MA143 Calculus 2

Term 2

Lecturer (2022-2023): Vassili Gelfreich

v.gelfreich@warwick.ac.uk

Version: 17:06, Monday 6th March, 2023

Contents

1 Limits and the derivative **2**
1.1 Limits . . . 3
1.2 The derivative . . . 9
1.3 The derivative machine . . . 10
1.4 The Mean Value Theorem . . . 13
1.5 Derivatives of inverses . . . 17
1.6 Cauchy's Mean Value Theorem . . . 19
1.7 L'Hôpital's rule . . . 20

2 Taylor's Theorem with remainder **24**

3 Power series **32**
3.1 Basic properties of power series . . . 32
3.2 The continuity of power series . . . 36
3.3 The differentiability of power series . . . 37
3.4 Taylor series . . . 40
3.5 The exponential . . . 41
3.6 The logarithm and powers . . . 44
3.7 The trigonometric functions . . . 46
3.8 Binomial series . . . 54
3.9 Example: a function which isn't the sum of its Taylor series . . . 56

4 The Riemann integral **59**
4.1 Motivation of the definition . . . 59
4.2 Riemann sums . . . 61
4.3 The construction of the Riemann integral . . . 64
4.4 Integrable functions . . . 68
4.5 Riemann integral and mesh size of partitions . . . 70
4.6 The basic properties of the integral . . . 71
4.7 The Fundamental Theorem of Calculus . . . 72
4.8 Methods of integration . . . 76
4.9 Improper integrals . . . 78

These notes are based on Analysis II lecture notes written by Prof.K.Ball.

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.

1 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 $h \to 0$. If the quotient approaches a limit, we define this to be the slope of the curve at $(x, f(x))$ or the derivative $f'(x)$.

1.1 Limits

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.

Definition 1.1 (**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 when x is close to c. Note that 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 1.2. *Let $f : \mathbb{R} \to \mathbb{R}$ 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 1.3.

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

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 1.4 (**Limits and continuity**). *If $f : I \to \mathbb{R}$ 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 will take for granted the following theorems.

Theorem 1.5 (**Continuous and sequential limits**). *If $f : I \backslash \{c\} \to \mathbb{R}$ 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. Exercise □

This theorem also provides a tool for checking that some limits do not exists. Consider the following examples.

Example 1.6. *Consider the function*

$$f(x) = \begin{cases} 1, & x \geq 0 \\ -1, & x < 0 \end{cases}$$

We claim that the limit of $f(x)$ as $x \to 0$ does not exists.

Proof. Indeed, suppose that there is a number L such that $f(x) \to L$ as $x \to 0$. Consider the sequence defined by $x_n = (-1)^n/n$. Since $x_n \to 0$, Theorem 1.5 implies that

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

On the other hand, the sequence $f(x_n) = (-1)^n$ diverges. The contradiction implies that L does not exists. □

Example 1.7. *Consider the function* $f(x) = \sin\frac{1}{x}$ *defined for all* $x \neq 0$. *We claim that the limit of* $f(x)$ *as* $x \to 0$ *does not exists.*

Proof. Indeed, suppose that there is a number L such that $f(x) \to L$ as $x \to 0$. Consider the sequence defined by $x_n = \frac{1}{\pi(n+1/2)}$. Since $x_n \to 0$, Theorem 1.5 implies that

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

On the other hand, the sequence $f(x_n) = \sin(\pi(n+1/2)) = (-1)^n$ diverges. The contradiction implies that L does not exists. □

We want to have a machine for calculating limits: 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.

An alternative approach is 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.

Theorem 1.8 (**Algebra of limits**)**.** *If* $f, g : I\backslash\{c\} \to \mathbb{R}$ *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 □

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 1.9 (**Infinite limits**). *If $f : I \setminus \{c\} \to \mathbb{R}$ is defined on an 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 1.10.

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

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

$$\frac{1}{x^2} > M.$$ □

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

Definition 1.11 (**Limits at infinity**). *If $f : \mathbb{R} \to \mathbb{R}$ 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 1.12.

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

Proof. Exercise. □

Example 1.13.

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

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

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

Given $\varepsilon > 0$, let $N = 2\varepsilon^{-1}$. Then for any $x > N$ we get

$$\left|\frac{x}{e^x} - 0\right| \leq \frac{2}{N} = \varepsilon.$$

Consequently $x\frac{x}{e^x} \to 0$ as $\to \infty$. □

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 1.14 (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 1.15. *Let $f : \mathbb{R} \to \mathbb{R}$ 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

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

There is also a one sided version of infinite limits.

Example 1.16.

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

1.2 The derivative

Definition 1.17 (**The derivative**). *Suppose $f : I \to \mathbb{R}$ 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 $f'(c)$.

We say that f is *differentiable* on I if it is differentiable at every point $c \in I$.

Example 1.18 (**The derivative of x**). *If $f(x) = x$ then f is differentiable at every point of $\mathbb{R}$ 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

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

Example 1.19 (**The derivative of x^2**). *If $f(x) = x^2$ then f is differentiable at every point of $\mathbb{R}$ 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

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

Example 1.20 (**The derivative of $1/x$**). *If $f(x) = 1/x$ then f is differentiable at every point of $\mathbb{R}$ 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

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

1.3 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 has several basic parts: the sum rule, the product rule, the chain rule, the inverse function theorem 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 1.21 (**Differentiability and continuity**). *If I is an open interval, $f : I \to \mathbb{R}$ 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. □

Example 1.22 (**continuous but not differentiable**). *For example, the function f given by $f(x) = |x|$ is continuous at 0 but not differentiable at 0.*

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

Theorem 1.23 (**The sum and product rules**). *Suppose $f, g : I \to \mathbb{R}$ 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.

Corollary 1.24 (**linearity of the derivative**). *If f and g are differentiable at a point c then for any $\alpha, \beta \in \mathbb{R}$ the function $\alpha f + \beta g$ is differentiable at c and*

$$(\alpha f + \beta g)'(c) = \alpha f'(c) + \beta g'(c).$$

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 1.25 (**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 $n = 1$ (for the purpose of this lemma assume that $x^0 = 1$ for all x including zero). 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

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

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 1.26 (**Chain rule**)**.** *Suppose I and J are open intervals, $f : I \to \mathbb{R}$ 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*

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

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 $g(c)$.

Proof. In order to prove the chain rule we have to investigate the ratio

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

as $x \to c$. If $g(x) \neq g(c)$ we can rewrite this ratio as

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

The second factor converges to $g'(c)$ as $x \to c$. The first factor looks a candidate for converging to $f'(g(c))$. 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 order to avoid this technical difficulty we note that for these values of x the ratio vanishes. Then we define the auxiliary function $h : I \to \mathbb{R}$ by

$$h(x) = \begin{cases} \dfrac{f(g(x)) - f(g(c))}{g(x) - g(c)} & \text{if } g(x) \neq g(c), \\ f'(g(c)) & \text{if } g(x) = g(c). \end{cases}$$

Then for any $x \neq c$ we get

$$\frac{f(g(x)) - f(g(c))}{x - c} = h(x) \cdot \frac{g(x) - g(c)}{x - c}.$$

The chain rule follows easily from this equality (take the limit $x \to c$) if we show that $\lim_{x \to c} h(x) = f'(g(c))$.

Let $\varepsilon > 0$. Since f is differentiable at $g(c)$, there exists $\delta_1 > 0$ such that if $0 < |y - g(c)| < \delta_1$ then

$$\left| \frac{f(y) - f(g(c))}{y - g(c)} - f'(g(c)) \right| < \varepsilon. \qquad (*)$$

Since g is differentiable at c it is also continuous at c. Consequently there exists $\delta_2 > 0$ such that $|x - c| < \delta_2$ implies $|g(x) - g(c)| < \delta_1$.

We conclude that if $0 < |x - c| < \delta_2$ then

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

Indeed, if $g(x) = g(c)$ then the inequality is satisfied as in this case $h(x) = f'(g(c))$, otherwise we can let $y = g(x)$ and use the inequality $(*)$.

Consequently, $h(x) \to f'(g(c))$ as required. □

1.4 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 1.27 (**Mean Value Theorem**). *Suppose $f : [a, b] \to \mathbb{R}$ 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 1.28 (**Functions with positive derivative**). *If $f : I \to \mathbb{R}$ 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 1.29 (**Functions with zero derivative**). *If $f : I \to \mathbb{R}$ 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 1.30 (**Uniqueness of solution to a differential equation**). *The only functions $f : \mathbb{R} \to \mathbb{R}$ 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 later). 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 $f(x) = e^x g(x) = Ae^x$. □

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

Theorem 1.31 (**Rolle**)**.** *Suppose $f : [a,b] \to \mathbb{R}$ 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 Calculus 1 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 $f:[a,b]\to\mathbb{R}$ 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 1.32 (Extrema and derivatives). *Suppose $f:[a,b]\to\mathbb{R}$ 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 $f''(c)\le 0$ which may be valuable information.)

If you have a function defined on an open interval, on $(0,\infty)$ or the whole of $\mathbb{R}$ 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 1.33 (maximum of a function). *Find the maximum of xe^{-x} on $\mathbb{R}$.*

Proof. 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}.

We have proved that $xe^{-x} \le e^{-1}$ for all x. □

1.5 Derivatives of inverses

In Calculus 1 it was proved 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 trigonometric functions are differentiable).

Example 1.34 (**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 1.35 (**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 1.36 (**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 1.37 (**Derivatives of inverses**)**.** *Let* $f : (a,b) \to \mathbb{R}$ *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
$$\lim_{u\to x}\frac{g(u)-g(x)}{u-x}=\lim_{v\to y}\frac{v-y}{f(v)-f(y)}=\frac{1}{f'(y)}=\frac{1}{f'(g(x))}.$$ □

1.6 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 1.38 (**Cauchy's Mean Value Theorem**)**.** *If $f,g:[a,b]\to\mathbb{R}$ 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 h defined by
$$h(x)=f(x)\big(g(b)-g(a)\big)-\big(f(b)-f(a)\big)g(x).$$
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

$$h'(x) = f'(x)\big(g(b) - g(a)\big) - \big(f(b) - f(a)\big)g'(x)$$

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

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

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

1.7 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

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

In particular this definition implies that

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

when x is close to c.

It is not surprising that in the more general situation, derivatives can often help to find a limit. 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

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

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

Theorem 1.39 (l'Hôpital's rule). *If $f,g : I \to \mathbb{R}$ 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. The definition of the limit requires the ratio to be defined on some interval around c (except perhaps at c itself). So g' is non-zero on an interval each side of c. As long as x is in the region around c where $g' \neq 0$, we can apply Cauchy's MVT which ensures that there is a point t (depending upon x) between c and x such that

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

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

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

where we used that $f(c) = g(c) = 0$. □

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

Theorem 1.40 (l'Hôpital's rule at infinity). *If $f, g : \mathbb{R} \to \mathbb{R}$ 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 1.41. $\lim_{x\to\infty} \frac{x}{e^x} = \lim_{x\to\infty} \frac{1}{e^x} = 0.$

2 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. For example, we can write

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

and use this approximation for $x = 1$. We get

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

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

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

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

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

So

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

Recalling the definition of the function g we get

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

which is exactly the statement of the theorem. □

Example 2.2. 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. 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 2.3. *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 $0 \leq x \leq \pi$ we have $0 < t < \pi$ and so $\sin t > 0$. □

Consider a function $f : I \to \mathbb{R}$ defined on an open interval I. Suppose that the function f is at least n-times differentiable. Then for any $a \in I$, we can define the Taylor polynomial

$$P_n(x) = f(a) + \frac{f'(a)}{1!}(x-a) + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n.$$

The coefficients of the Taylor polynomial are chosen to match the derivatives of the function f at the point a:

$$P_n(a) = f(a), \qquad P_n'(a) = f'(a), \quad \ldots \quad P_n^{(n)}(a) = f^{(n)}(a).$$

In other words the derivatives of f and P_n coincide up to the order n. The remainder

$$r_n(x) = f(x) - P_n(x)$$

describes the error of approximation of the function by its Taylor polynomial. Taylor's theorem with Lagrange remainder provides an expression for the remainder in terms of the derivative $f^{(n+1)}$.

Example 2.4 (**Taylor polynomial for a polynomial function**). *Let f be polynomial of degree m:*

$$f(x) = a_0 + a_1 x + \cdots + a_m x^m,$$

where $a_0, a_1, \ldots, a_m \in \mathbb{R}$ are coefficients of the polynomial. Then for any $n \geq m$ and any $a \in \mathbb{R}$

$$f(x) = P_n(x).$$

Indeed, the Lagrange expression for the remainder implies that for some c between x and a

$$f(x) - P_n(x) = \frac{f^{(n+1)}(c)}{(n+1)!}(x-a)^{n+1} = 0$$

as the derivative equals to zero.

Of course, only polynomial functions can coincide with their Taylor polynomials. On the other hand we can ask what accuracy can be achieved with the help of Taylor polynomials. Is it true that if we fix a and x, then larger n lead to more accurate approximations or, in other words, to smaller values of the remainder $r_n(x)$?

We can use the Lagrange expression for the remainder for this purpose, but in the next lecture we will see that an alternative expression sometimes provides a more accurate estimate.

Theorem 2.5 (a more general formula for the remainder). *Suppose that a function $f : I \to \mathbb{R}$ is $(n+1)$-times differentiable on the open interval I and $x, a \in I$ ($x \neq a$). Then for any differentiable function $g : I \to \mathbb{R}$ which has a non-zero derivative between x and a there is a point c between x and a such that the remainder of the Taylor formula*

$$r_n(x) = \frac{g(x) - g(a)}{g'(c)} \, f^{(n+1)}(c) \, \frac{(x-c)^n}{n!}.$$

Proof. Consider the function F defined by

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

Note that expression in the parenthesis coincides with the Taylor polynomial centred at a point t. Substituting $t = x$ and $t = a$ and recalling the definition of P_n we see that

$$F(x) = 0, \qquad F(a) = f(x) - P_n(x) = r_n(x).$$

Then we differentiate the function F and use the product rule for each term. Most of the sum cancels out:

$$\begin{aligned} F'(t) &= -\left(f(t) + \frac{f'(t)}{1!}(x-t) + \cdots + \frac{f^{(n)}(t)}{n!}(x-t)^n \right)' \\ &= -\Bigg(f'(t) - \frac{f'(t)}{0!} + \frac{f''(t)}{1!}(x-t) - \ldots \\ &\qquad\qquad - \frac{f^{(n)}(t)}{(n-1)!}(x-t)^{n-1} + \frac{f^{(n+1)}(t)}{n!}(x-t)^n \Bigg) \\ &= -\frac{f^{(n+1)}(t)}{n!}(x-t)^n. \end{aligned}$$

Now we apply Cauchy's Mean Value Theorem to the functions F and g. The theorem implies that there is a point c between x and a such that

$$\frac{F(x) - F(a)}{g(x) - g(a)} = \frac{F'(c)}{g'(c)}.$$

Substituting the expressions for the function F and its derivative F' we get

$$\frac{-r_n(x)}{g(x) - g(a)} = -\frac{f^{(n+1)}(c)}{n! g'(c)}(x-c)^n.$$

Multiplying this equality by $g(a) - g(x)$ we get the desired expression for r_n. $\square$

We can use the last theorem to obtain a second proof of Taylor's theorem with Lagrange remainder.

Corollary 2.6 (**Lagrange remainder**). *Under the assumption of the theorem there is a point c between a and x such that*

$$r_n(x) = \frac{f^{(n+1)}(c)}{(n+1)!}(x-a)^{n+1}.$$

Indeed, let $g(t) = (x-t)^{n+1}$. Then

$$g(x) - g(a) = -(x-a)^{n+1} \qquad \text{and} \qquad g'(c) = -(n+1)(x-c)^n.$$

Substituting these expressions we recover the Lagrange remainder:

$$\begin{aligned} r_n(x) &= \frac{g(x)-g(a)}{g'(c)n!} f^{(n+1)}(c)(x-c)^n = \frac{-(x-a)^{n+1}}{-(n+1)(x-c)^n n!} f^{(n+1)}(c)(x-c)^n \\ &= \frac{(x-a)^{n+1}}{(n+1)!} f^{(n+1)}(c). \end{aligned}$$

In the next lecture we will need another expression for the remainder.

Corollary 2.7 (**Cauchy remainder**). *Under the assumption of the theorem there is a point c between a and x such that*

$$r_n(x) = \frac{f^{(n+1)}(c)}{n!}(x-c)^n(x-a).$$

Indeed, let $g(t) = (x-t)$. Then

$$g(x) - g(a) = -(x-a) \qquad \text{and} \qquad g'(c) = -1.$$

Substituting these expression we get

$$r_n(x) = \frac{g(x)-g(a)}{g'(c)n!} f^{(n+1)}(c)(x-c)^n = \frac{(x-a)}{n!} f^{(n+1)}(c)(x-c)^n.$$

Cauchy reminder looks a bit more complex, but in our next example it will provide a more accurate estimate than the Lagrange remainder.

If a function is infinitely differentiable we can write a Taylor series instead of a Taylor polynomial. For example, for the logarithm and $a = 1$ we get

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

The ratio test of Calculus 1 can be used to show that the infinite series converge provided $|x| < 1$, and the series diverge for $|x| > 1$.

In our next example we will use the Taylor theorem with Cauchy remainder to prove that the logarithm is indeed equal to the sum of its Taylor series when x is near 1.

Example 2.8 (Taylor Series for log). Let f be the function defined by $f(x) = \log x$. We have not yet discussed the definition of the log function. So for the purpose of this example we assume that we know that $\log 1 = 0$ and $(\log x)' = x^{-1}$. Then we can differentiate the function any number of times:

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

Let us find the Taylor polynomial of degree n centred at $a = 1$. We have

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

Therefore

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

The Taylor's theorem with Cauchy remainder implies that there is a c_n between 1 and x such that

$$\log x - P_n(x) = r_n(x) = \frac{(x-1)}{n!} f^{(n+1)}(c_n)(x-c_n)^n = \frac{(x-1)}{c_n^{n+1}}(-1)^n(x-c_n)^n.$$

We rewrite the reminder in the form

$$r_n(x) = \frac{(x-1)}{c_n}\left(1 - \frac{x}{c_n}\right)^n.$$

We are going to show that if $0 < x < 2$ then $r_n(x) \to 0$ as $n \to \infty$.

We note that $1 - x/c$ is monotone increasing as a function of c for $c > 0$. Therefore if $0 < x < c < 1$ then

$$0 < 1 - \frac{x}{c} < 1 - x$$

and if $1 < c < x < 2$ then

$$1 - x < 1 - \frac{x}{c} < 0.$$

In both cases

$$\left|1 - \frac{x}{c}\right| \leq |1 - x|.$$

Since c_n is between 1 and x we get $c_n > \min\{1, x\} > 0$ and, consequently,

$$|r_n(x)| = \frac{|x-1|}{|c_n|}\left|1 - \frac{x}{c_n}\right|^n \leq \frac{|x-1|^{n+1}}{\min\{1,x\}} \to 0$$

provided $|x - 1| < 1$. The sandwich rule implies that $r_n(x) \to 0$ 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 $0 < x < 2$.

What about $x = 0$ and $x = 2$?

If we set $x = 0$ we get the series

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

From Calculus 1 you know that these series diverge to minus infinity.

If we set $x = 2$ then the series becomes

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

which converges by the alternating series theorem. The limit of the series equals $\log 2$, but our arguments do not imply this conclusion.

Finally we recall that the series diverge for $|x - 1| > 1$.

3 Power series

3.1 Basic properties of power series

A power series is 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.

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 3.1 (absolute convergence). *Let $\sum_0^\infty a_n x^n$ be a power series with $\sum_0^\infty a_n t^n$ convergent. Then*

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

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

Proof. Since $\sum_0^\infty a_n t^n$ 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 3.2 (Radius of convergence). *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 $|x| > R$.*

Definition 3.3. *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 the series converges for some $x \neq 0$.

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 $\mathbb{R}$.

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

$$R = \sup\left\{ |t| : \sum_0^\infty a_n t^n \text{ converges} \right\}.$$

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 3.4 (**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 3.5 (**The geometric series I**)**.** *The series $\sum_0^\infty x^n$ has radius of convergence $R = 1$.* Indeed, we know that the series converges if and only if $x \in (-1, 1)$.

Example 3.6 (**The geometric series II**)**.** *If p is real the series $\sum_0^\infty p^n x^n$ has radius of convergence $R = 1/|p|$.* Indeed, 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 3.7 (**The log series**)**.** *The series $\sum_1^\infty \frac{x^n}{n}$ has radius of convergence $R = 1$.* Indeed, 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 3.8. *The series $\sum_0^\infty n x^n$ has radius of convergence $R = 1$.*

For most interesting power series we can find the radius of convergence using the ratio test.

Example 3.9. *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 $x^2 \in (-1,1)$. However, there is another way to to do it which is a much more flexible method. If $x > 1$ then the terms do not tend to zero so the series diverges. If $0 < x < 1$ then

$$1+x^2+x^4+\cdots < 1+x+x^2+x^3+x^4+\cdots$$

which we already know converges. This means that the series converges if $0 < x < 1$ so the radius of convergence is 1. □

We have established that a power series $\sum_0^\infty a_n x^n$ has a radius of convergence R with $0 \le R \le +\infty$. We also know that

- if $|x| < R$, then the series converges absolutely
- if $|x| < R$, then $a_n x^n \to 0$ as $n \to \infty$
- if $|x| < R$, then $a_n x^n$ is bounded
- if $|x| > R$, then the series diverges
- if $|x| > R$, then the sequence of $a_n x^n$ is not bounded

These properties can help us to establish the radius of convergence for a given series.

It is also possible to consider power series centred at points other then 0. For example we know the geometric series

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

which converge for $|x| < 1$, i.e. the radius of convergence equals to one. On the other hand we can write

$$\begin{aligned}\frac{1}{1-x} &= \frac{1}{2-(x+1)} = \frac{1}{2}\frac{1}{1-\frac{x+1}{2}} \\ &= \frac{1}{2}\left(1+\frac{x+1}{2}+\left(\frac{x+1}{2}\right)^2+\cdots\right) = \frac{1}{2}+\frac{x+1}{4}+\frac{(x+1)^2}{8}+\cdots.\end{aligned}$$

This series converges if $\left|\frac{x+1}{2}\right| < 1$: in other words if x differs from -1 by at most 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 $x = 1$.

In general we can consider a series $\sum_0^\infty a_n(x-x_0)^n$ centred at a point x_0. Then if the series converges for $|x-x_0| < R$ and diverges for $|x-x_0| > R$, we say that R is the radius of convergence. We say that $R = 0$ if the series converges for $x = x_0$ only, and $R = \infty$ if the series converges for all x.

3.2 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 3.10 (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 $|y-x| < \delta_0$ then

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

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

$$\left|\sum_0^\infty a_n y^n - \sum_0^\infty a_n x^n\right| \le \left|\sum_{N+1}^\infty a_n y^n\right| + \left|\sum_0^N a_n y^n - \sum_0^N a_n x^n\right| + \left|\sum_{N+1}^\infty a_n x^n\right|$$
$$\le \sum_{N+1}^\infty |a_n|\,|y|^n + \left|\sum_0^N a_n y^n - \sum_0^N a_n x^n\right| + \sum_{N+1}^\infty |a_n|\,|x|^n < \varepsilon. \quad \square$$

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.

3.3 The differentiability of power series

Power series are differentiable inside the radius of convergence. Naturally the proof of this is more difficult than the proof of continuity. Suppose

$$f(x) = \sum_{n=0}^{\infty} a_n x^n$$

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 $\sum_{n=1}^{\infty} na_n x^{n-1}$ converges and

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

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

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

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.

Lemma 3.11. *Let $\sum a_n x^n$ be a power series with radius of convergence R. Then the series $\sum na_n x^{n-1}$ 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 $\sum_{n=1}^{\infty} |a_n| nx^{n-1}$ so the latter also converges. This means that $\sum na_n x^{n-1}$ converges absolutely as required. □

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 3.12 (**The differentiability of power series**). *Let $f(x) = \sum_0^\infty a_n x^n$ 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 such that $|x| < T < R$. 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

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

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 $0 < |y-x| < \delta_0$

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

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

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

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

We have proved that for any $\varepsilon>0$ there is $\delta>0$ such that whenever $0<|y-x|<\delta$

$$\left|\frac{f(y)-f(x)}{y-x} - \sum_{n=1}^{\infty} n a_n x^{n-1}\right| < \varepsilon.$$

Consequently the power series is differentiable at the point x and

$$f'(x) = \lim_{y\to x} \frac{f(y)-f(x)}{y-x} = \sum_{n=1}^{\infty} n a_n x^{n-1}. \qquad \square$$

Corollary 3.13. *A power series is infinitely differentiable on* $(-R,R)$*.*

3.4 Taylor series

Now we will discuss 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.

3.5 The exponential

We define the exponential function as a power series.

Definition 3.14 (The exponential). *If $x \in \mathbb{R}$ the series*

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

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

Notice that 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 $\mathbb{R}$ since it is a convergent power series.

Corollary 3.15 (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

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

□

We would like to check that the exponential function satisfies the characteristic property $e^{x+y} = e^x e^y$. 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.

Theorem 3.16 (**The characteristic property of the exponential**). *If $x, y \in \mathbb{R}$ 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. □

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 3.17 (**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

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

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 both inequalities are easy to establish for $x \geq 0$.

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

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

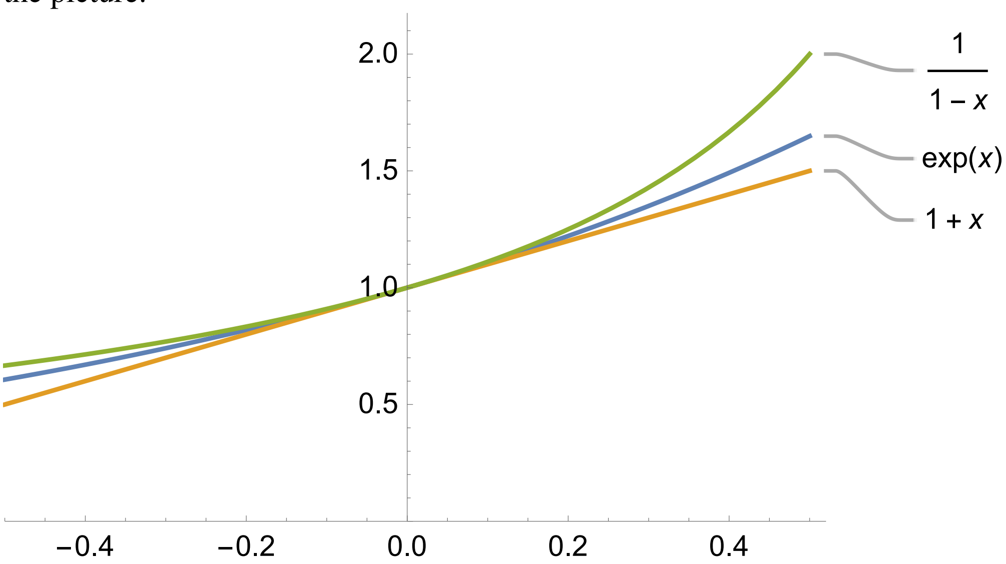
so the second inequality is now established for all $x < 1$.

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

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

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



Corollary 3.18 (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.

3.6 The logarithm and powers

We have seen that the exponential function maps $\mathbb{R}$ 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 3.19 (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. □

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.

Theorem 3.20 (**The derivative of** $\log$). *If $f : x \mapsto \log x$ then*

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

The inequalities we proved for the exponential immediately give us inequalities for the logarithm. The most crucial one is this:

Theorem 3.21 (**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 $(1, 0)$.

We know 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 3.22 (**Powers**). *If $x > 0$ and $p \in \mathbb{R}$ 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 $x \cdot x \cdot \ldots \cdot x$ of n copies of x.*

2. $x^{p+q} = x^p x^q$ *for all $x > 0$ and $p, q \in \mathbb{R}$.*

3. $\log(x^p) = p \log x$ *for $x > 0$ and $p \in \mathbb{R}$.*

4. $x^{pq} = (x^p)^q$ *for all $x > 0$ and $p, q \in \mathbb{R}$.*

5. $\exp(p) = e^p$ *for all $p \in \mathbb{R}$.*

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 $\exp(p)$. □

Example 3.23. *Let $a > 0$ and define $f : \mathbb{R} \to \mathbb{R}$ by $f(x) = a^x$. Then f is differentiable on $\mathbb{R}$ and $f'(x) = (\ln(a))a^x$.*

Proof. $f(x) = e^{x\ln(a)}$ and so f is differentiable at every $x \in \mathbb{R}$ by the chain rule and $f'(x) = \ln(a)e^{x\ln(a)} = (\ln(a))a^x$. □

Example 3.24. *Let $b \in \mathbb{R}$ and define $g : (0,\infty) \to \mathbb{R}$ by $g(x) = x^b$. Then g is differentiable on $(0,\infty)$ and $g'(x) = bx^{b-1}$.*

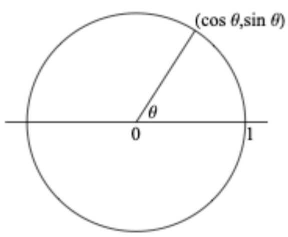
Proof. $g(x) = e^{b\ln(x)}$ and so g is differentiable at every $x > 0$ by the chain rule and $g'(x) = e^{b\ln(x)}\dfrac{b}{x} = x^b x^{-1} b = bx^{b-1}$. □

Example 3.25. *Define $h : (0,\infty) \to \mathbb{R}$ by $h(x) = x^x$. Then h is differentiable on $(0,\infty)$ and $h'(x) = x^x(\ln(x)+1)$.*

Proof. $h(x) = e^{x\ln(x)}$ so h is differentiable at every $x > 0$ by the chain rule and $h'(x) = e^{x\ln(x)}(\ln(x)+1) = x^x(\ln(x)+1)$. □

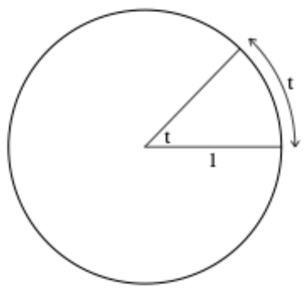
3.7 The trigonometric functions

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.

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.



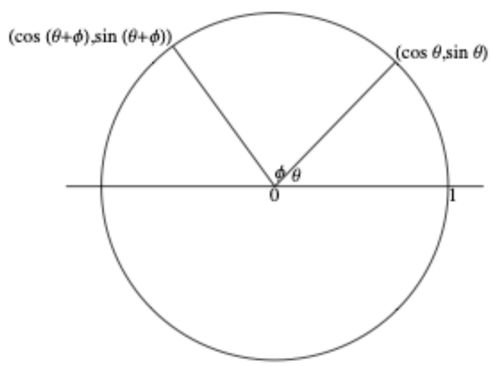
The size of the angle is the length of the arc it spans.

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 $(\cos t, \sin t)$.

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 $0.0174\ldots\cos x^{\circ}$: we get a funny multiple of cos rather than cos itself. The choice of radians removes this problem.

We use power series to define the cosine and sine functions.

Definition 3.26 (**Trigonometric functions**). *For $x \in \mathbb{R}$ we define*

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} + \cdots + (-1)^n \frac{x^{2n}}{(2n)!} + \dots$$
$$\sin x = x - \frac{x^3}{3!} + \cdots + (-1)^n \frac{x^{2n+1}}{(2n+1)!} + \dots$$

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. We also see that

$$\cos 0 = 1 \qquad \text{and} \qquad \sin 0 = 0.$$

From our general theory we know that both functions are differentiable. Differentiating the series we can check that

$$\frac{d}{dx}\cos x = -\sin x \qquad \text{and} \qquad \frac{d}{dx}\sin x = \cos x.$$

Let us establish properties of the trigonometric functions.

Theorem 3.27 (**The circular property**). *For all $x \in \mathbb{R}$*

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

Proof. Let $f(x) = \cos^2 x + \sin^2 x$. Obviously $f(0) = 1$. Compute the derivative

$$f'(x) = -2\cos x \sin x + 2\sin x \cos x = 0.$$

Consequently the function f is constant. Then $f(x) = 1$ for all x. □

Corollary 3.28. *For any real x*

$$|\cos(x)| \leq 1 \qquad \textit{and} \qquad |\sin x| \leq 1.$$

Theorem 3.29 (**The addition formulae for the trigonometric functions**). *For all real x and y*

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

Proof. For any fixed z consider the function

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

Differentiate the function:

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

Since $f'(x) = 0$ for all x, the function f is constant and we get $f(x) = f(0)$ for all x. Since

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

we conclude for any x and any z

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

We can let $z = x - y$ to get the first addition formula.

The proof of the second formula is similar and it is left as an exercise. □

The addition formulae with $x = y$ imply the expressions for doubled arguments:

$$\cos(2x) = \cos^2 x - \sin^2 x \qquad \text{and} \qquad \sin(2x) = 2 \sin x \cos x.$$

Now we are going to prove that the trigonometric functions are periodic. Of course you already know that the period equals to 2π, but this fact may look surprising since we have started our study of the trigonometric functions from power series.

The first step in the proof is to show that there is a point $x_0 \in (0,2)$ such that

$$\cos x_0 = 0.$$

From the definition we see that $\cos x \approx 1 - x^2/2$ and the right hand side is negative for $x = 2$. In order to conclude that $\cos 2$ is also negative we use the Taylor theorem with Lagrange remainder (with $f(x) = \cos x$, $a = 0$, $b = 2$ and $n = 4$) which implies that there is $t \in (0,2)$ such that

$$\cos 2 = 1 - \frac{2^2}{2} + \frac{2^4}{4!} \cos^{(iv)} t = -1 + \frac{2}{3} \cos t$$

(the fourth derivative of cos coincides with cos). Since $\cos t \leq 1$ we conclude that $\cos 2 < -2/3 < 0$. Now we use the Intermediate Value Theorem (IVT) of Calculus 1: cos is a continuous function such that $\cos 0 > 0$ and $\cos 2 < 0$, so there is $x_0 \in (0,2)$ such that $\cos x_0 = 0$.

The circular property implies that $\sin^2 x_0 = 1$. Using the formulae for doubled arguments we get

$$\begin{aligned}\cos 2x_0 &= \cos^2 x_0 - \sin^2 x_0 = -1,\\ \sin 2x_0 &= 2\sin x_0 \cos x_0 = 0,\\ \cos 4x_0 &= \cos^2 2x_0 - \sin^2 2x_0 = 1,\\ \sin 4x_0 &= 2\sin 2x_0 \cos 2x_0 = 0.\end{aligned}$$

Let $T = 4x_0$. We know that $T \in (0,8)$. The addition formulae imply that for any x

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

We have proved that the sine and cosine are periodic.

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.

There is a simple explanation of this fact in the language of geometry and motion. Suppose that t denotes time and study the motion of the point $r(t) = (\cos t, \sin t)$ on the plane. The circular property implies that the point $r(t)$ moves on the unit circle. Its velocity vector is given by the derivative

$$v(t) = \frac{dr(t)}{dt} = (-\sin t, \cos t).$$

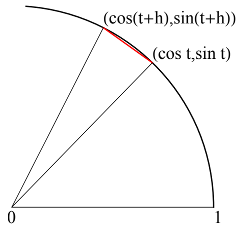
The speed of motion is the length of the vector $v(t)$. The circular property implies that

$$|v(t)| = \sqrt{(-\sin t)^2 + (\cos t)^2} = 1.$$

Therefore the point $r(t)$ moves along the circle with the unit speed. Consequently, the distance covered by the point over an interval of time of length t equals to t.

In particular, this argument also shows that the period $T = 2\pi$ as the time required to make a full round equals to the length of the unit circle.

It is also possible to prove that $L'(t) = 1$ for all t without making references to the knowledge from outside Calculus 1 and 2. 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 that for $h > 0$ (and no bigger than π)

$$\sqrt{(\cos(t+h) - \cos t)^2 + (\sin(t+h) - \sin t)^2} = 2\sin\left(\frac{h}{2}\right)$$

using the properties that we do know: the addition formulae. But

$$(\cos(t+h)-\cos t)^2+(\sin(t+h)-\sin t)^2$$

$$=\cos^2(t+h)-2\cos(t+h)\cos t+\cos^2 t+\sin^2(t+h)-2\sin(t+h)\sin t+\sin^2 t$$

$$=2-2\cos(t+h)\cos t-2\sin(t+h)\sin t$$

$$=2-2\cos(t+h-t)=2-2\cos h=4\sin^2(h/2)$$

by repeated use of the addition formulae. When h is positive (and less than 2π) then $\sin(h/2)$ is positive so the square root *is* $2\sin(h/2)$. 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.

Note that that we have earlier deduced the periodicity with the help of the addition formulae. Now we can write these arguments in a more familiar way. For example, as $\cos 2\pi=1$ and $\sin 2\pi=0$ the addition formula implies

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

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

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

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

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

3.8 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

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

provided $|x| < 1$. (The series will be infinite unless n is a non-negative integer.)

As we have the definition for non-integer powers we can study Taylor series for the function

$$f(x) = (1+x)^a$$

for any $a \in \mathbb{R}$. This function is defined for $x \geq -1$. In order to write down the Taylor series for this function we need to compute derivatives of $f(x) = e^{a\ln(1+x)}$:

$$\begin{aligned} f'(x) &= a(1+x)^{a-1}, \\ f''(x) &= a(a-1)(1+x)^{a-2}, \\ f'''(x) &= a(a-1)(a-2)x^{a-3}, \\ &\cdots \end{aligned}$$

Evaluating the function and its derivatives at $x = 0$ we get

$$\begin{aligned} f(0) &= 1, \\ f'(0) &= a, \\ f''(0) &= a(a-1), \\ f'''(0) &= a(a-1)(a-2) \\ &\cdots \end{aligned}$$

So the Taylor series of f about 0 is given by

$$1+ax+\frac{a(a-1)}{2!}x^2+\frac{a(a-1)(a-2)}{3!}x^3+...+\frac{a(a-1).....(a-n+1)}{n!}x^n+...$$

It is convenient to extend the definition of a binomial coefficient for $a \in \mathbb{R}$ and integer $n \geq 1$

$$\binom{a}{n} = \frac{a(a-1)\dots(a-n+1)}{n!}.$$

If $n = 0$, we define separately

$$\binom{a}{0} = 1.$$

Then the Taylor series takes the form which look similar to the binomial expansion

$$\sum_{n=0}^{\infty} \binom{a}{n} x^n.$$

If a is a positive integer, the sum contain only a finite number of non-zero terms and coincides with the binomial expansion. You can easily check that for $a = -1$ we get the formula for the geometrical progression

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

Let us study the Taylor expansion for non-integer values of a more carefully.

Theorem 3.30 (**Taylor series for the binomial expansion**)**.** *Let $a \in \mathbb{R}$, $a \notin \mathbb{N}_0$ and $f(x) = (1+x)^a$ for $x \geq -1$. The Taylor series for the function f,*

$$g(x) = \sum_{n=0}^{\infty} \binom{a}{n} x^n,$$

has radius of convergence 1 *and $f(x) = g(x)$ when $|x| < 1$.*

Proof. Note that if $a \neq 0$ and $a \notin \mathbb{N}$ then $\binom{a}{n} \neq 0$ for all $n \geq 0$. Then the ratio test gives

$$\left| \frac{\binom{a}{n+1} x^{n+1}}{\binom{a}{n} x^n} \right| = \left| \frac{a(a-1)\dots(a-n)}{(n+1)!} \frac{n!}{a(a-1)..(a-n+1)} \right| |x|$$
$$= \left| \frac{a-n}{n+1} \right| |x| = \left| \frac{a+1}{n+1} - 1 \right| |x| \quad \to \quad |x|$$

as $n \to \infty$. Therefore the radius of convergence is 1.

Inside the radius of convergence the power series are differentiable and we get

$$g'(x) = \sum_{n=1}^{\infty} n \binom{a}{n} x^{n-1}.$$

Since
$$(n+1)\binom{a}{n+1}+n\binom{a}{n}=a\binom{a}{n}$$
we get
$$(1+x)g'(x)=ag(x).$$
In order to check that the sum of the Taylor series $g(x)=f(x)$ for all $|x|<1$ we compute the derivative of the ratio using the ratio rule:
$$\left(\frac{g}{f}\right)'(x)=\frac{f(x)g'(x)-g(x)f'(x)}{[f(x)]^2}$$
$$=\frac{(1+x)^a g'(x)-g(x)a(1+x)^{a-1}}{(1+x)^{2a}}=\frac{(1+x)g'(x)-ag(x)}{(1+x)^{a+1}}=0.$$
Therefore $\frac{g(x)}{f(x)}$ is constant. For $x=0$ we have $\frac{g(0)}{f(0)}=\binom{a}{0}=1$. Consequently
$$(1+x)^a=\sum_{n=0}^{\infty}\binom{a}{n}x^n$$
for all $x\in(-1,1)$. □

3.9 Example: a function which isn't the sum of its Taylor series

Suppose that $f:\mathbb{R}\to\mathbb{R}$ is an infinitely differentiable function. Then we can write down the corresponding Taylor series around zero,
$$\sum_{k=0}^{\infty}\frac{f^{(k)}(0)}{k!}x^k.$$
If the series converges, it defines an infinitely differentiable function. In the previous lectures we have seen numerous examples of functions which coincide with the sum of the Taylor series inside the radius of convergence. Is it always true that the sum of the Taylor series coincides with the function?

Example 3.31 (**a function which isn't the sum of its Taylor series**). *The function $f:\mathbb{R}\to\mathbb{R}$ defined by*
$$f(x)=\begin{cases}0 & \text{if } x=0\\ e^{-\frac{1}{x^2}} & \text{if } x\neq 0\end{cases}$$

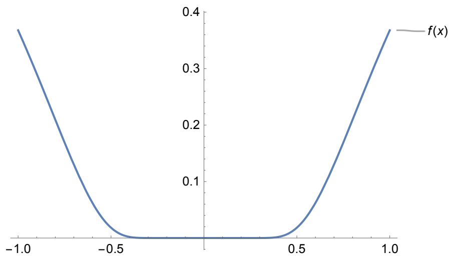
is infinitely differentiable on $\mathbb{R}$,

$$f^{(n)}(0) = 0 \qquad \textit{for all } n \geq 1.$$

Hence the Taylor series at 0,

$$\sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!} x^n = 0 + 0x + 0x^2 + \ldots$$

converges to 0 *for all* $x \in \mathbb{R}$. *Since* $f(x) > 0$ *for all* $x \neq 0$ *we see that the Taylor series is only equal to* $f(x)$ *when* $x = 0$.



Proof. In order to write down the Taylor series we have to find expressions for the derivatives of the function f. Let's compute the first three derivatives for $x \neq 0$:

$$f'(x) = \frac{2}{x^3} e^{-\frac{1}{x^2}}, \quad f''(x) = \left(\frac{4}{x^6} - \frac{6}{x^4}\right) e^{-\frac{1}{x^2}}, \quad f'''(x) = \left(\frac{8}{x^9} - \frac{36}{x^7} + \frac{24}{x^5}\right) e^{-\frac{1}{x^2}}.$$

We see the pattern:

$$f^{(n)}(x) = \left(\frac{2^n}{x^{3n}} + \ldots + (-1)^{n+1} \frac{(n+1)!}{x^{n+2}}\right) e^{-\frac{1}{x^2}}.$$

The induction can be used to prove that our guess is correct: for every $n \geq 1$ the are for some real numbers $a_{3n}, a_{3n-2}, ... a_{n+2}$ such that for $x \neq 0$

$$f^{(n)}(x) = \left(\frac{a_{3n}}{x^{3n}} + \frac{a_{3n-2}}{x^{3n-2}} + \ldots + \frac{a_{n+2}}{x^{n+2}}\right) e^{-\frac{1}{x^2}}.$$

Indeed, the formula holds for $n = 1$. Suppose it holds for some $n \in \mathbb{N}$. Then we use the product rule and the chain rule:

$$\begin{aligned} f^{(n+1)}(x) =& \frac{2}{x^3} e^{-\frac{1}{x^2}} \left[\frac{a_{3n}}{x^{3n}} + \frac{a_{3n-2}}{x^{3n-2}} + ... + \frac{a_{n+2}}{x^{n+2}}\right] \\ &+ e^{-\frac{1}{x^2}} \left[\frac{-3na_{3n}}{x^{3n+1}} - \frac{(3n-2)a_{3n-2}}{x^{3n-1}} - ... - \frac{(n+2)a_{n+2}}{x^{n+3}}\right] \\ =& e^{-\frac{1}{x^2}} \left[\frac{2a_{3n}}{x^{3(n+1)}} + \frac{2a_{3n-2} - 3na_{3n}}{x^{3n+1}} + ... - \frac{(n+2)a_{n+2}}{x^{n+3}}\right]. \end{aligned}$$

This expression shows us that $f^{(n+1)}$ can be written in the desired form.

Now we can check that f is also infinitely differentiable at zero and $f^{(n)}(0) = 0$ for all $n \in \mathbb{N}$. Using the definition of the derivative we get

$$f'(0) = \lim_{h\to 0} \frac{f(h) - f(0)}{h} = \lim_{h\to 0} \frac{e^{-1/h^2}}{h} = 0.$$

We continue by induction: suppose that $f^{(n)}(0) = 0$ for some $n \in \mathbb{N}$. Then

$$\begin{aligned} f^{(n+1)}(0) &= \lim_{x\to 0} \frac{f^{(n)}(h) - f^{(n)}(0)}{h} \\ &= \lim_{x\to 0} \left(\frac{a_{3n}}{x^{3n}} + \frac{a_{3n-2}}{x^{3n-2}} + \ldots + \frac{a_{n+2}}{x^{n+2}} \right) e^{-\frac{1}{x^2}} = 0. \end{aligned}$$

In this argument we used that

$$\lim_{x\to 0} x^{-k} e^{-x^{-2}} = 0$$

for every $k > 0$. A proof of this claim is left for you as an exercise.

We have proved that the Taylor series of f at 0,

$$f(0) + f'(0)x + \frac{f''(0)}{2!}x^2 + \frac{f'''(0)}{3!}x^3 + ... = 0 + 0 + ... = 0.$$

So the sum is 0 for all $x \in \mathbb{R}$ and is only equal to $f(x)$ when $x = 0$. □

4 The Riemann integral

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

4.1 Motivation of the definition

Suppose we consider a point which moves along a line. We denote the position of the point at time t by $x(t)$. Suppose that we know the velocity of the point at every time t,

$$v(t) = \frac{dx(t)}{dt}.$$

How can we find out the distance, $x(b) - x(a)$, travelled by the point in the period which starts at $t = a$ and finishes at $t = b$? In other words, is it possible to find a function starting from its derivative?

If the derivative is constant, i.e. $v(t) = v_0$ for all t, the answer is obvious:

$$x(b) - x(a) = v_0(b - a).$$

In general, we can use the Mean Value Theorem to show that there is a point $c \in [a,b]$ such that

$$x(b) - x(a) = v(c)(b - a).$$

We already discussed that the theorem does not provide us with knowledge of the position of the point c inside the interval. In order to decrease uncertainty in the position of the intermediate point we can cut the interval $[a,b]$ into n pieces by choosing some points $t_0, t_1, \ldots, t_n$ such that

$$a = t_0 < t_1 < \ldots < t_{n-1} < t_n = b.$$

Obviously, $x(a) = x(t_0)$ and $x(b) = x(t_n)$. Then we can write down the sum

$$x(b) - x(a) = x(t_n) - x(t_{n-1}) + x(t_{n-1}) - x(t_{n-2}) + \ldots + x(t_1) - x(t_0).$$

The equality holds as the majority of the terms cancel each other. The points t_k divide the interval into n smaller intervals. We can apply the Mean Value theorem on each one separately: for every k, $1 \le k \le n$, there is a point $c_k \in [t_{k-1}, t_k]$ such that $x(t_k) - x(t_{k-1}) = v(c_k)(t_k - t_{k-1})$. Substituting these expressions we get

$$x(b) - x(a) = v(c_n)(t_n - t_{n-1}) + v(c_{n-1})(t_{n-1} - t_{n-2}) + \ldots + v(c_1)(t_1 - t_0).$$

In order to save space we rewrite this sum in the compact form:

$$x(b) - x(a) = \sum_{k=1}^{n} v(c_k)(t_k - t_{k-1}).$$

The sum in the right hand side of this equality is an example of a Riemann sum. It is possible that the value of the sum is not too sensitive to the choice of c_k provided all intervals are sufficiently small. In this case replacing c_k by an arbitrarily chosen point in $[t_{k-1}, t_k]$ introduces only a small error. Let us formulate this idea more precisely.

Suppose that there is a number $I \in \mathbb{R}$ which has the following property: for any $\varepsilon > 0$ there is $\delta > 0$ such that for any choice of t_k,

$$a = t_0 < t_1 < \ldots < t_{n-1} < t_n = b,$$

and any choice of $c_k \in [t_{k-1}, t_k]$, if $t_k - t_{k-1} \leq \delta$ for all k then

$$\left| I - \sum_{k=1}^{n} v(c_k)(t_k - t_{k-1}) \right| < \varepsilon.$$

If this property holds, the function v is Riemann integrable on the interval $[a,b]$ and I is its Riemann integral over the interval $[a,b]$. We write

$$\int_a^b v(t)\, dt = I.$$

This property can be taken as a definition for the Riemann integral (one of several equivalent definitions used in textbooks). This description of the integral involves a sum, called a Riemann sum, which is defined by the finite sequence of points t_k. These points define a mesh on $[a,b]$, and a selection of points $c_k \in [t_{k-1}, t_k]$. When the mesh size (=max distance between t_{k-1} and t_k) is small, the Riemann sum is close to the Riemann integral.

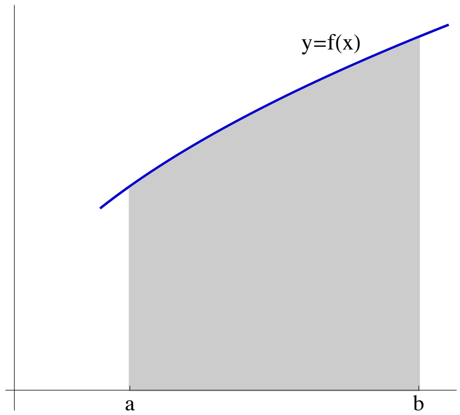
Let us come back to the motion of the point and suppose that the function $v(t) = x'(t)$ is integrable. We proved that for any finite sequence of t_k, there is a choice of c_k such that the Riemann sum equals to $x(b) - x(a)$. If the mesh size is less than δ from the definition we can conclude that $|I - (x(b) - x(a))| < \varepsilon$. Since ε is arbitrary we conclude that $I = x(b) - x(a)$. We have proved that

$$x(b) - x(a) = \int_a^b x'(t)\, dt.$$

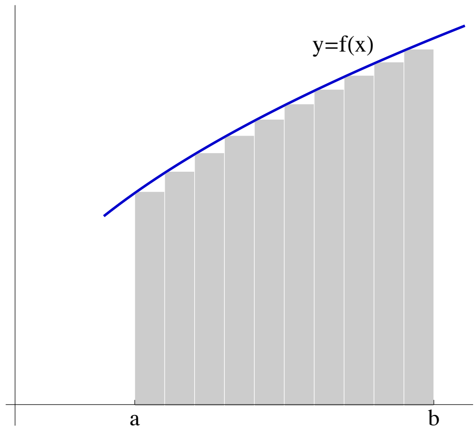
This equality can be interpreted that, in certain sense, the integral is the inverse of the derivative.

4.2 Riemann sums

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.



We calculate the total area of the rectangles and as the number of pieces increases, the total should approach the area we want.

Let n be the number of pieces. Then $h = \dfrac{b-a}{n}$ is the length of each piece. The endpoints of the rectangles are located at the points

$$x_k = a + kh$$

with $0 \le k \le n$.

In order to estimate the area under the graph we need to decide on the height for each of the rectangles.

The left hand rule suggests to use the value of f on the left hand side end of the corresponding interval. Then the approximate value of the area is given by

$$L_n(f) = f(x_0)h + f(x_1)h + f(x_2)h + \ldots + f(x_{n-1})h.$$

This sum is an example of Riemann sum. Since $x_k - x_{k-1} = h$ for every k we can write

$$L_n(f) = \sum_{k=1}^{n} f(x_{k-1})(x_k - x_{k-1}).$$

If f is Riemann integrable we can conclude that

$$\int_a^b f(x)\,dx = \lim_{n\to\infty} L_n(f).$$

The right hand rule suggests to use the value of f on the right hand side end of the corresponding interval. Then the approximate value of the area is given by

$$R_n(f) = f(x_1)h + f(x_2)h + \ldots + f(x_{n-1})h + f(x_n)h.$$

Therefore we see the second example of a Riemann sum,

$$R_n(f) = \sum_{k=1}^{n} f(x_k)(x_k - x_{k-1}).$$

We note that the difference between the areas defined by the left and right hand rules converges to zero as n increases:

$$R_n(f) - L_n(f) = f(x_n)h - f(x_0)h = \frac{(f(b) - f(a))(b-a)}{n} \to 0$$

as $n \to 0$. Therefore these two sequences converge or diverge simultaneously. If the sequences converge their limits coincide.

Let us consider some examples.

Example 4.1. *Let $a = 0$, $b = 1$ and consider the constant function $f(x) = 1$ for all x. Then it is easy to check that*

$$L_n(f) = R_n(f) = 1.$$

We conclude that

$$\int_0^1 1\,dx = 1.$$

Of course, the area under the graph also equals to one.

Example 4.2. *Let $a = 0$, $b = 1$ and consider the function $f(x) = x$ for all x. Then we note that $h = \frac{b-1}{n} = \frac{1}{n}$ and $x_k = kh = \frac{k}{n}$. Applying the right rule we get*

$$R_n(f) = \sum_{k=1}^{n} f(x_k)h = \sum_{k=1}^{n} \frac{k}{n^2} = \frac{n(n+1)}{2n^2} = \frac{1}{2} + \frac{1}{2n} \to \frac{1}{2}$$

as $n \to \infty$. We conclude that

$$\int_0^1 x\,dx = \frac{1}{2}.$$

We note that he area under the graph of f equals to one half.

Example 4.3. *Let $a = 0$, $b = 1$ and consider the function*

$$f(x) = \begin{cases} 0, & x \in \mathbb{Q} \\ 1, & x \notin \mathbb{Q} \end{cases}$$

Then $h = \frac{1}{n}$ and $x_k = \frac{k}{n}$. Since all x_k are rational we see that

$$R_n(f) = \sum_{k=1}^{n} f(x_k)h = 0.$$

Obviously the sequence of $R_n(f)$ converges to zero. Nevertheless the function f is not integrable as the definition of integrability requires that all Riemann sums, $\sum_1^n f(c_k)h$, converge to the same limit I. This property is not satisfied. Indeed, for example, if we choose irrational $c_k \in [x_{k-1}, x_k]$ for every k, we get the Riemann sum equal to 1 for every n.

We arrive to the following conclusion: from the definition of integrability we know that if a function $f : [a,b] \to \mathbb{R}$ is integrable then its integral is the limit of the Riemann sums obtained with the help of the right (or left) hand rule:

$$\int_a^b f(x)\,dx = \lim_{n\to\infty} R_n(f).$$

On the other hand, the existence of this limit does not automatically imply that the function is integrable. We can say that the integrability property gives the precise meaning to the notion of the area under the graph. If a function is integrable, the area is well defined and we can use Riemann sums to approximate the area.

In the theory of Riemann integral we consider Riemann sums of the form

$$\sum_{k=1}^{n} f(c_k)(x_k - x_{k-1})$$

where $c_k \in [x_{k-1}, x_k]$. Each term of this sum represents the area of a rectangle placed on the interval $[x_{k-1}, x_k]$. The height of the rectangle $f(c_k)$ depends on the choice of the point c_k and is allowed to take any value between the minimum and the maximum values of the function on this interval. If we choose the minimal values, the rectangles are located under the graph of f. If we choose the maximal values, the rectangles fully cover the graph. These two choices sandwich the area under the graph. For an integrable function, as the spacing between points x_k decreases both choices approach the area we want.

In order to establish integrability of a function it is sufficient to restrict the study to equally spaced points x_k. On the other hand the introduction of arbitrarily spaced points gives more flexibility to the theory and can be used to simplify some proofs. Moreover, Riemann sums can be used to evaluate an integral numerically. A well optimized program can reach the desired accuracy faster by using smaller number of points if it chooses positions of x_k wisely. Think for example about integrating the function $f(x) = \sin(1/x)$ in an interval located near zero.

4.3 The construction of the Riemann integral

In this section we are going to discuss an alternative approach to the definition of the Riemann integral. This definition is equivalent to the definition discussed in the previous sections. The advantage of this approach is linked to the absence of limits in the definitions which makes proofs of integrability easier. In the lectures we will not discuss the proofs in details. The parts which are not formally covered in Calculus 2 are typed in blue.

A partition P of the interval $[a,b]$ is 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.

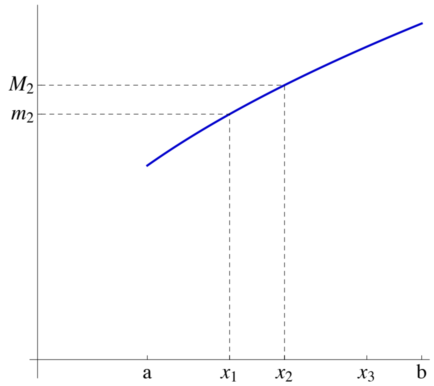
Now suppose that $f : [a,b] \to \mathbb{R}$ is a **bounded** function. For each i let

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

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.



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 4.4 (**Upper and lower sums**)**.** *Let $f : [a,b] \to \mathbb{R}$ 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*

$$U(f,P) = \sum_1^n M_i\,(x_i - x_{i-1}) \qquad \textit{and} \qquad L(f,P) = \sum_1^n m_i\,(x_i - x_{i-1})$$

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

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

Example 4.5. *Let $f : [0,1] \to [0,1]$ be given by $f(x) = x$ for each x. For a partition $P = \{x_0, x_1, \dots, x_n\}$ the lower sum is*

$$\sum_1^n x_{i-1}(x_i - x_{i-1}).$$

Without using facts about area and integrals show that this is less than $1/2$.

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

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

and

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

Then for any partition P

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

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

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

We can now take the sup and inf.

Definition 4.7 (**Upper and lower integrals**)**.** *Let $f : [a,b] \to \mathbb{R}$ 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 4.8 (**The Riemann integral**)**.** *Let* $f:[a,b]\to\mathbb{R}$ *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 4.9 (**A function that is not integrable**)**.** *Let* $f:[0,1]\to\mathbb{R}$ *be given by*

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

Then every lower sum of f is 0 and every upper sum is 1. So f is not integrable.

A slightly more careful analysis shows that any lower sum is smaller than any upper sum: for any two partitions P_1 and P_2

$$L(f,P_1)\le U(f,P_2).$$

It follows that the lower integral is not larger than the upper integral:

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

In the next section this inequality will help us to prove that continuous functions are integrable.

4.4 Integrable functions

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 4.10 (**Uniform continuity**). *If $f : [a,b] \to \mathbb{R}$ 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 4.11 (**Integrability of continuous functions**). *If $f : [a,b] \to \mathbb{R}$ is continuous then it is integrable.*

Proof. From Calculus 1 you know that any continuous function defined on a closed interval is bounded so we need to check that the lower and upper integrals are equal. 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 $[a,b]$ with each gap $x_i - x_{i-1}$ less than δ. Let M_i and m_i be the maximum and the minimum values of the function f on $[x_{i-1}, x_i]$. Then for each i

$$M_i - m_i \leq \frac{\varepsilon}{b-a}.$$

Therefore

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

Recalling the definition of the upper and lower integrals we conclude that

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

It follows that for any $\varepsilon > 0$

$$0 \le \overline{\int} f - \underline{\int} f \le \varepsilon.$$

We conclude that the upper and lower integrals are equal and, consequently, the function f is integrable. ☐

We have proved that on a closed interval the class of integrable functions includes all continuous functions. On the other hand an integrable function is not necessary continuous. For example let us prove that any monotone (increasing or decreasing) function is integrable.

Theorem 4.12 (Integrability of monotone functions). *If $f : [a,b] \to \mathbb{R}$ is bounded and either increasing or decreasing, then it is integrable.*

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

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

Now suppose we take a partition into n equal intervals. Then

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

Recalling the definition of the upper and lower integrals we conclude that

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

It follows that for any $n \in \mathbb{N}$

$$0 \le \overline{\int} f - \underline{\int} f \le U(f,P) - L(f,P) \le \frac{(b-a)(f(b) - f(a))}{n}.$$

We conclude that the upper and lower integrals are equal and, consequently, the function f is integrable. ☐

The proofs show that both for continuous and monotone functions it is enough to consider partitions into equal pieces.

We have establish that on a closed and bounded interval every continuous function and every monotone function is integrable. It is convenient to have a toolbox which can be used to check that a function is integrable without direct analysis of upper and lower sums.

Theorem 4.13 (**properties of integrable functions**). *If $f, g : [a,b] \to \mathbb{R}$ are integrable then*

1. *$f + g$ is integrable,*
2. *αf is integrable for any $\alpha \in \mathbb{R}$,*
3. *$|f|$ is integrable,*
4. *$f \cdot g$ is integrable,*
5. *if $\phi : \mathbb{R} \to \mathbb{R}$ is continuous then $\phi \circ f$ is integrable on $[a,b]$.*

The proof of this theorem is placed in the appendix. If you read this text for an exam revision it is not necessary to study that section.

4.5 Riemann integral and mesh size of partitions

The mesh size of a partition is the length of the longest interval in the partition. For a partition $P = \{x_0, x_1, \ldots, x_n\}$ we write

$$\|P\| = \max_{1 \le k \le n} |x_k - x_{k-1}|.$$

If a function is integrable then we can use a partition with sufficiently small mesh size to approximate the integral with arbitrary precision.

Theorem 4.14 (**Small mesh**). *If $f : [a,b] \to \mathbb{R}$ is integrable, then for any $\varepsilon > 0$ there is $\delta > 0$ such that for any partition P with $\|P\| < \delta$*

$$\left| U(f,P) - \int_a^b f(x)dx \right| < \varepsilon$$

and similarly for the lower sums.

Corollary 4.15. *If $f : [a,b] \to \mathbb{R}$ is integrable and we pick a sequence of partitions P_n whose mesh sizes tend to 0 then*

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

similarly for the lower sums and for any sequence of Riemann sums based on the partitions P_n.

In particular, any Riemann integral can be obtained as the limit of Riemann sums defined with the help of the right hand rule (or left hand rule).

Corollary 4.16. *If $f : [a,b] \to \mathbb{R}$ is integrable then*

$$\int_a^b f(x)\,dx = \lim_{n\to\infty} \frac{b-a}{n} \sum_{k=1}^{n} f(a+k(b-a)/n).$$

This corollary implies that we can approximate the integral of an integrable function using partitions with equally spaced points. Then we can use the known properties of limits to establish properties of integrals.

4.6 The basic properties of the integral

Theorem 4.17 (**Linearity of the integral**)**.** *If $f,g : [a,b] \to \mathbb{R}$ are integrable and $\lambda \in \mathbb{R}$ then λf and $f+g$ are integrable. Moreover*

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

and

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

Proof. The proof of integrability of λf and $f+g$ can be found in the appendix. After the integrability is established the second claim of the theorem follows easily from the algebra of limits (as Riemann sums are linear). □

Another important property of the integral is monotonicity: if $f \le g$ at each point then $\int f \le \int g$.

Theorem 4.18 (**Monotonicity of the integral**)**.** *If $f,g : [a,b] \to \mathbb{R}$ are integrable and $f(x) \le g(x)$ for all $x \in [a,b]$ then*

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

An important consequence of what we have done so far is a version of the triangle inequality for integrals rather than sums.

Corollary 4.19 (). *If $f : [a,b] \to \mathbb{R}$ is integrable then $|f|$ is integrable and*

$$\left| \int_a^b f(x)\,dx \right| \leq \int_a^b |f(x)|\,dx.$$

We have one principle for integrals that has no analogue for derivatives.

Theorem 4.20 (**The addition of ranges**). *Let $f : [a,c] \to \mathbb{R}$ be bounded and $a < b < c$. Then f is integrable on $[a,c]$ if and only if it is integrable on $[a,b]$ and $[b,c]$.*

If f is integrable on $[a,c]$ then

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

Proof. (Sketch) We will assume the first part of the theorem without proof. Then we know that f is integrable on $[a,b]$, $[b,c]$ and $[a,c]$. Take any partitions Q of $[a,b]$ and S of $[b,c]$. Together Q and S define a partition P of $[a,c]$. Moreover, $U(f,P) = U(f,Q) + U(f,S)$. If we decrease the mesh size of P towards 0, the left hand side converges to $\int_a^c f(x)\,dx$ while the right hand side converges to $\int_a^b f(x)\,dx + \int_b^c f(x)\,dx$. □

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.

4.7 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 a function F is differentiable, then F' is integrable and

$$F(x) = F(a) + \int_a^x F'(t)dt \ ?$$

  No, because the derivative might be unbounded, hence non-integrable. (Can you find an example?)

- 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 the function F defined by

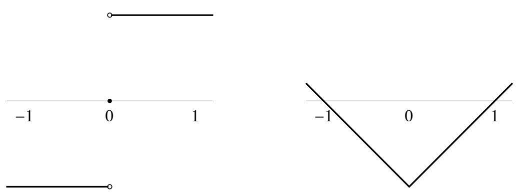
$$F(x) = \int_a^x f(t)\,dt$$

  is differentiable with derivative $F' = f$?

  No. For example, if $f : [-1,1] \to \mathbb{R}$ is given by

$$f(x) = \begin{cases} -1, & \text{if } -1 \leq x \leq 0 \\ 1, & \text{if } \ 0 < x \leq 1 \end{cases}$$

  then the integral $\int_{-1}^x f(x)\,dx = |x| - 1$ is not differentiable at 0.



Theorem 4.21 (**The Fundamental Theorem of Calculus I**)**.** *Let $f : [a,b] \to \mathbb{R}$ 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 both in $[a,b]$ with $h>0$ then

$$F(x+h)-F(x)=\int_x^{x+h} f(t)\,dt.$$

Integrable functions are bounded. Let M be a bound for f: so $|f(x)|\le M$ for all $x\in[a,b]$. Then

$$|F(x+h)-F(x)|=\left|\int_x^{x+h} f(t)\,dt\right|\le\int_x^{x+h}|f(t)|\,dt\le Mh.$$

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 $0<h<\delta$

$$\begin{aligned}\left|\frac{F(u+h)-F(u)}{h}-f(u)\right| &= \left|\frac{1}{h}\int_u^{u+h} f(t)\,dt-\frac{1}{h}\int_u^{u+h} f(u)\,dt\right|\\ &= \left|\frac{1}{h}\int_u^{u+h}(f(t)-f(u))\,dt\right|\\ &\le \frac{1}{h}\int_u^{u+h}|f(t)-f(u)|\,dt\\ &\le \frac{1}{h}\int_u^{u+h}\varepsilon\,dt=\varepsilon.\end{aligned}$$

Similarly for $h<0$. Hence $F'(u)$ exists and equals $f(u)$. □

Now suppose we have a nice function like $f(t)=t^2$. The FTC tells us that the function

$$F(x)=\int_0^x t^2\,dt$$

satisfies $F'(u)=u^2$ for all u and $F(0)=0$. Earlier we proved that $(x^3/3)'=x^2$. Then by the MVT we know that F is of the form $F(x)=x^3/3+C$. Substituting $x=0$ we get that $C=0$ and consequently $F(x)=x^3/3$.

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 4.22 (The Fundamental Theorem of Calculus II). *Let $F:[a,b]\to\mathbb{R}$ be continuous on $[a,b]$ and differentiable on (a,b) with $F'=f$.* **If f is Riemann integrable** *then*

$$\int_a^b f(t)\,dt = F(b)-F(a).$$

In other words we don't need to assume that f is continuous, merely that it is integrable.

Proof. It suffices to show that for each partition P

$$L(f,P)\le F(b)-F(a)\le U(f,P)$$

because if f is integrable the upper and lower sums get pushed together at $F(b)-F(a)$. This is where we use the integrability of f.

Let $P=\{x_0,x_1,\dots,x_n\}$ 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

$$f(c_i)(x_i-x_{i-1})=F'(c_i)(x_i-x_{i-1})=F(x_i)-F(x_{i-1}).$$

Hence for each i

$$m_i(x_i-x_{i-1})\le F(x_i)-F(x_{i-1})\le M_i(x_i-x_{i-1}).$$

Summing over all i gives

$$L(f,P)\le F(b)-F(a)\le U(f,P).$$ □

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.

We can now integrate the usual functions: polynomials, exponentials and rational functions in partial fractions. In particular we have

$$\log x=\int_1^x \frac{1}{t}\,dt.$$

4.8 Methods of integration

We can also check the integrated versions of the product and chain rules: integration by parts and integration by substitution.

Theorem 4.23 (Integration by parts). *Suppose $f, g : [a,b] \to \mathbb{R}$ 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 4.24 (Integration by substitution). *Suppose $u : [a,b] \to \mathbb{R}$ 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.

4.9 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 4.25 (Improper integrals). *Suppose $f : [a,b] \to \mathbb{R}$ 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 4.26 (**Improper integrals**). *Suppose $f : [a,\infty) \to \mathbb{R}$ 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 4.27 (**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.

We have an analogue of the comparison test.

Theorem 4.28 (**Comparison test for improper integrals**)**.** *Suppose $f,g:[a,\infty)\to\mathbb{R}$ are integrable on each interval $[a,b]$, that $|f(x)|\le g(x)$ for all $x\ge a$ 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

$$b \mapsto \int_a^b f(x)\,dx. \qquad \square$$

Example 4.29. *For each $\lambda > 0$*

$$\int_0^\infty e^{-\lambda x}\,dx = \frac{1}{\lambda}.$$

Proof. Exercise. $\square$

Example 4.30.

$$\int_0^\infty e^{-x} x^p\,dx$$

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

$$e^{-x} x^p \le K e^{-x/2}$$

because

$$e^{-x/2} x^p \to 0$$

as $x \to \infty$. We already saw that

$$\int_1^\infty e^{-x/2}\,dx$$

converges. $\square$

Handout: Uniform continuity

Calculus 2 (2022-20223)

In this handout we are going to discuss the concept of uniform continuity. In Calculus 1 you studied the definition of a continuous function. Assume that $a < b$ are two real numbers and consider a function $f : [a, b] \to \mathbb{R}$. Read carefully the following definitions.

Definition 1. The function f is called *continuous at a point* $x \in [a, b]$ if for every $\epsilon > 0$ there is $\delta > 0$ such that

$$\text{if } |y - x| < \delta \text{ and } y \in [a, b] \text{ then } |f(y) - f(x)| < \epsilon.$$

Definition 2. The function f is called *continuous on* $[a, b]$ if it is continuous at every point $x \in [a, b]$.

Definition 3. The function f is called *uniformly continuous on* $[a, b]$ if for every $\epsilon > 0$ there is $\delta > 0$ such that

$$\text{if } |y - x| < \delta \text{ and } x, y \in [a, b] \text{ then } |f(y) - f(x)| < \epsilon.$$

There is a subtle difference between Definitions 2 and 3.

The following claim is easy to prove — try to do it. If f is uniformly continuous on $[a, b]$ then it is continuous at every point $x \in [a, b]$ and, consequently, it is continuous on $[a, b]$. Consequently, every uniformly continuous function is automatically continuous.

Is it true that every continuous function is uniformly continuous?

If you try to answer this question yourself, you will soon see that the answer is not obvious. Although Definitions 1+2 and 3 look very similar they are not identical: Definition 2 picks a point x and then chooses a δ according to Definition 1, while Definition 3 uses a single δ for all x.

Exercise 1. Show that the function $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^2$ is continuous but not uniformly continuous.

Exercise 2. Show that the function $f : (0, 1) \to \mathbb{R}$ defined by $f(x) = 1/x$ is continuous but not uniformly continuous.

Exercise 3. Show that the function $f : (0, 1) \to \mathbb{R}$ defined by $f(x) = \sin(1/x)$ is continuous but not uniformly continuous.

Note that in Exercises 1 and 2 the function is not bounded, while in Exercise 3 the function is bounded. We see that a continuous function is not necessarily uniformly continuous.

Note that in Example 1 the domain of f is not bounded and in Examples 2 and 3 the domain of f is an open interval $(0, 1)$.

On the other hand, on a closed bounded interval, a function is continuous if (and only if) it is uniformly continuous. Let use state this claim in the form of a theorem.

Theorem (**uniform continuity**). *If a function $f : [a, b] \to \mathbb{R}$ is continuous on a closed and bounded interval $[a, b]$, then it is uniformly continuous on $[a, b]$:*

for every $\epsilon > 0$ there is $\delta > 0$ such that $|f(y) - f(x)| < \epsilon$ for any $x, y \in [a, b]$ with $|y - x| < \delta$.

Proof. Suppose that a function $f : [a, b] \to \mathbb{R}$ is not uniformly continuous on $[a, b]$. Then there is $\epsilon_0 > 0$ such that for any $\delta > 0$ there are two points $x, y \in [a, b]$ such that $|y - x| < \delta$ and $|f(y) - f(x)| \geq \epsilon_0$.

In particular we can choose any $n \in \mathbb{N}$ and let $\delta = \frac{1}{n}$. In this way for every n we find two points $x_n, y_n \in [a, b]$ such that

$$|y_n - x_n| < \tfrac{1}{n} \qquad \text{and} \qquad |f(y_n) - f(x_n)| \geq \epsilon_0.$$

Now we recall the Bolzano Weierstrass Theorem of Calculus 1 (p.34 of Lecture Notes by Roger Tribe): every bounded sequence (x_n) has a convergent subsequence. Obviously (x_n) is bounded as $a \leq x_n \leq b$. Thus there is a subsequence (x_{n_k}) and a point $c \in [a, b]$ such that

$$x_{n_k} \to c \quad \text{as } k \to \infty.$$

Then we note that we can use the triangle inequality to show that

$$|y_{n_k} - c| \leq |y_{n_k} - x_{n_k}| + |x_{n_k} - c| \leq \frac{1}{n_k} + |x_{n_k} - c| \to 0 \quad \text{as } k \to \infty.$$

Consequently,

$$y_{n_k} \to c \quad \text{as } k \to \infty.$$

We see that both x_{n_k} and y_{n_k} converge to the same limit c. On the other hand for every k

$$|f(x_{n_k}) - f(y_{n_k})| \geq \epsilon_0 > 0.$$

Consequently, the sequences $(f(x_{n_k}))$ and $(f(y_{n_k}))$ do not converge to the same limit, $f(c)$. With the help of the theorem on sequential continuity (Calculus 1, Lecture Notes by Roger Tribe, p.58), we conclude that f is not continuous at $c \in [a, b]$.

We supposed that f is not uniformly continuous on $[a, b]$ and proved that f is not continuous of $[a, b]$. Consequently, every continuous function on $[a, b]$ has be uniformly continuous on $[a, b]$. $\square$

The claim that every uniformly continuous function is continuous is an easy exercise.