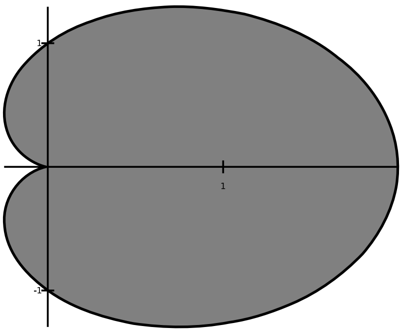
▷ Find the volume of a sphere.

▷

$$\begin{aligned} v &= \int \mathrm{d}v \\ &= \int_{\theta=0}^{\pi} \int_{r=0}^{r=b} \int_{\phi=0}^{2\pi} r^2 \sin\theta \,\mathrm{d}\phi\,\mathrm{d}r\,\mathrm{d}\theta \\ &= 2\pi \int_{\theta=0}^{\pi} \int_{r=0}^{r=b} r^2 \sin\theta \,\mathrm{d}r\,\mathrm{d}\theta \\ &= 2\pi \cdot \frac{b^3}{3} \int_{\theta=0}^{\pi} \sin\theta\,\mathrm{d}\theta \\ &= \frac{4\pi b^3}{3} \end{aligned}$$

Problems

1 Pascal's snail (named after Étienne Pascal, father of Blaise Pascal) is the shape shown in the figure, defined by $R = b(1 + \cos\theta)$ in polar coordinates.
(a) Make a rough visual estimate of its area from the figure.
(b) Find its area exactly, and check against your result from part a.
(c) Show that your answer has the right units. [Thompson, 1919]



Problem 1: Pascal's snail with $b = 1$.

2 A cone with a curved base is defined by $r \leq b$ and $\theta \leq \pi/4$ in spherical coordinates.
(a) Find its volume.
(b) Show that your answer has the right units.

3 Find the moment of inertia of a sphere for rotation about an axis passing through its center.

4 A jump-rope swinging in circles has the shape of a sine function. Find the volume enclosed by the swinging rope, in terms of the radius b of the circle at the rope's fattest point, and the straight-line distance ℓ between the ends.

5 A curvy-sided cone is defined in cylindrical coordinates by $0 \leq z \leq h$ and $R \leq kz^2$. (a) What units are implied for the constant k? (b) Find the volume of the shape. (c) Check that your answer to b has the right units.

6 The discovery of nuclear fission was originally explained by modeling the atomic nucleus as a drop of liquid. Like a water balloon, the drop could spin or vibrate, and if the motion became sufficiently violent, the drop could split in half — undergo fission. It was later learned that even the nuclei in matter under ordinary conditions are often not spherical but deformed, typically with an elongated ellipsoidal shape like an American football. One simple way of describing such a shape is with the equation

$$r \leq b[1 + c(\cos^2\theta - k)] \qquad ,$$

where $c = 0$ for a sphere, $c > 0$ for an elongated shape, and $c < 0$ for a flattened one. Usually for nuclei in ordinary matter, c ranges from about 0 to +0.2. The constant k is introduced because without it, a change in c would entail not just a change in the shape of the nucleus, but a change in its volume as well. Observations show, on the contrary, that the nuclear fluid is highly incompressible, just like ordinary water, so the volume of the nucleus is not expected to change significantly, even in violent processes like fission. Calculate the volume of the nucleus, throwing away terms of order c^2 or higher, and show that $k = 1/3$ is required in order to keep the volume constant.

7 This problem is a continuation of problem 6, and assumes the result of that problem is already known. The nucleus ^{168}Er has the type of elongated ellipsoidal shape described in that problem, with $c > 0$. Its mass is 2.8×10^{-25} kg, it is observed to have a moment of inertia of 2.62×10^{-54} $\text{kg} \cdot \text{m}^2$ for end-over-end rotation, and its shape is believed to be described by $b \approx 6 \times 10^{-15}$ m and $c \approx 0.2$. Assuming that it rotated rigidly, the usual equation for the moment of inertia could be applicable, but it may rotate more like a water balloon, in which case its moment of inertia would be significantly less because not all the mass would actually flow. Test which type of rotation it is by calculating its moment of inertia for end-over-end rotation and comparing with the observed moment of inertia. ⋆

A Detours

Formal definition of the tangent line

Let (a, b) be a point on the graph of the function $x(t)$. A line $\ell(t)$ through this point is said not to cut through the graph if there exists some real number $d > 0$ such that $x(t) - \ell(t)$ has the same sign for all t between $a - d$ and $a + d$. The line is said to be the tangent line at this point if it is the only line through this point that doesn't cut through the graph.

As an exception, there are cases in which the function is smooth and well-behaved throughout a certain region, but has no tangent line according to this definition at one particular point. For example, the function $x(t) = t^3$ has tangent lines everywhere except at $t = 0$, which is an inflection point (p. 18). In such cases, we fill in the "gap tooth" in the derivative function in the obvious way.

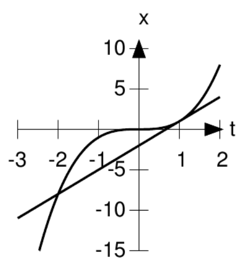
For an alternative definition that avoids the "gap tooth" issue, at the expense of some complication in the definition, see *Calculus Unlimited*, by Jerrold Marsden and Alan Weinstein, `http://resolver.caltech.edu/CaltechBOOK:1981.001`.

Derivatives of polynomials

Some ideas in this proof are due to Michael Livshits.

We want to prove that the derivative of t^k is kt^{k-1}. It suffices to prove that the derivative equals k when evaluated at $t = 1$, since we can then apply the kind of scaling argument[1] used on page 14 to show that the derivative of $t^2/2$ was t. The proposed tangent line at $(1, 1)$ has the equation $\ell = k(t - 1) + 1$, so what we need to prove is that the polynomial $t^k - [k(t-1)+1]$ is greater than or equal to zero throughout some region around $t = 1$.

Figure a shows a typical case. The graph of $3(t - 1) + 1$ lies entirely below the graph of t^3 in a large region. It does pop back up above it at $t = -2$, but that's far away, and the definition of the tangent line only requires that some region around $(1, 1)$ be free of such crossing points. In fact, a little experimentation shows that these crossings occur only for odd k, and always for $t < 0$. This suggests that we ought to aim for a general proof that there are no crossings for $t \geq 0$.



a / The graphs of t^3 and $3(t - 1) + 1$.

Suppose that such a crossing happens at the point (t, t^k). Then the slope of the line $\ell(t)$ is k, so we must have

$$\frac{t^k - 1}{t - 1} = k \qquad .$$

The left-hand side is the quotient of two polynomials, and we expect it to divide without a remainder, because $t^k - 1$ equals zero at $t = 1$, and

[1] Scaling fails in the special case of $t = 0$ and odd k, so we have to fill in the "gap tooth" as mentioned in the preceding section.

therefore it must have $t-1$ as a factor. If we try the example of $k=3$, we find that the quotient has the very simple form t^2+t+1, i.e., a polynomial of order $k-1$ whose coefficients are all equal to 1. We can easily verify that this works for all k, by checking the multiplication

$$t^k - 1 = (t-1)(t^k + t^{k-1} + \ldots + 1) \qquad ,$$

in which all the terms in the expansion of the right-hand side cancel except for t^k and -1. Let's refer to the quotient as $Q(t) = t^k + t^{k-1} + \ldots + 1$. How can we get $Q(t) = k$? Clearly we have a solution for $t=1$, since there are k terms, each equal to 1. For $t>1$, all the terms except the constant one are greater than 1, so there can't be any solution. For $0 \le t < 1$, all the terms except the constant one are positive and less than 1, so again there can't be any solution. This completes the proof that there are no crossings for $t \ge 0$, which establishes the desired result.

Details of the proof of the derivative of the sine function

Some ideas in this proof are due to Jerome Keisler (see references, p. 171).

On page 28, I computed the derivative of $\sin t$ to be $\cos t$ as follows:

$$\begin{aligned} \mathrm{d}x &= \sin(t + \mathrm{d}t) - \sin t \qquad , \\ &= \sin t\ \cos \mathrm{d}t \\ &\quad + \cos t\ \sin \mathrm{d}t - \sin t \\ &= \cos t\ \mathrm{d}t + \ldots \qquad . \end{aligned}$$

We want to prove prove that the error "..." introduced by the small-angle approximations really is of order $\mathrm{d}t^2$.

A quick and dirty way to check whether this is likely to be true is to use Inf to calculate $\sin(t+\mathrm{d}t)$ at some specific value of t. For example, at $t=1$ we have this result:

```
: sin(1+d)
  (0.84147)+(0.54030)d
  +(-0.42074)d^2+(-0.09006)d^3
  +(0.03506)d^4
```

The small-angle approximations give $\sin(1+d) \approx \sin 1 + (\cos 1)d$. The coefficients of the first two terms of the exact result are, as expected

$\sin(1) = 0.84147$ and $\cos(1) = 0.5403\ldots$, so although the small-angle approximations have introduced some errors, they involve only higher powers of $\mathrm{d}t$, as claimed.

The demonstration with Inf has two shortcomings. One is that it only works for $t = 1$, but we need to prove that the result for all values of t. That doesn't mean that the check for $t = 1$ was useless. Even though a general mathematical statement about all numbers can never be *proved* by demonstrating specific examples for which it succeeds, a single counterexample suffices to *disprove* it. The check for $t = 1$ was worth doing, because if the first term had come out to be 0.88888, it would have immediately disproved our claim, thereby saving us from wasting hours attempting to prove something that wasn't true.

The other problem is that I've never explained how Inf calculates this kind of thing. The answer is that it uses something called a Taylor series, discussed in section 6.4. Using Inf here without knowing yet how Taylor series work is like using your calculator as a "black box" to extract the square root of $\sqrt{2}$ without knowing how it does it. Not knowing the inner workings of the black box makes the demonstration less than satisfying.

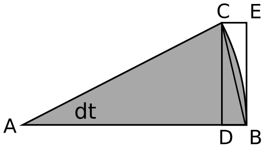
In any case, this preliminary check makes it sound like it's reasonable to go on and try to produce a real proof. We have

$$\sin(t + \mathrm{d}t) = \sin t + \cos t\,\mathrm{d}t - E \qquad ,$$

where the error E introduced by the approximations is

$$\begin{aligned} E = & \sin t(1 - \cos \mathrm{d}t) \\ & + \cos t(\mathrm{d}t - \sin \mathrm{d}t) \qquad . \end{aligned}$$

Let the radius of the circle in figure b be one, so AD is $\cos \mathrm{d}t$ and CD is



b / Geometrical interpretation of the error term.

$\sin \mathrm{d}t$. The area of the shaded pie slice is $\mathrm{d}t/2$, and the area of triangle ABC is $\sin \mathrm{d}t/2$, so the error made in the approximation $\sin \mathrm{d}t \approx \mathrm{d}t$

equals twice the area of the dish shape formed by line BC and arc BC. Therefore $\mathrm{d}t - \sin \mathrm{d}t$ is less than the area of rectangle CEBD. But CEBD has both an infinitesimal width and an infinitesimal height, so this error is of no more than order $\mathrm{d}t^2$.

For the approximation $\cos \mathrm{d}t \approx 1$, the error (represented by BD) is $1 - \cos \mathrm{d}t = 1 - \sqrt{1 - \sin^2 \mathrm{d}t}$, which is less than $1 - \sqrt{1 - \mathrm{d}t^2}$, since $\sin \mathrm{d}t < \mathrm{d}t$. Therefore this error is of order $\mathrm{d}t^2$.

Formal statement of the transfer principle

On page 33, I gave an informal description of the transfer principle. The idea being expressed was that the phrases "for any" and "there exists" can only be used in phrases like "for any real number x" and "there exists a real number y such that..." The transfer principle does not apply to statements like "there exists an integer x such that..." or even "there exists a subset of the real numbers such that..."

The way to state the transfer principle more rigorously is to get rid of the ambiguities of the English language by restricting ourselves to a well-defined language of mathematical symbols. This language has symbols $\forall$ and $\exists$, meaning "for all" and "there exists," and these are called quantifiers. A quantifier is always immediately followed by a variable, and then by a statement involving that variable. For example, suppose we want to say that a number greater than 1 exists. We can write the statement $\exists x\ x > 1$, read as "there exists a number x such that x is greater than 1." We don't actually need to say "there exists a number x *in the set of real numbers* such that ...," because our intention here is to make statements that can be translated back and forth between the reals and the hyperreals. In fact, we forbid this type of explicit reference to the domain to which the quantifiers apply. This restriction is described technically by saying that we're only allowing *first-order logic.*

Quantifiers can be nested. For example, I can state the commutativity of addition as $\forall x \forall y\ x + y = y + x$, and the existence of additive inverses as $\forall x \exists y\ x + y = 0$.

After the quantifier and the variable, we have some mathematical assertion, in which we're allowed to use the symbols $=$, $>$, $\times$ and $+$ for the basic operations of arithmetic, and also parentheses and the logical operators $\neg$, $\wedge$ and $\vee$ for "not," "and," and "or." Although we will often find it convenient to use other symbols, such as 0, 1, $-$, $/$, $\leq$,

$\neq$, etc., these are not strictly necesary. We use them only as a way of making the formulas more readable, with the understanding that they could be translated into the more basic symbols. For instance, I can restate $\exists x\ x > 1$ as $\exists x \exists y \forall z\ yz = z \wedge x > y$. The number y ends up just being a name for 1, because it's the only number that will always satisfy $yz = z$.

Finally, these statements need to satisfy certain syntactic rules. For example, we can't have a string of symbols like $x + \times y$, because the operators $+$ and $\times$ are supposed to have numbers on both sides.

A finite string of symbols satisfying all the above rules is called a well-formed formula (wff) in first-order logic.

The transfer principle states that a wff is true on the real numbers if and only if it is true on the hyperreal numbers.

If you look in an elementary algebra textbook at the statement of all the elementary axioms of the real number system, such as commutativity of multiplication, associativity of addition, and so on, you'll see that they can all be expressed in terms of first-order logic, and therefore you can use them when manipulating hyperreal numbers. However, it's not possible to fully characterize the real number system without giving at least some further axioms that cannot be expressed in first order. There is more than one way to set up these additional axioms, but for example one common axiom to use is the Archimedean principle, which states that there is no number that is greater than 1, greater than $1 + 1$, greater than $1 + 1 + 1$, and so on. If we try to express this as a well-formed formula in first order logic, one attempt would be $\neg\exists x\ x > 1 \quad \wedge \quad x > 1 + 1 \quad \wedge \quad x > 1 + 1 + 1 \ldots$, where the $\ldots$ indicates that the string of symbols would have to go on forever. This doesn't work because a well-formed formula has to be a *finite* string of symbols. Another attempt would be $\exists x \forall n \in \mathbb{N}\ x > n$, where $\mathbb{N}$ means the set of integers. This one also fails to be a wff in first-order logic, because in first-order logic we're not allowed to explicitly refer to the domain of a quantifier. We conclude that the transfer principle does not necessarily apply to the Archimedean principle, and in fact the Archimedean principle is not true on the hyperreals, because they include numbers that are infinite.

Now that we have a thorough and rigorous understanding of what the transfer principle says, the next obvious question is why we should believe that it's true. This is discussed in the following section.

Is the transfer principle true?

The preceding section stated the transfer principle in rigorous language. But why should we believe that it's true?

One approach would be to begin deducing things about the hyperreals, and see if we can deduce a contradiction. As a starting point, we can use the axioms of elementary algebra, because the transfer principle tells us that those apply to the hyperreals as well. Since we also assume that the Archimedean principle does *not* hold for the hyperreals, we can also base our reasoning on that, and therefore many of the things we can prove will be things that are true for the hyperreals, but false for the reals. This is essentially what mathematicians started doing immediately after Newton and Leibniz invented the calculus, and they were immediately successful in producing contradictions. However, they weren't using formally defined logical systems, and they hadn't stated anything as specific and rigorous as the transfer principle. In particular, they didn't understand the need for anything like our restriction of the transfer principle to first-order logic. If we could reach a contradiction based on the more modern, rigorous statement of the transfer principle, that would be a different matter. It would tell us that one of two things was true: either (1) the hyperreal number system lacks logical self-consistency, or (2) both the hyperreals and the reals lack self-consistency.

Abraham Robinson proved, however, around 1960 that the reals and the hyperreals have the same level of consistency: one is self-consistent if and only if the other is. In other words, if the hyperreals harbor a ticking logical time bomb, so do the reals. Since most mathematicians don't lose much sleep worrying about a lack of self-consistency in the real number system, this is generally taken as meaning that infinitesimals have been rehabilitated. In fact, it gives them an even higher level of respectability than they had in the era of Gauss and Euler, when they were widely used, but mathematicians knew a valid style of proof involving infinitesimals only because they'd slowly developed the right "Spidey sense."

But how in the world could Robinson have proved such a thing? It seems like a daunting task. There is an infinite number of possible logical trains of argument in mathematics. How could he have demonstrated, with a stroke of a pen, that *none* of them could ever lead to a contradiction (unless it indicated a contradiction lurking in the real number system as well)? Obviously it's not possible to check them all explicitly.

The way modern logicians prove such things is usually by using *models*.

For an easy example of a model, consider Euclidean geometry. Euclid believed that the following four postulates[2] were all self-evident:

1. Let the following be postulated: to draw a straight line from any point to any point.

2. To extend a finite straight line continuously in a straight line.

3. To describe a circle with any center and radius.

4. That all right angles are equal to one another.

These postulates, which today we would call "axioms," played the same role with respect to Euclidean geometry that the elementary axioms of arithmetic play for the real number system.

Euclid also found that he needed a fifth postulate in order to prove many of his most important theorems, such as the Pythagorean theorem. I'll state a different axiom that turns out to be equivalent to it:

5. *Playfair's version of the parallel postulate:* Given any infinite line L, and any point P not on that line, there exists a unique infinite line through P that never crosses L.

The ancients believed this to be less obviously self-evident than the first four, partly because if you were given the two lines, it could theoretically take an infinite amount of time to inspect them and verify that they never crossed, even at some very distant point. Euclid avoided even mentioning infinite lines in postulates 1-4, and he considered postulate 5 to be so much less intuitively appealing in comparison that he organized the *Elements* so that the first 28 propositions were those that could be proved without resorting to it. Continuing the analogy with the reals and hyperreals, the parallel postulate plays the role of the Archimedean principle: a statement about infinity that we don't feel quite so sure about.

For centuries, geometers tried to prove the parallel postulate from the first five. The trouble with this kind of thing was that it could be difficult to tell what was a valid proof and what wasn't. The postulates were written in an ambiguous human language, not a formal logical system. As an example of the kind of confusion that could result, suppose we assume the following postulate, $5'$, in place of 5:

[2] modified slightly by me from a translation by T.L. Heath, 1925

$5'$: Given any infinite line L, and any point P not on that line, every infinite line through P crosses L.

Postulate $5'$ plays the role for noneuclidean geometry that the negation of the Archimedean principle plays for the hyperreals. It tells us we're not in Kansas anymore. If a geometer can start from postulates 1-4 and $5'$ and arrive at a contradiction, then he's made significant progress toward proving that postulate 5 has to be true based on postulates 1-4. (He would also have to disprove another version of the postulate, in which there is more than one parallel through P.) For centuries, there have been reasonable-sounding arguments that seemed to give such a contradiction. For instance, it was proved that a geometry with $5'$ in it was one in which distances were limited to some finite maximum. This would appear to contradict postulate 3, since there would be a limit on the radius of a circle. But there's plenty of room for disagreement here, because the ancient Greeks didn't have any notion of a set of real numbers. For them, the thing we would call a number was simply a finite straight line (line segment) with a certain length. If postulate 3 says that we can make a circle given any radius, it's reasonable to interpret that as a statement that *given any finite straight line* as the specification of the radius, we can make the circle. There is then no contradiction, because the too-long radius can't be specified in the first place. This muddle is similar to the kind of confusion that reigned for centuries after Newton: did infinitesimals lead to contradictions?

In the 19th century, Lobachevsky and Bolyai came up with a version of Euclid's axioms that was more rigorously defined, and that was carefully engineered to avoid the kinds of contradictions that had previously been discovered in noneuclidean geometry. This is analogous to the invention of the transfer principle and the realization that the restriction to first-order logic was necessary. Lobachevsky and Bolyai slaved away for year after year proving new results in noneuclidean geometry, wondering whether they would ever reach a contradiction. Eventually they started to doubt that there were ever going to be contradictions, and finally they proved that the contradictions didn't exist.

The technique for proving consistency was to make a *model* of the noneuclidean system. Consider geometry done on the surface of a sphere. The word "line" in the axioms now has to be understood as referring to a great circle, i.e., one with the same radius as the sphere. The parallel postulate fails, because parallels don't exist: every great circle intersects every other great circle. One modification has to be made to the model in order to make it consistent with the first postulate. The constructions described in Euclid's postulates are tacitly assumed to be unique (and in more rigorous formulations are explicitly stated to be so). We want there to be a unique line defined by any two distinct points. This works fine on the sphere as long as the points aren't too far apart, but it fails if the points are antipodes, i.e., they lie at opposite sides of the sphere. For example, every line of longitude on the Earth's surface passes through both poles. The solution to this problem is to modify what we mean by "point." Points at each other's antipodes are considered to be the *same point*. (Or, equivalently, we can do geometry on a hemisphere, but agree that when we go off one edge, we "wrap around" to the opposite side.)

This spherical model obeys all the postulates of this particular system of noneuclidean geometry. But consider now that we constructed it *inside* a surrounding three-dimensional space in which the parallel postulate does hold. Now suppose we keep on proving theorems in this system of noneuclidean geometry, filling up page after page with proofs using words like "line," which we mentally associate with great circles on a certain sphere — and eventually we reach a contradiction. But now we can go back through our proofs, and in every place where the word "line" occurs we can cross it out with a red pencil and put in "great circle on this particular sphere." It would now be a proof about *Euclidean* geometry, and the contradiction would prove that *Euclidean* geometry lacked self-consistency. We therefore arrive at the result that if noneuclidean geometry is inconsistent, so is Euclidean geometry. Since nobody believes that Euclidean geometry is inconsistent, this is considered the moral equivalent of proving noneuclidean geometry to be consistent.

If you've been keeping the system of analogies in mind as you read this story, it should be clear what's coming next. If we want to prove that the hyperreals have the same consistency as the reals, we just have to construct a *model* of the hyperreals using the reals. This is done in detail elsewhere (see Stroyan and Mathforum.org in the references, p. 171). I'll just sketch the general idea. A hyperreal number is represented by an infinite sequence of real numbers. For example, the sequence

$$7, 7, 7, 7, \ldots$$

would be the hyperreal version of the number 7. A sequence like

$$1, 2, 3, \ldots$$

represents an infinite number, while

$$1, \frac{1}{2}, \frac{1}{3}, \ldots$$

is infinitesimal. All the arithmetic operations are defined by applying them to the corresponding members of the sequences. For example, the sum of the 7, 7, 7, ... sequence and the 1, 2, 3, ... sequence would be 8, 9, 10, ..., which we interpret as a somewhat larger infinite number.

The big problem in this approach is how to compare hyperreals, because a comparison like $<$ is supposed to give an answer that is either true or false. It's not supposed to give a hyperreal number as the result.

It's clear that 8, 9, 10, ... is greater than 1, 1, 1, ..., because every member of the first sequence is greater than every member of the second one. But is 8, 9, 10, ... greater than 9, 9, 9, ...? We want the answer to be "yes," because we're thinking of the first one as an infinite number and the second one as the ordinary finite number 9. The first sequence is indeed greater than the second at almost every one of the infinite number of places at which they could be compared. The only place where it loses the contest is at the very first position, and the only spot where we get a tie is the second one. Essentially the idea is that we want to define a concept of what happens "almost everywhere" on some infinite list. If one thing happens in an infinite number of places and something else only happens at some finite number of spots, then the definition of "almost everywhere" is clear. What's harder is a comparison of something like these two sequences:

$$2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, \ldots$$

and

$$1, 3, 1, 1, 3, 1, 1, 1, 3, 1, 1, 1, 1, 3, \ldots$$

where the second sequence has longer and longer runs of ones interspersed between the threes. The two sequences are never equal at any position, so clearly they can't be considered to be equal as hyperreal numbers. But there is an infinite number of spots in which the first sequence is greater than the second, and likewise an infinite number in which it's less. It seems as though there are more in which it's greater, so we probably want to define the second sequence as being a hyperreal number that's less than 2. The problem is that it can be very difficult to write down an acceptable definition of this "almost everywhere" notion. The answer is very technical, and I won't go into it here, but it can be done. Because two sequences could be equal almost everywhere, we end up having to define a hyperreal number not as a particular sequence but as a *set* of sequences that are equal to each other almost everywhere.

With the construction of this model, it is possible to prove that the hyperreals have the same level of consistency as the reals.

The transfer principle applied to functions

On page 34, I told you not to worry about whether it was legitimate to apply familiar functions like x^2, $\sqrt{x}$, $\sin x$, $\cos x$, and e^x to hyperreal numbers. But since you're reading this, you're obviously in need of more reassurance.

For some of these functions, the transfer principle straightforwardly guarantees that they work for hyperreals, have all the familiar properties, and can be computed in the same way. For example, the following statement is in a suitable form to have the transfer principle applied to it: *For any real number x, $x \cdot x \geq 0$.* Changing "real" to "hyperreal," we find out that the square of a hyperreal number is greater than or equal to zero, just like the square of a real number. Writing it as x^2 or calling it a square is just a matter of notation and terminology. The same applies to this statement: *For any real number $x \geq 0$, there exists a real number y such that $y^2 = x$.* Applying the transfer function to it tells us that square roots can be defined for the hyperreals as well.

There's a problem, however, when we get to functions like $\sin x$ and e^x. If you look up the definition of the sine function in a trigonometry textbook, it will be defined geometrically, as the ratio of the lengths of two sides of a certain triangle. The transfer principle doesn't apply to geometry, only to arithmetic. It's not even obvious intuitively that it makes sense to define a sine function on the hyperreals. In an application like the differentiation of the sine function on page 28, we only had to take sines of hyperreal numbers that were infinitesimally close to real numbers, but if the sine is going to be a full-fledged function defined on the hyperreals, then we should be allowed, for example, to take the sine of an infinite number. What would that mean? If you take the sine of a number like a million or a billion on your calculator, you just get some apparently random result between -1 and 1. The sine function wiggles back and forth indefinitely as x gets bigger and bigger, never settling down to any specific limiting value. Apparently we could have $\sin H = 1$ for a particular infinite H, and then $\sin(H + \pi/2) = 0$, $\sin(H + \pi) = -1$, ...

It turns out that the moral equivalent of the transfer function can indeed be applied to any function on the reals, yielding a function that is in some sense its natural "big brother" on the the hyperreals, but the consequences can be either disturbing or exhilirating depending on your tastes. For example, consider the function $[x]$ that takes a real number x and rounds it down to the greatest integer that is less than or equal to to x, e.g., $[3] = 3$, and $[\pi] = 3$. This function, like any other real function, can be extended to the hyperreals, and that means that we can define the *hyperintegers*, the set of hyperreals that satisfy $[x] = x$. The hyperintegers include the integers as a subset, but they also include infinite numbers. This is likely to seem magical, or even unreasonable, if we come at the hyperreals from a purely axiomatic point of view. The extension of functions to the hyperreals seems much more natural in view of the construction of the hyperreals in terms of sequences given in the preceding section. For example, the sequence $1.3, 2.3, 3.3, 4.3, 5.3, \ldots$ represents an infinite number. If we apply the $[x]$ function to it, we get $1, 2, 3, 4, 5, \ldots$, which is an infinite integer.

Proof of the chain rule

In the statement of the chain rule on page 37, I followed my usual custom of writing derivatives as $\mathrm{d}y/\mathrm{d}x$, when actually the derivative is the standard part, $\mathrm{st}(\mathrm{d}y/\mathrm{d}x)$. In more rigorous notation, the chain rule should be stated like this:

$$\mathrm{st}\left(\frac{\mathrm{d}z}{\mathrm{d}x}\right) = \mathrm{st}\left(\frac{\mathrm{d}z}{\mathrm{d}y}\right)\mathrm{st}\left(\frac{\mathrm{d}y}{\mathrm{d}x}\right) \qquad .$$

The transfer principle allows us to rewrite the left-hand side as $\mathrm{st}[(\mathrm{d}z/\mathrm{d}y)(\mathrm{d}y/\mathrm{d}x)]$, and then we can get the desired result using the identity $\mathrm{st}(ab) = \mathrm{st}(a)\mathrm{st}(b)$.

Derivative of e^x

All of the reasoning on page 38 would have applied equally well to any other exponential function with a different base, such as 2^x or 10^x. Those functions would have different values of c, so if we want to determine the value of c for the base-e case, we need to bring in the definition of e, or of the exponential function e^x, somehow.

We can take the definition of e^x to be

$$e^x = \lim_{n\to\infty}\left(1 + \frac{x}{n}\right)^n \qquad .$$

The idea behind this relation is similar to the idea of compound interest. If the interest rate is 10%, compounded annually, then $x = 0.1$, and the balance grows by a factor $(1 + x) = 1.1$ in one year. If, instead, we want to compound the interest monthly, we can set the monthly interest rate to 0.1/12, and then the growth of the balance over a year is $(1+x/12)^{12} = 1.1047$, which is slightly larger because the interest from the earlier months itself accrues interest in the later months. Continuing this limiting process, we find $e^{1.1} = 1.1052$.

If n is large, then we have a good approximation to the base-e exponential, so let's differentiate this finite-n approximation and try to find an approximation to the derivative of e^x. The chain rule tells is that the derivative of $(1 + x/n)^n$ is the derivative of the raising-to-the-nth-power function, multiplied by the derivative of the inside stuff, $\mathrm{d}(1 + x/n)/\mathrm{d}x = 1/n$. We then have

$$\begin{aligned}\frac{\mathrm{d}\left(1+\frac{x}{n}\right)^n}{\mathrm{d}x} &= \left[n\left(1+\frac{x}{n}\right)^{n-1}\right]\cdot\frac{1}{n}\\ &= \left(1+\frac{x}{n}\right)^{n-1} \qquad .\end{aligned}$$

But evaluating this at $x = 0$ simply gives 1, so at $x = 0$, the approximation to the derivative is exactly 1 for all values of n — it's not even necessary to imagine going to larger and larger values of n. This establishes that $c = 1$, so we have

$$\frac{\mathrm{d}e^x}{\mathrm{d}x} = e^x$$

for all values of x.

Proof of the fundamental theorem of calculus

There are three parts to the proof: (1) Take the equation that states the fundamental theorem, differentiate both sides with respect to b, and show that they're equal. (2) Show that continuous functions with equal derivatives must be essentially the same function, except for an additive constant. (3) Show that the constant in question is zero.

1. By the definition of the indefinite integral, the derivative of $x(b)-x(a)$ with respect to b equals $\dot{x}(b)$. We have to establish that this equals the

following:

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}b}\int_a^b \dot{x}(t)\mathrm{d}t &= \operatorname{st}\frac{1}{\mathrm{d}b}\left[\int_a^{b+\mathrm{d}b}\dot{x}(t)\mathrm{d}t - \int_a^b \dot{x}(t)\mathrm{d}t\right] \\
&= \operatorname{st}\frac{1}{\mathrm{d}b}\int_b^{b+\mathrm{d}b}\dot{x}(t)\mathrm{d}t \\
&= \operatorname{st}\frac{1}{\mathrm{d}b}\lim_{H\to\infty}\sum_{i=0}^{H}\dot{x}(b\ +\ i\ \mathrm{d}b/H)\frac{\mathrm{d}b}{H} \\
&= \operatorname{st}\lim_{H\to\infty}\frac{1}{H}\sum_{i=0}^{H}\dot{x}(b\ +\ i\ \mathrm{d}b/H)
\end{aligned}
$$

Since $\dot{x}$ is continuous, all the values of $\dot{x}$ occurring inside the sum can differ only infinitesimally from $\dot{x}(b)$. Therefore the quantity inside the limit differs only infinitesimally from $\dot{x}(b)$, and the standard part of its limit must be $\dot{x}(b)$.[3]

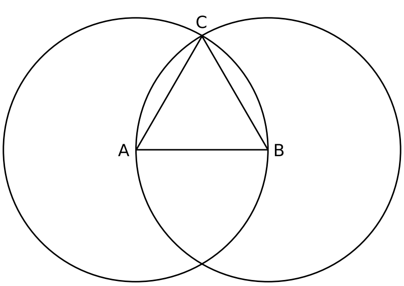
2. Suppose f and g are two continuous functions whose derivatives are equal. Then $d = f - g$ is a continuous function whose derivative is zero. But the only continuous function with a derivative of zero is a constant, so f and g differ by at most an additive constant.

3. I've established that the derivatives with respect to b of $x(b) - x(a)$ and $\int_a^b \dot{x}\mathrm{d}t$ are the same, so they differ by at most an additive constant. But at $b = a$, they're both zero, so the constant must be zero.

[3]If you don't want to use infinitesimals, then you can express the derivative as a limit, and in the final step of the argument use the mean value theorem, introduced later in the chapter.

The intermediate value theorem

On page 47 I asserted that the intermediate value theorem was really more a statement about the (real or hyperreal) number system than about functions. For insight, consider figure c, which is a geometrical construction that constitutes the proof of the very first proposition in Euclid's celebrated *Elements*. The proposition to be proved is that given a line segment AB, it is possible to construct an equilateral triangle with AB as its base. The proof is by construction; that is, Euclid doesn't just give a logical argument that convinces us the triangle must exist, he actually demonstrates how to construct it. First we draw a circle with center A and radius AB, which his third postulate says we can do. Then we draw another circle with the same radius, but centered at B. Pick one of the intersections of the circles and call it C. Construct the line segments AC and BC (postulate 1). Then AC equals AB by the definition of the circle, and likewise BC equals AB. Euclid also has an axiom that things equal to the same thing are equal to one another, so it follows that AC equals BC, and therefore the triangle is equilateral.



c / A proof from Euclid's *Elements*.

It seems like a model of mathematical rigor, but there's a flaw in the reasoning, which is that he assumes without justififcation that the circles do have a point in common. To see that this is not as secure an assumption as it seems, consider the usual Cartesian representation of plane geometry in terms of coordinates (x, y). Usually we assume that x and y are real numbers. What if we instead do our Cartesian geometry using rational numbers as coordinates? Euclid's five postulates are all consistent with this. For example, circles do exist. Let $\text{A} = (0, 0)$ and $\text{B} = (1, 0)$. Then there are infinitely many pairs of rational numbers in the set that satisfies the definition of the circle centered at A. Examples include $(3/5, 4/5)$ and $(-7/25, 24/25)$. The circle is also continuous in the sense that if I specify a point on it such as $(-7/25, 24/25)$, and a distance that I'm allowed to make as small as I please, say 10^{-6}, then other points exist on the circle within that distance of the given point. However, the intersection assumed by Euclid's proof doesn't exist. It would lie at $(1/2, \sqrt{3}/2)$, but $\sqrt{3}$ doesn't exist in the rational number system.

In exactly the same way, we can construct counterexamples to the intermediate value theorem if the underlying system of numbers doesn't have the same properties as the real numbers. For example, let $y = x^2$. Then y is a continuous function, on the interval from 0 to 1, but if we take the rational numbers as our foundation, then there is no x for which $y = 1/2$. The solution would be $x = 1/\sqrt{2}$, which doesn't exist in the rational number system. Notice the similarity between this problem and the one in Euclid's proof. In both cases we have curves that cut one another without having an intersection. In the present example, the curves are the graphs of the functions $y = x^2$ and $y = 1/2$.

The interpretation is that the real numbers are in some sense more densely packed than the rationals, and with two thousand years worth of hindsight, we can see that Euclid should have included a sixth postulate that expressed this density property. One possible way of stating such a postulate is the following. Let L be a ray, and O its endpoint. We think of O as the origin of the positive number line. Let P and Q be sets of points on L such that every point in P is closer to O than every point in Q. Then there exists some point Z on L such that Z lies at least as far from O as every point in P, but no farther than any point in Q. Technically this property is known as *completeness*. As an example, let $\mathrm{P} = \{x|x^2 < 2\}$ and $\mathrm{Q} = \{x|x^2 \geq 2\}$. Then the point Z would have to be $\sqrt{2}$, which shows that the rationals are not complete. The reals are complete, and the completeness axiom can serve as one of the fundamental axioms of the real numbers.

Note that the axiom refers to *sets* P and Q, and says that a certain fact is true for any choice of those sets; it therefore isn't the type of proposition that is covered by the transfer principle, and in fact it fails for the hyperreals, as we can see if P is the set of all infinitesimals and Q the positive real numbers.

Here is a skeletal proof of the intermediate value theorem, in which I'll make some simplifying assumptions and leave out some cases. We want to prove that if y is a continuous real-valued function on the real interval from a to b, and if y takes on values y_1 and y_2 at certain points within this interval, then for any y_3 between y_1 and y_2, there is some real x in the interval for which $y(x) = y_3$. I'll assume the case in which $x_1 < x_2$ and $y_1 < y_2$. Define sets of real numbers $\mathrm{P} = \{x | y \leq y_3\}$, and let $\mathrm{Q} = \{x | y \geq y_3\}$. For simplicity, I'll assume that every member of P is less than or equal to every member of Q, which happens, for example, if the function $y(x)$ is always increasing on the interval $[a, b]$. If P and Q intersect, then the theorem holds. Suppose instead that P and Q do not intersect. Using the completeness axiom, there exists some real x which is greater than or equal to every element of P and less than or equal to every element of Q. Suppose x belongs to P. Then the following statement is in the right form for the transfer principle to apply to it: for any number $x' > x$, $y(x') > y_3$. We can conclude that the statement is also true for the hyperreals, so that if $\mathrm{d}x$ is a positive infinitesimal and $x' = x + \mathrm{d}x$, we have $y(x) < y_3$, but $y(x + \mathrm{d}x) > y_3$. Then by continuity, $y(x) - y(x + \mathrm{d}x)$ is infinitesimal. But $y(x) < y_3$ and $y(x + \mathrm{d}x) > y_3$, so the standard part of $y(x)$ must equal y_3. By assumption y takes on real values for real arguments, so $y(x) = y_3$. The same reasoning applies if x belongs to Q, and since x must belong either to P or to Q, the result is proved.

For an alternative proof of the intermediate value theorem by an entirely different technique, see Keisler (references, p. 171).

As a side issue, we could ask whether there is anything like the intermediate value theorem that can be applied to functions on the hyperreals. Our definition of continuity on page 45 explicitly states that it only applies to real functions. Even if we could apply the definition to a function on the hyperreals, the proof given above would fail, since the hyperreals lack the completeness property. As a counterexample, let ϵ be some positive infinitesimal, and define a function y such that $y = -\epsilon$ when $\mathrm{st}(x) \leq 0$ and $y = \epsilon$ everywhere else. If we insist on applying the definition of continuity to this function, it appears to be continuous, so it violates the intermediate value theorem. Note, however, that the way this function is defined is different from the way we usually define functions on the hyperreals. Usually we define a function on the reals, say $y = x^2$, in language to which the transfer principle applies, and then we use the transfer principle to reason about the function's analog on the hyperreals. For instance, the function $y = x^2$ has the property that $y \geq 0$ everywhere, and the transfer principle guarantees that that's also true if we take $y = x^2$ as the definition of a function on the hyperreals. For functions defined in this way, the intermediate value theorem makes a statement that the transfer principle applies to, and it is therefore true for the hyperreal version of the function as well.

Proof of the extreme value theorem

The extreme value theorem was stated on page 49. Before we can prove it, we need to establish some preliminaries, which turn out to be interesting for their own sake.

Definition: Let C be a subset of the real numbers whose definition can be expressed in the type of language to which the transfer principle applies. Then C is *compact* if for every hyperreal number x satisfying the definition of C, the standard part of x exists and is a member of C.

To understand the content of this definition, we need to look at the two ways in which a set could fail to satisfy it.

First, suppose U is defined by $x \geq 0$. Then there are positive infinite hyperreal numbers that satisfy the definition, and their standard part is not defined, so U is not compact. The reason U is not compact is that it is unbounded.

Second, let V be defined by $0 \leq x < 1$. Then if dx is a positive infinitesimal, $1-dx$ satisfies the definition of V, but its standard part is 1, which is not in V, so V is not compact. The set V has boundary points at 0 and 1, and the reason it is not compact is that it doesn't contain its right-hand boundary point. A boundary point is a real number which is infinitesimally close to some points inside the set, and also to some other points that are on the outside.

We therefore arrive at the following alternative characterization of the notion of a compact set, whose proof is straightforward.

Theorem: A set is compact if and only if it is bounded and contains all of its boundary points.

Intuitively, the reason compact sets are interesting is that if you're standing inside a compact set and start taking steps in a certain direction, without ever turning around, you're guaranteed to approach some point in the set as a limit. (You might step over some gaps that aren't included in the set.) If the set was unbounded, you could just walk forever at a constant speed. If the set didn't contain its boundary point, then you could asymptotically approach the boundary, but the goal you were approaching wouldn't be a member of the set.

The following theorem turns out to be the most difficult part of the discussion.

Theorem: A compact set contains its maximum and minimum.
Proof: Let C be a compact set. We know it's bounded, so let M be the

set of all real numbers that are greater than any member of C. By the completeness property of the real numbers, there is some real number x between C and M. Let $^*C$ be the set of hyperreal numbers that satisfies the same definition that C does.

Every real x' greater than x fails to satisfy the condition that defines C, and by the transfer principle the same must be true if x' is any hyperreal, so if dx is a positive infinitesimal, $x + dx$ must be outside of $^*C$.

But now consider $x - dx$. The following statement holds for the reals: there is no number $x' < x$ that is greater than every member of C. By the transfer principle, we find that there is some hyperreal number q in $^*C$ that is greater than $x - dx$. But the standard part of q must equal x, for otherwise $\operatorname{st} q$ would be a member of C that was greater than x. Therefore x is a boundary point of C, and since C is compact, x is a member of C. We conclude C contains its maximum. A similar argument shows that C contains its minimum, so the theorem is proved.

There were two subtle things about this proof. The first was that we ended up constructing the set of hyperreals $^*C$, which was the hyperreal "big brother" of the real set C. This is exactly the sort of thing that the transfer principle does *not* guarantee we can do. However, if you look back through the proof, you can see that $^*C$ is used only as a notational convenience. Rather than talking about whether a certain number was a member of $^*C$, we could have referred, more cumbersomely, to whether or not it satisfied the condition that had originally been used to define C. The price we paid for this was a slight loss of generality. There are so many different sets of real numbers that they can't possibly all have explicit definitions that can be written down on a piece of paper. However, there is very little reason to be interested in studying the properties of a set that we were never able to define in the first place. The other subtlety was that we had to construct the auxiliary point $x - dx$, but there was not much we could actually say about $x - dx$ itself. In particular, it might or might not have been a member of C. For example, if C is defined by the condition $x = 0$, then $^*C$ likewise contains only the single element 0, and $x - dx$ is not a member of $^*C$. But if C is defined by $0 \leq x \leq 1$, then $x - dx$ is a member of $^*C$.

The original goal was to prove the extreme value theorem, which is a statement about continuous functions, but so far we haven't said anything about functions.

Lemma: Let f be a real function defined on a set of points C. Let D be

the image of C, i.e., the set of all values $f(x)$ that occur for some x in C. Then if f is continous and C is compact, D is compact as well. In other words, continuous functions take compact sets to compact sets.
Proof: Let $y = f(x)$ be any hyperreal output corresponding to a hyperreal input x in $^*C$. We need to prove that the standard part of y exists, and is a member of D. Since C is compact, the standard part of x exists and is a member of C. But then by continuity y differs only infinitesimally from $f(\text{st}x)$, which is real, so $\text{st}y = f(\text{st}x)$ is defined and is a member of D.

We are now ready to prove the extreme value theorem, in a version slightly more general than the one originally given on page 49.

The extreme value theorem: Any continuous function on a compact set achieves a maximum and minimum value, and does so at specific points in the set.

Proof: Let f be continuous, and let C be the compact set on which we seek its maximum and minimum. Then the image D as defined in the lemma above is compact. Therefore D contains its maximum and minimum values.

Proof of the mean value theorem

Suppose that the mean value theorem is violated. Let L be the set of all x in the interval from a to b such that $y(x) < \bar{y}$, and likewise let M be the set with $y(x) > \bar{y}$. If the theorem is violated, then the union of these two sets covers the entire interval from a to b. Neither one can be empty; if, for example, M was empty, then we would have $y < \bar{y}$ everywhere and also $\int_a^b y = \int_a^b \bar{y}$, but it follows directly from the definition of the definite integral that when one function is less than another, its integral is also less than the other's. Since y takes on values less than and greater than $\bar{y}$, it follows from the intermediate value theorem that y takes on the value $\bar{y}$ somewhere (intuitively, at a boundary between L and M).

Proof of the fundamental theorem of algebra

We start with the following lemma, which is intuitively obvious, because polynomials don't have asymptotes. Its proof is given after the proof of the main theorem.

Lemma: For any polynomial $P(z)$ in the complex plane, its magnitude $|P(z)|$ achieves its minimum value at some specific point z_o.

The fundamental theorem of algebra: In the complex number system, a nonzero nth-order polynomial has exactly n roots, i.e., it can be factored into the form $P(z) = (z-a_1)(z-a_2)\ldots(z-a_n)$, where the a_i are complex numbers.

Proof: The proofs in the cases of $n = 0$ and 1 are trivial, so our strategy is to reduce higher-n cases to lower ones. If an nth-degree polynomial P has at least one root, a, then we can always reduce it to a polynomial of degree $n-1$ by dividing it by $(z-a)$. Therefore the theorem is proved by induction provided that we can show that every polynomial of degree greater than zero has at least one root.

Suppose, on the contrary, that there is an nth order polynomial $P(z)$, with $n > 0$, that has no roots at all. Then by the lemma $|P|$ achieves its minimum value at some point z_o. To make things more simple and concrete, we can construct another polynomial $Q(z) = P(z+z_o)/P(z_o)$, so that $|Q|$ has a minimum value of 1, achieved at $Q(0) = 1$. This means that Q's constant term is 1. What about its other terms? Let $Q(z) = 1 + c_1 z + \ldots + c_n z^n$. Suppose c_1 was nonzero. Then for infinitesimally small values of z, the terms of order z^2 and higher would be negligible, and we could make $Q(z)$ be a real number less than one by an appropriate choice of z's argument. Therefore c_1 must be zero. But that means that if c_2 is nonzero, then for infinitesimally small z, the z^2 term dominates the z^3 and higher terms, and again this would allow us to make $Q(z)$ be real and less than one for appropriately chosen values of z. Continuing this process, we find that $Q(z)$ has no terms at all beyond the constant term, i.e., $Q(z) = 1$. This contradicts the assumption that n was greater than zero, so we've proved by contradiction that there is no P with the properties claimed.

Uninteresting proof of the lemma: Let $M(r)$ be the minimum value of $|P(z)|$ on the disk defined by $|z| \le r$. We first prove that $M(r)$ can't asymptotically approach a minimum as r approaches infinity. Suppose to the contrary: for every r, there is some $r' > r$ with $M(r') < M(r)$. Then by the transfer principle, the same would have to be true for hyperreal values of r. But it's clear that if r is infinite, the lower-order terms of P will be infinitesimally small compared to the highest-order term, and therefore $M(r)$ is infinite for infinite values of r, which is a contradiction, since by construction M is decreasing, and finite for finite r. We can therefore conclude by the extreme value theorem that M achieves its minimum for some specific value of r. The least such r describes a circle $|z| = r$ in the complex plane, and the minimum of $|P|$ on this circle must be the same as its global minimum. Applying the extreme value function to $|P(z)|$ as a function of arg z on the interval $0 \leq \arg z \leq 2\pi$, we establish the desired result.

B Answers and solutions

Answers to Self-Checks

Answers to self-checks for chapter 3

page 72, self-check 1:

The area under the curve from 130 to 135 cm is about 3/4 of a rectangle. The area from 135 to 140 cm is about 1.5 rectangles. The number of people in the second range is about twice as much. We could have converted these to actual probabilities (1 rectangle = 5 cm $\times$ 0.005 cm^{-1} = 0.025), but that would have been pointless, because we were just going to compare the two areas.

Answers to self-checks for chapter 5

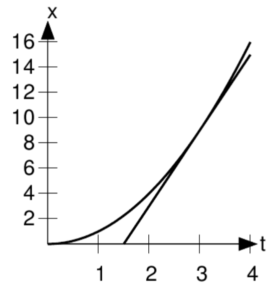
page 104, self-check 1: Say we're looking for $u = \sqrt{z}$, i.e., we want a number u that, multiplied by itself, equals z. Multiplication multiplies the magnitudes, so the magnitude of u can be found by taking the square root of the magnitude of z. Since multiplication also adds the arguments of the numbers, squaring a number doubles its argument. Therefore we can simply divide the argument of z by two to find the argument of u. This results in one of the square roots of z. There is another one, which is $-u$, since $(-u)^2$ is the same as u^2. This may seem a little odd: if u was chosen so that doubling its argument gave the argument of z, then how can the same be true for $-u$? Well for example, suppose the argument of z is $4\,^\circ$. Then $\arg u = 2\,^\circ$, and $\arg(-u) = 182\,^\circ$. Doubling 182 gives 364, which is actually a synonym for 4 degrees.

Solutions to homework problems

Solutions for chapter 1

page 22, problem 1:

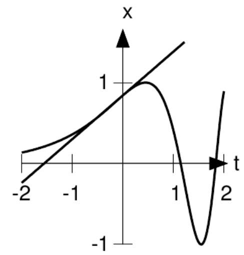
The tangent line has to pass through the point (3,9), and it also seems, at least approximately, to pass through (1.5,0). This gives it a slope of $(9-0)/(3-1.5) = 9/1.5 = 6$, and that's exactly what $2t$ is at $t = 3$.



a / Problem 1.

page 22, problem 2:

The tangent line has to pass through the point $(0, \sin(e^0)) = (0, 0.84)$, and it also seems, at least approximately, to pass through (-1.6,0). This gives it a slope of $(0.84-0)/(0-(-1.6)) = 0.84/1.6 = 0.53$. The more accurate result given in the problem can be found using the methods of chapter 2.



b / Problem 2.

page 22, problem 3:
The derivative is a rate of change, so the derivatives of the constants 1 and 7, which don't change, are clearly zero. The derivative can be interpreted geometrically as the slope of the tangent line, and since the functions t and $7t$ *are* lines, their derivatives are simply their slopes, 1, and 7. All of these could also have been found using the formula that says the derivative of t^k is kt^{k-1}, but it wasn't really necessary to get that fancy. To find the derivative of t^2, we can use the formula, which gives $2t$. One of the properties of the derivative is that multiplying a function by a constant multiplies its derivative by the same constant, so the derivative of $7t^2$ must be $(7)(2t) = 14t$. By similar reasoning, the derivatives of t^3 and $7t^3$ are $3t^2$ and $21t^2$, respectively.

page 22, problem 4:

One of the properties of the derivative is that the derivative of a sum is the sum of the derivatives, so we can get this by adding up the derivatives of $3t^7$, $-4t^2$, and 6. The derivatives of the three terms are $21t^6$, $-8t$, and 0, so the derivative of the whole thing is $21t^6 - 8t$.

page 22, problem 5:

This is exactly like problem 4, except that instead of explicit numerical constants like 3 and -4, this problem involves symbolic constants a, b, and c. The result is $2at + bt$.

page 22, problem 6:

The first thing that comes to mind is $3t$. Its graph would be a line with a slope of 3, passing through the origin. Any other line with a slope of 3 would work too, e.g., $3t + 1$.

page 22, problem 7:

Differentiation lowers the power of a monomial by one, so to get something with an exponent of 7, we need to differentiate something with an exponent of 8. The derivative of t^8 would be $8t^7$, which is eight times too big, so we really need $(t^8/8)$. As in problem 6, any other function that differed by an additive constant would also work, e.g., $(t^8/8) + 1$.

page 22, problem 8:

This is just like problem 7, but we need something whose derivative is three times bigger. Since multiplying by a constant multiplies the derivative by the same constant, the way to accomplish this is to take the answer to problem 7, and multiply by three. A possible answer is $(3/8)t^8$, or that function plus any constant.

page 22, problem 9:

This is just a slight generalization of problem 8. Since the derivative of a sum is the sum of the derivatives, we just need to handle each term individually, and then add up the results. The answer is $(3/8)t^8 - (4/3)t^3 + 6t$, or that function plus any constant.

page 22, problem 10:

The function $v = (4/3)\pi(ct)^3$ looks scary and complicated, but it's nothing more than a constant multiplied by t^3, if we rewrite it as $v = \left[(4/3)\pi c^3\right] t^3$. The whole thing in square brackets is simply one big constant, which just comes along for the ride when we differentiate. The result is $\dot{v} = \left[(4/3)\pi c^3\right](3t^2)$, or, simplifying, $\dot{v} = \left(4\pi c^3\right) t^2$. (For further physical insight, we can factor this as $\left[4\pi(ct)^2\right] c$, where ct is the radius of the expanding sphere, and the part in brackets is the sphere's surface area.)

For purposes of checking the units, we can ignore the unitless constant 4π, which just leaves c^3t^2. This has units of (meters per second)^3(seconds)^2, which works out to be cubic meters per second. That makes sense, because it tells us how quickly a volume is increasing over time.

page 22, problem 11:
This is similar to problem 10, in that it looks scary, but we can rewrite it as a simple monomial, $K = (1/2)mv^2 = (1/2)m(at)^2 = (ma^2/2)t^2$. The derivative is $(ma^2/2)(2t) = ma^2t$. The car needs more and more power to accelerate as its speed increases.

To check the units, we just need to show that the expression ma^2t has units that are like those of the original expression for K, but divided by seconds, since it's a rate of change of K over time. This indeed works out, since the only change in the factors that aren't unitless is the reduction of the powet of t from 2 to 1.

page 22, problem 12:

The area is $a = \ell^2 = (1+\alpha T)^2\ell_o^2$. To make this into something we know how to differentiate, we need to square out the expression involving T, and make it into something that is expressed explicitly as a polynomial:

$$a = \ell_o^2 + 2\ell_o^2\alpha T + \ell_o^2\alpha^2 T^2$$

Now this is just like problem 5, except that the constants superficially look more complicated. The result is

$$\begin{aligned}\dot{a} &= 2\ell_o^2\alpha + 2\ell_o^2\alpha^2 T \\ &= 2\ell_o^2\left(\alpha + \alpha^2 T\right) \qquad .\end{aligned}$$

We expect the units of the result to be area per unit temperature, e.g., degrees per square meter. This is a little tricky, because we have to figure out what units are implied for the constant α. Since the question talks about $1 + \alpha T$, apparently the quantity αT is unitless. (The 1 is unitless, and you can't add things that have different units.) Therefore the units of α must be "per degree," or inverse degrees. It wouldn't make sense to add α and $\alpha^2 T$ unless they had the same units (and you can check for yourself that they do), so the whole thing inside the parentheses must have units of inverse degrees. Multiplying by the ℓ_o^2 in front, we have units of area per degree, which is what we expected.

page 23, problem 13:

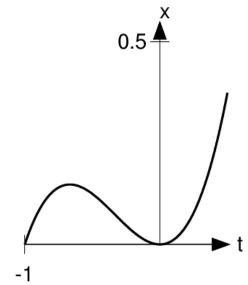
The first derivative is $6t^2 - 1$. Going again, the answer is $12t$.

page 23, problem 14:

The first derivative is $3t^2+2t$, and the second is $6t+2$. Setting this equal to zero and solving for t, we find $t = -1/3$. Looking at the graph, it does look like the concavity is down for $t < -1/3$, and up for $t > -1/3$.

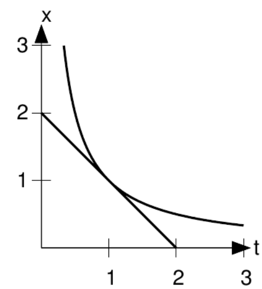
page 23, problem 15:

I chose $k = -1$, and $t = 1$. In other words, I'm going to check the slope of the function $x = t^{-1} = 1/r$ at $t = 1$, and see whether it really equals



c / Problem 14.

$kt^{k-1} = -1$. Before even doing the graph, I note that the sign makes sense: the function $1/t$ is decreasing for $t > 0$, so its slope should indeed be negative.



d / Problem 15.

The tangent line seems to connect the points (0,2) and (2,0), so its slope does indeed look like it's -1.

The problem asked us to consider the logical meaning of the two possible outcomes. If the slope had been significantly different from -1 given the accuracy of our result, the conclusion would have been that it was incorrect to extend the rule to negative values of k. Although our example did come out consistent with the rule, that doesn't prove the rule in general. An example can disprove a conjecture, but can't prove it. Of course, if we tried lots and lots of examples, and they all worked, our confidence in the conjecture would be increased.

page 23, problem 16:

A minimum would occur where the derivative was zero. First we rewrite the function in a form that we know how to differentiate:

$$E(r) = ka^{12}r^{-12} - 2ka^6r^{-6}$$

We're told to have faith that the derivative of t^k is kt^{k-1} even for $k < 0$, so

$$\begin{aligned} 0 &= \dot{E} \\ &= -12ka^{12}r^{-13} + 12ka^6r^{-7} \end{aligned}$$

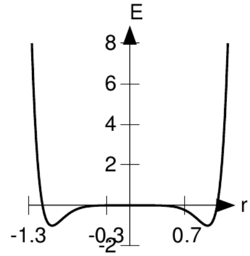
To simplify, we divide both sides by $12k$. The left side was already zero, so it keeps being zero.

$$\begin{aligned} 0 &= -a^{12}r^{-13} + a^6r^{-7} \\ a^{12}r^{-13} &= a^6r^{-7} \\ a^{12} &= a^6r^6 \\ a^6 &= r^6 \\ r &= \pm a \end{aligned}$$

To check that this is a minimum, not a maximum or a point of inflection, one method is to construct a graph. The constants a and k are irrelevant to this issue. Changing a just rescales the horizontal r axis, and changing k does the same for the vertical E axis. That means we can arbitrarily set $a = 1$ and $k = 1$, and construct the graph shown in the figure. The points $r = \pm a$ are now simply $r = \pm 1$. From the graph, we can see that they're clearly minima. Physically, the minimum at $r = -a$ can be interpreted as the same physical configuration of the molecule, but with the positions of the atoms reversed. It makes sense that $r = -a$ behaves the same as $r = a$, since physically the behavior of the system has to be symmetric, regardless of whether we view it from in front or from behind.

The other method of checking that $r = a$ is a minimum is to take the second derivative. As before, the values of a and k are irrelevant, and can be set to 1. We then have

$$\begin{aligned} \dot{E} &= -12r^{-13} + 12r^{-7} \\ \ddot{E} &= 156r^{-14} - 84r^{-8} \qquad . \end{aligned}$$



e / Problem 16.

Plugging in $r = \pm 1$, we get a positive result, which confirms that the concavity is upward.

page 23, problem 17:

Since polynomials don't have kinks or endpoints in their graphs, the maxima and minima must be points where the derivative is zero. Differentiation bumps down all the powers of a polynomial by one, so the derivative of a third-order polynomial is a second-order polynomial. A second-order polynomial can have at most two real roots (values of t for which it equals zero), which are given by the quadratic formula. (If the number inside the square root in the quadratic formula is zero or negative, there could be less than two real roots.) That means a third-order polynomial can have at most two maxima or minima.

page 23, problem 18:

Considering V as a function of h, with b treated as a constant, we have for the slope of its graph

$$\dot{V} = \frac{e_V}{e_h} \qquad ,$$

so

$$\begin{aligned} e_V &= \dot{V} \cdot e_h \\ &= \frac{1}{3} b e_h \end{aligned}$$

page 23, problem 19:

Thinking of the rocket's height as a function of time, we can see that goal is to measure the function at its maximum. The derivative is zero at the maximum, so the error incurred due to timing is approximately zero. She should not worry about the timing error too much. Other factors are likely to be more important, e.g., the rocket may not rise exactly vertically above the launchpad.

Solutions for chapter 2

page 58, problem 1:

$$\begin{aligned}\frac{\mathrm{d}x}{\mathrm{d}t} &= \frac{(t+\mathrm{d}t)^4 - t^4}{\mathrm{d}t} \\ &= \frac{4t^3\,\mathrm{d}t + 6t^2\mathrm{d}t^2 + 4t\,\mathrm{d}t^3 + \mathrm{d}t^4}{\mathrm{d}t} \\ &= 4t^3 + \ldots \qquad ,\end{aligned}$$

where ... indicates infinitesimal terms. The derivative is the standard part of this, which is $4t^3$.

page 58, problem 2:

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \frac{\cos(t+\mathrm{d}t) - \cos t}{\mathrm{d}t}$$

The identity $\cos(\alpha+\beta) = \cos\alpha\cos\beta - \sin\alpha\sin\beta$ then gives

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \frac{\cos t\ \cos\mathrm{d}t - \sin t\ \sin\mathrm{d}t - \cos t}{\mathrm{d}t} \qquad .$$

The small-angle approximations $\cos\mathrm{d}t \approx 1$ and $\sin\mathrm{d}t \approx \mathrm{d}t$ result in

$$\begin{aligned}\frac{\mathrm{d}x}{\mathrm{d}t} &= \frac{-\sin t\,\mathrm{d}t}{\mathrm{d}t} \\ &= -\sin t \qquad .\end{aligned}$$

page 58, problem 3:

H	$\sqrt{H+1} - \sqrt{H-1}$
1000	.032
$1000,000$	0.0010
$1000,000,000$	0.00032

The result is getting smaller and smaller, so it seems reasonable to guess that if H is infinite, the expression gives an infinitesimal result.

page 58, problem 4:

$\mathrm{d}x$	$\sqrt{\mathrm{d}x}$
.1	.32
.001	.032
.00001	.0032

The square root is getting smaller, but is not getting smaller as fast as the number itself. In proportion to the original number, the square root is actually getting *bigger*. It looks like $\sqrt{\mathrm{d}x}$ is infinitesimal, but it's still infinitely big compared to $\mathrm{d}x$. This makes sense, because $\sqrt{\mathrm{d}x}$ equals $\mathrm{d}x^{1/2}$. we already knew that $\mathrm{d}x^0$, which equals 1, was infinitely big compared to $\mathrm{d}x^1$, which equals $\mathrm{d}x$. In the hierarchy of infinitesimals, $\mathrm{d}x^{1/2}$ fits in between $\mathrm{d}x^0$ and $\mathrm{d}x^1$.

page 58, problem 5:
Statements (a)-(d), and (f)-(g) are all valid for the hyperreals, because they meet the test of being directly translatable, without having to interpret the meaning of things like particular subsets of the reals in the context of the hyperreals.

Statement (e), however, refers to the rational numbers, a particular subset of the reals, that that means that it can't be mindlessly translated into a statement about the hyperreals, unless we had figured out a way to translate the set of rational numbers into some corresponding subset of the hyperreal numbers like the hyperrationals! This is not the type of statement that the transfer principle deals with. The statement is not true if we try to change "real" to "hyperreal" while leaving "rational" alone; for example, it's not true that there's a rational number that lies between the hyperreal numbers 0 and $0 + \mathrm{d}x$, where $\mathrm{d}x$ is infinitesimal.

page 59, problem 8: This would be a horrible problem if we had to expand this as a polynomial with 101 terms, as in chapter 1! But now we know the chain rule, so it's easy. The derivative is

$$\left[100(2x+3)^{99}\right][2] \qquad ,$$

where the first factor in brackets is the derivative of the function on the outside, and the second one is the derivative of the "inside stuff." Simplifying a little, the answer is $200(2x + 3)^{99}$.

page 59, problem 9:
Applying the product rule, we get

$$(x+1)^{99}(x+2)^{200} + (x+1)^{100}(x+2)^{199} \qquad .$$

(The chain rule was also required, but in a trivial way — for both of the factors, the derivative of the "inside stuff" was one.)

page 59, problem 10:
The derivative of e^{7x} is $e^{7x} \cdot 7$, where the first factor is the derivative of the outside stuff (the derivative of a base-e exponential is just the same thing), and the second factor is the derivative of the inside stuff. This would normally be written as $7e^{7x}$.

The derivative of the second function is $e^{e^x} e^x$, with the second exponential factor coming from the chain rule.

page 59, problem 11:
We need to put together three different ideas here: (1) When a function to be differentiated is multiplied by a constant, the constant just comes along for the ride. (2) The derivative of the sine is the cosine. (3) We need to use the chain rule. The result is $-ab\cos(bx + c)$.

page 59, problem 12:
If we just wanted to fine the integral of $\sin x$, the answer would be $-\cos x$ (or $-\cos x$ plus an arbitrary constant), since the derivative would be $-(-\sin x)$, which would take us back to the original function. The obvious thing to guess for the integral of $a\sin(bx + c)$ would therefore be $-a\cos(bx+c)$, which almost works, but not quite. The derivative of this function would be $ab\sin(bx + c)$, with the pesky factor of b coming from the chain rule. Therefore what we really wanted was the function $-(a/b)\cos(bx + c)$.

page 59, problem 14:
To find a maximum, we take the derivative and set it equal to zero. The whole factor of $2v^2/g$ in front is just one big constant, so it comes along for the ride. To differentiate the factor of $\sin\theta \ \cos\theta$, we need to use the chain rule, plus the fact that the derivative of sin is cos, and the derivative of cos is $-\sin$.

$$
\begin{aligned}
0 &= \frac{2v^2}{g}(\cos\theta \ \cos\theta + \sin\theta(-\sin\theta)) \\
0 &= \cos^2\theta - \sin^2\theta \\
\cos\theta &= \pm\sin\theta
\end{aligned}
$$

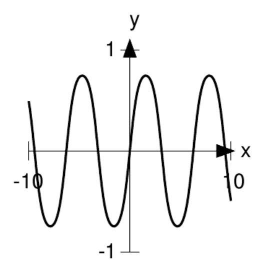
We're interested in angles between, 0 and 90 degrees, for which both the sine and the cosine are positive, so

$$\begin{aligned} \cos\theta &= \sin\theta \\ \tan\theta &= 1 \\ \theta &= 45\,^\circ \qquad . \end{aligned}$$

To check that this is really a maximum, not a minimum or an inflection point, we could resort to the second derivative test, but we know the graph of $R(\theta)$ is zero at $\theta = 0$ and $\theta = 90\,^\circ$, and positive in between, so this must be a maximum.

page 59, problem 15:
Taking the derivative and setting it equal to zero, we have $(e^x - e^{-x})/2 = 0$, so $e^x = e^{-x}$, which occurs only at $x = 0$. The second derivative is $(e^x + e^{-x})/2$ (the same as the original function), which is positive for all x, so the function is everywhere concave up, and this is a minimum.

page 59, problem 16:
There are no kinks, endpoints, etc., so extrema will occur only in places where the derivative is zero. Applying the chain rule, we find the derivative to be $\cos(\sin(\sin x))\cos(\sin x)\cos x$. This will be zero if any of the three factors is zero. We have $\cos u = 0$ only when $|u| \ge \pi/2$, and $\pi/2$ is greater than 1, so it's not possible for either of the first two factors to equal zero. The derivative will therefore equal zero if and only if $\cos x = 0$, which happens in the same places where the derivative of $\sin x$ is zero, at $x = \pi/2 + \pi n$, where n is an integer.



f / Problem 16.

This essentially completes the required demonstration, but there is one more technical issue, which is that it's conceivable that some of these could be points of inflection. Constructing a graph of $\sin(\sin(\sin x))$ gives us the necessary insight to see that this can't be the case. The function essentially looks like the sine function, but its extrema have been "shaved down" a little, giving them slightly flatter tips that don't quite extend out to ± 1. It's therefore fairly clear that these aren't points of inflection. To prove this more rigorously, we could take the second derivative and show that it was nonzero at the places where the first derivative is zero. That would be messy. A less tedious argument is as follows. We can tell from its formula that the function is *periodic*, i.e., it has the property that $f(x+\ell) = f(x)$, for $\ell = 2\pi$. This follows because the innermost sine function is periodic, and the outer layers only depend on the result of the inner layer. Therefore all the points of the form $\pi/2 + 2\pi n$ have the same behavior. Either they're all maxima or they're all points of inflection. But clearly a function can't oscillate back and forth without having any maxima at all, so they must all be maxima. A similar argument applies to the minima.

page 60, problem 17:
(a) As suggested, let $c = \sqrt{g/A}$, so that $d = A \ln \cosh ct = A \ln\left(e^{ct} + e^{-ct}\right)$. Applying the chain rule, the velocity is

$$A \frac{ce^{ct} - ce^{-ct}}{\cosh ct} \qquad .$$

(b) The expression can be rewritten as $Ac \tanh ct$.
(c) For large t, the e^{-ct} terms become negligible, so the velocity is $Ace^{ct}/e^{ct} = Ac$. (d) From the original expression, A must have units of distance, since the logarithm is unitless. Also, since ct occurs inside a function, ct must be unitless, which means that c has units of inverse time. The answers to parts b and c get their units from the factors of Ac, which have units of distance multiplied by inverse time, or velocity.

page 60, problem 18:
Since I've advocated not memorizing the quotient rule, I'll do this one from first principles, using the product rule.

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}\theta} \tan\theta \\
&\quad = \frac{\mathrm{d}}{\mathrm{d}\theta}\left(\frac{\sin\theta}{\cos\theta}\right) \\
&\quad = \frac{\mathrm{d}}{\mathrm{d}\theta}\left[\sin\theta\,(\cos\theta)^{-1}\right] \\
&\quad = \cos\theta\,(\cos\theta)^{-1} + (\sin\theta)(-1)(\cos\theta)^{-2}(-\sin\theta) \\
&\quad = 1 + \tan^2\theta
\end{aligned}$$

(Using a trig identity, this can also be rewritten as $\sec^2\theta$.)

page 60, problem 19:
Reexpressing $\sqrt[3]{x}$ as $x^{1/3}$, the derivative is $(1/3)x^{-2/3}$.

page 60, problem 20:
(a) Using the chain rule, the derivative of $(x^2+1)^{1/2}$ is $(1/2)(x^2+1)^{-1/2}(2x) = x(x^2+1)^{-1/2}$.
(b) This is the same as a, except that the 1 is replaced with an a^2, so the answer is $x(x^2+a^2)^{-1/2}$. The idea would be that a has the same units as x.
(c) This can be rewritten as $(a+x)^{-1/2}$, giving a derivative of $(-1/2)(a+x)^{-3/2}$.
(d) This is similar to c, but we pick up a factor of $-2x$ from the chain rule, making the result $ax(a-x^2)^{-3/2}$.

page 60, problem 21:
By the chain rule, the result is $2/(2t+1)$.

page 60, problem 22:
Using the product rule, we have

$$\left(\frac{\mathrm{d}}{\mathrm{d}x}3\right)\sin x + 3\left(\frac{\mathrm{d}}{\mathrm{d}x}\sin x\right) \qquad ,$$

but the derivative of a constant is zero, so the first term goes away, and we get $3\cos x$, which is what we would have had just from the usual method of treating multiplicative constants.

page 60, problem 23:

```
N(Gamma(2))
   1
N(Gamma(2.00001))
   1.0000042278
N( (1.0000042278-1)/(.00001) )
   0.4227799998
```

Probably only the first few digits of this are reliable.

page 60, problem 24:
The area and volume are

$$A = 2\pi r\ell + 2\pi r^2$$

and

$$V = \pi r^2 \ell \qquad .$$

The strategy is to use the equation for A, which is a constant, to eliminate the variable ℓ, and then maximize V in terms of r.

$$\ell = (A - 2\pi r^2)/2\pi r$$

Substituting this expression for ℓ back into the equation for V,

$$V = \frac{1}{2} r A - \pi r^3 \qquad .$$

To maximize this with respect to r, we take the derivative and set it equal to zero.

$$\begin{gathered} 0 = \frac{1}{2} A - 3\pi r^2 \\ A = 6\pi r^2 \\ \ell = (6\pi r^2 - 2\pi r^2)/2\pi r \\ \ell = 2r \end{gathered}$$

In other words, the length should be the same as the diameter.

page 61, problem 25:
(a) We can break the expression down into three factors: the constant $m/2$ in front, the nonrelativistic velocity dependence v^2, and the relativistic correction factor $(1 - v^2/c^2)^{-1/2}$. Rather than substituting in at for v, it's a little less messy to calculate $\mathrm{d}K/\mathrm{d}t = (\mathrm{d}K/\mathrm{d}v)(\mathrm{d}v/\mathrm{d}t) = a\mathrm{d}K/\mathrm{d}v$. Using the product rule, we have

$$\begin{aligned} \frac{\mathrm{d}K}{\mathrm{d}t} &= a \cdot \frac{1}{2} m \left[2v \left(1 - \tfrac{v^2}{c^2}\right)^{-1/2} \right. \\ &\left. + v^2 \cdot \left(-\tfrac{1}{2}\right) \left(1 - \tfrac{v^2}{c^2}\right)^{-3/2} \left(-\tfrac{2v}{c^2}\right) \right] \\ &= m a^2 t \left[\left(1 - \tfrac{v^2}{c^2}\right)^{-1/2} \right. \\ &\left. + \frac{v^2}{2c^2} \left(1 - \tfrac{v^2}{c^2}\right)^{-3/2} \right] \end{aligned}$$

(b) The expression ma^2t is the nonrelativistic (classical) result, and has the correct units of kinetic energy divided by time. The factor in square

brackets is the relativistic correction, which is unitless.
(c) As v gets closer and closer to c, the expression $1 - v^2/c^2$ approaches zero, so both the terms in the relativistic correction blow up to positive infinity.

page 61, problem 26:
We already know it works for positive x, so we only need to check it for negative x. For negative values of x, the chain rule tells us that the derivative is $1/|x|$, multiplied by -1, since $\mathrm{d}|x|/\mathrm{d}x = -1$. This gives $-1/|x|$, which is the same as $1/x$, since x is assumed negative.

page 61, problem 27:
Let $f = \mathrm{d}x^k/\mathrm{d}x$ be the unknown function. Then

$$\begin{aligned}1 &= \frac{\mathrm{d}x}{\mathrm{d}x} \\ &= \frac{\mathrm{d}}{\mathrm{d}x}\left(x^k x^{-k+1}\right) \\ &= fx^{-k+1} + x^k(-k+1)x^{-k} \qquad ,\end{aligned}$$

where we can use the ordinary rule for derivatives of powers on x^{-k+1}, since $-k+1$ is positive. Solving for f, we have the desired result.

page 61, problem 29:

(a) The Weierstrass definition requires that if we're given a particular ϵ, and we be able to find a δ so small that $f(x) + g(x)$ differs from $F + G$ by at most ϵ for $|x - a| < \delta$. But the Weierstrass definition also tells us that given $\epsilon/2$, we can find a δ such that f differs from F by at most $\epsilon/2$, and likewise for g and G. The amount by which $f + g$ differs from $F + G$ is then at most $\epsilon/2 + \epsilon/2$, which completes the proof.

(b) Let $\mathrm{d}x$ be infinitesimal. Then the definition of the limit in terms of infinitesimals says that the standard part of $f(a + \mathrm{d}x)$ differs at most infinitesimally from F, and likewise for g and G. This means that $f + g$ differs from $F + G$ by the sum of two infinitesimals, which is itself an infinitesimal, and therefore the standard part of $f+g$ evaluated at $x+\mathrm{d}x$ equals $F + G$, satisfying the definition.

page 62, problem 33:

The normal definition of a repeating decimal such as $0.999\ldots$ is that it is the *limit* of the sequence 0.9, 0.99, ..., and the limit is a real number, by definition. $0.999\ldots$ equals 1. However, there is an intuition that the limiting process 0.9, 0.99, ... "never quite gets there." This intuition can, in fact, be formalized in the construction described beginning on page 127; we can define a hyperreal number based on the sequence 0.9, 0.99, ..., and it is a number infinitesimally less than one. This is not, however, the normal way of defining the symbol $0.999\ldots$, and we probably wouldn't want to change the definition so that it was. If it was, then $0.333\ldots$ would not equal $1/3$.

page 62, problem 34:

Converting these into Leibniz notation, we find

$$\frac{\mathrm{d}f}{\mathrm{d}x} = \frac{\mathrm{d}g}{\mathrm{d}h}$$

and

$$\frac{\mathrm{d}f}{\mathrm{d}x} = \frac{\mathrm{d}g}{\mathrm{d}h} \cdot h \qquad .$$

To prove something is not true in general, it suffices to find one counterexample. Suppose that g and h are both unitless, and x has units of seconds. The value of f is defined by the output of g, so f must also be unitless. Since f is unitless, $\mathrm{d}f/\mathrm{d}x$ has units of inverse seconds ("per second"). But this doesn't match the units of either of the proposed expressions, because they're both unitless. The correct chain rule, however, works. In the equation

$$\frac{\mathrm{d}f}{\mathrm{d}x} = \frac{\mathrm{d}g}{\mathrm{d}h} \cdot \frac{\mathrm{d}h}{\mathrm{d}x} \qquad ,$$

the right-hand side consists of a unitless factor multiplied by a factor with units of inverse seconds, so its units are inverse seconds, matching the left-hand side.

page 62, problem 35:

We can make life a lot easier by observing that the function $s(f)$ will be maximized when the expression inside the square root is minimized. Also, since f is squared every time it occurs, we can change to a variable $x = f^2$, and then once the optimal value of x is found we can take its square root in order to find the optimal f. The function to be optimized is then

$$a(x - f_\mathrm{o}^2)^2 + bx \qquad .$$

Differentiating this and setting the derivative equal to zero, we find

$$2a(x - f_\mathrm{o}^2) + b = 0 \qquad ,$$

which results in $x = f_\mathrm{o}^2 - b/2a$, or

$$f = \sqrt{f_\mathrm{o}^2 - b/2a} \qquad ,$$

(choosing the positive root, since f represents a frequencies, and frequencies are positive by definition). Note that the quantity inside the square root involves the square of a frequency, but then we take its square root, so the units of the result turn out to be frequency, which makes sense. We can see that if b is small, the second term is small, and the maximum occurs very nearly at f_o.

There is one subtle issue that was glossed over above, which is that the graph on page 62 shows *two* extrema: a minimum at $f = 0$ and a maximum at $f > 0$. What happened to the $f = 0$ minimum? The issue is that I was a little sloppy with the change of variables. Let I stand for the quantity inside the square root in the original expression for s. Then by the chain rule,

$$\frac{\mathrm{d}s}{\mathrm{d}f} = \frac{\mathrm{d}s}{\mathrm{d}I} \cdot \frac{\mathrm{d}I}{\mathrm{d}x} \cdot \frac{\mathrm{d}x}{\mathrm{d}f}.$$

We looked for the place where $\mathrm{d}I/\mathrm{d}x$ was zero, but $\mathrm{d}s/\mathrm{d}f$ could also be zero if one of the other factors was zero. This is what happens at $f = 0$, where $\mathrm{d}x/\mathrm{d}f = 0$.

Solutions for chapter 6

page 100, problem 1:

We can define the sequence $f(n)$ as converging to ℓ if the following is true: for any real number ϵ, there exists an integer N such that for all n greater than N, the value of f lies within the range from $\ell - \epsilon$ to $\ell + \epsilon$.

page 100, problem 2:

(a) The convergence of the series is defined in terms of the convergence of its partial sums, which are 1, 0, 1, 0, ... In the notation used in the definition given in the solution to problem 1 above, suppose we pick $\epsilon = 1/4$. Then there is clearly no way to choose any numbers ℓ and N that would satisfy the definition, for regardless of N, ℓ would have to be both greater than 3/4 and less than 1/4 in order to agree with the zeroes and ones that occur beyond the Nth member of the sequence.

(b) As remarked on page 92, the axioms of the real number system, such as associativity, only deal with finite sums, not infinite ones. To see that absurd conclusions result from attempting to apply them to infinite sums, consider that by the same type of argument we could group the sum as $1 + (-1 + 1) + (-1 + 1) + \ldots$, which would equal 1.

page 100, problem 3:

The quantity x^n can be reexpressed as $e^{n \ln x}$, where $\ln x$ is negative by hypothesis. The integral of this exponential *with respect to* n is a similar exponential with a constant factor in front, and this converges as n approaches infinity.

page 100, problem 4:

(a) Applying the integral test, we find that the integral of $1/x^2$ is $-1/x$, which converges as x approaches infinity, so the series converges as well.

(b) This is an alternating series whose terms approach zero, so it converges. However, the terms get small extremely slowly, so an extraordinarily large number of terms would be required in order to get any kind of decent approximation to the sum. In fact, it is impossible to carry out a straightforward numerical evaluation of this sum because it would require such an enormous number of terms that the rounding errors would overwhelm the result.

(c) Split the sum into two sums, one for the 1103 term and one for the $26390k$. The ratio of the two factorials is always less than 4^{4k}, so discarding constant factors, the first sum is less than a geometric series with $x = (4/396)^4 < 1$, and must therefore converge. The second sum is less than a series of the form kx^k. This one also converges, by the integral test. (It has to be integrated with respect to k, not x, and the integration can be done by parts.) Since both separate sums converge, the entire sum converges. This bizarre-looking expression was formulated and shown to equal $1/\pi$ by the self-taught genius Srinivasa Ramanujan (1887-1920).

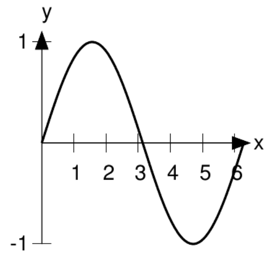
Solutions for chapter 3

page 75, problem 1:

```
a := 0;
b := 1;
H := 1000;
dt := (b-a)/H;
sum := 0;
t := a;
While (t<=b) [
  sum := N(sum+Exp(x^2)*dt);
  t := N(t+dt);
];
Echo(sum);
```

The result is 1.46.

page 75, problem 2:
The derivative of the cosine is minus the sine, so to get a function whose



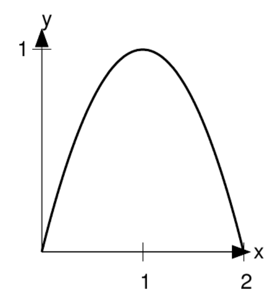
g / Problem 2.

derivative is the sine, we need minus the cosine.

$$\begin{aligned}\int_0^{2\pi} \sin x \, \mathrm{d}x \\ &= \left(-\cos x\right)\Big|_0^{2\pi} \\ &= (-\cos 2\pi) - (-\cos 0) \\ &= (-1) - (-1) \\ &= 0\end{aligned}$$

As shown in figure g, the graph has equal amounts of area above and below the x axis. The area below the axis counts as negative area, so the total is zero.

page 75, problem 3:



h / Problem 3.

The rectangular area of the graph is 2, and the area under the curve fills a little more than half of that, so let's guess 1.4.

$$\begin{aligned}\int_0^2 -x^2+2x &= \left(-\frac{1}{3}x^3+x^2\right)\Big|_0^2 \\ &= (-8/3+4)-(0) \\ &= 4/3\end{aligned}$$

This is roughly what we were expecting from our visual estimate.

page 75, problem 4:
Over this interval, the value of the sin function varies from 0 to 1, and it spends more time above 1/2 than below it, so we expect the average to be somewhat greater than 1/2. The exact result is

$$\begin{aligned}\overline{\sin} &= \frac{1}{\pi-0}\int_0^\pi \sin x\,\mathrm{d}x \\ &= \frac{1}{\pi}\left(-\cos x\right)\Big|_0^\pi \\ &= \frac{1}{\pi}[-\cos\pi-(-\cos 0)] \\ &= \frac{2}{\pi} \qquad ,\end{aligned}$$

which is, as expected, somewhat more than 1/2.

page 75, problem 5:
Consider a function $y(x)$ defined on the interval from $x = 0$ to 2 like this:

$$y(x) = \begin{cases} -1 & \text{if } 0 \le x \le 1 \\ 1 & \text{if } 1 < x \le 2 \end{cases}$$

The mean value of y is zero, but y never equals zero.

page 75, problem 6:
Let $\dot{x}$ be defined as

$$\dot{x}(t) = \begin{cases} 0 & \text{if } t < 0 \\ 1 & \text{if } t \ge 0 \end{cases}$$

Integrating this function up to t gives

$$x(t) = \begin{cases} 0 & \text{if } t \le 0 \\ t & \text{if } t \ge 0 \end{cases}$$

The derivative of x at $t = 0$ is undefined, and therefore integration followed by differentiation doesn't recover the original function $\dot{x}$.

C Photo Credits

Except as specifically noted below or in a parenthetical credit in the caption of a figure, all the illustrations in this book are under my own copyright, and are copyleft licensed under the same license as the rest of the book.

In some cases it's clear from the date that the figure is public domain, but I don't know the name of the artist or photographer; I would be grateful to anyone who could help me to give proper credit. I have assumed that images that come from U.S. government web pages are copyright-free, since products of federal agencies fall into the public domain. I've included some public-domain paintings; photographic reproductions of them are not copyrightable in the U.S. (Bridgeman Art Library, Ltd. v. Corel Corp., 36 F. Supp. 2d 191, S.D.N.Y. 1999).

cover: Daniel Schwen, 2004; GFDL licensed **10** *Gauss:* C.A. Jensen (1792-1870). **12** *Newton:* Godfrey Kneller, 1702. **25** *Leibniz:* Bernhard Christoph Francke, 1700. **20** *Baseketball photo:* Wikimedia Commons user Reisio, public domain. **30** *Berkeley:* public domain?. **31** *Robinson:* source: www-groups.dcs.st-and.ac.uk, copyright status unknown. **106** *Euler:* Emanuel Handmann, 1753. **113** *tightrope walker:* public domain, since Blondin died in 1897.

D References and Further Reading

Further Reading

The amount of high-quality material on elementary calculus available for free online these days is an embarrassment of riches, so most of my suggestions for reading are online. I'll refer to books in this section only by the surname of the first author; the references section below tells you where to find the book online or in print.

The reader who wants to learn more about the hyperreal system might want to start with Stroyan and the Mathforum.org article. For more depth, one could next read the relevant parts of Keisler. The standard (difficult) treatise on the subject is Robinson.

Given sufficient ingenuity, it's possible to develop a surprisingly large amount of the machinery of calculus without using limits *or* infinitesimals. Two examples of such treatments that are freely available online are Marsden and Livshits. Marsden gives a geometrical definition of the derivative similar to the one used in ch. 1 of this book, but in my opinion his efforts to develop a sufficient body of techniques without limits or infinitesimals end up bogging down in complicated formulations that have the same flavor as the Weierstrass definition of the limit and are just as complicated. Livshits treats differentiation of rational functions as division of functions.

Tall gives an interesting construction of a number system that is smaller than the hyperreals, but easier to construct explicitly, and sufficient to handle calculus involving analytic functions.

References

Keisler, J., *Elementary Calculus: An Approach Using Infinitesimals*, `www.math.wisc.edu/~keisler/calc.html`

Livshits, Michael, `mathfoolery.org/calculus.html`

Marsden and Weinstein, *Calculus Unlimited*, www.cds.caltech.edu/~marsden/books/Calculus_Unlimited.html

Mathforum.org, *Nonstandard Analysis and the Hyperreals*, http://mathforum.org/dr.math/faq/analysis_hyperreals.html

Robinson, A., *Non-Standard Analysis*, Princeton University Press

Stroyan, K., *A Brief Introduction to Infinitesimal Calculus*, www.math.uiowa.edu/~stroyan/InfsmlCalculus/InfsmlCalc.htm

Tall, D., *Looking at graphs through infinitesimal microscopes, windows and telescopes*, Mathematical Gazette, 64, 22-49, http://www.warwick.ac.uk/staff/David.Tall/downloads.html

E Reference

E.1 Review

Algebra

Quadratic equation:

The solutions of $ax^2+bx+c=0$ are $x=\frac{-b\pm\sqrt{b^2-4ac}}{2a}$.

Logarithms and exponentials:

$$\ln(ab)=\ln a+\ln b$$

$$e^{a+b}=e^a e^b$$

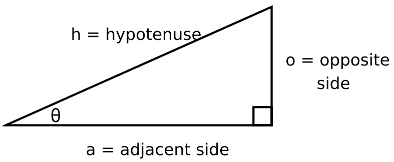
$$\ln e^x=e^{\ln x}=x$$

$$\ln(a^b)=b\ln a$$

Geometry, area, and volume

area of a triangle of base b and height h	$=\frac{1}{2}bh$
circumference of a circle of radius r	$=2\pi r$
area of a circle of radius r	$=\pi r^2$
surface area of a sphere of radius r	$=4\pi r^2$
volume of a sphere of radius r	$=\frac{4}{3}\pi r^3$

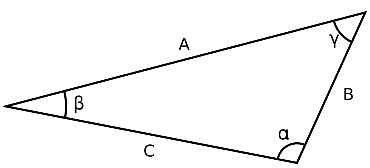
Trigonometry with a right triangle



$\sin\theta=o/h \quad \cos\theta=a/h \quad \tan\theta=o/a$

Pythagorean theorem: $h^2=a^2+o^2$

Trigonometry with any triangle



Law of Sines:

$$\frac{\sin\alpha}{A}=\frac{\sin\beta}{B}=\frac{\sin\gamma}{C}$$

Law of Cosines:

$$C^2=A^2+B^2-2AB\cos\gamma$$

E.2 Hyperbolic functions

$$\sinh x=\frac{e^x-e^{-x}}{2}$$
$$\cosh x=\frac{e^x+e^{-x}}{2}$$
$$\tanh x=\frac{\sinh x}{\cosh x}$$

E.3 Calculus

Let f and g be functions of x, and let c be a constant.

Linearity of the derivative:

$$\frac{\mathrm{d}}{\mathrm{d}x}(cf) = c\frac{\mathrm{d}f}{\mathrm{d}x}$$

$$\frac{\mathrm{d}}{\mathrm{d}x}(f+g) = \frac{\mathrm{d}f}{\mathrm{d}x} + \frac{\mathrm{d}g}{\mathrm{d}x}$$

Rules for differentiation

The chain rule:

$$\frac{\mathrm{d}}{\mathrm{d}x}f(g(x)) = f'(g(x))g'(x)$$

Derivatives of products and quotients:

$$\frac{\mathrm{d}}{\mathrm{d}x}(fg) = \frac{\mathrm{d}f}{\mathrm{d}x}g + \frac{\mathrm{d}g}{\mathrm{d}x}f$$

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{f}{g}\right) = \frac{f'}{g} - \frac{fg'}{g^2}$$

Integral calculus

The fundamental theorem of calculus:

$$\int \frac{\mathrm{d}f}{\mathrm{d}x}\mathrm{d}x = f$$

Linearity of the integral:

$$\int cf(x)\mathrm{d}x = c\int f(x)\mathrm{d}x$$

$$\int [f(x)+g(x)] = \int f(x)\mathrm{d}x + \int g(x)\mathrm{d}x$$

Integration by parts:

$$\int f\mathrm{d}g = fg - \int g\mathrm{d}f$$

Table of integrals

$$\int x^m\,\mathrm{d}x = \frac{1}{m+1}x^{m+1} + c,\ m \neq -1$$

$$\int \frac{\mathrm{d}x}{x} = \ln|x| + c$$

$$\int \sin x\,\mathrm{d}x = -\cos x + c$$

$$\int \cos x\,\mathrm{d}x = \sin x + c$$

$$\int e^x\,\mathrm{d}x = e^x + c$$

$$\int \ln x\,\mathrm{d}x = x\ln x - x + c$$

$$\int \frac{\mathrm{d}x}{1+x^2} = \tan^{-1} x + c$$

$$\int \frac{\mathrm{d}x}{\sqrt{1-x^2}} = \sin^{-1} x + c$$

$$\int \cosh x\,\mathrm{d}x = \sinh x + c$$

$$\int \sinh x\,\mathrm{d}x = \cosh x + c$$

$$\int \tan x\,\mathrm{d}x = -\ln|\cos x| + c$$

$$\int \cot x\,\mathrm{d}x = \ln|\sin x| + c$$

$$\int \sec x\,\mathrm{d}x = \ln|\sec x + \tan x| + c$$

$$\int \sec^2 x\,\mathrm{d}x = \tan x + c$$

$$\int \csc^2 x\,\mathrm{d}x = -\cot x + c$$

Index