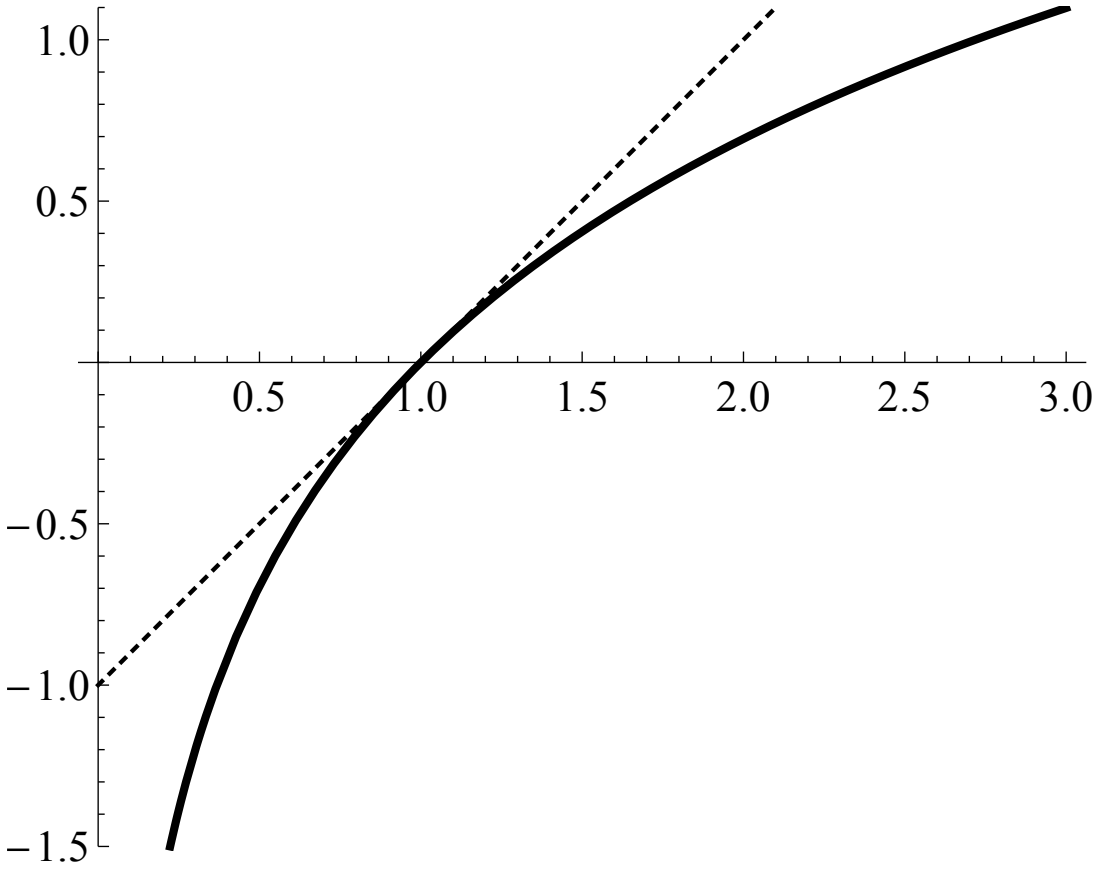
Analysis II: MA139

Keith Ball

index by Daniel Ueltschi



Contents

Power series I **3**
Basic properties of power series . . . 3
The continuity of power series . . . 7
The exponential . . . 8
Inequalities for the exponential . . . 12
The logarithm and powers . . . 13

Limits and the derivative **16**
Limits . . . 17
One sided limits . . . 19
The derivative . . . 22
The derivative machine . . . 24
The Mean Value Theorem . . . 28
Derivatives of inverses . . . 33

Power series II **36**
The differentiability of power series . . . 36
The trigonometric functions . . . 42
The complex exponential . . . 48
The tangent . . . 49

Taylor's Theorem **51**
Cauchy's Mean Value Theorem . . . 51
L'Hôpital's rule . . . 52
Taylor's Theorem with remainder . . . 55

The Riemann integral **64**
The construction . . . 65
Uniform continuity . . . 71
Basic properties . . . 74
The fundamental theorem . . . 81
Improper integrals . . . 87
The Gamma function . . . 91

The radius of convergence formula **93**

Chapter 1. Power series I

A power series is a actually a family of series. Given a sequence of coefficients $(a_0, a_1, \ldots)$ we look at the series

$$\sum_{n=0}^{\infty} a_n\, x^n$$

for each possible value of x. (More generally we consider series

$$\sum_{n=0}^{\infty} a_n\, (x - x_0)^n$$

in which we centre the series at a number x_0 rather than at 0. We will develop the theory just for the first type: everything transfers immediately to the more general case.)

We think of a power series as a kind of infinite polynomial in which the (a_n) are coefficients. The series may or may not converge depending upon the value of x. Where it does converge it defines a function of x and we shall see that the function will automatically be continuous and differentiable. Many of the most important functions in mathematics can be written as power series: in particular the exponential function.

Basic properties of power series

We want to know the values of x for which a power series $\sum_0^\infty a_n\, x^n$ converges. Of course the answer will depend upon the specific a_n, but there is a fundamental feature of the answer that is common to all power series. The set of values where a power series converges will be an interval (possibly infinite) of the real line. The idea is quite simple: if $\sum a_n t^n$ converges and we choose x to be smaller than t in size then the terms $a_n x^n$ will be much smaller than the terms $a_n t^n$ so we should expect $\sum a_n\, x^n$ to converge as well.

We would like to prove this just by quoting the Comparison Test for series but that doesn't quite work because the series $\sum a_n\, t^n$ might not converge absolutely. However we can get around this problem because $a_n x^n$ is not just smaller than $a_n t^n$ but is *very much* smaller when n is large.

Theorem (**Radius of convergence I**). *Let* $\sum_0^\infty a_n x^n$ *be a power series with* $\sum a_n\, t^n$ *convergent. Then*

$$\sum_0^\infty a_n x^n$$

converges absolutely for all x *with* $|x| < |t|$.

Proof Since $\sum a_n\, t^n$ converges we know that $a_n\, t^n \to 0$ as $n \to \infty$ and so the sequence is bounded. There is some M for which $|a_n t^n| < M$ for all n. Now

$$\begin{aligned}
\sum_0^N |a_n x^n| &= \sum_0^N |a_n t^n| \left|\frac{x}{t}\right|^n \\
&\leq M \sum_0^N \left|\frac{x}{t}\right|^n \\
&\leq M \sum_0^\infty \left|\frac{x}{t}\right|^n \\
&= M \frac{1}{1 - |x|/|t|}
\end{aligned}$$

Hence $\sum_0^\infty |a_n x^n| < \infty$ and the series converges absolutely. □

Using this principle we can introduce what is called the radius of convergence.

Theorem (**Radius of convergence II**). *Let* $\sum_0^\infty a_n x^n$ *be a power series. One of the following holds.*

- *The series converges only if* $x = 0$.
- *The series converges for all real numbers* x.
- *There is a positive number* R *with the property that the series converges if* $|x| < R$ *and diverges if* $|x| > R$.

In the third case the number R *is called the* **radius of convergence**. *In the first case we say that the radius of convergence is* 0 *and in the second that it is* ∞.

Proof Obviously the series does converge if $x = 0$. If the first option does not hold than it converges for some other values of x.

The set of x for which it converges might be unbounded: it might contain arbitrarily large numbers. In that case the power series will converge absolutely for all real x and defines a function on $\mathbf{R}$.

Otherwise the set might contain some numbers other than 0 but be bounded. In that case we let

$$R = \sup\{|t| : \sum a_n t^n \text{ converges}\}.$$

Then by our previous theorem the series converges absolutely whenever $|x| < R$. It does not converge if $|x| > R$ by definition of R. □

As an immediate corollary we get an observation that will be useful later.

Corollary (**Absolute series**). *Let $\sum_0^\infty a_n x^n$ be a power series with radius of convergence R. Then $\sum_0^\infty |a_n| x^n$ also has radius of convergence R.*

Example (**The geometric series I**). *The series $\sum_0^\infty x^n$ has radius of convergence $R = 1$.*

Proof We know that the series converges if and only if $x \in (-1, 1)$. □

Example (**The geometric series II**). *If p is real the series $\sum_0^\infty p^n x^n$ has radius of convergence $R = 1/|p|$.*

Proof We know that the series converges if and only if $px \in (-1, 1)$ which is the same as saying that $x \in (-1/|p|, 1/|p|)$. □

Example (**The log series**). *The series $\sum_1^\infty \frac{x^n}{n}$ has radius of convergence $R = 1$.*

Proof From homework we know that the series converges if and only if $x \in [-1, 1)$. □

Notice that although we have put in the extra factor $1/n$ in front of x^n we have not changed the radius of convergence. The point is that if we increase x beyond 1, its powers x^n increase exponentially fast and this swamps the effect of the factor $1/n$. However, the extra factor of $1/n$ does make the series converge at $x = -1$ but not absolutely.

Example. *The series $\sum_1^\infty nx^n$ has radius of convergence $R = 1$.*

Proof From homework we know that the series converges if and only if $x \in (-1, 1)$. □

For most interesting power series we can find the radius of convergence using the ratio test. There is a formula for the radius of convergence which works for all series which appears at the end of these notes. As a taste let us look at another example:

Example. *The series* $1 + 0x + x^2 + 0x^3 + x^4 + \cdots$ *has radius of convergence* $R = 1$.

Proof The series is

$$1 + x^2 + x^4 + x^6 + \cdots$$

which is a geometric series with ratio x^2. So it converges if and only if $x^2 \in (-1, 1)$. □

Another way to do it which is a much more flexible method is as follows.

Corollary (Comparison of power series)**.** *Suppose the power series* $\sum a_n x^n$ *has radius of convergence* R *and the sequence* (b_n) *satisfies* $|b_n| \leq |a_n|$ *for every* n*. Then the series* $\sum b_n x^n$ *has radius of convergence at least* R*.*

Proof If $|x| < R$ then $\sum |a_n|\,|x|^n$ converges. By the comparison test, so does

$$\sum_0^\infty |b_n|\,|x|^n.$$

□

In particular if you throw out some terms you can't make the RoC smaller.

We remarked that it is possible to consider power series centred at points other then 0. For example

$$\frac{1}{1-x} = \frac{1}{2-(x+1)} = \frac{1}{2}\,\frac{1}{1-(x+1)/2}$$

$$= \frac{1}{2}\left(1 + \frac{x+1}{2} + \left(\frac{x+1}{2}\right)^2 + \cdots\right) = \frac{1}{2} + \frac{x+1}{4} + \frac{(x+1)^2}{8} + \cdots.$$

This converges if $\left|\frac{x+1}{2}\right| < 1$: in other words if x differs from -1 by less than 2. The series is centred at -1 and has radius of convergence 2. This means that it converges on $(-3, 1)$. Notice that it converges "as far as it possibly can". The function has an asymptote at $x = 1$ so the series cannot represent the function at $x = 1$.

The continuity of power series

As was remarked several times, many of the most important functions in mathematics are given by power series so we want to know that such functions have nice properties: that they are continuous for example.

Theorem (**Continuity of power series**). *Let* $\sum_0^\infty a_n x^n$ *with radius of convergence* R. *Then the function*

$$x \mapsto \sum_0^\infty a_n x^n$$

is continuous on the interval $(-R, R)$.

We already saw some examples in which the power series converges at one or both ends of the interval $[-R, R]$. If so it makes sense to ask whether the function is continuous on this larger set. Perhaps surprisingly this is a bit more subtle than the statement above. There are some quite slick proofs of the continuity of power series. We will take a very down to earth approach. Inside the radius of convergence a power series can be approximated by a polynomial: and polynomials are continuous.

Proof Suppose $-R < x < R$. We want to show that the function is continuous at x. Choose a number T with $|x| < T < R$. Then the series $\sum |a_n|\, T^n$ converges, so for each $\varepsilon > 0$ there is some number N for which

$$\sum_{N+1}^\infty |a_n|\, T^n < \varepsilon/3.$$

Now if $|y - x| < T - |x|$ we will have $|y| < T$ as well as $|x| < T$. Hence

$$\sum_{N+1}^\infty |a_n|\, |x|^n < \varepsilon/3 \quad \text{and} \quad \sum_{N+1}^\infty |a_n|\, |y|^n < \varepsilon/3.$$

The partial sum $\sum_0^N a_n\, y^n$ is a polynomial in y and polynomials are continuous so there is some $\delta_0 > 0$ with the property that if $|y - x| < \delta_0$

$$\left| \sum_0^N a_n\, y^n - \sum_0^N a_n\, x^n \right| < \varepsilon/3.$$

Therefore if we choose δ to be the smaller of δ_0 and $T - |x|$ then if $|y - x| < \delta$ we get

$$\left| \sum_0^\infty a_n\, y^n - \sum_0^\infty a_n\, x^n \right| \leq \left| \sum_{N+1}^\infty a_n\, y^n \right| + \left| \sum_0^N a_n\, y^n - \sum_0^N a_n\, x^n \right| + \left| \sum_{N+1}^\infty a_n\, x^n \right|$$

$$\leq \sum_{N+1}^{\infty} |a_n|\,|y|^n + \left|\sum_{0}^{N} a_n\,y^n - \sum_{0}^{N} a_n\,x^n\right| + \sum_{N+1}^{\infty} |a_n|\,|x|^n < \varepsilon.$$

□

Armed with these properties of power series we can begin to investigate the standard functions such as the exponential. It will turn out that power series are not only continuous but also differentiable inside the radius of convergence. The derivative makes it easier to demonstrate the crucial properties of standard functions but we shall begin with the exponential before we cover derivatives so as to see some concrete applications of our theory.

The exponential

As explained in the introduction we shall define the exponential function as a power series.

Definition (**The exponential**). *If* $x \in \mathbf{R}$ *the series*

$$1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \cdots + \frac{x^n}{n!} + \cdots$$

converges. We call the sum $\exp x$.

Proof The ratio of successive terms of the series is

$$\frac{x^{n+1}}{(n+1)!}\frac{n!}{x^n} = \frac{x}{n+1} \to 0$$

so the series converges by the ratio test. □

We know that the function $x \mapsto \exp x$ is continuous on $\mathbf{R}$ since it is a convergent power series. We would like to check that it satisfies the characteristic property $e^{x+y} = e^x e^y$. This will be easy to check once we have derivatives later in the course but it is very instructive to see how to do it directly from the series definition.

The idea is this. When we multiply e^x by e^y we have to expand the product of two brackets

$$\left(1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \cdots\right)\left(1 + y + \frac{y^2}{2} + \frac{y^3}{6} + \cdots\right)$$

which means we have to consider all possible products of one term from the first bracket and one term from the second bracket (and then add all of these products together). It is natural to arrange these products into a grid

	1	x	$x^2/2$	$x^3/6$	$\dots$
1	1	x	$x^2/2$	$x^3/6$	$\dots$
y	y	xy	$x^2y/2$	$\dots$	
$y^2/2$	$y^2/2$	$xy^2/2$	$\dots$		
$y^3/6$	$y^3/6$	$\dots$			
$\vdots$	$\vdots$				

We now want to add all the terms in this grid. One way to make sure that we include everything is to add them diagonally.

$$1 + (x + y) + (x^2/2 + xy + y^2/2) + (x^3/6 + x^2y/2 + xy^2/2 + y^3/6) + \cdots$$

Now let us collect each diagonal sum over a common denominator. The first is 1 and the second is $x + y$. The third is

$$\frac{x^2 + 2xy + y^2}{2}$$

and the fourth is

$$\frac{x^3 + 3x^2y + 3xy^2 + y^3}{6}.$$

We immediately see that we have the binomial expansions of $(x + y)^2$ and $(x + y)^3$. So the whole sum looks like

$$1 + (x + y) + \frac{(x + y)^2}{2} + \frac{(x + y)^3}{6} + \cdots$$

and this is $\exp(x + y)$.

Because we have infinite sums we need an argument to justify this. We shall compare the sum of a certain number of the diagonals with the product of partial sums of the series for e^x and e^y.

Theorem (**The characteristic property of the exponential**). *If* $x, y \in \mathbf{R}$ *then*

$$\exp(x+y) = \exp(x)\exp(y).$$

Proof By the algebra of limits it suffices to check that

$$\sum_0^{2m} \frac{(x+y)^k}{k!} - \left(\sum_0^m \frac{x^i}{i!}\right)\left(\sum_0^m \frac{y^j}{j!}\right) \to 0$$

because the first term converges to e^{x+y} and the second to $e^x e^y$. By the binomial theorem, for each k,

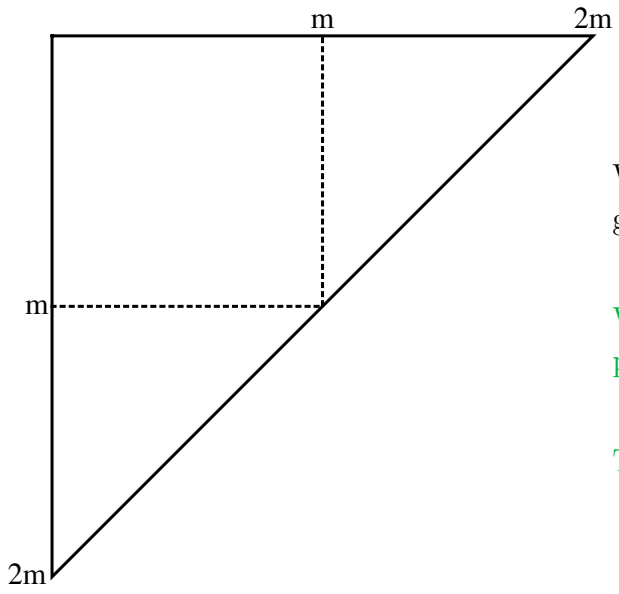
$$(x+y)^k = \sum_{i=0}^k \frac{k!}{i!(k-i)!} x^i y^{k-i} = \sum_{i+j=k} k! \frac{x^i}{i!}\frac{y^j}{j!}.$$

So

$$\sum_0^{2m} \frac{(x+y)^k}{k!} - \left(\sum_0^m \frac{x^i}{i!}\right)\left(\sum_0^m \frac{y^j}{j!}\right)$$

$$= \sum_{i+j\le 2m} \frac{x^i}{i!}\frac{y^j}{j!} - \sum_{i\le m, j\le m} \frac{x^i}{i!}\frac{y^j}{j!}.$$

We are considering the terms of the form $\frac{x^i}{i!}\frac{y^j}{j!}$ within a triangle but with the terms in a square removed.



We are left with terms in two smaller triangles.

We want to prove that the sums in these two pieces are small.

The key is that all these terms are "far out".

These sums are

$$\sum_{i\geq m+1, i+j\leq 2m} \frac{x^i}{i!}\frac{y^j}{j!}$$

and a similar one with the restrictions on i and j interchanged. By the triangle inequality the absolute value of the sum is at most

$$\sum_{i\geq m+1, i+j\leq 2m} \frac{|x|^i}{i!}\frac{|y|^j}{j!}.$$

Once we have made all the terms positive we can include other terms without decreasing the sum. So the sum of the terms in the smaller triangle has absolute value at most

$$\sum_{i\geq m+1, i+j\leq 2m} \frac{|x|^i}{i!}\frac{|y|^j}{j!} \leq \sum_{i\geq m+1, j\geq 0} \frac{|x|^i}{i!}\frac{|y|^j}{j!}$$

and the latter is

$$\left(\sum_{i\geq m+1} \frac{|x|^i}{i!}\right)\left(\sum_{j=0}^{\infty} \frac{|y|^j}{j!}\right).$$

The first factor tends to 0 as $m \to \infty$ while the second is equal to $e^{|y|}$. □

Theorem (**The characteristic property of the exponential**). *If $x, y \in \mathbf{R}$ then*

$$\exp(x+y) = \exp(x)\exp(y).$$

The characteristic property makes us feel comfortable writing $\exp(x)$ as e^x. We have $e^{x+y} = e^x e^y$ and in particular we have $e^{-u} = 1/e^u$ for all real u.

The argument given above can be used to show that power series can be multiplied in the obvious way inside their intervals of convergence.

Theorem (**The product of power series**). *If $\sum a_n x^n$ and $\sum b_n x^n$ converge for x in the interval $(-R, R)$ then so does the series $\sum_{n=0}^{\infty}\left(\sum_{k=0}^{n} a_k b_{n-k}\right)x^n$ and for all such x*

$$\sum_{n=0}^{\infty}\left(\sum_{k=0}^{n} a_k b_{n-k}\right)x^n = \left(\sum_{i=0}^{\infty} a_i x^i\right)\left(\sum_{i=0}^{\infty} b_j x^j\right).$$

Proof Exercise. □

Inequalities for the exponential

For many purposes it is important to have some estimates for the exponential in terms of simpler functions. To begin with let us observe that if x is positive it is clear that $e^x = 1 + x + x^2/2 + \cdots$ is positive. For negative x it is not immediate from the power series that e^x is positive but it follows from the fact that $e^{-u} = 1/e^u$ and this is positive if u is positive. The most useful inequalities are the following.

Theorem (**Inequalities for the exponential**)**.** *The following estimates hold for the exponential function:*

1. $1 + x \leq e^x$ *for all real* x
2. $e^x \leq 1/(1-x)$ *if* $x < 1$.

Proof If $x \geq 0$ then

$$e^x = 1 + x + \frac{x^2}{2} + \cdots \geq 1 + x$$

and if $0 \leq x < 1$

$$e^x = 1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \cdots \leq 1 + x + x^2 + x^3 + \cdots = \frac{1}{1-x}.$$

So the inequalities are easy to establish if $x \geq 0$. To obtain them for negative x we use the characteristic property of the exponential much as we did to prove positivity.

Suppose $x = -u$ is negative. We know that $e^u \geq 1 + u$ and hence $e^{-x} \geq 1 - x$. But this implies that

$$\frac{1}{1-x} \geq e^x$$

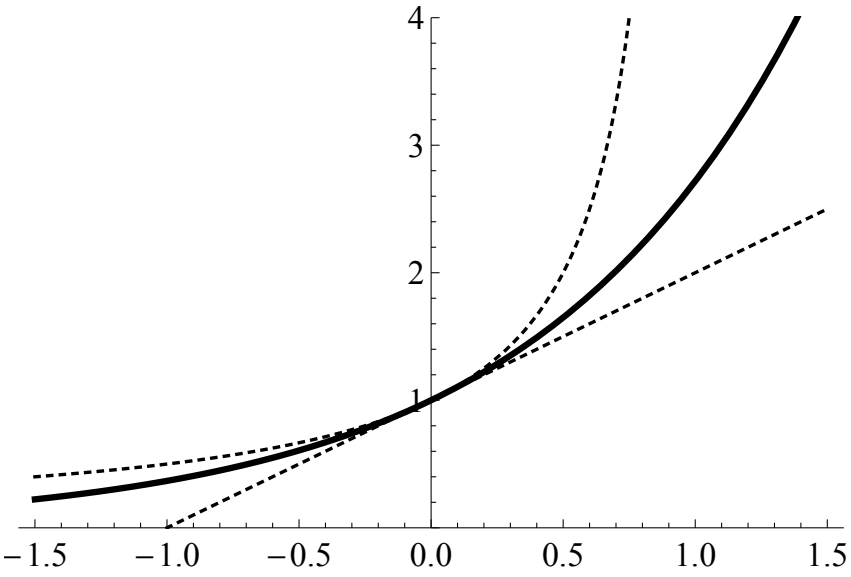
so the second inequality is now established for all $x < 1$.

If $x \leq -1$ then $1 + x \leq 0$ whereas $e^x > 0$ so $e^x \geq 1 + x$. It remains to prove the first inequality for $-1 < x < 0$. If $x = -u$ then $0 < u < 1$ and so $e^u \leq 1/(1-u)$. This says that $e^{-x} \leq 1/(1+x)$ and this implies that

$$1 + x \leq e^x.$$

□

These two inequalities sandwich the exponential rather nicely near 0 as shown in the picture.



Corollary (**The exponential increases**)**.** *The exponential function is strictly increasing and its range is* $(0, \infty)$.

Proof Suppose $x < y$. Then

$$e^y = e^{y-x}e^x \geq (1 + y - x)e^x > e^x.$$

This shows that the exponential is strictly increasing.

Since $e^x \geq 1 + x$ the exponential takes arbitrarily large values (at large x). Since $e^{-x} = 1/e^x$ the exponential also takes values arbitrarily close to 0 (at large negative x). By the IVT the exponential takes all positive values. □

Since we know that the exponential increases and is continuous we know that it has an inverse, the logarithm, which will be the topic of the next section.

The logarithm and powers

We have seen that the exponential function maps $\mathbf{R}$ onto $(0, \infty)$ and is continuous and strictly increasing. So we know by the IVT and its corollaries that the exponential has a continuous inverse defined on $(0, \infty)$: the natural logarithm.

Theorem (**The logarithm**). *There is a continuous strictly increasing function* $x \mapsto \log x$ *defined on* $(0, \infty)$ *satisfying*

$$e^{\log x} = x$$

for all positive x *and*

$$\log(e^y) = y$$

for all real y. *We have that for all positive* u *and* v,

$$\log(uv) = \log u + \log v.$$

Proof We just need to check the last assertion. But

$$e^{\log u + \log v} = e^{\log u} e^{\log v} = uv.$$

Applying log to both sides we get what we want. □

It was proved last term that each positive number has a positive square root. We would like to know that we can take other non-integer powers of positive numbers. The simplest way to define these is using the logarithm and exponential.

Definition (**Powers**). *If* $x > 0$ *and* $p \in \mathbf{R}$ *we define*

$$x^p = \exp(p \log x).$$

We have the usual rules

1. *If* n *is a positive integer then* x^n *as defined here is indeed the product* $x.x. \ldots .x$ *of* n *copies of* x.
2. $x^{p+q} = x^p x^q$ *for all* $x > 0$ *and* $p, q \in \mathbf{R}$.
3. $\log(x^p) = p \log x$ *for* $x > 0$ *and* $p \in \mathbf{R}$.
4. $x^{pq} = (x^p)^q$ *for all* $x > 0$ *and* $p, q \in \mathbf{R}$.
5. $\exp(p) = e^p$ *for all* $p \in \mathbf{R}$.

Notice that the last statement is not something we could have proved earlier because we did not have a definition of powers with which to make sense of the p^{th} power of e. We shall prove the last one and leave the rest as an exercise.

Proof By definition of the power,

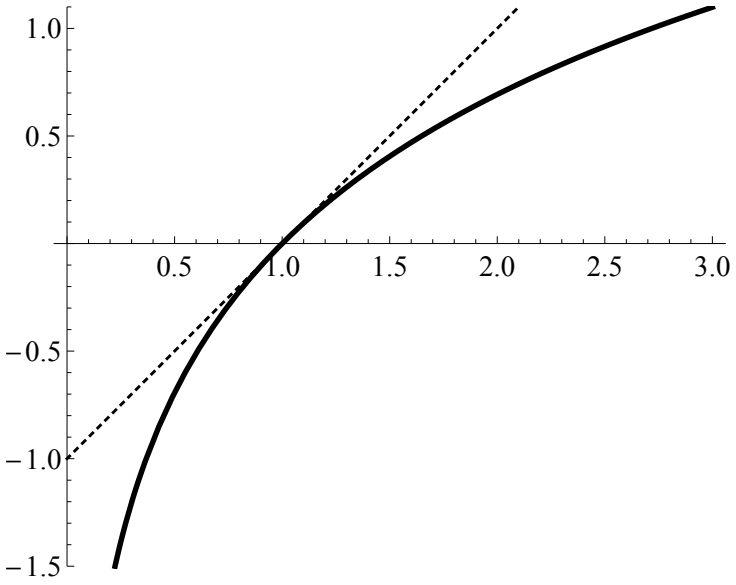
$$e^p = \exp(p \log(e)).$$

Now we know that $\log(e) = 1$ so the second expression is $\exp(p)$. □

The inequalities we proved for the exponential immediately give us inequalities for the logarithm. You are asked to demonstrate these in the Homework. The most crucial one is this:

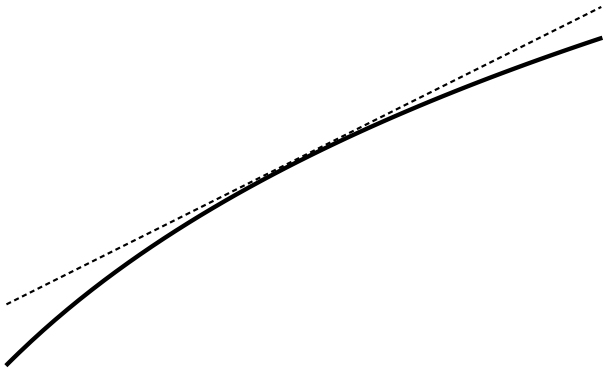
Theorem (**The tangent to the logarithm**). *If* $x > 0$ *then* $\log x \leq x - 1$.

This says that the graph of $y = \log x$ lies below the line which is its tangent at the point $(1, 0)$.

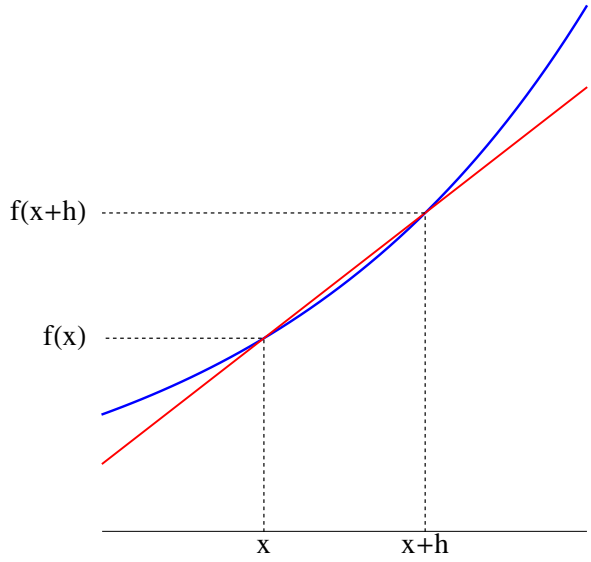


Chapter 2. Limits and the derivative

In this chapter we shall develop the basic theory of derivatives. The problem that we start with is to find the instantaneous slope of a curve: the slope of its tangent line at a point $(x, f(x))$ on the curve $y = f(x)$.



The geometric picture of our method is this: we consider the point $(x, f(x))$ and a nearby point $(x + h, f(x + h))$ and look at the chord joining the two points.



The slope of this chord is

$$\frac{f(x+h)-f(x)}{h}.$$

We now ask what happens to this slope as the nearby point gets closer and closer to the point we care about: what happens as $h \to 0$. We ask whether the quotient $\frac{f(x+h)-f(x)}{h}$ approaches a limit. If so we define this to be the slope of the curve at $(x, f(x))$ or the derivative $f'(x)$.

In order to carry out this process and analyse it we need to have a definition of the limit in question. This will be the first part of the chapter.

Limits

Definition (**Limits of functions**)**.** *Let I be an open interval, $c \in I$ and f a real valued function defined on I except possibly at c. We say that*

$$\lim_{x \to c} f(x) = L$$

if for every $\varepsilon > 0$ there is a number $\delta > 0$ so that if $0 < |x - c| < \delta$ then

$$|f(x) - L| < \varepsilon.$$

Thus we can guarantee that $f(x)$ is close to L by insisting that x is close to c, but we exclude the possibility that $x = c$: we don't care what f does at c itself nor even whether f is defined at c. The reason for this freedom is that we want to discuss limits like

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

in which the function is not defined at $h = 0$.

Let's look at a couple of examples.

Example. *Let $f : \mathbf{R} \to \mathbf{R}$ be defined by*

$$f(x) = \begin{cases} 1 & \text{if } x = 0 \\ 0 & \text{if } x \neq 0 \end{cases}$$

Then

$$\lim_{x \to 0} f(x) = 0.$$

Example.

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

Proof As long as $x\neq 1$

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

and as $x\to 1$ this approaches 2. □

Strictly speaking the last statement is something we haven't yet proved: we shall do so now. By comparing the definitions of continuity and limits we can immediately prove the following lemma.

Lemma (**Limits and continuity**). *If $f: I\to \mathbf{R}$ is defined on the open interval I and $c\in I$ then f is continuous at c if and only if*

$$\lim_{x\to c} f(x)=f(c).$$

Proof Exercise. □

The last line really contains two pieces of information: that the limit exists and that it equals $f(c)$. In the previous example the function $x\mapsto x+1$ is continuous everywhere and hence at $c=1$ so we get $\lim_{x\to 1}(x+1)=1+1=2$.

We want to have a machine for calculating limits like the continuity machine: for example we want to know that

$$\lim_{x\to c}(f(x)+g(x))=\lim_{x\to c}f(x)+\lim_{x\to c}g(x)$$

and

$$\lim_{x\to c}f(x)g(x)=\left(\lim_{x\to c}f(x)\right)\left(\lim_{x\to c}g(x)\right)$$

whenever f and g have limits at c. One way to prove these would be to relate limits of functions to limits of sequences just as we did for continuity. This will be an exercise in the homework.

An alternative approach would be to use a slightly weird trick to deduce what we want from what we already know about continuity. The trick is that if $\lim_{x\to c}f(x)=L$ we can define a new function

$$\tilde{f}(x)=\begin{cases} f(x) & \text{if } x\neq c\\ L & \text{if } x=c\end{cases}$$

and this new function will be continuous at c. Then we just use the rules for continuity to deduce the same rules for limits.

We will take for granted the following theorems.

Theorem (**Continuous and sequential limits**). *If $f : I\backslash\{c\} \to \mathbf{R}$ is defined on the interval I except at $c \in I$ then*

$$\lim_{x\to c} f(x) = L$$

if and only if for every sequence (x_n) in $I\backslash\{c\}$ with $x_n \to c$ we have

$$f(x_n) \to L.$$

Proof Homework □

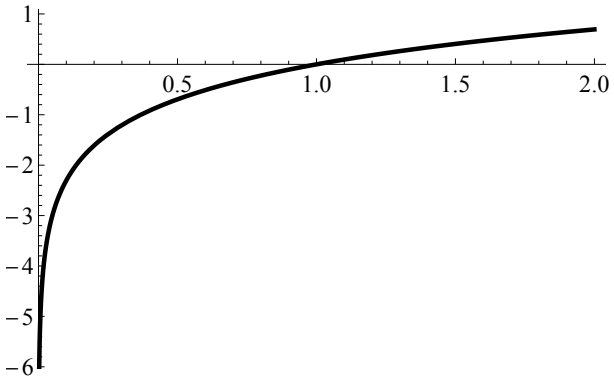
Theorem (**Algebra of limits**). *If $f, g : I\backslash\{c\} \to \mathbf{R}$ are defined on the interval I except at $c \in I$ and $\lim_{x\to c} f(x)$ and $\lim_{x\to c} g(x)$ exist then*

1. $\lim_{x\to c}(f(x) + g(x)) = \lim_{x\to c} f(x) + \lim_{x\to c} g(x)$
2. $\lim_{x\to c} f(x)g(x) = \lim_{x\to c} f(x) \lim_{x\to c} g(x)$
3. *if* $\lim_{x\to c} g(x) \neq 0$ *then*
$$\lim_{x\to c} \frac{f(x)}{g(x)} = \frac{\lim_{x\to c} f(x)}{\lim_{x\to c} g(x)}.$$

Proof Exercise □

One sided limits

It often happens that we wish to understand the behaviour of $f(x)$ as approaches the end of the interval where f is defined. For example we want to know what happens to $\log x$ as x approaches 0 from the right.



So we define one sided limits.

Definition (**One sided limits**). *Let f a real valued function defined on the open interval (c, d). We say that*

$$\lim_{x \to c^+} f(x) = L$$

if for every $\varepsilon > 0$ there is a number $\delta > 0$ so that if $c < x < c + \delta$ then

$$|f(x) - L| < \varepsilon.$$

We read the expression as "The limit of $f(x)$ as x approaches c from the right is L".

We define the limit from the left

$$\lim_{x \to c^-} f(x)$$

similarly.

Example. *Let $f : \mathbf{R} \to \mathbf{R}$ be defined by*

$$f(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1/2 & \text{if } x = 0 \\ 1 & \text{if } x > 0 \end{cases}$$

Then

$$\lim_{x \to 0^+} f(x) = 1$$

and

$$\lim_{x \to 0^-} f(x) = 0.$$

At this point we digress slightly to discuss limits at infinity and infinite limits. At times it is useful to have a notation indicating that a function behaves something like $1/x^2$ as $x \to 0$.

Definition (**Infinite limits**). *If $f : I\backslash\{c\} \to \mathbf{R}$ is defined on an open interval I except perhaps at $c \in I$ we write*

$$\lim_{x\to c} f(x) = \infty$$

if for every $M > 0$ there is a $\delta > 0$ so that if $0 < |x - c| < \delta$ then $f(x) > M$. The limit $-\infty$

$$\lim_{x\to c} f(x) = -\infty$$

is defined similarly.

The limit is infinite if we can make $f(x)$ as large as we please by insisting that x is close to c (but not equal to c).

Example.

$$\lim_{x\to 0} \frac{1}{x^2} = \infty.$$

Proof Given $M > 0$ choose $\delta = 1/\sqrt{M}$. Then if $0 < |x| < \delta = 1/\sqrt{M}$ we have $0 < x^2 < 1/M$ and hence

$$\frac{1}{x^2} > M.$$

□

There is also a one sided version.

Example.

$$\lim_{x\to 0^+} \frac{1}{x} = \infty, \qquad \lim_{x\to 0^-} \frac{1}{x} = -\infty.$$

We also on occasions wish to study the behaviour of functions as the variable becomes large.

Definition (**Limits at infinity**). *If $f : \mathbf{R} \to \mathbf{R}$ we write*

$$\lim_{x\to\infty} f(x) = L$$

if for every $\varepsilon > 0$ there is an N so that if $x > N$ then $|f(x) - L| < \varepsilon$.

This looks very much like the definition of convergence of a sequence. The only difference is that we now consider arbitrary real x instead of just natural numbers n. The algebra of limits applies equally well to limits at infinity.

Example.

$$\lim_{x\to\infty} \frac{1}{x} = 0.$$

Proof Exercise.

Example.

$$\lim_{x\to\infty} \frac{x}{e^x} = 0.$$

Proof If $x > 0$ we have $e^x \geq 1 + x + x^2/2 > x^2/2$. So

$$0 < \frac{x}{e^x} < \frac{x}{x^2/2} = \frac{2}{x} \to 0$$

as $x \to \infty$. □

The derivative

The idea of calculating the slope of a curve existed before Newton and Leibniz and the derivatives of certain functions were already known: in particular the derivatives of the monomials $x \mapsto x^n$ for positive integers n. It was also known how to calculate areas under certain curves. Three key points made up the invention of what we call calculus.

The first key point was the creation of a derivative "machine" to enable us to calculate derivatives of all the standard functions: polynomials, rational functions, the exponential and trigonometric functions and anything we can build by adding multiplying or composing these functions.

The second key point of calculus is the recognition that the slope of a curve, the derivative, can be considered as a function and then differentiated again. For Newton this was crucial since his aim was to derive Kepler's Laws of planetary motion from the inverse square law of gravitation. The inverse square law tells you the *acceleration* of a planet: the second derivative of its position.

The third key point of calculus was the realisation that differentiation and integration are opposites of one another. This made it possible to simplify the process of integration enormously by using the rules for derivatives.

Definition (The derivative). *Suppose $f : I \to \mathbf{R}$ is defined on the open interval I and $c \in I$. We say that f is differentiable at c if*

$$\lim_{h \to 0} \frac{f(c+h) - f(c)}{h}$$

exists. If so we call the limit $f'(c)$.

Example (The derivative of x). *If $f(x) = x$ then f is differentiable at every point of $\mathbf{R}$ and $f'(c) = 1$ for all c.*

Proof For every c and $h \neq 0$

$$\frac{f(c+h) - f(c)}{h} = \frac{c+h-c}{h} = 1$$

and so

$$\lim_{h \to 0} \frac{f(c+h) - f(c)}{h} = \lim_{h \to 0} 1 = 1.$$

□

Example (The derivative of x^2). *If $f(x) = x^2$ then f is differentiable at every point of $\mathbf{R}$ and $f'(c) = 2c$ for all c.*

Proof For every c and $h \neq 0$

$$\frac{f(c+h) - f(c)}{h} = \frac{(c+h)^2 - c^2}{h} = \frac{c^2 + 2ch + h^2 - c^2}{h} = 2c + h$$

and so

$$\lim_{h \to 0} \frac{f(c+h) - f(c)}{h} = \lim_{h \to 0} (2c + h) = 2c.$$

□

Example (The derivative of $1/x$). *If $f(x) = 1/x$ then f is differentiable at every point of $\mathbf{R}$ except 0 and $f'(c) = -1/c^2$ for all $c \neq 0$.*

Proof For every $c \neq 0$ and h satisfying $0 < |h| < |c|$

$$\frac{f(c+h)-f(c)}{h} = \frac{1/(c+h)-1/c}{h} = \frac{c-(c+h)}{h(c+h)c} = \frac{-1}{(c+h)c}$$

and so

$$\lim_{h\to 0} \frac{f(c+h)-f(c)}{h} = \lim_{h\to 0} \frac{-1}{(c+h)c} = \frac{-1}{c^2}.$$

□

The derivative machine

Obviously we don't want to carry out this process for complicated functions so we need to build a machine to do it for us. The derivative machine is like a food mixer. The mixer has 3 basic parts: the motor, jug and lid (say) and separate blades to handle different foods. The derivative machine has 3 basic parts: the sum rule, the product rule and the chain rule and special rules to handle powers, the exponential and the trigonometric functions.

In order to start building the machine we need to know that if a function is differentiable at a point c then it is continuous there. To make the picture clearer let us start by observing that we can rewrite the derivative

$$f'(c) = \lim_{x\to c} \frac{f(x)-f(c)}{x-c}$$

since if we put $x = c+h$ and $h \to 0$ we have $x \to c$.

Lemma (**Differentiability and continuity**). *If I is an open interval, $f : I \to \mathbf{R}$ is differentiable at $c \in I$ then f is continuous at c.*

Proof We know that

$$\frac{f(x)-f(c)}{x-c} \to f'(c)$$

as $x \to c$. Hence

$$f(x)-f(c) = \frac{f(x)-f(c)}{x-c}(x-c) \to f'(c).0 = 0$$

as $x \to c$ which implies that $f(x) \to f(c)$ as required. □

The sum and product rules follow easily from the algebra of limits.

Theorem (**The sum and product rules**). *Suppose $f, g : I \to \mathbf{R}$ are defined on the open interval I and are differentiable at $c \in I$. Then $f + g$ and fg are differentiable at c and*

$$(f + g)'(c) = f'(c) + g'(c)$$

and

$$(fg)'(c) = f'(c)g(c) + f(c)g'(c).$$

Proof The first will be left as an exercise. For the second

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

The algebra of limits tells us that as $x \to c$ this expression approaches

$$f'(c)g(c) + f(c)g'(c).$$

□

Note that to use the algebra of limits for the product rule we needed to use the fact that $g(x) \to g(c)$ as $x \to c$: in other words, the continuity of g at c.

We are now in a position to prove the differentiability of all polynomials. To begin with let's observe that if f is a constant function then its derivative is clearly 0. By the product rule or just a direct check we can see that if we multiply a function f by a constant C then we also multiply the derivative by C. Using the sum rule we can then handle all polynomials as long as we check the derivative of each power $x \mapsto x^n$. We shall do this by induction.

Lemma (**The derivatives of the monomials**). *If n is a positive integer then the derivative of $x \mapsto x^n$ is $x \mapsto nx^{n-1}$.*

Proof We already saw this for $n = 1$. Assume inductively that we have the result for $f(x) = x^n$. Then $x^{n+1} = xf(x)$ so by the product rule its derivative is

$$1.f(x) + xf'(x) = x^n + xnx^{n-1} = (n + 1)x^n$$

completing the inductive step. □

At school you learned the quotient rule for derivatives. I have not included it in the machine since it follows from the product rule, the chain rule and the derivative of the function $x \mapsto 1/x$.

The third part of the derivative machine is the chain rule. If we form the composition of two functions f and g

$$x \mapsto f(g(x))$$

we want to know that we can differentiate it and obtain the correct formula for the derivative.

Theorem (**Chain rule**)**.** *Suppose I and J are open intervals, $f : I \to \mathbf{R}$ and $g : J \to I$, that g is differentiable at c and f is differentiable at $g(c)$. Then the composition $f \circ g$ is differentiable at c and*

$$(f \circ g)'(c) = f'(g(c)) \,.\, g'(c).$$

This may look a bit different from the chain rule that you know but in fact it is exactly the one that you are accustomed to using: let's see why. Let's take $g(x) = 1 + x^2$ and $f(u) = u^3$. Then the composition is $f \circ g : x \mapsto (1 + x^2)^3$. We have $g'(x) = 2x$ and $f'(u) = 3u^2$. According to the theorem the derivative of the composition is

$$f'(g(c)) \,.\, g'(c) = 3g(c)^2 \,.\, g'(c) = 3 \left(1 + c^2\right)^2 \; 2c = \cdots .$$

Note that the derivative of f is evaluated at $g(c)$: the only place that makes any sense because f is defined on an interval containing $g(c)$. The chain rule is a bit harder to prove than the other two rules. How could we try to prove the theorem? We want to investigate the ratio

$$\frac{f(g(x)) - f(g(c))}{x - c}$$

as $x \to c$. The obvious thing to do is to write this as

$$\frac{f(g(x)) - f(g(c))}{g(x) - g(c)} \times \frac{g(x) - g(c)}{x - c}.$$

The second factor converges to $g'(c)$ as $x \to c$.

For the first factor we note that $g(x) \to g(c)$ as $x \to c$ so the limit of the first factor is therefore

$$\lim_{g(x) \to g(c)} \frac{f(g(x)) - f(g(c))}{g(x) - g(c)}$$

which appears to be $f'(g(c))$. So we appear to have proved that

$$(f \circ g)'(c) = f'(g(c))g'(c).$$

There is a problem however. The quantity $g(x) - g(c)$ could be equal to zero for lots of values of x: perhaps even all of them. In that case we can't divide by this quantity. For this reason we need to start with a new way to express the differentiability of a function.

Lemma (**Local linearisation**)**.** *Suppose I is an open interval, $f : I \to \mathbf{R}$ and $c \in I$. Then f is differentiable at c if and only if there is a number A and a function ε with the properties that for all x*

$$f(x) - f(c) = A(x - c) + \varepsilon(x)(x - c),$$

$\varepsilon(c) = 0$ and ε is continuous at c: ($\varepsilon(x) \to 0$ as $x \to c$). If this happens $A = f'(c)$.

This is sometimes called the Weierstrass-Caratheodory criterion. The lemma says that if x is close to c then $f(x)$ is approximately given by the linear function

$$x \mapsto f(c) + f'(c)(x - c)$$

which you recognise as the first Taylor approximation to f.

Proof If the condition holds then

$$\frac{f(x) - f(c)}{x - c} = A + \varepsilon(x)$$

and this approaches A as $x \to c$. Hence f is differentiable with derivative $f'(c) = A$.

On the other hand suppose f is differentiable at c. Set $A = f'(c)$ and define ε as follows

$$\varepsilon(x) = \begin{cases} \frac{f(x)-f(c)}{x-c} - A & \text{if } x \neq c \\ 0 & \text{if } x = c. \end{cases}$$

If $x \neq c$ then $f(x) - f(c) = A(x - c) + \varepsilon(x)(x - c)$ holds because of the way ε is defined, while if $x = c$ the formula is obvious. To check that $\varepsilon(x) \to 0$ as $x \to c$ observe that

$$\lim_{x \to c} \varepsilon(x) = \lim_{x \to c} \left(\frac{f(x) - f(c)}{x - c} - f'(c) \right) = 0.$$

□

We are now ready to prove the chain rule.

Proof (**of the chain rule**) f is differentiable at $g(c)$ so for all y

$$f(y) - f(g(c)) = f'(g(c))(y - g(c)) + \varepsilon(y)(y - g(c)).$$

where $\varepsilon(g(c)) = 0$ and ε is continuous at $g(c)$. Hence

$$f(g(x)) - f(g(c)) = f'(g(c))(g(x) - g(c)) + \varepsilon(g(x))(g(x) - g(c)).$$

Consequently if $x \neq c$

$$\frac{f(g(x)) - f(g(c))}{x - c} = f'(g(c))\frac{g(x) - g(c)}{x - c} + \varepsilon(g(x))\frac{g(x) - g(c)}{x - c}.$$

As $x \to c$,

$$\frac{g(x) - g(c)}{x - c} \to g'(c)$$

while $\varepsilon \circ g$ is continuous at c so $\varepsilon(g(x)) \to \varepsilon(g(c)) = 0$. Hence

$$\frac{f(g(x)) - f(g(c))}{x - c} \to f'(g(c)).g'(c)$$

as required. □

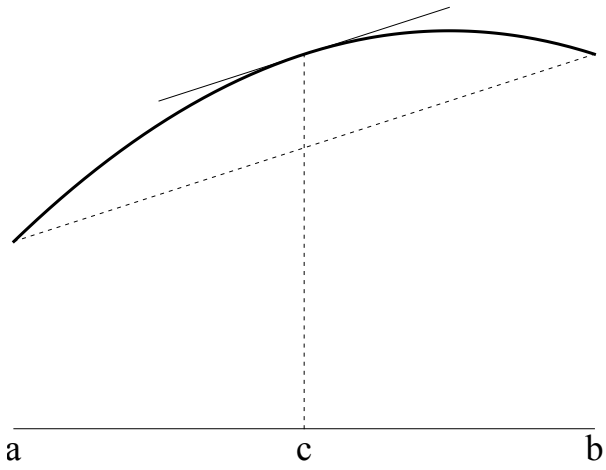
The Mean Value Theorem

In this section we prove one of the most useful facts about derivatives: the so-called Mean Value Theorem (MVT). It provides a way to relate the values of a function to the values of its derivative.

Theorem (**Mean Value Theorem**)**.** *Suppose $f : [a, b] \to \mathbf{R}$ is continuous on the closed interval $[a, b]$ and differentiable on the open interval (a, b). Then there is a point $c \in (a, b)$ where*

$$f'(c) = \frac{f(b) - f(a)}{b - a}.$$

The theorem says that between any pair of points a and b, there is a third point c where the slope of the curve $y = f(x)$ is equal to the slope of the chord joining $(a, f(a))$ and $(b, f(b))$. Geometrically this is intuitively obvious as can be seen from the picture below.



Another way to interpret the theorem, that explains its name, is this. Suppose you drive a distance of 30 miles in one hour. Then your average (or mean) speed for the trip is 30mph. The theorem says that at some point in your trip your speed will be exactly 30mph: at some point the needle of your speedometer will point to 30.

To see why the MVT might be useful let's deduce some immediate consequences.

Corollary (**Functions with positive derivative**). *If $f : I \to \mathbf{R}$ is differentiable on the open interval I and $f'(x) > 0$ for all x in the interval then f is strictly increasing on the interval.*

Proof If there were two points a and b with $a < b$ but $f(a) \geq f(b)$ then we could find a point c where

$$f'(c) = \frac{f(b) - f(a)}{b - a} \leq 0$$

contradicting the hypothesis. □

Corollary (**Functions with zero derivative**). *If $f : I \to \mathbf{R}$ is differentiable on the open interval I and $f'(x) = 0$ for all x in the interval then f is constant on the interval.*

Proof Exercise. □

The MVT has many uses in the theory of integration and the study of differential equations. The previous corollary can be regarded as the statement that the only solutions of the differential equation $f'(x) = 0$ are the constant functions.

In a similar way we can obtain uniqueness of solutions to other differential equations.

Example (**Uniqueness of solution to a Diff. Eq. I**). *The only functions* $f : \mathbf{R} \to \mathbf{R}$ *satisfying*

$$f'(x) = f(x)$$

are the functions $f(x) = Ae^x$ *for some constant* A.

Proof We shall assume that e^x is differentiable and is its own derivative (a fact that will be proved in the next chapter). Suppose f is such a solution and let $g(x) = e^{-x}f(x)$. Then by the chain rule we have

$$g'(x) = -e^{-x}f(x) + e^{-x}f'(x) = e^{-x}(-f(x) + f'(x)) = 0.$$

By the corollary g is a constant function with value A (say). Then $f(x) = e^x g(x) = Ae^x$. □

Example (**Uniqueness of solution to a Diff. Eq. II**). *The only functions* $f : (0, \infty) \to \mathbf{R}$ *satisfying*

$$f'(x) + \frac{1}{x}f(x) = 2$$

are the functions $f(x) = A/x + x$ *for some constant* A.

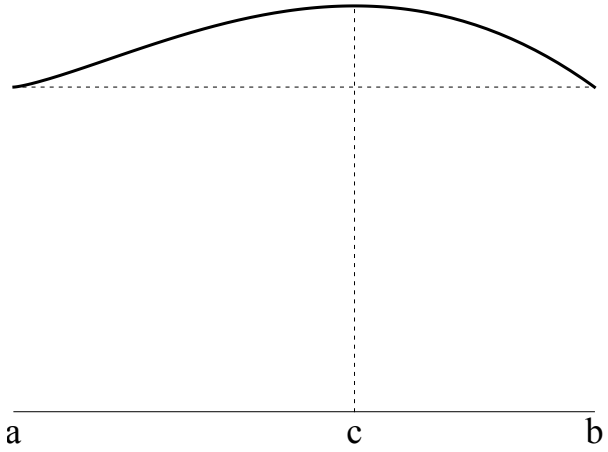
Proof Let $g(x) = xf(x) - x^2$ for positive x. Then

$$g'(x) = xf'(x) + f(x) - 2x = x\left(f'(x) + \frac{1}{x}f(x) - 2\right) = 0.$$

So g is constant A (say), and the formula for f follows. □

In order to prove the MVT we shall start by proving the special case in which the slope is 0.

Theorem (**Rolle**). *Suppose* $f : [a, b] \to \mathbf{R}$ *is continuous on the closed interval* $[a, b]$ *and differentiable on the open interval* (a, b) *and that* $f(a) = f(b)$. *Then there is a point* c *in the open interval where* $f'(c) = 0$.



Proof If f is constant on the interval then its derivative is zero everywhere. If not it takes values different from $f(a) = f(b)$. Assume it is somewhere larger than $f(a)$.

Since f is continuous on the closed interval it attains its maximum value at some point c and this cannot be a or b: so c lies in (a, b). If $x > c$ then $f(x) - f(c) \leq 0$ while $x - c > 0$ so the ratio

$$\frac{f(x) - f(c)}{x - c} \leq 0.$$

So $f'(c)$ is a limit of non-positive values and so is not positive.

On the other hand if $x < c$ then $f(x) - f(c) \leq 0$ while $x - c < 0$ so the ratio

$$\frac{f(x) - f(c)}{x - c} \geq 0.$$

So $f'(c)$ is a limit of non-negative values and so is not negative. Therefore $f'(c) = 0$. □

Notice that we found the point c without having any formula for f or its derivative. We simply used a fact which has nothing to do with derivatives: that f attains its maximum. This fact was proved in the Analysis I course using the Bolzano-Weierstrass Theorem which does not construct the point in any computable way.

We now come to the proof of the MVT. We shall modify the function f by subtracting a linear function whose slope is $\frac{f(b)-f(a)}{b-a}$. This new function will take the same values at the two ends and so have a point with zero slope by Rolle's Theorem. But at this point the slope of f must be the same as the slope of the linear function that we subtracted from it.

Proof (**of the MVT**) Consider the function given by

$$g(x) = f(x) - x\frac{f(b) - f(a)}{b - a}.$$

Then

$$g(b) - g(a) = f(b) - f(a) - (b - a)\frac{f(b) - f(a)}{b - a} = 0.$$

So by Rolle's Theorem there is a point c where $g'(c) = 0$. But this implies

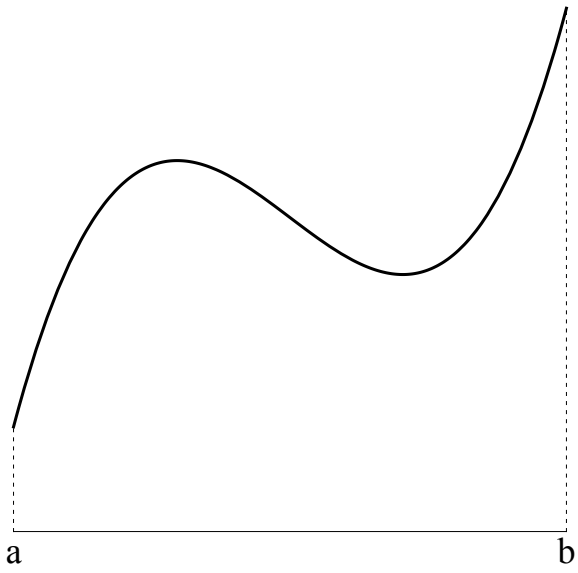
$$f'(c) = \frac{f(b) - f(a)}{b - a}.$$

□

In the proof of Rolle's Theorem we used the fact that if a differentiable function $f : [a, b] \to \mathbf{R}$ attains its maximum at a point of the open interval (a, b) then its derivative must be zero at that point. This principle underlies the familiar method for finding maxima and minima. For a function on a closed interval the cleanest statement of the principle is this.

Theorem (**Extrema and derivatives**)**.** *Suppose $f : [a, b] \to \mathbf{R}$ is continuous and that it is differentiable on the open interval. Then f attains its maximum and minimum either at points in (a, b) where $f' = 0$ or at one of the ends a or b.*

The second possibility can of course happen.



Finding where $f' = 0$ doesn't tell you where the maximum is but it narrows down the options very considerably. Normally there will only be a few points where $f' = 0$ and so

you only have to check those and the ends. Once you have found all these possibilities (a, b and some points $x_1, x_2, \ldots$ where the derivative vanishes) the most reliable way to find the maximum is just to calculate the values $f(a)$, $f(b)$, $f(x_1)$, $f(x_2)$ and so on. You then just check which one is greatest and which is least.

It is customary in elementary calculus texts to suggest the use of the second derivative to try to find maxima and minima. This method is sometimes useful but has two drawbacks:

- If $f'(c) = 0$ and $f''(c) < 0$ it tells you that c is a *local* maximum but the function may have several local maxima many of which are not the maximum.
- If $f''(c) = 0$ then it tells you nothing.

(On the other hand, the idea can often be useful in theoretical situations in the other direction. If the maximum occurs at c in the open interval then you can conclude that $f'(c) = 0$ and $f''(c) \leq 0$ which may be valuable information.)

If you have a function defined on an open interval, on $(0, \infty)$ or the whole of $\mathbf{R}$ there are a number of options. Sometimes it is easy to see that the function has its maximum in a certain closed interval and then use the closed interval version. Sometimes it is better to use the derivative to see that the function increases up to the maximum and decreases after it.

Example. *Find the maximum of* xe^{-x} *on* $\mathbf{R}$.

The derivative is $x \mapsto (1-x)e^{-x}$ which is positive if $x < 1$ and negative if $x > 1$. By the MVT the function increases until $x = 1$ and then decreases. So the maximum occurs at $x = 1$ where the function is equal to e^{-1}. □

As a consequence we have proved that $xe^{-x} \leq e^{-1}$ for all x. Actually we already knew this.

$$x = 1 + x - 1 \leq e^{x-1}.$$

Derivatives of inverses

We saw earlier that if f is a continuous strictly increasing function then it has a continuous inverse. An obvious question is whether the inverse is differentiable whenever the original function is differentiable and how the derivatives are related.

Actually the second question is not too hard to answer. Suppose f and g are inverses and we know that they are differentiable: then

$$f(g(x)) = x$$

on the domain of g. We can differentiate this equation using the chain rule to get

$$f'(g(x))g'(x) = 1$$

and thus we conclude that

$$g'(x) = \frac{1}{f'(g(x))}.$$

Notice that the derivative of f is evaluated at $g(x)$ which looks a bit complicated but as we discussed when looking at the chain rule, it is the only thing that makes sense. It also gives the right answers. Let's see how this works in practice (ignoring for the moment the fact that we don't yet know that the exponential and trig. functions are differentiable).

Example (**The derivative of $x \mapsto \sqrt{x}$**). *Let $f : x \mapsto x^2$ be the squaring function and g the square root. Then*

$$g'(x) = \frac{1}{f'(g(x))} = \frac{1}{2g(x)} = \frac{1}{2\sqrt{x}}.$$

Example (**The derivative of** $\log$). *Let f be the exponential function and g the logarithm. Then*

$$g'(x) = \frac{1}{f'(g(x))} = \frac{1}{\exp(\log x)} = \frac{1}{x}.$$

Example (**The derivative of** $\sin^{-1}$). *Let $f : [-\pi/2, \pi/2] \to [-1, 1]$ be the function $x \mapsto \sin x$ and g the inverse sine. Then*

$$g'(x) = \frac{1}{f'(g(x))} = \frac{1}{\cos(\sin^{-1} x)}.$$

This is not the usual way you are accustomed to writing the derivative. In the homework you are asked to confirm that it is $\frac{1}{\sqrt{1-x^2}}$. Now let's prove the general theorem.

Theorem (**Derivatives of inverses**). *Let $f : (a, b) \to \mathbf{R}$ be differentiable with positive derivative. Then $g = f^{-1}$ is differentiable and*

$$g'(x) = \frac{1}{f'(g(x))}.$$

Proof Since f has positive derivative it is continuous and strictly increasing. Therefore it has a continuous inverse. Let (c, d) be the range of f, let x be in the interval (c, d) and $g(x) = y$. We want to calculate

$$\lim_{u \to x} \frac{g(u) - g(x)}{u - x}.$$

Let $v = g(u)$ so that $u = f(v)$. Then the quotient is

$$\frac{v - y}{f(v) - f(y)}.$$

As $u \to x$ we know that $v = g(u) \to y = g(x)$ because g is continuous at x. So we want to calculate

$$\lim_{v \to y} \frac{v - y}{f(v) - f(y)}.$$

We know that

$$\lim_{v \to y} \frac{f(v) - f(y)}{v - y} = f'(y)$$

and that this limit is positive. So by the properties of limits we have

$$\lim_{u \to x} \frac{g(u) - g(x)}{u - x} = \lim_{v \to y} \frac{v - y}{f(v) - f(y)} = \frac{1}{f'(y)} = \frac{1}{f'(g(x))}.$$

□

Chapter 3. Power series II

The differentiability of power series

We start this chapter by checking that power series are differentiable inside the radius of convergence. Naturally the proof of this is more difficult than the proof of continuity. Suppose $f(x) = \sum a_n x^n$ is a power series with radius of convergence R. We want to show that for $|x| < R$ the derivative exists and

$$f'(x) = \sum_{n=1}^{\infty} n a_n x^{n-1}.$$

In other words we want to know that we can differentiate the series term by term as if it were a polynomial. The sum rule doesn't tell us that we can, because an infinite sum is not the same as a finite one.

We want to show that if $|x| < R$, the sum $\sum_{n=1}^{\infty} n a_n \, x^{n-1}$ converges and

$$\frac{\sum_{n=0}^{\infty} a_n y^n - \sum_{n=0}^{\infty} a_n x^n}{y - x} \to \sum_{n=1}^{\infty} n a_n x^{n-1}$$

as $y \to x$. The left side is

$$\sum_{n=0}^{\infty} a_n \frac{y^n - x^n}{y - x} = \sum_{n=1}^{\infty} a_n (y^{n-1} + y^{n-2}x + \cdots + x^{n-1}).$$

This looks promising because if y is close to x the sum

$$y^{n-1} + y^{n-2}x + \cdots + x^{n-1}$$

is close to nx^{n-1} because we know that polynomials are continuous.

We want to conclude that

$$\sum_{n=1}^{\infty} a_n (y^{n-1} + y^{n-2}x + \cdots + x^{n-1}) \to \sum_{n=1}^{\infty} n a_n x^{n-1}.$$

The trouble is that because we have an infinite sum we want the two things to be close together for every n at the same time and this is not something that follows from what we already know. Let's start gently.

Lemma (**The differentiability of power series I**). *Let $\sum a_n x^n$ be a power series with radius of convergence R. Then the series $\sum n a_n x^{n-1}$ has the same radius of convergence.*

Proof We know that the absolute series $\sum |a_n| x^n$ has the same radius of convergence as $\sum a_n x^n$. Now if $0 < x < R$ choose y with $x < y < R$. Then $\sum |a_n| x^n$ and $\sum |a_n| y^n$ both converge and hence so does

$$\sum_{n=0}^{\infty} |a_n| \frac{y^n - x^n}{y - x} = \sum_{n=1}^{\infty} |a_n| (y^{n-1} + y^{n-2}x + \cdots + x^{n-1}).$$

But the last sum is larger than $\sum_{n=1}^{\infty} |a_n| n x^{n-1}$ so the latter also converges. This means that $\sum n a_n x^{n-1}$ converges absolutely as required.

To check that the radius of convergence of the derivative is actually the same and not larger one can use the comparison principle for power series from Chapter 1. □

We now move on to the full theorem and use the existence of the derivative to control the limits we are trying to evaluate.

Theorem (**The differentiability of power series II**). *Let $f(x) = \sum a_n x^n$ be a power series with radius of convergence R. Then f is differentiable on $(-R, R)$ and*

$$f'(x) = \sum_{n=1}^{\infty} n a_n \, x^{n-1}.$$

Proof Choose T with $|x| < T < R$. We know from the lemma that the series $\sum n |a_n| T^{n-1}$ converges so given $\varepsilon > 0$ there is a number N so that

$$\sum_{n=N+1}^{\infty} n |a_n| T^{n-1} < \frac{\varepsilon}{3}.$$

Now if $0 < |y - x| < T - |x|$ we have $|y| < T$ as well as $|x| < T$ and so

$$\left| \sum_{n=N+1}^{\infty} n a_n \, x^{n-1} \right| \leq \sum_{n=N+1}^{\infty} n |a_n| \, |x|^{n-1} < \frac{\varepsilon}{3}$$

and also

$$\begin{aligned}\left|\sum_{N+1}^{\infty} a_n \frac{y^n - x^n}{y - x}\right| &= \left|\sum_{N+1}^{\infty} a_n (y^{n-1} + y^{n-2}x + \cdots + x^{n-1})\right| \\ &\leq \sum_{N+1}^{\infty} |a_n| \left(|y|^{n-1} + \cdots + |x|^{n-1}\right) \\ &\leq \sum_{N+1}^{\infty} n|a_n|\, T^{n-1} < \frac{\varepsilon}{3}.\end{aligned}$$

The sum

$$\sum_{n=1}^{N} a_n (y^{n-1} + y^{n-2}x + \cdots x^{n-1})$$

is a polynomial in y whose value at x is $\sum_1^N n a_n\, x^{n-1}$ so there is a $\delta_0 > 0$ with the property that if $0 < |y - x| < \delta_0$

$$\left|\sum_{1}^{N} a_n \frac{y^n - x^n}{y - x} - \sum_{1}^{N} n a_n\, x^{n-1}\right|$$

$$= \left|\sum_{1}^{N} a_n (y^{n-1} + y^{n-2}x + \cdots x^{n-1}) - \sum_{1}^{N} n a_n\, x^{n-1}\right| < \frac{\varepsilon}{3}.$$

So if we choose δ to be the smaller of δ_0 and $T - |x|$ then whenever $0 < |y - x| < \delta$

$$\left|\sum_{1}^{\infty} a_n \frac{y^n - x^n}{y - x} - \sum_{1}^{\infty} n a_n\, x^{n-1}\right|$$

$$\leq \left|\sum_{N+1}^{\infty} a_n \frac{y^n - x^n}{y - x}\right| + \left|\sum_{1}^{N} a_n \frac{y^n - x^n}{y - x} - \sum_{1}^{N} n a_n\, x^{n-1}\right| + \left|\sum_{N+1}^{\infty} n a_n x^{n-1}\right| < \varepsilon.$$

□

Corollary (**The derivative of the exponential**).

$$\exp'(x) = \exp(x).$$

Proof

$$\exp(x) = 1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \cdots.$$

We can differentiate term by term to get

$$\exp'(x) = 0 + 1 + x + \frac{x^2}{2} + \cdots = \exp(x).$$

□

As promised we can now give a simple proof of the characteristic property of the exponential.

Corollary (**The characteristic property of the exponential**)**.** *If x and y are real numbers then* $\exp(x+y) = \exp(x)\exp(y)$.

Proof For a fixed number z consider the function

$$x \mapsto \exp(x)\exp(z-x).$$

We may differentiate this with respect to x using the product rule and the chain rule to get

$$\exp(x)\exp(z-x) - \exp(x)\exp(z-x) = 0.$$

By the MVT the function is constant. At $x = 0$ the function is $\exp z$ so we know that for all x

$$\exp(x)\exp(z-x) = \exp(z).$$

Now if we set $z = x + y$ we get the conclusion we want. □

We now know that the exponential function is differentiable, has the correct derivative and satisfies the characteristic property. Using what we did earlier on the derivatives of inverses we can also conclude that the logarithm has the correct derivative.

Corollary (**The derivative of** log)**.** *If $f : x \mapsto \log x$ then*

$$f'(x) = \frac{1}{x}.$$

Example (**Power series solution of a Diff Eq. I**)**.** *Let f be defined on $(-1, 1)$ by the series*

$$f(x) = \sum_{n=1}^{\infty} \frac{x^n}{n^2}.$$

Then f satisfies the differential equation

$$xf''(x) + f'(x) = \frac{1}{1-x}.$$

Proof The series has radius of convergence 1 as you saw in the homework.

$$f'(x) = \sum_{n=1}^{\infty} \frac{x^{n-1}}{n}$$

$$f''(x) = \sum_{n=1}^{\infty} \frac{(n-1)x^{n-2}}{n}$$

and so

$$xf''(x) = \sum_{n=1}^{\infty} \frac{(n-1)x^{n-1}}{n}.$$

So

$$xf''(x) + f'(x) = \sum_{n=1}^{\infty} \frac{nx^{n-1}}{n} = \sum_{n=1}^{\infty} x^{n-1} = \frac{1}{1-x}.$$

□

Example (**Power series solution of a Diff Eq. II**). *Consider the differential equation*

$$xy'' + y' + xy = 0.$$

Suppose there were a solution given by a power series

$$y = \sum_{0}^{\infty} a_n x^n.$$

Then

$$xy'' = \sum_{2}^{\infty} n(n-1)a_n x^{n-1} \quad \text{and} \quad y' = \sum_{1}^{\infty} n a_n x^{n-1}$$

while

$$xy = \sum_{2}^{\infty} a_{n-2} x^{n-1}.$$

So

$$a_1 + \sum_{2}^{\infty} (n^2 a_n + a_{n-2}) x^{n-1} = 0.$$

We can look at the coefficients of each power and set them to 0

$$a_1, \qquad a_0 + 4a_2, \qquad a_1 + 9a_3, \qquad a_2 + 16a_4, \qquad \ldots.$$

We can make these all zero by choosing all the odd numbered a_i to be zero but then take

$$a_0 = 1, \qquad a_2 = -\frac{1}{4}, \qquad a_4 = \frac{1}{64}, \qquad \dots$$

and more generally

$$a_{2m} = \frac{(-1)^m}{4^m (m!)^2}.$$

It is quite easy to see that if we define a function J_0 by

$$J_0(x) = \sum_0^\infty \frac{(-1)^m}{4^m (m!)^2} x^{2m}$$

then the series has infinite radius of convergence and the function satisfies the differential equation.

$$xy'' + y' + xy = 0.$$

This function is called the Bessel function of order zero. Bessel functions turn up in the modelling of the vibration of a drumhead.

The last item in this section will explain how we can determine the coefficients of a power series if we know the function. Suppose I know that

$$f(x) = \sum_{n=0}^\infty a_n x^n = a_0 + a_1 x + a_2 x^2 + \cdots$$

on the interval $(-R, R)$. From the coefficients I can build the function by carrying out the sum. How can I go the other way? If I know the function f how can I find its coefficients?

The first one is easy $a_0 = f(0)$. The next one, a_1 is not so obvious but we know that

$$f'(x) = a_1 + 2a_2 x + 3a_3 x^2 + \cdots$$

and so we can calculate a_1 by differentiating f: we have $a_1 = f'(0)$.

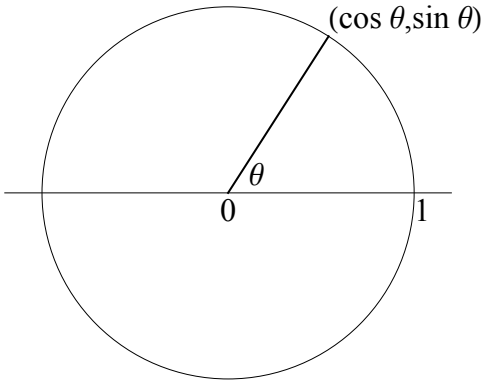
We can continue in this way

$$\begin{aligned} a_0 &= f(0) \\ a_1 &= f'(0) \\ a_2 &= \frac{f''(0)}{2} \\ &\vdots \end{aligned}$$

You will recognise these numbers as the coefficients in the Taylor expansion for f to which we shall return in the last chapter.

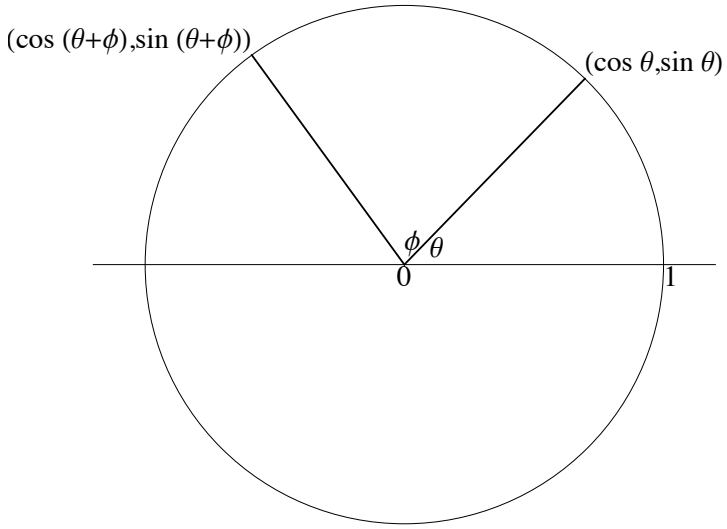
The trigonometric functions

For the purposes of this course we shall define the trigonometric functions sine and cosine by power series and then check that they have the properties we expect them to have. Our aim will be to show that for each θ the point $(\cos\theta, \sin\theta)$ lies on the circle of radius one with centre 0 and that the radius through this point makes an angle θ with the horizontal.

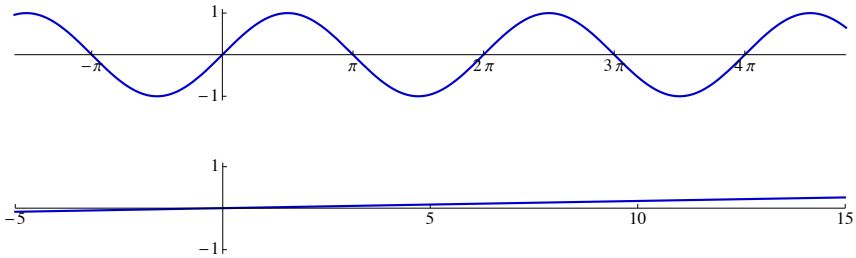


The first statement is just that $\cos^2\theta + \sin^2\theta = 1$. For the second we need to be clear that we are measuring angle in radians so let us recall what that means.

The Babylonians were originally responsible for the division of the circle into 360 equal parts: what we now call degrees. Measuring angle in degrees has the property that adding two angles corresponds to the geometric process of rotating through one angle followed



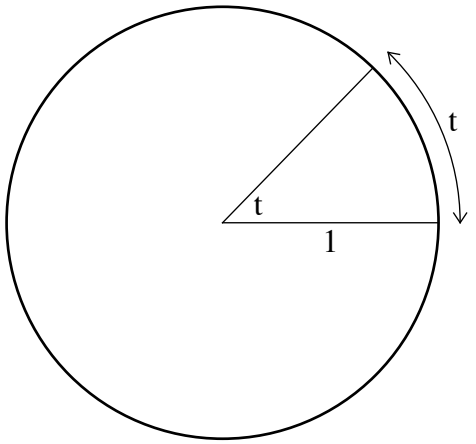
by another.
This is a desirable property (also possessed by radians). However, degrees have a property that is far from desirable. Below is the graph of $y = \sin x^\circ$ drawn with the same scale on



both axes.

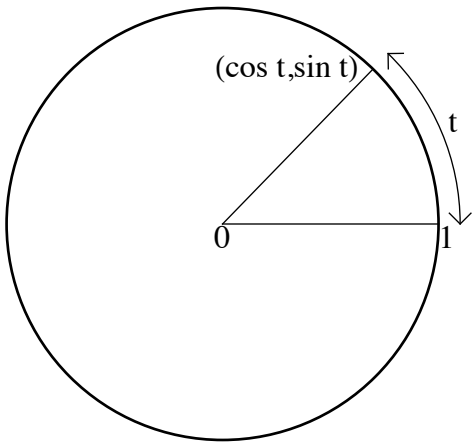
The slope of the graph at $x = 0$ is $\pi/180 \approx 0.0174533$. So if we differentiate the function we get $0.0174... \cos x^\circ$: we get a funny multiple of cos rather than cos itself. The choice of radians removes this problem.

How do we measure the size of an angle in radians? We draw the circle of radius 1 and then for a given angle we use the length of the circular arc that it spans as the measure of the angle.



The size of the angle is the length of the arc it spans: simple.

So the second thing we will need to check is that for each t, the point on the circle whose distance from the horizontal measured around the circle is t, has coordinates $(\cos t, \sin t)$.



Definition (**The trig functions**). *For $x \in \mathbf{R}$ we define*

$$\cos x = 1 - \frac{x^2}{2} + \frac{x^4}{24} - \cdots + (-1)^k \frac{x^{2k}}{(2k)!} + \cdots$$

$$\sin x = x - \frac{x^3}{6} + \frac{x^5}{120} - \cdots + (-1)^k \frac{x^{2k+1}}{(2k+1)!} + \cdots$$

It is easy to check that these series converge everywhere by the ratio test. Cosine is obviously an even function and sine is obviously odd. From our general theory we know

that both functions are differentiable and we can check that

$$\frac{d}{dx}\cos x = -\sin x$$

and

$$\frac{d}{dx}\sin x = \cos x.$$

Using these derivatives we can check the addition formulae quite easily.

Theorem (**The addition formulae for the trig functions**). *For all real x and y*

$$\begin{aligned}\cos(x+y) &= \cos x\cos y - \sin x\sin y\\ \sin(x+y) &= \sin x\cos y + \cos x\sin y\end{aligned}$$

Proof We shall check the first: the second is similar. For a fixed z let

$$f(x) = \cos x\cos(z-x) - \sin x\sin(z-x).$$

Using the derivatives above it is easy to check that $f'(x) = 0$ for all x and so f is constant. When $x = 0$ we have

$$f(0) = \cos 0\cos z - \sin 0\sin z = \cos z.$$

Hence $f(x) = \cos z$ for all x. Now if we set $z = x + y$ we get

$$\cos x\cos y - \sin x\sin y = f(x) = \cos z = \cos(x+y).$$

□

This immediately shows us that the point $(\cos x, \sin x)$ lies on the circle for each x.

Corollary (**The circular property**). *For all real x*

$$\cos^2 x + \sin^2 x = 1.$$

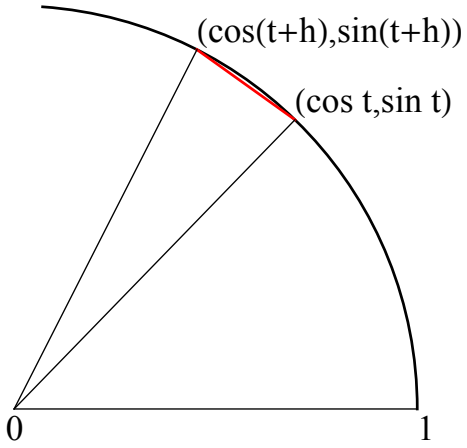
Proof Set $y = -x$ in the addition formula for cos. We know that $\cos(-x) = \cos x$ and $\sin(-x) = -\sin x$ and so we get

$$1 = \cos 0 = \cos(x-x) = \cos x\cos(-x) - \sin x\sin(-x) = \cos^2 x + \sin^2 x.$$

□

The remaining thing we need to check is that as t increases the point $(\cos t, \sin t)$ traces out the circle at rate 1. Let $L(t)$ be the length of the circular arc from $(1, 0)$ to the point $(\cos t, \sin t)$. We want to show that $L(t) = t$ for each t. By the MVT it suffices to show that $L'(t) = 1$ at each point.

Consider the point $(\cos t, \sin t)$ and a nearby point $(\cos(t + h), \sin(t + h))$.



When h is very small, the (straight line) distance between these two points is approximately the same as the length of the circular arc between them. The straight line distance is

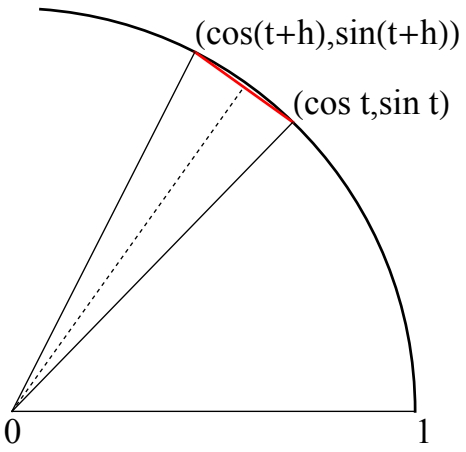
$$\sqrt{(\cos(t + h) - \cos t)^2 + (\sin(t + h) - \sin t)^2}.$$

So our aim is to show that

$$\lim_{h \to 0^+} \frac{1}{h} \sqrt{(\cos(t + h) - \cos t)^2 + (\sin(t + h) - \sin t)^2} = 1$$

for every t.

If we knew that our power series *do* give the point at the correct angle then we could use geometry to calculate the length:



The length is supposed to be $2\sin(h/2)$. That would be good because then

$$\lim_{h\to 0^+} \frac{1}{h}\sqrt{(\cos(t+h)-\cos t)^2+(\sin(t+h)-\sin t)^2}$$
$$= \lim_{h\to 0^+} \frac{2\sin(h/2)}{h} = \lim_{h\to 0^+} \frac{\sin(h/2)}{h/2} = \lim_{p\to 0^+} \frac{\sin p}{p} = 1$$

because the last expression is just the derivative of sin at 0.

However we don't yet know that our functions cosine and sine correspond to the geometry of the circle. So we have to check the limit using the properties that we do know.

$$\lim_{h\to 0^+} \frac{1}{h}\sqrt{(\cos(t+h)-\cos t)^2+(\sin(t+h)-\sin t)^2}$$
$$= \lim_{h\to 0^+} \sqrt{\left(\frac{\cos(t+h)-\cos t}{h}\right)^2+\left(\frac{\sin(t+h)-\sin t}{h}\right)^2}$$
$$= \sqrt{\left(\lim_{h\to 0^+} \frac{\cos(t+h)-\cos t}{h}\right)^2+\left(\lim_{h\to 0^+} \frac{\sin(t+h)-\sin t}{h}\right)^2}$$
$$= \sqrt{(-\sin t)^2+(\cos t)^2} = 1.$$

This completes the proof.

We have now checked that our power series do produce the x and y coordinates of the correct point on the circle. If we use the symbol π to denote half the circumference of the circle then we know that $\cos\pi = -1$, $\sin\pi = 0$ and so on. We can also see that the trig functions are periodic with period 2π although we could deduce that directly from the addition formulae once we know that $\cos 2\pi = 1$ and $\sin 2\pi = 0$. For example

$$\cos(x + 2\pi) = \cos x \cos 2\pi - \sin x \sin 2\pi = \cos x.$$

There are a number of special values that one has to know.

$$\begin{array}{rclrcl} \sin 0 & = & 0 & \cos 0 & = & 1 \\ \sin \pi/6 & = & 1/2 & \cos \pi/6 & = & \sqrt{3}/2 \\ \sin \pi/4 & = & 1/\sqrt{2} & \cos \pi/4 & = & 1/\sqrt{2} \\ \sin \pi/3 & = & \sqrt{3}/2 & \cos \pi/3 & = & 1/2 \end{array}$$

Interestingly, the existence of the series for sine and cosine predates the invention of calculus by a couple of hundred years. As far as we can tell they were discovered by Madhava. His work does not itself survive, but later scholars in the Kerala school were clear that he was the original discoverer.

The complex exponential

The proof of the addition formulae for the trig functions looks very much like the MVT proof of the characteristic property of the exponential and the addition formulae have the same "shape". The cosine of the sum $\cos(x + y)$ is a product of the cosines of x and y together with a product of sines. This is not coincidence. If we introduce the complex number i whose square is -1 we can write

$$e^{it} = 1 + it - \frac{t^2}{2} - i\frac{t^3}{6} + \frac{t^4}{24} + \cdots = \cos t + i \sin t.$$

This formula

$$e^{it} = \cos t + i \sin t$$

linking the exponential and trig functions was described by Feynman as "our jewel". Now we have

$$\begin{aligned}\cos(x+y)+i\sin(x+y) &= e^{i(x+y)}=e^{ix}e^{iy}\\ &= (\cos x+i\sin x)(\cos y+i\sin y).\end{aligned}$$

The last can be expanded as

$$\cos x\cos y-\sin x\sin y+i(\sin x\cos y+\cos x\sin y)$$

from which we can read off the addition formulae for cosine and sine.

In a similar way we can relate the derivatives of cos and sin to the derivative of the exponential function.

We chose the exponential function and we chose to measure angle in radians in order to make the derivatives work out right. We can now see that these two choices are really the same choice. The formula

$$e^{it}=\cos t+i\sin t$$

only works if the angles are in radians and we use the correct exponential.

The tangent

Once we have defined cos and sin we can define the tangent. We know that $\cos x=0$ whenever x is an odd multiple of $\pi/2$. For all other points we can define

$$\tan x=\frac{\sin x}{\cos x}.$$

Using our knowledge of the derivatives of cos and sin we can find the derivative of tan. In the HW you are asked to find the derivative of the inverse, $\tan^{-1}$ and from this to find a power series for $\tan^{-1}$. You are also asked to derive the addition formula for tan from those for cos and sin.

An obvious question: what is the power series for $\tan x$ for x close to 0? Assuming that the power series exists you can find it by repeatedly differentiating tan. Here are the first few terms

$$x+\frac{x^3}{3}+\frac{2x^5}{15}+\frac{17x^7}{315}+\frac{62x^9}{2835}+\frac{1382x^{11}}{155925}+\cdots.$$

The numerator 1382 has a prime factor 691. This suggests that you cannot write down a simple formula for the coefficients: and indeed you can't. They are related to values of what is known as the ζ function.

Chapter 4. Taylor's Theorem

Cauchy's Mean Value Theorem

We start this section with a more general form of the MVT known as Cauchy's MVT. The MVT says that under suitable conditions there is a point t between a and b with

$$f'(t) = \frac{f(b) - f(a)}{b - a}.$$

This can be interpreted as saying that if g is the function given by $g(x) = x$ then

$$\frac{f'(t)}{g'(t)} = \frac{f(b) - f(a)}{g(b) - g(a)}$$

since $g'(t) = 1$ for every t. Cauchy's MVT says that the same thing holds for any differentiable function g for which the statement makes sense. This theorem will be of interest to us solely in order to prove something called l'Hôpital's Rule for finding limits.

Theorem (**Cauchy's Mean Value Theorem**)**.** *If $f, g : [a, b] \to \mathbf{R}$ are continuous, are differentiable on (a, b) and $g'(t) \neq 0$ for t between a and b then there is a point t where*

$$\frac{f'(t)}{g'(t)} = \frac{f(b) - f(a)}{g(b) - g(a)}.$$

Proof Consider the function

$$x \mapsto h(x) = f(x)(g(b) - g(a)) - (f(b) - f(a))g(x).$$

At $x = a$ the value of the function is

$$h(a) = f(a)g(b) - f(a)g(a) - f(b)g(a) + f(a)g(a) = f(a)g(b) - f(b)g(a).$$

Similarly

$$h(b) = f(b)g(b) - f(b)g(a) - f(b)g(b) + f(a)g(b) = f(a)g(b) - f(b)g(a).$$

So $h(b) = h(a)$ and by Rolle's Theorem there is a point t between a and b where $h'(t) = 0$.

Thus we have

$$h(x) = f(x)(g(b) - g(a)) - (f(b) - f(a))g(x)$$

and there is a point t where $h'(t) = 0$. But this means that

$$f'(t)(g(b) - g(a)) = (f(b) - f(a))g'(t).$$

Since g' is non-zero on (a, b) Rolle's Theorem applied to g shows that $g(b) - g(a) \neq 0$ as well and we can rearrange to get the conclusion of the theorem. □

L'Hôpital's rule

It often happens that we wish to calculate limits such as

$$\lim_{x\to 0} \frac{\sin x}{1-\sqrt{1-x}}$$

in which both the numerator and denominator converge to 0. So we can't calculate the limit just by substituting $x = 0$. An obvious example that we have already considered and understood is the definition of $f'(c)$

$$\lim_{x\to c} \frac{f(x)-f(c)}{x-c}.$$

It is not surprising that in the more general situation, derivatives can often provide an answer. In the example above

$$\lim_{x\to 0} \frac{\sin x}{1-\sqrt{1-x}}$$

we know that the derivative of sin at 0 is 1 and therefore that

$$\sin x \approx x$$

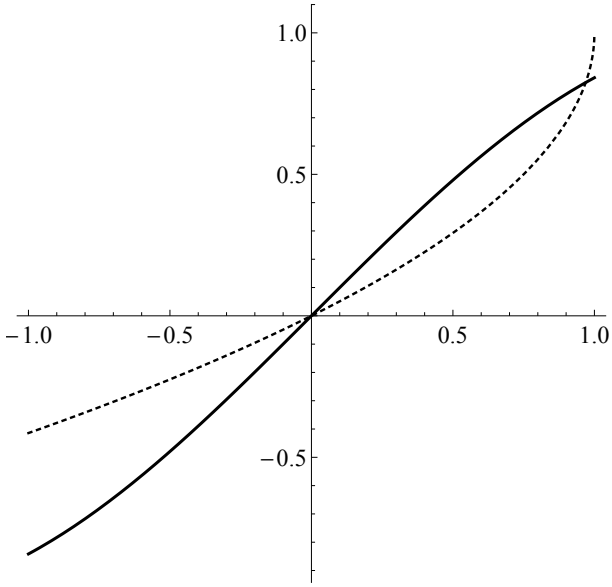
when x is close to 0. The derivative of $x \mapsto \sqrt{1-x}$ at 0 is $-1/2$. Therefore

$$\sqrt{1-x} \approx 1 - x/2$$

when x is close to 0. This in turn means that $1-\sqrt{1-x} \approx x/2$ and so it looks as though

$$\frac{\sin x}{1-\sqrt{1-x}} \approx \frac{x}{x/2} = 2.$$

The figure shows the graphs of $\sin x$ (solid) and $1-\sqrt{1-x}$ (dashed). Both pass through the origin. The slope of sin at the origin is 1. The slope of the other graph is 1/2. So close to the origin the value of the sine graph is about twice as large as that of the other graph.



We can make this argument rigorous and streamline the calculation: the upshot is a principle known as l'Hôpital's rule.

Theorem (**l'Hôpital's rule**)**.** *If $f, g : I \to \mathbf{R}$ are differentiable on the open interval I containing c and $f(c) = g(c) = 0$ then*

$$\lim_{x \to c} \frac{f(x)}{g(x)} = \lim_{x \to c} \frac{f'(x)}{g'(x)}$$

provided the second limit exists.

In the example above we take

$$f(x) = \sin x \quad \text{for which} \quad f'(x) = \cos x$$

and

$$g(x) = 1 - (1 - x)^{1/2} \quad \text{for which} \quad g'(x) = 1/2(1 - x)^{-1/2}.$$

In order to compute the limit

$$\lim_{x \to 0} \frac{\sin x}{1 - \sqrt{1 - x}}$$

we consider the ratio of the derivatives

$$\lim_{x \to 0} \frac{\cos x}{1/2(1 - x)^{-1/2}} = \lim_{x \to 0} \frac{2 \cos x}{(1 - x)^{-1/2}} = 2.$$

Therefore according to l'Hôpital

$$\lim_{x\to 0} \frac{\sin x}{1-\sqrt{1-x}} = 2.$$

The advantage of expressing the theorem in the way we did, with a limit on the right, is that it may be possible to use it repeatedly. If we find that $f'(c) = g'(c) = 0$ then we cannot calculate the limit of the derivatives just by substitution as we did with the example. But we can apply the theorem a second time to get

$$\lim_{x\to c} \frac{f''(x)}{g''(x)}$$

provided that the functions are twice differentiable.

For example, suppose we want

$$\lim_{x\to 0} \frac{1-\cos x}{x^2}.$$

The numerator and denominator both approach 0 as $x \to 0$. Differentiating both with respect to x we are led to consider the limit

$$\lim_{x\to 0} \frac{\sin x}{2x}.$$

The numerator and denominator of this fraction also approach 0 as $x \to 0$ so we can differentiate again and consider

$$\lim_{x\to 0} \frac{\cos x}{2} = \frac{1}{2}.$$

Proof (**of l'Hôpital's rule**). Suppose that

$$\lim_{x\to c} \frac{f'(x)}{g'(x)}$$

does indeed exist. Then it cannot be that $g'(x) = 0$ at a sequence of points converging to c. So there is some interval around c on which g' is non-zero (except perhaps at c itself). So g' is non-zero on an interval each side of c. This enables us to apply Cauchy's MVT. Because $f(c) = g(c) = 0$

$$\lim_{x\to c} \frac{f(x)}{g(x)} = \lim_{x\to c} \frac{f(x)-f(c)}{g(x)-g(c)}.$$

As long as x is in the region around c where $g' \neq 0$ Cauchy's MVT ensures that there is a point t (depending upon x) between c and x where

$$\frac{f(x)-f(c)}{g(x)-g(c)} = \frac{f'(t)}{g'(t)}.$$

As $x \to c$ the corresponding t is forced to approach c as well and so

$$\frac{f(x)}{g(x)} = \frac{f(x)-f(c)}{g(x)-g(c)} \to \lim_{t\to c} \frac{f'(t)}{g'(t)}.$$

□

There is also a version of l'Hôpital at infinity.

Theorem (**l'Hôpital's rule at infinity**). *If $f, g : \mathbf{R} \to \mathbf{R}$ are differentiable and*

$$\lim_{x\to\infty} f(x) = 0 \quad \textit{and} \quad \lim_{x\to\infty} g(x) = 0$$

or

$$\lim_{x\to\infty} f(x) = \infty \quad \textit{and} \quad \lim_{x\to\infty} g(x) = \infty$$

then

$$\lim_{x\to\infty} \frac{f(x)}{g(x)} = \lim_{x\to\infty} \frac{f'(x)}{g'(x)}$$

provided the second limit exists.

Example.

$$\lim_{x\to\infty} \frac{x}{e^x} = \lim_{x\to\infty} \frac{1}{e^x} = 0.$$

l'Hôpital's rule is included in this chapter because it takes the form of an extended mean value theorem. However the main topic of the section will be Taylor's Theorem.

Taylor's Theorem with remainder

At school you met Taylor expansions. You saw a statement something along the lines of

$$f(x) \approx f(a) + f'(a)(x-a) + \frac{f''(a)}{2}(x-a)^2 + \cdots .$$

For most purposes we need something a bit more precise. It isn't clear from the statement above how good the approximation is. Let us recall the example from the introduction.

$$(1+x)^{1/2} \approx 1 + \frac{x}{2} - \frac{x^2}{8}.$$

If $x = 1$ for example we can check that the error is

$$\sqrt{2} - (1 + 1/2 - 1/8) = \sqrt{2} - 11/8 = 0.0392....$$

But this depends upon knowing the value $\sqrt{2}$ which is the number we are trying to approximate. Can we estimate the error without actually calculating the thing we want to approximate? The error will depend in a complicated way on the particular function that we are approximating. So we will have to express the error in terms of the function: the aim is to find an expression which we can (usually) estimate.

The first such expression is contained in the following theorem.

Theorem (**Taylor's Theorem, Lagrange Remainder**). *If $f : I \to \mathbf{R}$ is n times differentiable on the open interval I containing a and b then*

$$\begin{aligned} f(b) &= f(a) + f'(a)(b-a) + \frac{f''(a)}{2}(b-a)^2 + \cdots \\ &\quad + \frac{f^{(n-1)}(a)}{(n-1)!}(b-a)^{n-1} + \frac{f^{(n)}(t)}{n!}(b-a)^n \end{aligned}$$

for some point t between a and b.

The number t depends upon the function f as well as upon a and b so we don't have any way of determining what it is in general. The error is given in terms of the n^{th} derivative of the function. If we can calculate this derivative we may be able to show that it can't be too big anywhere between a and b and so it won't matter that we don't know the exact value of t.

We will give two proofs of this theorem. The first is the "natural" proof. The second is a trick proof which allows us to prove several different versions of the theorem which are useful for different purposes. The first proof is an extension of our original proof for the MVT in which we modified f by a linear function. In this argument we shall modify f (to get a new function h) by a polynomial of degree n.

Proof The function g given by

$$g(x) = f(x) - \left(f(a) + f'(a)(x-a) + \cdots + \frac{f^{(n-1)}(a)}{(n-1)!}(x-a)^{n-1} \right)$$

satisfies $g(a) = 0$, $g'(a) = 0$ and so on up to $g^{(n-1)}(a) = 0$. It also satisfies $g^{(n)}(x) = f^{(n)}(x)$ for all x because f and g differ by a polynomial of degree only $n - 1$.

If we put

$$h(x) = g(x) - g(b)\frac{(x-a)^n}{(b-a)^n}$$

then h also has its first $n-1$ derivatives vanishing at a but in addition it satisfies $h(b) = 0$.

We now proceed inductively. Since $h(b) = h(a) = 0$ there is a point t_1 in (a, b) where $h'(t_1) = 0$ by Rolle's Theorem. Since $h'(t_1) = h'(a) = 0$ there is a point t_2 in (a, t_1) with $h''(t_2) = 0$. Continuing in this way we eventually get a point $t = t_n$ where $h^{(n)}(t) = 0$. In terms of g this says that

$$g^{(n)}(t) = g(b)\frac{n!}{(b-a)^n}.$$

$$g(x) = f(x) - \left(f(a) + f'(a)(x-a) + \cdots + \frac{f^{(n-1)}(a)}{(n-1)!}(x-a)^{n-1}\right)$$

so

$$g(b) = \frac{g^{(n)}(t)}{n!}(b-a)^n = \frac{f^{(n)}(t)}{n!}(b-a)^n$$

or

$$f(b) - \left(f(a) + f'(a)(b-a) + \cdots + \frac{f^{(n-1)}(a)}{(n-1)!}(b-a)^{n-1}\right) = \frac{f^{(n)}(t)}{n!}(b-a)^n$$

which is exactly the statement of the theorem. □

Example.

Let us try to estimate the error in the Taylor approximation at 0

$$(1+x)^{1/2} \approx 1 + \frac{x}{2} - \frac{x^2}{8}$$

for $x = 1/2$ say. If $f(u) = (1+u)^{1/2}$ then

$$\begin{aligned} f'(u) &= 1/2(1+u)^{-1/2} \\ f''(u) &= -1/4(1+u)^{-3/2} \\ f'''(u) &= 3/8(1+u)^{-5/2} \end{aligned}$$

so when u is zero we get

$$\begin{aligned} f(0) &= 1 \\ f'(0) &= 1/2 \\ f''(0) &= -1/4. \end{aligned}$$

The theorem tells us that

$$(1+x)^{1/2} = f(0) + f'(0)x + \frac{f''(0)}{2}x^2 + \frac{f'''(t)}{6}x^3 = 1 + \frac{x}{2} - \frac{x^2}{8} + \frac{f'''(t)}{6}x^3$$

for some t between 0 and x. Now

$$f'''(u) = 3/8(1+u)^{-5/2}$$

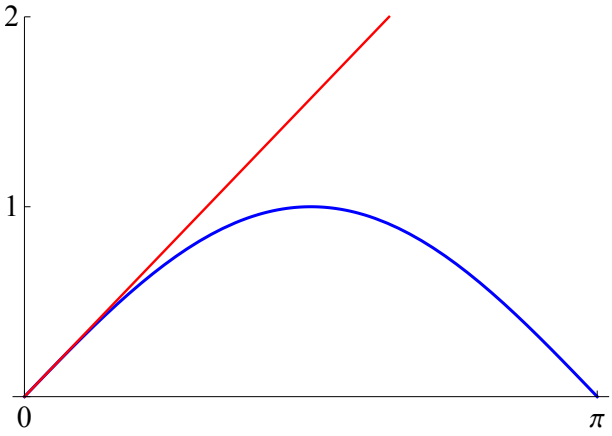
so when $x = 1/2$ the error is

$$\frac{f'''(t)}{6}x^3 = \frac{3/8(1+t)^{-5/2}}{6}\left(\frac{1}{2}\right)^3 = \frac{1}{16 \times 8(1+t)^{5/2}} = \frac{1}{128(1+t)^{5/2}}.$$

We don't know the value of t but we do know that it lies between 0 and $1/2$. So $1+t > 1$ and hence the error cannot be more than $1/128$. In fact the error is about 0.006 so our estimate $1/128 = 0.0078...$ is quite good.

Taylor's Theorem can be used to prove inequalities like the ones we have seen before: $e^x \geq 1 + x$ and so on.

Example. *If* $0 \leq x \leq \pi$ *then* $\sin x \leq x$.



Proof Let $f(x) = \sin x$. We have $f'(x) = \cos x$ and $f''(x) = -\sin x$. So

$$f(0) = 0, \quad f'(0) = 1$$

and by Taylor's Theorem with $n = 2$

$$f(x) = f(0) + f'(0)x + \frac{f''(t)}{2}x^2.$$

This says

$$\sin x = 0 + x - \frac{\sin t}{2}x^2$$

for some t between 0 and x. As long as $0 \le x \le \pi$ we have $0 < t < \pi$ and so $\sin t > 0$. □

In the Homework you showed that for $|x| < 1$

$$-\log(1-x) = x + \frac{x^2}{2} + \frac{x^3}{3} + \cdots$$

which can be rewritten as a series for the logarithm near 1:

$$\log x = (x-1) - \frac{(x-1)^2}{2} + \cdots .$$

In our next example we shall derive this using Taylor's Theorem.

Example (Taylor Series for log)**.**

Let f be the function $f : x \mapsto \log x$. Then we know

$$\begin{aligned} f'(x) &= x^{-1} \\ f''(x) &= -x^{-2} \\ f'''(x) &= 2x^{-3} \\ &\vdots \\ f^{(n)}(x) &= (-1)^{n-1}\,(n-1)!\,x^{-n}. \end{aligned}$$

Let us find the Taylor approximations at $x = 1$.

$$\begin{aligned} f(1) &= 0 \\ f'(1) &= 1 \\ f''(1) &= -1 \\ f'''(1) &= 2 \\ &\vdots \\ f^{(n)}(1) &= (-1)^{n-1}\,(n-1)!. \end{aligned}$$

The n^{th} Taylor formula at 1 is therefore

$$\begin{aligned}\log x &= 0 + (x-1) - \frac{(x-1)^2}{2} + \cdots + (-1)^{n-2}\frac{(x-1)^{n-1}}{n-1} \\ &\quad +(-1)^{n-1}t^{-n}\frac{(x-1)^n}{n} \\ &= (x-1) - \frac{(x-1)^2}{2} + \cdots + (-1)^{n-2}\frac{(x-1)^{n-1}}{n-1} \\ &\quad +(-1)^{n-1}\frac{1}{n}\left(\frac{x-1}{t}\right)^n\end{aligned}$$

where t is a number between 1 and x.

If $1 \le x \le 2$ then $0 < x-1 \le 1 \le t$. Hence the error term is at most $\frac{1}{n}$ which tends to 0 as $n \to \infty$. This means that we can take a limit and conclude that we have an infinite series for $\log x$

$$\log x = (x-1) - \frac{(x-1)^2}{2} + \frac{(x-1)^3}{3} - \cdots$$

as long as $1 \le x \le 2$. If we set $x = 2$ then the series becomes

$$1 - \frac{1}{2} + \frac{1}{3} - \cdots$$

which converges by the alternating series theorem. In the homework you will see another proof that it converges to $\log 2$ which the earlier proof didn't give.

The same argument works with a bit more care if $1/2 \le x < 1$. But if $x < 1/2$ then it might go wrong: for example if $x = 1/3$ and $t = 1/2$ then

$$\frac{x-1}{t} = -\frac{4}{3}$$

and so the powers of this number increase very fast. It turns out that the series does converge to the logarithm for $0 < x < 1/2$ but the Lagrange Remainder is the wrong tool for proving this. (To put it another way, the error term is in fact small and so the number t that actually gives the remainder will be bigger than $x-1$ but the theorem doesn't tell us this.)

Cauchy came up with a different form of the remainder that works nicely in some cases. We will deduce it from the following version of Taylor's Theorem.

Theorem (**Taylor's Theorem, Trick Proof**). *If $f : I \to \mathbf{R}$ is n times differentiable on the open interval I containing a and b and $0 \leq k \leq n-1$ then*

$$\begin{aligned} f(b) &= f(a) + f'(a)(b-a) + \cdots + \frac{f^{(n-1)}(a)}{(n-1)!}(b-a)^{n-1} \\ &\quad + \frac{f^{(n)}(t)}{(n-1)!(n-k)}(b-t)^k(b-a)^{n-k} \end{aligned}$$

for some point t between a and b.

Proof Let R be the remainder for which we are trying to find a formula

$$R = f(b) - \left(f(a) + f'(a)(b-a) + \cdots + \frac{f^{(n-1)}(a)}{(n-1)!}(b-a)^{n-1} \right).$$

This time we define

$$h(x) = f(b) - f(x) - f'(x)(b-x) - \cdots - \frac{f^{(n-1)}(x)}{(n-1)!}(b-x)^{n-1} - R\frac{(b-x)^{n-k}}{(b-a)^{n-k}}.$$

Obviously $h(b) = 0$. Also $h(a) = 0$ because of the definition of R. So by Rolle's Theorem there is a number t between a and b where $h'(t) = 0$. So

$$\begin{aligned} h'(x) &= -f'(x) + f'(x) - f''(x)(b-x) + \frac{f''(x)}{2}2(b-x) - \frac{f'''(x)}{2}(b-x)^2 \\ &\quad + \cdots - \frac{f^{(n)}(x)}{(n-1)!}(b-x)^{n-1} + (n-k)R\frac{(b-x)^{n-k-1}}{(b-a)^{n-k}} \\ &= -\frac{f^{(n)}(x)}{(n-1)!}(b-x)^{n-1} + (n-k)R\frac{(b-x)^{n-k-1}}{(b-a)^{n-k}} \end{aligned}$$

where for each term we use the product rule, and most of the sum cancels out. If we now substitute t and set the expression equal to 0 we get

$$R = \frac{f^{(n)}(t)}{(n-1)!(n-k)}(b-t)^k(b-a)^{n-k}$$

which is the statement we wanted. □

The Lagrange remainder form of Taylor's Theorem follows immediately by setting $k = 0$. The most useful other consequence of the theorem is the case $k = n-1$ which is known as Taylor's Theorem with Cauchy remainder.

Theorem (**Taylor's Theorem, Cauchy Remainder**). *If $f : I \to \mathbf{R}$ is n times differentiable on the open interval I containing a and b then*

$$\begin{aligned} f(b) &= f(a) + f'(a)(b-a) + \cdots + \frac{f^{(n-1)}(a)}{(n-1)!}(b-a)^{n-1} \\ &\quad + \frac{f^{(n)}(t)}{(n-1)!}(b-t)^{n-1}(b-a) \end{aligned}$$

for some point t between a and b.

The point is that we have replaced $(b-a)^n$ by $(b-t)^{n-1}(b-a)$ and t is closer to b than a is.

Let us return to the logarithm and try to check the error for $0 < x < 1$. We have $f(x) = \log x$ and we set $a = 1$ and $b = x$. The Taylor formula is

$$\begin{aligned} \log x &= (x-1) - \frac{(x-1)^2}{2} + \cdots + (-1)^{n-2}\frac{(x-1)^{n-1}}{n-1} \\ &\quad + (-1)^{n-1}\frac{(x-t)^{n-1}(x-1)}{t^n} \end{aligned}$$

We now have an error term which apart from $(-1)^{n-1}$ is

$$\frac{(x-t)^{n-1}(x-1)}{t^n} = \frac{x-1}{t}\left(\frac{x-t}{t}\right)^{n-1} = \frac{x-1}{t}(-1)^{n-1}\left(1-\frac{x}{t}\right)^{n-1}$$

where $0 < x < t < 1$. Since $x < t < 1$ we have that $0 < 1 - x/t < 1 - x$ and so the error tends to zero as $n \to \infty$ as fast as the exponential $(1-x)^{n-1}$. Therefore the infinite series converges for $0 < x < 1$.

Theorem (**The log series**). *If $0 < x \leq 2$*

$$\log x = (x-1) - \frac{(x-1)^2}{2} + \frac{(x-1)^3}{3} - \cdots .$$

Equivalently, if $-1 \leq t < 1$

$$-\log(1-t) = t + \frac{t^2}{2} + \frac{t^3}{3} + \cdots .$$

The second form tends to be the one we use most often. You already saw a different proof of this in the homework (except in the case $t = -1$).

Example (**Binomial Series**)**.**

We know the Binomial Theorem which allows us to expand a power

$$(1+x)^n = 1 + nx + \frac{n(n-1)}{2}x^2 + \cdots + x^n$$

if n is a non-negative whole number. We also know that

$$(1+x)^{-1} = 1 - x + x^2 - x^3 + \cdots$$

provided $-1 < x < 1$.

It is not altogether surprising that we can find a series expansion for every power

$$(1+x)^s = 1 + sx + \frac{s(s-1)}{2}x^2 + \frac{s(s-1)(s-2)}{6}x^3 + \cdots$$

provided $|x| < 1$. (The series will be infinite unless n is a non-negative integer.) In the homework you are asked to prove this (without using Taylor series).

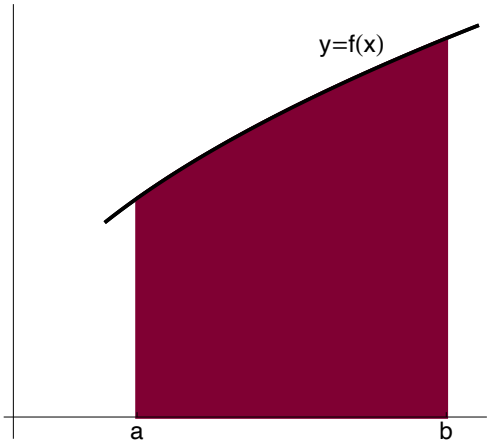
Instead you can employ a programme that we have used several times.

- You define a function by a power series.
- You check that the series converges.
- You show that the function satisfies a differential equation.
- You use the MVT to prove that the function has the properties you want.

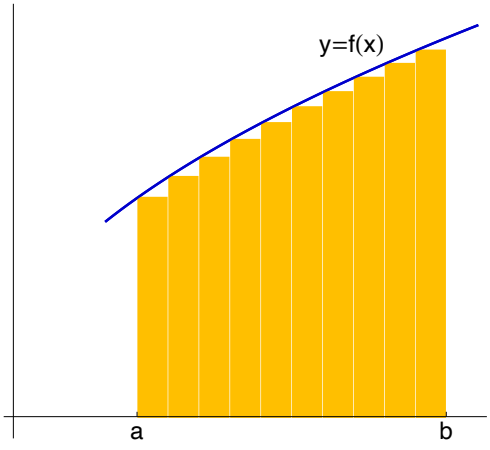
Chapter 5. The Riemann integral

The last chapter will be devoted to the construction of the integral, its relationship to derivatives and some applications.

The geometric picture is familiar to you. We have a function $f : [a, b] \to [0, \infty)$ and we want to calculate the area under the curve $y = f(x)$.



The simplest thing to try is to cut up the interval $[a, b]$ into equal pieces and place a rectangle on each piece, underneath the curve, which just touches the curve.



We calculate the total area of the rectangles and as the number of pieces increases, the total should approach the area we want. This is essentially the Newton construction.

There are a couple of problems with this approach. Firstly it isn't very easy to see that the total areas of the rectangles *do* approach a limit as the number of intervals increases. Secondly, with this definition you can find a pair of functions f and g with

$$\int (f+g) \neq \int f + \int g.$$

We shall use something that avoids this problem and is a bit more flexible.

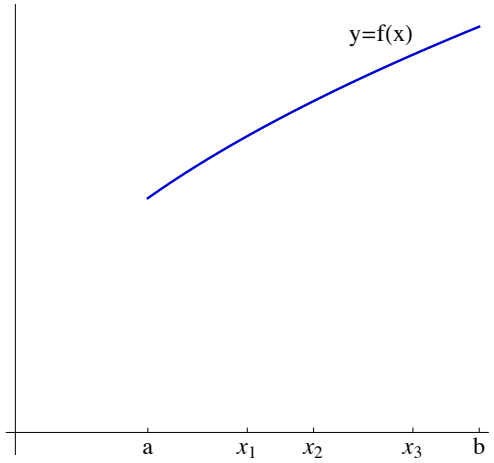
The construction

A partition P of the interval $[a,b]$ will be a finite sequence of numbers

$$a = x_0 < x_1 < \cdots < x_n = b$$

the first and last of which are the end points.

These points divide the interval into n pieces. For example if $n = 4$.



Now suppose that $f : [a,b] \to \mathbf{R}$ is a **bounded** function. For each i let

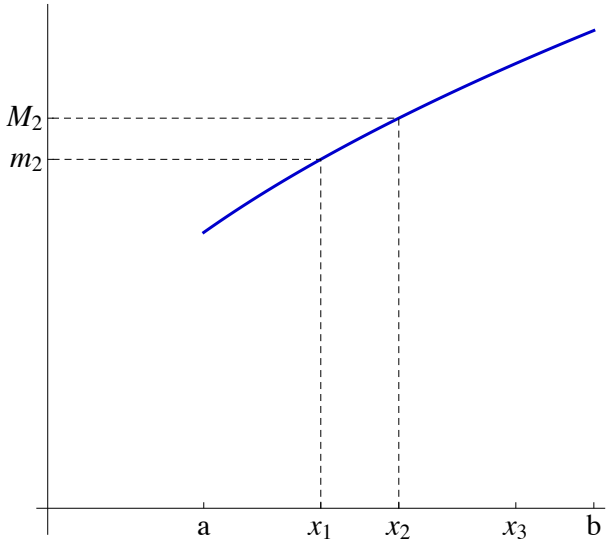
$$m_i = \inf\{f(x) : x_{i-1} \leq x \leq x_i\}$$

and

$$M_i = \sup\{f(x) : x_{i-1} \leq x \leq x_i\}$$

be the "lowest" and "highest" values that f takes on the i^{th} piece.

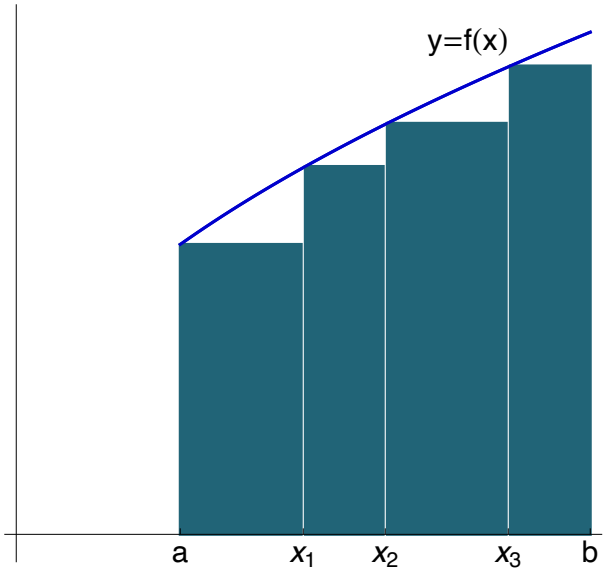
We identify the inf and sup of f on each interval.



This tells us the height of a rectangle below the curve based on the interval and the height of one above the curve.

The total area of the rectangles below the curve is

$$\sum_{1}^{n} m_i \, (x_i - x_{i-1}).$$



The total area of those above is

$$\sum_{1}^{n} M_i \, (x_i - x_{i-1}).$$

Definition (Upper and lower sums)**.** *Let $f : [a, b] \to \mathbf{R}$ be bounded and $P = \{x_0, x_1, \ldots, x_n\}$ be a partition of $[a, b]$. The upper and lower Riemann sums of the function f with respect to P are*

$$U(f, P) = \sum_{1}^{n} M_i \, (x_i - x_{i-1}) \qquad \text{and} \qquad L(f, P) = \sum_{1}^{n} m_i \, (x_i - x_{i-1})$$

respectively, where for each i

$$m_i = \inf\{f(x) : x_{i-1} \leq x \leq x_i\}$$

and

$$M_i = \sup\{f(x) : x_{i-1} \leq x \leq x_i\}.$$

Exercise Suppose $f : [a, b] \to \mathbf{R}$ is constant: say $f(x) = K$ for each x. Show that for every partition P

$$U(f, P) = L(f, P) = K(b - a).$$

This is similar to what you looked at in 'A'-level, except that now the pieces on which we build the rectangles are not necessarily of equal length and we sandwich the integral between the upper and lower sums.

We now ask "How large can we make the lower sums" and "how small can we make the upper sums"? How much can we push upwards on the function from below and downwards from above? We take the sup of the lower sums and the inf of the upper sums. To do so we need to know that the lower sums are bounded above and similarly that the upper sums are bounded below.

Lemma (**The upper sum is bigger than the lower**)**.** *Suppose* $f : [a, b] \to \mathbf{R}$ *is bounded,*

$$m = \inf\{f(x) : a \leq x \leq b\}$$

and

$$M = \sup\{f(x) : a \leq x \leq b\}.$$

Then for any partition P

$$m\,(b - a) \leq L(f, P) \leq U(f, P) \leq M\,(b - a).$$

Proof Clearly for each i, we have $m \leq m_i \leq M_i \leq M$ and so

$$m\,(b - a) \leq L(f, P) \leq U(f, P) \leq M\,(b - a).$$

□

We can now take the sup and inf.

Definition (**Upper and lower integrals**). *Let $f : [a, b] \to \mathbf{R}$ be bounded. The upper and lower Riemann integrals of the function f are*

$$\overline{\int} f = \inf_P U(f, P)$$

and

$$\underline{\int} f = \sup_P L(f, P)$$

where the sup and inf are taken over all partitions of the interval $[a, b]$.

Definition (**The Riemann integral**). *Let $f : [a, b] \to \mathbf{R}$ be bounded. Then f is said to be Riemann integrable if*

$$\overline{\int} f = \underline{\int} f$$

and in this case we write

$$\int_a^b f(x)\, dx$$

for the common value.

The point of the word "if" is that the upper and lower integrals might not be the same and in that case we don't define the integral.

Example (**A function that is not integrable**). *Let $f : [0, 1] \to \mathbf{R}$ be given by*

$$f(x) = \begin{cases} 1 & \text{if } x \in \mathbf{Q} \\ \\ 0 & \text{if } x \notin \mathbf{Q}. \end{cases}$$

Then every lower sum of f is 0 and every upper sum is 1. So f is not integrable. It is functions of this sort that cause the problem mentioned earlier $\int(f+g) \neq \int f + \int g$. By insisting that the upper and lower integrals have to be equal we prevent this problem from occurring.

By taking the sup of the lower sums and the inf of the upper sums we also avoid the problem of checking whether there is a limit. However, the definition as it stands looks disgusting. The family of all partitions is fantastically complicated. It is hugely big and has a complicated structure. How can we hope to compare the lower sum for one partition with the upper (or even the lower) sum for another? We can't unless...

Definition (**Refinements**). *If P and Q are partitions of an interval $[a,b]$ then Q is said to be a refinement of P if every point of P belongs to Q.*

So we get Q by starting with P and cutting its intervals into more pieces.

Lemma (**More refined partitions are better**). *Suppose $f : [a,b] \to \mathbf{R}$ is bounded, P and Q are partitions of $[a,b]$ and Q refines P. Then*

$$L(f,P) \leq L(f,Q) \leq U(f,Q) \leq U(f,P).$$

Proof We will just check the lower sums: the upper ones are similar. Suppose I is an interval in P and that Q breaks it into intervals $J_1, J_2, \ldots, J_m$. Then the infimum of f on I will be at most the infimum on each J_j. So the sums in $L(f,Q)$ based on the J_j have a total at least as big as the sum in $L(f,P)$. □

If P and Q are partitions of an interval $[a,b]$ then we can find a **common refinement** of P and Q by including all points of P and Q.

Lemma (**All upper sums are bigger than all lower sums**). *Suppose $f : [a,b] \to \mathbf{R}$ is bounded and P and Q are partitions of $[a,b]$.*

Then

$$L(f,P) \leq U(f,Q).$$

Proof Choose R to be a refinement of both P and Q. Then

$$L(f,P) \leq L(f,R) \leq U(f,R) \leq U(f,Q).$$

□

At this point you have enough information to compute some integrals. For example if you look at the function $x \mapsto x^2$ on $[0,1]$ you can compute upper and lower sums using equal intervals quite easily. You can check that as the number of intervals goes to infinity both sums approach $1/3$. **HW**

This means for example that

$$\underline{\int} f \geq \frac{1}{3}$$

but to check that the supremum of the lower sums over all partitions is not more than $1/3$ you need the upper bound provided by the previous lemma or the next one.

Corollary (**The upper integral is bigger than the lower**)**.** *Suppose* $f : [a,b] \to \mathbf{R}$ *is bounded. Then*

$$\underline{\int} f \leq \overline{\int} f.$$

Proof Exercise. □

We now create the crucial test that will enable us to prove that functions (such as continuous functions) are integrable.

Lemma (**The integrability condition**)**.** *Suppose* $f : [a,b] \to \mathbf{R}$ *is bounded. Then* f *is Riemann integrable if (and only if) for every* $\varepsilon > 0$ *we can find a partition* P *of* $[a,b]$ *with*

$$U(f,P) - L(f,P) < \varepsilon.$$

Proof We check the easy direction first. Suppose that the condition holds. Then for any partition P

$$L(f,P) \leq \underline{\int} f$$

and

$$\overline{\int} f \leq U(f, P).$$

Therefore

$$\overline{\int} f - \underline{\int} f \leq U(f, P) - L(f, P).$$

Since the right side can be made smaller than any positive ε the left side must be 0.

For the other direction suppose f is integrable. Then

$$\int f = \overline{\int} f = \inf_Q U(f, Q)$$

so we can choose a partition Q_1 with

$$U(f, Q_1) < \int f + \frac{\varepsilon}{2}.$$

Similarly we can choose Q_2 so that

$$L(f, Q_2) > \int f - \frac{\varepsilon}{2}.$$

Now choose P to be a common refinement of Q_1 and Q_2. It will satisfy both inequalities and hence

$$U(f, P) - L(f, P) < \varepsilon.$$

□

Note that common refinements did the work in proving this criterion.

Before proving some of the basic properties of the integral we need to return to continuous functions.

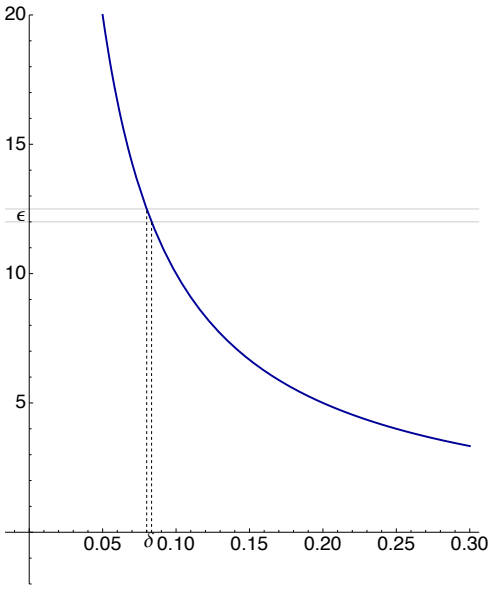
Uniform continuity

A function $f : I \to \mathbf{R}$ is continuous on the interval I if for every $c \in I$ and for every $\varepsilon > 0$, there is a $\delta > 0$ so that if $x \in I$ and $|x - c| < \delta$ then

$$|f(x) - f(c)| < \varepsilon.$$

The value of δ depends upon ϵ and also upon c. On an open interval it might be that you cannot use the same δ for every c.

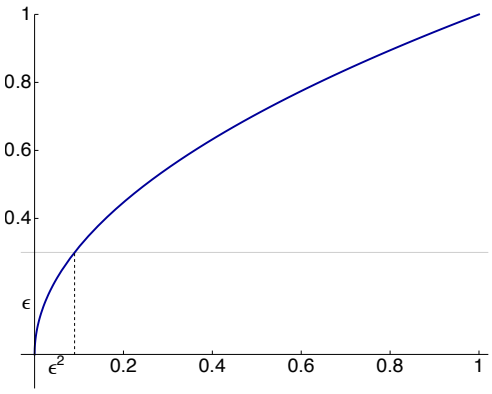
For example if $f : (0,1) \to \mathbf{R}$ is given by $f(x) = 1/x$ then if c is close to 0 you need a very small δ.



Definition (**Uniform continuity**)**.** *A function $f : I \to \mathbf{R}$ defined on an interval I is said to be* uniformly continuous *on I if for every $\varepsilon > 0$, there is a $\delta > 0$ so that if $x, y \in I$ and $|x - y| < \delta$ then*

$$|f(x) - f(y)| < \varepsilon.$$

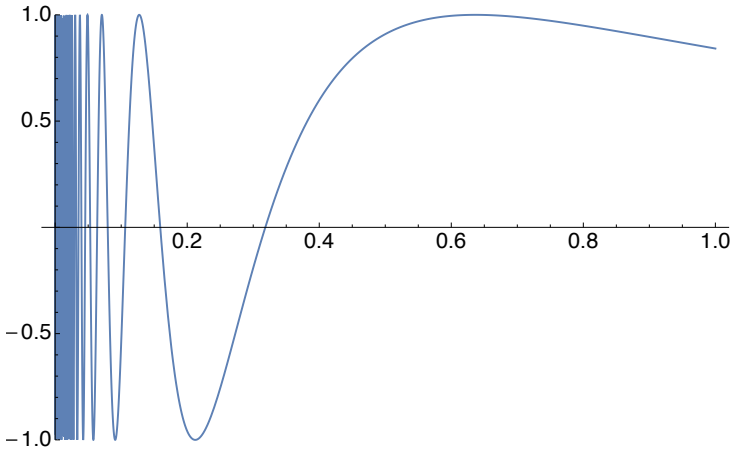
The same δ works for every pair x and y wherever they are in I. The square root function has an infinite derivative at 0 but there is a δ that works:



If $|x-y| < \varepsilon^2$ then $|\sqrt{x} - \sqrt{y}| < \varepsilon$.

Proof Exercise. □

The "perfect" example of a function on $(0,1)$ which is **not** uniformly continuous is $x \mapsto \sin 1/x$.



The function jumps a fixed amount on arbitrarily small intervals. By now you probably can guess that this cannot happen on a closed bounded interval: any continuous function on a closed bounded interval is uniformly continuous.

Theorem (**Uniform continuity**)**.** *Let $f : [a,b] \to \mathbf{R}$ be continuous. Then f is* uniformly *continuous on the interval.*

Proof Suppose that f is not uniformly continuous. Then for some ε and every n we can find points x_n and y_n with

$$|x_n - y_n| < 1/n$$

but

$$|f(x_n) - f(y_n)| \geq \varepsilon.$$

$$|x_n - y_n| < 1/n \quad \text{but} \quad |f(x_n) - f(y_n)| \geq \varepsilon.$$

Now choose a subsequence x_{n_k} which converges to x say. Since the interval $[a,b]$ is closed we must have $x \in [a,b]$. By the first condition, $y_{n_k} \to x$ as well. By continuity

$$f(x_{n_k}) \to f(x) \quad \text{and} \quad f(y_{n_k}) \to f(x).$$

But this means that

$$f(x_{n_k}) - f(y_{n_k}) \to 0.$$

The basic properties of the integral

Our first goal will be to check that continuous functions are integrable. Given such a function we want to find a partition on which the upper and lower sums are almost the same. So we want to cut into intervals on which f doesn't vary much. We have a tool that does it for us: uniform continuity.

Theorem (**Uniform continuity rewritten**). *If $f : [a, b] \to \mathbf{R}$ is continuous then for any $\varepsilon > 0$ we can find δ so that if $|x - y| < \delta$ then*

$$|f(x) - f(y)| < \varepsilon.$$

The point is that the number δ does not depend upon x or y. It works for all pairs in the interval. Now partition the interval into pieces of length less than δ. Then on any piece the function changes by less than ε.

Theorem (**Integrability of continuous functions**). *If $f : [a, b] \to \mathbf{R}$ is continuous then it is integrable.*

Proof The function is certainly bounded so we just need to show that the integrability condition is satisfied. Given $\varepsilon > 0$ choose δ so that if $x, y \in [a, b]$ satisfy

$$|x - y| < \delta$$

then

$$|f(x) - f(y)| < \frac{\varepsilon}{b - a}.$$

Now let P be a partition of the interval with each gap $x_i - x_{i-1}$ less than δ. Then for each i

$$\begin{aligned} M_i - m_i &= \sup\{f(x) : x_{i-1} \le x \le x_i\} - \inf\{f(x) : x_{i-1} \le x \le x_i\} \\ &\le \frac{\varepsilon}{b - a}. \end{aligned}$$

Therefore

$$\begin{aligned} U(f, P) - L(f, P) &= \sum_1^n (M_i - m_i)\,(x_i - x_{i-1}) \\ &\le \frac{\varepsilon}{b - a} \sum_1^n (x_i - x_{i-1}) = \varepsilon. \end{aligned}$$

□

The proof shows that for a continuous function it would be enough to consider partitions into equal pieces. If we use n equal pieces and $1/n < \delta$ then we get the estimate we want. Why do we use more general partitions? Wait and see...

Interestingly enough we can quite easily prove that monotone (increasing or decreasing) functions are integrable.

Theorem (**Integrability of monotone functions**). *If $f : [a, b] \to \mathbf{R}$ is bounded and either increasing or decreasing, then it is integrable.*

Proof Suppose f is increasing. Then for any partition, on each interval $M_i = f(x_i)$ and $m_i = f(x_{i-1})$. The maximum occurs at the right and the minimum at the left. Therefore

$$U(f, P) - L(f, P) = \sum_1^n (f(x_i) - f(x_{i-1}))\,(x_i - x_{i-1})\,.$$

Now suppose we take a partition into n equal intervals.

$$\begin{aligned} U(f, P) - L(f, P) &= \sum_1^n (f(x_i) - f(x_{i-1}))\,(x_i - x_{i-1}) \\ &= \frac{b-a}{n} \sum_1^n (f(x_i) - f(x_{i-1})) \\ &= \frac{b-a}{n}\,(f(x_n) - f(x_0)) \\ &= \frac{b-a}{n}\,(f(b) - f(a))\,. \end{aligned}$$

This approaches 0 as $n \to \infty$. □

It is beginning to look as though we could compute Riemann integrals just by looking at partitions into equal pieces. The **mesh size** of a partition is the length of the longest interval in the partition. In fact...

Theorem (**Small mesh**). *If $f : [a, b] \to \mathbf{R}$ is integrable and we pick a sequence of partitions P_n whose mesh sizes tend to 0 then*

$$U(f, P_n) \to \int_a^b f(x)\,dx$$

and similarly for the lower sums.

Seriously! Why are not just looking at partitions into equal pieces?

We would now like to check that sums and products of integrable functions are integrable and that the integral is linear

$$\int_a^b \lambda f(x)\,dx = \lambda \int_a^b f(x)\,dx$$

and

$$\int_a^b (f(x)+g(x))\,dx = \int_a^b f(x)\,dx + \int_a^b g(x)\,dx.$$

The first will be left as an exercise. If $\lambda \geq 0$ it is obvious. For negative λ you swap sup and inf so it is slightly more subtle.

The second may be trickier than you think. You would like to say that on each interval the sup for $f+g$ is the sum of the sups for f and g. In fact what we have is this.

Lemma. *For any functions f and g on an interval I*

$$\sup_I (f+g) \leq \sup_I f + \sup_I g$$

but the two may not be equal.

Proof For any x in the interval $f(x) \leq \sup f$ and $g(x) \leq \sup g$ so

$$f(x)+g(x) \leq \sup_I f + \sup_I g.$$

So

$$\sup_{x\in I} (f(x)+g(x)) \leq \sup_I f + \sup_I g.$$

To see that the two sides might not be equal consider $f : x \mapsto x$ and $g : x \mapsto 1-x$ on the interval $[0,1]$. In this case

$$\sup f = \sup g = \sup(f+g) = 1.$$

□

Theorem (**Linearity of the integral**). *If $f, g : [a, b] \to \mathbf{R}$ are integrable and $\lambda \in \mathbf{R}$ then λf and $f + g$ are integrable and*

$$\int_a^b \lambda f(x)\,dx = \lambda \int_a^b f(x)\,dx$$

and

$$\int_a^b (f(x) + g(x))\,dx = \int_a^b f(x)\,dx + \int_a^b g(x)\,dx.$$

Proof We shall prove the second and leave the first as an exercise. By the lemma we immediately get that for any bounded $f, g : [a, b] \to \mathbf{R}$ and any partition P

$$U(f + g, P) \leq U(f, P) + U(g, P)$$

and the reverse for the lower sums. We would like to prove that

$$\overline{\int}(f + g) \leq \overline{\int} f + \overline{\int} g = \underline{\int} f + \underline{\int} g \leq \underline{\int}(f + g)$$

since that would force the upper and lower integrals of $f + g$ to be equal, and equal to $\int f + \int g$.

Given $\varepsilon > 0$ choose P and Q so that

$$U(f, P) \leq \overline{\int} f + \frac{\varepsilon}{2}$$

and

$$U(g, Q) \leq \overline{\int} g + \frac{\varepsilon}{2}.$$

Now if R is a common refinement of P and Q we have

$$\begin{aligned} \overline{\int}(f + g) \leq\ U(f + g, R) &\leq U(f, R) + U(g, R) \\ &\leq U(f, P) + U(g, Q) \\ &\leq \overline{\int} f + \overline{\int} g + \varepsilon. \end{aligned}$$

Since $\varepsilon > 0$ was arbitrary we have

$$\overline{\int}(f + g) \leq \overline{\int} f + \overline{\int} g.$$

□

Another important property of the integral is monotonicity: if $f \leq g$ at each point then $\int f \leq \int g$. This is much easier than additivity and is left as an exercise.

Theorem (**Monotonicity of the integral**). *If $f, g : [a, b] \to \mathbf{R}$ are integrable and for all $x \in [a, b]$ we have $f(x) \leq g(x)$ then and*

$$\int_a^b f(x)\,dx \leq \int_a^b g(x)\,dx.$$

Checking that products work is also a bit trickier than you might think. If you can show that f^2 is integrable then you are home because

$$fg = \frac{1}{2}\left((f+g)^2 - f^2 - g^2\right).$$

Since f^2 *is* a product of two integrable functions we had better be able to prove it integrable. This is not very different from proving that $\phi \circ f$ is integrable for an arbitrary continuous function ϕ.

Theorem (**Continuous function of an integrable function**). *If $f : [a, b] \to \mathbf{R}$ is integrable and $\phi : \mathbf{R} \to \mathbf{R}$ is continuous then $\phi \circ f$ is integrable on $[a, b]$.*

Proof f is a bounded so ϕ restricted to the image $f([a, b])$ is bounded and uniformly continuous. Suppose $|\phi(u)| \leq K$ if $u \in f([a, b])$ and given $\varepsilon > 0$ choose δ so that if $u, v \in f([a, b])$ and $|u - v| < \delta$ then

$$|\phi(u) - \phi(v)| < \varepsilon.$$

Given $\eta > 0$ depending upon ε, δ and K we can choose a partition $P = \{x_0, x_1, \ldots, x_n\}$ of $[a, b]$ for which

$$U(f, P) - L(f, P) < \eta.$$

On most intervals of this partition the function f will not change by more than δ, so $\phi \circ f$ won't change by more than ε. On the other intervals $\phi \circ f$ won't change by more than $2K$ whatever f does.

For each i consider M_i and m_i, the sup and inf of f on the interval $[x_{i-1}, x_i]$. Let B be the set of indices for which $M_i - m_i \geq \delta$ and G be the set of the other indices. Then

$$\begin{aligned} \eta &> \sum_1^n (M_i - m_i)(x_i - x_{i-1}) \\ &\geq \sum_{i \in B} (M_i - m_i)(x_i - x_{i-1}) \\ &\geq \delta \sum_{i \in B} (x_i - x_{i-1}). \end{aligned}$$

The total length of the bad intervals is at most η/δ.

For each i let N_i and n_i be the sup and inf of $\phi \circ f$ on the i^{th} interval. Now

$U(\phi \circ f, P) - L(\phi \circ f, P)$

$$\begin{aligned} &= \sum_1^n (N_i - n_i)(x_i - x_{i-1}) \\ &= \sum_{i \in G} (N_i - n_i)(x_i - x_{i-1}) + \sum_{i \in B} (N_i - n_i)(x_i - x_{i-1}) \\ &\leq \varepsilon \sum_{i \in G} (x_i - x_{i-1}) + 2K \sum_{i \in B} (x_i - x_{i-1}) \\ &\leq \varepsilon(b - a) + 2K \frac{\eta}{\delta}. \end{aligned}$$

If we choose $\eta = \delta\varepsilon/(2K)$ we get

$$U(\phi \circ f, P) - L(\phi \circ f, P) \leq \varepsilon(b - a + 1)$$

and this can be made as small as we wish by choosing ε small.

So the integrability condition is satisfied and $\phi \circ f$ is integrable. □

Corollary (**The integrability of products**). *If $f, g : [a, b] \to \mathbf{R}$ are integrable then $f.g$ is integrable on $[a, b]$.*

Of course we have no formula for the integral of a product in terms of the integrals of the factors.

An important consequence of what we have done so far is a version of the triangle inequality for integrals rather than sums.

Corollary (**The triangle inequality**). *If $f : [a, b] \to \mathbf{R}$ is integrable then $|f|$ is integrable and*

$$\left| \int_a^b f(x)\, dx \right| \leq \int_a^b |f(x)|\, dx.$$

The only real issue is to prove that $|f|$ is integrable but we can take $\phi(x) = |x|$ in the "Continuous function of an integrable function" theorem. Once we have that then because $f \leq |f|$ and also $-f \leq |f|$ the montonicity theorem gives

$$\int_a^b f(x)\, dx \leq \int_a^b |f(x)|\, dx \qquad \text{and} \qquad -\int_a^b f(x)\, dx \leq \int_a^b |f(x)|\, dx$$

which gives the inequality.

We have one principle for integrals that has no analogue for derivatives. This is really the first place we get a payoff from having arbitrary partitions.

Corollary (**The addition of ranges**). *Let $f : [a, c] \to \mathbf{R}$ be bounded and $a < b < c$. Then f is integrable on $[a, c]$ if and only if it is integrable on $[a, b]$ and $[b, c]$ and if so*

$$\int_a^c f(x)\, dx = \int_a^b f(x)\, dx + \int_b^c f(x)\, dx.$$

Proof (Sketch) Given $\varepsilon > 0$ we can choose a partition P of $[a, c]$ with

$$U(f, P) - L(f, P) < \varepsilon.$$

We can now refine P by including the point b and then regard the new partition as a partition of $[a, b]$ together with a partition of $[b, c]$. The upper sums for these partitions will add up to the upper sum for the refinement (and similarly the lower sums).

On the other hand if we start with partitions of $[a, b]$ and $[b, c]$ we can put them together to form a partition of $[a, c]$. □

If we were restricting ourselves to partitions into equal intervals then we would find things very irritating if for example $a = 0$, $b = 1$ and $c = \sqrt{2}$. Because of the irrationality of $\sqrt{2}$ it isn't possible to find partitions of the two subintervals into equal parts that give equal parts for the whole interval. For almost all other issues it wouldn't make much difference if we used equal parts. But note that we are using the sup and inf of upper and lower sums rather than asking about a *limit* as the number of pieces goes to ∞. Once you use

sup and inf you need the common refinement idea and at that point having equal intervals isn't much simpler than having arbitrary ones.

We now come to the machinery that made it possible to integrate some functions easily. As you know there are many standard functions whose integrals cannot be written as standard functions

$$\int \frac{1}{\log x}\,dx, \qquad \int e^{-x^2}\,dx, \qquad \int \frac{1}{\sqrt{1-\alpha^2\sin^2\theta}}\,d\theta.$$

At school: "you can differentiate everything but integrate almost nothing".
At university: "you can differentiate almost nothing and integrate a lot."

But you can differentiate the standard functions and write down the derivatives whereas you can't write down the integrals of the standard functions.

The Fundamental Theorem of Calculus

We want to show that integration and differentiation are "opposites" of one another so that we can calculate integrals by un-differentiating.

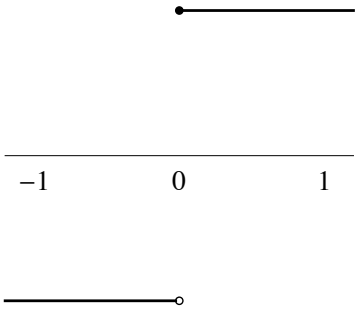
Can we prove that if F is differentiable, then F' is integrable and gives F? No, because the derivative might be unbounded. **HW**

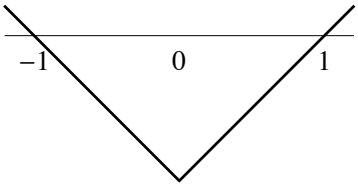
Can we prove that if F is differentiable with bounded derivative, then F' is integrable and gives F? No, for subtle reasons.

Can we prove that if f is integrable then

$$x \mapsto \int_a^x f(t)\,dt$$

is differentiable with derivative f. No. If $f : [-1,1] \to \mathbf{R}$ is -1 on $[-1,0)$ and 1 on $[0,1]$ then the integral is $x \mapsto |x| - 1$.





Theorem (**The Fundamental Theorem of Calculus I**). *Let $f : [a, b] \to \mathbf{R}$ be integrable. Then the function*

$$F : x \mapsto \int_a^x f(t)\,dt$$

is continuous. If f (the integrand) is continuous at a point $u \in (a, b)$ then F is differentiable at that point and

$$F'(u) = f(u).$$

Proof If x and $x + h$ are in $[a, b]$ with $h > 0$ then

$$F(x + h) - F(x) = \int_x^{x+h} f(t)\,dt.$$

Let M be a bound for f: so $|f(x)| \leq M$ for all $x \in [a, b]$. Then

$$|F(x + h) - F(x)| = \left| \int_x^{x+h} f(t)\,dt \right| \leq \int_x^{x+h} |f(t)|\,dt \leq Mh.$$

Similarly for $|F(x - h) - F(x)|$. So F is continuous.

Now if f is continuous at u then given $\varepsilon > 0$ choose $\delta > 0$ so that if $|t - u| < \delta$ we have $|f(t) - f(u)| < \varepsilon$. Then if $0 < h < \delta$

$$\begin{aligned}
\left| \frac{F(u + h) - F(u)}{h} - f(u) \right| &= \left| \frac{1}{h} \int_u^{u+h} f(t)\,dt - \frac{1}{h} \int_u^{u+h} f(u)\,dt \right| \\
&= \left| \frac{1}{h} \int_u^{u+h} (f(t) - f(u))\,dt \right| \\
&\leq \frac{1}{h} \int_u^{u+h} |f(t) - f(u)|\,dt \\
&\leq \frac{1}{h} \int_u^{u+h} \varepsilon\,dt = \varepsilon.
\end{aligned}$$

Similarly for $h < 0$. Hence $F'(u)$ exists and equals $f(u)$. □

Now suppose we have a nice function like $f(t) = t^2$. We know that

$$F(x) = \int_0^x t^2 \, dt$$

satisfies $F'(u) = u^2$ for all u and $F(0) = 0$. By the MVT we know that F is of the form $F(x) = x^3/3 + C$ and then by the second condition we get $F(x) = x^3/3$.

More generally FTC I and MVT show that if F is differentiable and $F' = f$ is continuous then

$$\int_a^b f(t) \, dt = F(b) - F(a).$$

Actually something a bit stronger is true. We rarely need it but it is very instructive.

Theorem (**The Fundamental Theorem of Calculus II**)**.** *Let $F : [a, b] \to \mathbf{R}$ be continuous on $[a, b]$ and differentiable on (a, b) with $F' = f$. Then if f is* **Riemann integrable** *we have*

$$\int_a^b f(t) \, dt = F(b) - F(a).$$

In other words we don't need to assume that f is continuous, merely that it is integrable. Earlier I remarked that we do need some condition. Strictly speaking I need f to be defined on the closed interval $[a, b]$ to make sense of the hypothesis but we can extend f by defining it to be 0 (say) at a and b without affecting the value of the integral. See HW 10 Q 7.

Proof It suffices to show that for each partition P

$$L(f, P) \leq F(b) - F(a) \leq U(f, P)$$

because we then get

$$\underline{\int} f \leq F(b) - F(a) \leq \overline{\int} f$$

and if f is integrable the upper and lower integrals are equal. This is where we use the integrability of f.

Let $P = \{x_0, x_1, \ldots, x_n\}$ be a partition of $[a, b]$. For each i, the function F is continuous on $[x_{i-1}, x_i]$ and differentiable on (x_{i-1}, x_i) so by the MVT there is a point $c_i \in (x_{i-1}, x_i)$ with

$$F(x_i) - F(x_{i-1}) = F'(c_i)(x_i - x_{i-1}) = f(c_i)(x_i - x_{i-1}).$$

Hence for each i

$$m_i(x_i - x_{i-1}) \leq F(x_i) - F(x_{i-1}) \leq M_i(x_i - x_{i-1}).$$

Summing over all i gives

$$L(f, P) \leq F(b) - F(a) \leq U(f, P).$$

□

Note that the proof is actually shorter than the one for continuous functions but the one for continuous functions tells us more: that integrals of continuous functions **are** in fact differentiable.

Note also that both arguments (continuous using FTC I and integrable using FTC II) use the sledge hammer: MVT. This is the theorem that relates derivatives to the values of the function. It's not surprising you need it to prove

$$\int_a^b F'(t)\, dt = F(b) - F(a).$$

We can now integrate the usual functions: polynomials, exponentials and rational functions in partial fractions. In particular we have

$$\log x = \int_1^x \frac{1}{t}\, dt.$$

We can also check the integrated versions of the product and chain rules: integration by parts and integration by substitution.

Theorem (**Integration by parts**). *Suppose $f, g : [a, b] \to \mathbf{R}$ are differentiable on an open interval including $[a, b]$ and that f' and g' are integrable on $[a, b]$. Then*

$$\int_a^b f(x)g'(x)\, dx = f(b)g(b) - f(a)g(a) - \int_a^b f'(x)g(x)\, dx.$$

Proof By the product rule $(fg)' = f'g + fg'$ and each term is integrable. So

$$\int_a^b f(x)g'(x)\,dx + \int_a^b f'(x)g(x)\,dx = \int_a^b (fg)'(x)\,dx = f(b)g(b) - f(a)g(a).$$

□

In order to state the natural form of integration by substitution we need to adopt a convention. If $b > a$ then

$$\int_b^a f = -\int_a^b f.$$

This fits with what we already proved because by the addition of ranges formula

$$\int_b^a f + \int_a^b f = \int_b^b f = 0.$$

The FTC works fine with this convention.

Theorem (**Integration by substitution**)**.** *Suppose $u : [a,b] \to \mathbf{R}$ is differentiable on an open interval including $[a,b]$ and that u' is integrable on $[a,b]$. Suppose that f is a continuous function on the bounded set $u([a,b])$. Then*

$$\int_a^b f(u(x))u'(x)\,dx = \int_{u(a)}^{u(b)} f(t)\,dt.$$

Proof For each x in $u([a,b])$ define

$$F(x) = \int_{u(a)}^{x} f(t)\,dt.$$

By FTC I we know that F is differentiable and $F'(x) = f(x)$ for each x. By the chain rule we have

$$\frac{d}{dx}F(u(x)) = F'(u(x))u'(x) = f(u(x))u'(x).$$

The function $f \circ u$ is continuous and u' is integrable so this derivative is integrable and by FTC II we have

$$\begin{aligned}
\int_a^b f(u(x))u'(x)\,dx &= F(u(b)) - F(u(a)) \\
&= \int_{u(a)}^{u(b)} f(t)\,dt - \int_{u(a)}^{u(a)} f(t)\,dt \\
&= \int_{u(a)}^{u(b)} f(t)\,dt.
\end{aligned}$$

□

At school you used integration by substitution to evaluate things like

$$\int_0^x \frac{1}{\sqrt{1-t^2}}\,dt$$

and integration by parts to evaluate

$$\int_1^x t\,\log t\,dt.$$

But this is not the main mathematical value of these theorems: after all, you could just look up the integrals.

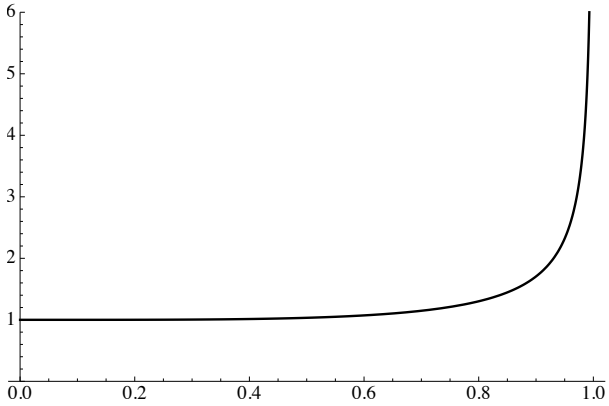
Often, substitution enables us to rewrite integrals that we *cannot* evaluate, in more useful forms. We shall have one example.

Suppose we want to estimate

$$\int_0^1 \frac{1}{\sqrt{1-x^4}}\,dx.$$

This isn't actually Riemann integrable because the function isn't bounded but we shall explain how to get around that in the next section.

We might like to use the trapezium rule to estimate the integral. But the function is unbounded!



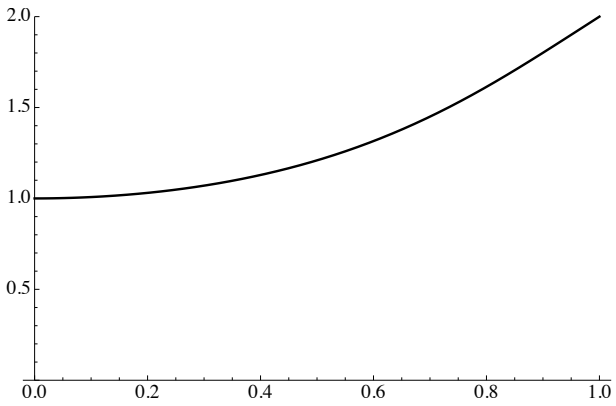
Make the substitution $x = 1 - u^2$.

$$\begin{aligned}\int_0^1 \frac{1}{\sqrt{1-x^4}}\,dx &= \int_0^1 \frac{1}{\sqrt{1-(1-u^2)^4}}2u\,du \\ &= \int_0^1 \frac{1}{\sqrt{4u^2-6u^4+4u^6-u^8}}2u\,du \\ &= \int_0^1 \frac{2}{\sqrt{4-6u^2+4u^4-u^6}}\,du.\end{aligned}$$

We want to estimate the integral

$$\int_0^1 \frac{2}{\sqrt{4-6u^2+4u^4-u^6}}\,du.$$

The integrand is now well-behaved



Thus we have used substitution on an integral that we cannot evaluate exactly, to put it into a more tractable form.

Improper integrals

How do we handle integrals like

$$\int_0^1 \frac{1}{\sqrt{x}}\,dx \qquad \text{or} \qquad \int_1^\infty \frac{1}{x^2}\,dx$$

in which the integrand or the range of integration is unbounded?

We take a limit.

Definition (**Improper integrals**). *Suppose $f : [a, b] \to \mathbf{R}$ is integrable on each subinterval $[c, b]$. We say that f is improperly Riemann integrable on $[a, b]$ if*

$$\lim_{c \to a^+} \int_c^b f(x)\, dx$$

exists and in that case we call the limit

$$\int_a^b f(x)\, dx$$

and say that the latter improper integral converges.

We do the same at the top end of an interval and the same for a half infinite interval:

Definition (**Improper integrals**). *Suppose $f : [a, \infty) \to \mathbf{R}$ is integrable on each subinterval $[a, b]$. We say that f is improperly Riemann integrable on $[a, \infty)$ if*

$$\lim_{b \to \infty} \int_a^b f(x)\, dx$$

exists and in that case we call the limit

$$\int_a^\infty f(x)\, dx$$

and say that the latter improper integral converges.

Warning We define improper integrals like

$$\int_{-\infty}^\infty f(x)\, dx$$

as

$$\int_{-\infty}^0 f(x)\, dx + \int_0^\infty f(x)\, dx.$$

In other words we insist that each half converges separately. We want to avoid things like

$$\int_{-K}^K x\, dx = 0$$

leading to a "cancellation of infinities".

Example (Integrals of powers). *If* $p > -1$ *and* $x > 0$

$$\int_0^x t^p \, dt = \frac{x^{p+1}}{p+1}.$$

If $p < -1$ *and* $x > 0$

$$\int_x^\infty t^p \, dt = -\frac{x^{p+1}}{p+1}.$$

Proof Let's do the first.

$$\lim_{c \to 0^+} \int_c^x t^p \, dt = \lim_{c \to 0^+} \frac{x^{p+1} - c^{p+1}}{p+1} = \frac{x^{p+1}}{p+1}.$$

□

Several of our previous theorems do not hold for improper integrals. $x \mapsto 1/\sqrt{x}$ is improperly integrable on $[0, 1]$ but its square $1/x$ is not.
It is possible for

$$\int_1^\infty f(x) \, dx$$

to converge but not

$$\int_1^\infty |f(x)| \, dx$$

in the same way as a convergent series may not be absolutely convergent. **HW**

We have an analogue of the comparison test.

Theorem (Comparison test for improper integrals). *Suppose* $f, g : [a, \infty) \to \mathbf{R}$ *are integrable on each interval* $[a, b]$*, that* $|f(x)| \leq g(x)$ *for all* $x \geq a$ *and that*

$$\int_a^\infty g(x) \, dx$$

converges. Then

$$\int_a^\infty f(x) \, dx$$

converges.

A similar statement works for improper integrals as one of the limits of integration approaches a number (rather than ∞).

Proof The functions $|f|$ and $f + |f|$ are integrable on each interval $[a, b]$. For each b

$$\int_a^b |f(x)|\,dx \le \int_a^b g(x)\,dx \le \int_a^\infty g(x)\,dx.$$

and

$$\int_a^b (f(x) + |f(x)|)\,dx \le 2\int_a^b g(x)\,dx \le 2\int_a^\infty g(x)\,dx.$$

Both functions $|f|$ and $f + |f|$ are non-negative so the functions

$$b \mapsto \int_a^b |f(x)|\,dx$$

and

$$b \mapsto \int_a^b (f(x) + |f(x)|)\,dx$$

are bounded increasing functions of b. So both have limits as $b \to \infty$ and hence so does their difference

$$b \mapsto \int_a^b f(x)\,dx.$$

□

Example. *For each* $\lambda > 0$

$$\int_0^\infty e^{-\lambda x}\,dx = \frac{1}{\lambda}.$$

Proof Exercise. □

Example.

$$\int_0^\infty e^{-x}\,x^p\,dx$$

converges for all $p > -1$.

Proof Remember we must handle the two ends separately (at least if $p < 0$ so that x^p is unbounded). For the left hand end,

$$\int_0^1 e^{-x}\, x^p\, dx$$

converges for all $p > -1$ because the function is dominated by x^p.

For the right hand end

$$\int_1^\infty e^{-x}\, x^p\, dx,$$

there is a constant K for which

$$e^{-x}x^p \leq Ke^{-x/2}$$

because

$$e^{-x/2}x^p \to 0$$

as $x \to \infty$. We already saw that

$$\int_1^\infty e^{-x/2}\, dx$$

converges. □

The Gamma function

In the previous section we saw that we can make sense of the integrals

$$\int_0^\infty e^{-x}\, x^{s-1}\, dx$$

for every $s > 0$. It is easy to show by integration by parts that for each non-negative integer n

$$\int_0^\infty e^{-x}\, x^n\, dx = n!$$

This will be **HW**. If we define

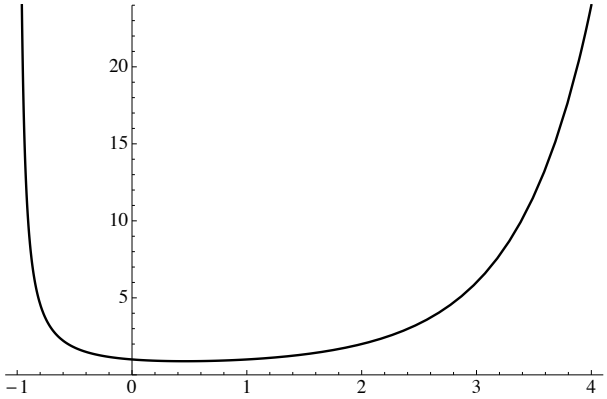
$$\Gamma(s) = \int_0^\infty e^{-x}\, x^{s-1}\, dx$$

for $s > 0$ we then have

$$\Gamma(n+1) = n!$$

for each non-negative integer n.

We thus have an extension of the factorial function to the interval $(-1, \infty)$.



So $\Gamma(s)$ is like $(s-1)!$ It is quite easy to check that for all s,

$$\Gamma(s+1) = s\,\Gamma(s).$$

There is an alternative representation found by Gauss which is in many ways the most natural form.

Theorem (**Gauss' definition of the gamma function**)**.** *For each* $s > 0$

$$\Gamma(s) = \lim_{n\to\infty} \frac{n^s\, n!}{s(s+1)\cdots(s+n)}.$$

Proof (outline) We want to show that

$$\frac{\Gamma(s)s(s+1)\cdots(s+n)}{n!\, n^s} \to 1$$

as $n \to \infty$. This says

$$\frac{\Gamma(s+n+1)}{n!\, n^s} \to 1$$

or in other words that

$$\frac{1}{n!\, n^s} \int_0^\infty e^{-x}\, x^{s+n}\, dx \to 1.$$

We want that

$$\int_0^\infty \frac{e^{-x}\,x^n}{n!}\,\left(\frac{x}{n}\right)^s\,dx \to 1.$$

We make the substitution $x = nu$ and ask

$$\int_0^\infty \frac{n^{n+1}e^{-nu}u^n}{n!}\,u^s\,du \to 1?$$

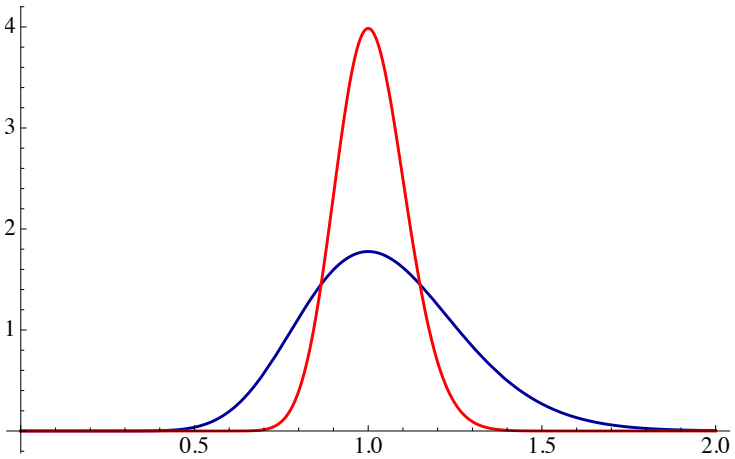
or

$$\int_0^\infty f_n(u)u^s\,du \to 1?$$

where

$$f_n(u) = \frac{n^{n+1}e^{-nu}u^n}{n!}.$$

The integral of f_n is 1. What does the function look like?



The graph shows f_{20} and f_{100}. □

Appendix. The radius of convergence formula

In this last section we shall produce a formula for the radius of convergence of a general power series. In order to do so we shall need to recall the definition of the lim sup. If (x_n)

is a sequence of real numbers which is bounded above we can examine the suprema of the tails

$$u_m = \sup \{x_n : n \geq m\}.$$

The further out the starting point m the fewer terms we are looking at. So these suprema decrease. So either they converge to some number L or they decrease to $-\infty$. In the first case we say that

$$\limsup x_n = L$$

and in the second that

$$\limsup x_n = -\infty.$$

If the sequence is *not* bounded above we write

$$\limsup x_n = \infty.$$

Whatever the value of the lim sup it can be characterised in the following way. If $\limsup x_n = L$ then for every $t < L$ there are infinitely many terms of the sequence above t, while if $t > L$ there is some point beyond which all terms are below t. Thus the lim sup tells you how high the sequence reaches over and over again.

Now suppose that we have a power series $\sum a_n x^n$ and consider the sequence whose terms are $|a_n|^{1/n}$. Let its lim sup be L and note that L is non-negative (or ∞) because the terms are non-negative. Suppose $|x| > 1/L$ and hence $1/|x| < L$. Then for infinitely many values of n

$$|a_n|^{1/n} > 1/|x|$$

and hence

$$|a_n| > 1/|x|^n.$$

So for infinitely many values of n the terms of the power series satisfy $|a_n x^n| > 1$. So the terms do not tend to zero and the series diverges.

On the other hand suppose $|x| < 1/L$. Choose a number t with $|x| < 1/t < 1/L$. Since $t > L$ we know that from some point onward

$$|a_n|^{1/n} < t$$

and hence

$$|a_n| < t^n.$$

From this point on the terms of the series satisfy $|a_n x^n| < (t|x|)^n$. We chose t in such a way that $t|x| < 1$ so the power series has terms which are dominated by a convergent geometric series and hence it converges.

The upshot is that the series diverges if $|x| > 1/L$ and converges if $|x| < 1/L$. So the radius of convergence is $1/L$. The formula we get is sometimes called the Hadamard formula.

Theorem (**The radius of convergence formula**). *The power series $\sum a_n x^n$ has radius of convergence*

$$\frac{1}{\limsup_{n\to\infty} |a_n|^{1/n}}.$$

Example. *The series $\sum 2^n x^n$ has radius of convergence $1/2$.*

Example. *The series $\sum x^{k^2}$ has radius of convergence 1.*

Proof The coefficients are given by

$$a_n = \begin{cases} 1 & \text{if } n \text{ is a square} \\ \\ 0 & \text{if not.} \end{cases}$$

Then

$$|a_n|^{1/n} = \begin{cases} 1 & \text{if } n \text{ is a square} \\ \\ 0 & \text{if not.} \end{cases}$$

So

$$\limsup_{n\to\infty} |a_n|^{1/n} = 1.$$

□

It is worth noting that the Power series I Lemma showing that $\sum n a_n x^{n-1}$ has the same radius of convergence as $\sum a_n x^n$ follows easily from the theorem above. As an exercise, check it by showing that

$$\limsup_{n\to\infty} |a_n|^{1/n} = \limsup_{n\to\infty} |n a_n|^{1/n}.$$

Index

Cauchy mean value theorem, 51
Cauchy remainder, 62
chain rule, 26
cosine, 44

derivative, 23
    chain rule, 26
    of exponential, 38
    of inverses, 34
    of logarithm, 39
    of power series, 37
    product rule, 25

exponential, 8
    derivative, 38
    inequalities, 12
    property, 10, 39

fundamental theorem of calculus, 82, 83

geometric series, 5

integral
    improper, 88
    Riemann, 64, 68
    triangle inequality, 80
    uniform continuity, 74
integration by parts, 84
integration by substitution, 85

l'Hôpital's rule, 53
Lagrange remainder, 56
limit
    one sided, 20
limits
    algebra of, 19
linearisation, 27

logarithm, 14

mean value theorem, 28
    Cauchy, 51

power series, 3
    derivative, 37
    product of, 11
powers, 14
product rule, 25

radius of convergence, 3, 95
Riemann integral, 64, 68
Rolle theorem, 30

series
    geometric, 5
    radius of convergence, 3
sine, 44

tangent, 49
Taylor theorem
    with remainder, 56, 62
theorem
    Cauchy mean value, 51
    fundamental th. of calculus, 82, 83
    l'Hôpital, 53
    mean value, 28
    Rolle, 30
    Taylor with remainder, 56, 62
trigonometric functions, 44

uniform continuity, 71