MATH2017 Real Analysis, 2022/2023

Contact Details

Lecturer:	Professor Martin Speight
Office:	Mathematics 9.11
E-mail:	j.m.speight@leeds.ac.uk
Web:	https://cp1lump.github.io/

All handouts will be made available on the module homepage, which may be accessed via Minerva.

Assessment arrangements

There will be a $2\frac{1}{2}$ hour exam during the May/June exam period, counting 85% towards your final module grade.

There will also be 5 homework assignments, collectively counting 15% towards your final grade, due in **by 5pm** on the following dates:

- Monday 20 February 2023
- Monday 6 March 2023
- Monday 20 March 2023
- Monday 3 April 2023
- Monday 15 May 2023

These are Mondays in even-numbered teaching weeks, starting in week 4 (and the Monday immediately after teaching finishes). Submission will be electronic, via Gradescope: upload a scanned PDF of your work using the Gradescope link on the module's Minerva page. It is your responsibility to correctly tag your upload so that the graders are guided to the relevant part of your script for each question. If you do not, graders can reasonably assume that you have not answered that question and give you a mark of 0. You are advised to leave at least one hour to complete the upload process. Help and advice on the upload process can be found on the Gradescope website.

Each homework assignment will be supported by two workshop sessions. **It is very important that you attend these.**

Late submissions

Late homework submissions will attract a 2 point penalty (out of a total of 10) for each day, or part thereof, that they are late. Model solutions to the homework assignments will be posted shortly after 5pm on the Wednesday following the submission deadline.

Submissions after this point will not be accepted. If you have a valid reason for missing a homework deadline (for example, illness), please inform the Taught Students' Office. You will be granted an exemption from that homework.

Directed reading list

There are two recommended texts for this course:

- J.M. Speight, "A sequential introduction to real analysis," World Scientific, New Jersey, (2016).
- Robert G. Bartle, and Donald R. Sherbert, "Introduction to real analysis," 4th ed., Wiley, New Jersey, (2011).

Both of these texts are for background reading only. All the material studied will be covered thoroughly in lectures.

Contents

What are we going to study? Why?

Differential calculus is arguably the single most powerful and influential mathematical idea ever invented. Its applications pervade the natural sciences, and it is essential to most applied mathematics and much pure mathematics too (for example, in my own research specialism, Differential Geometry). For scientists in general, it's OK to use calculus as a computational method, without worrying overmuch about what the objects being calculated (derivatives, integrals, limits) really mean. We, however, are *mathematicians*, and we hold ourselves to a much higher standard. For us, it is essential that every object we use has a precise *definition*, and that general facts about, and relationships between, these objects are *proved*. It is not good enough simply to appeal to "common sense," or claim that a given assertion is "obvious": **just because something is "obvious" doesn't mean it's true!**

Example 0.1 Say we have a function $f : \mathbb{R} \to \mathbb{R}$ which is differentiable everywhere and has $f'(a) > 0$ at some particular point a. Then obviously, f is a strictly increasing function on some neighbourhood of a, right? That is, there exists $\delta > 0$ (possibly extremely small) such that, for all $x, y \in (a - \delta, a + \delta)$, if $x < y$ then $f(x) < f(y)$. Well, actually, no: this doesn't follow, and we will construct a counterexample shortly (Example 4.13 if you're impatient).

If something as "obvious" as this isn't necessarily true, how do we know that *any* of the rules of calculus, which you've informally derived and used profusely (the chain rule, product rule, quotient rule, fundamental theorem of the calculus etc. etc.), are reliable? We don't, until we have *proved* them, and we can't *prove* theorems about derivatives and integrals until we have properly *defined* derivatives and integrals.

The object of this module is to do just this: provide a rigorous development of calculus for functions $f : \mathbb{R} \to \mathbb{R}$, with all terms precisely defined, and all assertions proved. The central object in calculus is the *derivative*. By now, you should be well aware that this is actually a *limit*:

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

So our first task will be to precisely define limits of functions. Luckily, you've already thought carefully about limits, at least in one context: limits of convergent *sequences*. We will begin, therefore, by reminding ourselves what exactly it means to say that a real sequence *converges* to a limit, then see how the basic underlying idea can be adapted to deal with limits of functions.

Chapter 1

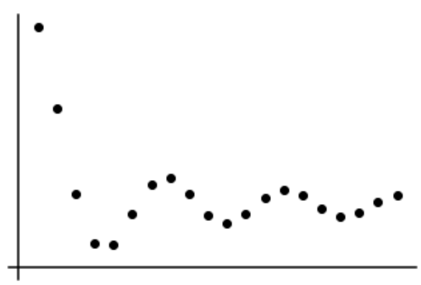
Limits of sequences

1.1 The limit of a convergent sequence

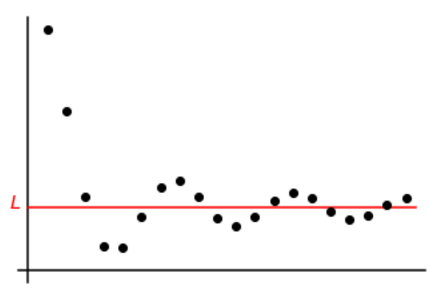
Recall the key definition from MATH1026 Sets, Sequences and Series (also covered in MATH1055 Numbers and Vectors), that of the *limit* of a *convergent* real sequence:

Definition 1.1 A real sequence (a_n) **converges to the limit** L if, for each $\varepsilon > 0$, there exists $N \in \mathbb{Z}^+$ such that, for all $n \geq N$, $|a_n - L| < \varepsilon$. The shorthand for this is $a_n \to L$ or $\lim_{n\to\infty} a_n = L$.

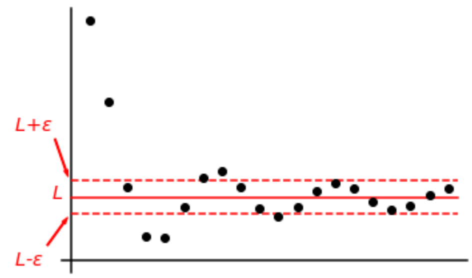
We can give Definition 1.1 a geometric interpretation. Imagine plotting the graph of the sequence (a_n):



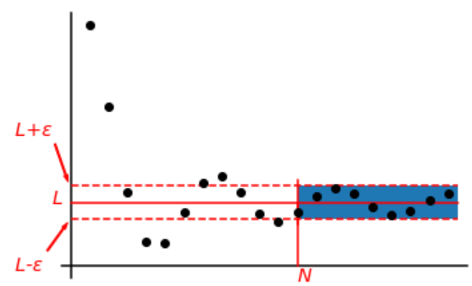
Now draw the horizontal line $y = L$, where L is our claimed limit:



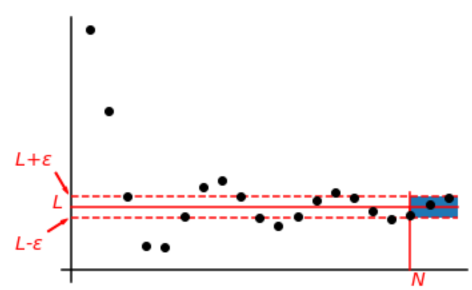
Then the statement that (a_n) converges to L means that, if we draw a horizontal strip of width 2ε centred on the line $y = L$, then, no matter how small we choose $\varepsilon > 0$, there is a vertical line $x = N$ to the right of which all points on the graph lie in the horizontal strip. For example, if we choose this value of ε:



then this value of N (or any N larger than this) will do:

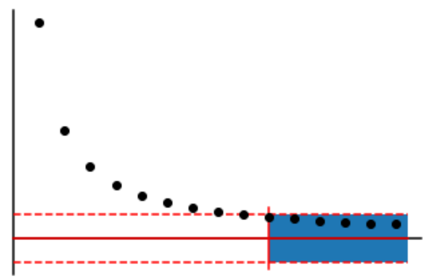


If we choose a smaller value of ε, we need to make N larger:



But no matter how small we make ε, there is always a point on the graph to the right of which all points lie in the narrow strip.

Pictures can be a powerful aid to understanding in Real Analysis, and I strongly encourage you to develop the habit of drawing them to illustrate definitions or to organize your thoughts at the beginning of a problem or proof. It's not hard, for example, to draw a picture that convinces us that the simple sequence $1/n$ converges to 0:



Of course, it's also vital that you can convert the picture into a proof:

Example 1.2 Claim: $a_n = 1/n \to 0$.

Proof: Let $\varepsilon \in (0, \infty)$ be given. Then there exists $N \in \mathbb{Z}^+$ such that $N > 1/\varepsilon$ (because the set $\mathbb{Z}^+$ is *unbounded above*). But then, for all $n \geq N$,

$$|a_n - 0| = \frac{1}{n} \leq \frac{1}{N} < \varepsilon.$$

$\square$

In the proof above, we showed that, for any prescribed $\varepsilon > 0$, there does indeed exist a positive integer N such that, for all $n \geq N$, $|a_n - 0| < \varepsilon$ (namely, any $N > 1/\varepsilon$ will do). A proof like this is called a proof "from first principles", or an "ε—N proof." The key step was to estimate (that is, bound above) the quantity $|a_n - L|$.

Example 1.3 Give a direct ε—N proof that $x_n = 2n/(n+1) \to 2$.

Solution Rough work: we begin by examining the quantity $|x_n - 2|$:

$$\begin{aligned} |x_n - 2| &= \left|\frac{2n}{n+1} - 2\right| \\ &= \left|\frac{2n - 2n - 2}{n+1}\right| = \frac{2}{n+1} \\ &< \frac{2}{n} \end{aligned}$$

OK. So, for a given constant $\varepsilon > 0$, how big should we make n to ensure that $|x_n - 2| < \varepsilon$? Answer: making n bigger than $\boxed{\frac{2}{\varepsilon}}$ will do. We're now ready to write out our proof.

Proof: Let $\varepsilon \in (0, \infty)$ be given. Then there exists $N \in \mathbb{Z}^+$ such that

$$N > \frac{2}{\varepsilon} \qquad .$$

Now, for all $n \geq N$,

$$\begin{aligned} |x_n - 2| &= \left|\frac{2n}{n+1} - 2\right| \\ &= \left|\frac{2n - 2n - 2}{n+1}\right| = \frac{2}{n+1} \\ &< \frac{2}{n} \\ &\leq \frac{2}{N} \\ &< \varepsilon. \end{aligned}$$

□

Many fundamental properties of convergent sequences follow quickly from Definition 1.1.

Proposition 1.4 (Uniqueness) *The limit of a convergent sequence is unique.*

Proof: Assume, towards a contradiction, that $x_n \to L_1$ and $x_n \to L_2$ where $L_1 \neq L_2$. Let $\varepsilon = |L_1 - L_2|/2 > 0$. Since $x_n \to L_1$, there exists $N_1 \in \mathbb{Z}^+$ such that, for all $n \geq N_1$, $|x_n - L_1| < \varepsilon$. Since $x_n \to L_2$, there exists $N_2 \in \mathbb{Z}^+$ such that, for all $n \geq N_2$, $|x_n - L_2| < \varepsilon$. Now let n be any positive integer greater than both N_1 and N_2 (for example, $n = N_1 + N_2$ will do). Then

$$\begin{aligned} 2\varepsilon &= |L_1 - L_2| = |(x_n - L_2) - (x_n - L_1)| \\ &\leq |x_n - L_2| + |x_n - L_1| \qquad \text{(by the triangle inequality)} \\ &< |x_n - L_2| + \varepsilon \qquad \text{(since } n > N_1\text{)} \\ &< \varepsilon + \varepsilon \qquad \text{(since } n > N_2\text{)} \end{aligned}$$

Hence, $2\varepsilon < 2\varepsilon$, a contradiction. It follows that our original assumption that x_n converges to two different limits is false. □

Definition 1.5 A sequence (a_n) is **bounded above** if there exists $M \in \mathbb{R}$ such that, for all $n \in \mathbb{Z}^+$, $a_n \leq M$. Any such M is called an **upper bound** on (a_n). A sequence (a_n) is **bounded below** if there exists $K \in \mathbb{R}$ such that, for all $n \in \mathbb{Z}^+$, $a_n \geq K$. Any such K is called a **lower bound** on (a_n). The sequence is **bounded** if it bounded above and below.

Proposition 1.6 (Boundedness) *If (a_n) converges then (a_n) is bounded (that is, there exists $K \in \mathbb{R}$ such that, for all $n \in \mathbb{Z}^+$, $|a_n| \leq K$).*

Proof: Exercise. □

Note the converse of Proposition 1.6 is *false*. E.g. $a_n = (-1)^n$ is bounded, but not convergent. If a sequence does not converge, we say it **diverges** (or is **divergent**). So $(-1)^n$ diverges. Don't confuse the concepts of divergence and unboundedness.

Proposition 1.7 (Limits preserve non-strict inequalities) *If $a_n \to L$ and, for all $n \in \mathbb{Z}^+$, $K \leq a_n \leq M$, then $K \leq L \leq M$.*

Proof: Assume, towards a contradiction, that $L \notin [K, M]$. Then $L < K$ or $L > M$. We handle each of these two cases separately.

If $L < K$, then let $\varepsilon = K - L > 0$. Since $a_n \to L$, there exists $N \in \mathbb{Z}^+$ such that, for all $n \geq N$, $|a_n - L| < \varepsilon$, and hence $L - \varepsilon < a_n < L + \varepsilon$. In particular, $a_N < L + \varepsilon = L + (K - L) = K$. But *all* terms a_n of the sequence satisfy $a_n \geq K$, a contradiction. Hence $L \geq K$.

I leave the case where $L > M$ as an exercise. □

Another way of stating this is that, if $K \leq a_n \leq M$ for all $n \in \mathbb{Z}^+$ then

$$K \leq \lim_{n\to\infty} a_n \leq M,$$

assuming the limit exists. So limits of sequences preserve non-strict inequalities. We will make very frequent use of this fact, so it's important you understand and can prove it.

What if we know that a_n actually obeys *strict* inequalities,

$$K < a_n < M \quad ?$$

Can we conclude that

$$K < \lim_{n\to\infty} a_n < M \quad ?$$

Sadly, no, and Example 1.2 provides a counterexample: $0 < 1/n < 2$ for all $n \in \mathbb{Z}^+$, but its limit $L = 0$ does not satisfy $0 < L < 2$.

Proposition 1.8 (The Squeeze Rule) *Assume $a_n \leq b_n \leq c_n$, $a_n \to L$ and $c_n \to L$. Then $b_n \to L$.*

Proof: Let $\varepsilon \in (0, \infty)$ be given. We must show that there exists $N \in \mathbb{Z}^+$ such that, for all $n \geq N$, $|b_n - L| < \varepsilon$.

Since $a_n \to L$, there exists $N_1 \in \mathbb{Z}^+$ such that, for all $n \geq N_1$, $|a_n - L| < \varepsilon$, and hence $a_n > L - \varepsilon$.

Since $c_n \to L$, there exists $N_2 \in \mathbb{Z}^+$ such that, for all $n \geq N_2$, $|c_n - L| < \varepsilon$, and hence $c_n < L + \varepsilon$.

Hence, for all $n \geq N = \max\{N_1, N_2\}$,

$$b_n \geq a_n > L - \varepsilon,$$

since $n \geq N_1$, and

$$b_n \leq c_n < L + \varepsilon,$$

since $n \geq N_2$. So, for all $n \geq N$, $|b_n - L| < \varepsilon$. $\square$

So if a sequence (b_n) is "squeezed" between two sequences which both converge to a common limit L, (b_n) converges to L also. We will use this fact very frequently.

Proposition 1.9 (The Algebra of Limits) *If $a_n \to A$ and $b_n \to B$ then*

(i) $a_n + b_n \to A + B$.

(ii) $a_n b_n \to AB$.

(iii) $a_n/b_n \to A/B$, *provided* $B \neq 0$ *and, for all* $n \in \mathbb{Z}^+$, $b_n \neq 0$.

Proof: I leave the proofs of parts 1 and 3 as exercises, and present a proof of part 2.

Let $\varepsilon \in (0, \infty)$ be given. We must show that there exists $N \in \mathbb{Z}^+$ such that, for all $n \geq N$,

$$|a_n b_n - AB| < \varepsilon.$$

We begin by noting that

$$\begin{aligned} |a_n b_n - AB| &= |a_n b_n - a_n B + a_n B - AB| \\ &\leq |a_n||b_n - B| + |B||a_n - A| \end{aligned}$$

Since $a_n \to A$, there exists $N_1 \in \mathbb{Z}^+$ such that, for all $n \geq N_1$, $|a_n - A| < 1$, and hence, $a_n < A + 1 \leq |A| + 1$ and $a_n > A - 1 \geq -|A| - 1$, so $-a_n \leq |A| + 1$. Hence, for all $n \geq N_1$, $|a_n| < |A| + 1$.

Since $b_n \to B$, there exists $N_2 \in \mathbb{Z}^+$ such that, for all $n \geq N_2$,

$$|b_n - B| < \frac{\varepsilon}{2(|A| + 1)}.$$

(Here we're just applying the definition of convergence to A for the particular choice of positive real number $\varepsilon' = \varepsilon/(2|A| + 2)$.)

Since $a_n \to A$, there exists $N_3 \in \mathbb{Z}^+$ such that, for all $n \geq N_3$,

$$|a_n - A| < \frac{\varepsilon}{2(|B| + 1)}.$$

(Here we're just applying the definition of convergence to B for the particular choice of positive real number $\varepsilon'' = \varepsilon/(2|B| + 2)$.)

Now, let $N = \max\{N_1, N_2, N_3\}$. Then, for all $n \geq N$,

$$\begin{aligned}
|a_n b_n - AB| &\leq |a_n||b_n - B| + |B||a_n - A| \\
&\leq (|A|+1)|b_n - B| + |B||a_n - A| \qquad \text{(since } n \geq N_1\text{)} \\
&< (|A|+1)\frac{\varepsilon}{2(|A|+1)} + |B||a_n - A| \qquad \text{(since } n \geq N_2\text{)} \\
&< \frac{\varepsilon}{2} + |B|\frac{\varepsilon}{2(|B|+1)} \qquad \text{(since } n \geq N_3\text{)} \\
&< \varepsilon
\end{aligned}$$

□

Informally speaking, a subsequence (b_k) of a sequence (a_n) is a sequence all of whose terms occur in (a_n), in the correct order. More precisely:

Definition 1.10 (b_k) is a **subsequence** of (a_n) if there exists a strictly increasing sequence of positive integers (n_k) such that $b_k = a_{n_k}$.

For example $b_k = 1/2^k$ is a subsequence of $a_n = 1/n$: take $n_k = 2^k$. So is $c_k = 1/(k+50)$: take $n_k = k + 50$. However $d_k = k/(k+1)$ is not a subsequence of (a_n) because, for example, $b_2 = 2/3$ is not a term in (a_n). Neither is the sequence $(e_k) = (1, 1/2, 1, 1/3, 1, 1/4, \ldots)$. All the terms of (e_k) do appear in (a_n), but not in the correct order.

The key fact about subsequences is the following:

Proposition 1.11 *If $a_n \to L$ and (b_k) is a subsequence of (a_n), then $b_k \to L$.*

Proof: Exercise. □

So to prove that a sequence *diverges* it's enough to show that it has a subsequence that diverges, or to find two subsequences which converge to different limits (why? Look at Proposition 1.4).

Example 1.12 The sequence $x_n = (-1)^n$ diverges.

Proof: Assume, towards a contradiction, that $x_n \to L$. The even subsequence $y_n = x_{2n} = 1$ which converges to 1. But, by Proposition 1.11, $y_n \to L$. Limits are unique, by Proposition 1.4, so $L = 1$. But the odd subsequence $z_n = x_{2n-1} = -1$ converges to -1. Again, by Proposition 1.11, $z_n \to L$, and limits are unique, by Proposition 1.4, so $L = -1$. Hence $1 = -1$, a contradiction. □

So far, all the facts we've established about limits of sequences follow very directly from Definition 1.1. The following two results are rather deeper. To prove them, one needs to appeal to a fundamental defining property of the set of real numbers called *completeness*. We will return to this later in the module, when we study the theory of integration, but for the time being you should just take them on trust (or, if you studied MATH1026 last year, consult your notes from there). The first result is called the Montone Convergence Theorem, and to state it we need a preliminary definition:

Definition 1.13 A sequence (a_n) is **increasing** if $a_{n+1} \geq a_n$ for all $n \in \mathbb{Z}^+$. It is **decreasing** if $a_{n+1} \leq a_n$ for all $n \in \mathbb{Z}^+$. It is **monotonic** (or **monotone**) if it is increasing or decreasing.

So $(1/n)$ is decreasing, $(2n + 1)$ is increasing, and hence both are monotonic. The constant sequence $(1, 1, 1, \ldots)$ is both increasing and decreasing , hence monotonic. The sequence $(-1, 1, -1, 1, -1, 1, \ldots)$ is neither increasing nor decreasing, so is not monotonic.

Theorem 1.14 (Montone Convergence Theorem) *If (a_n) is bounded and monotonic, then (a_n) converges.*

Finally we have the wonderfully named Bolzano-Weierstrass[1] Theorem:

Theorem 1.15 (Bolzano-Weierstrass Theorem) *Every bounded real sequence has a convergent subsequence.*

In fact, the Bolzano-Weierstrass Theorem follows quite easily from the Monontone Convergence theorem: we just show that *every* sequence has a monotonic subsequence. If the sequence is bounded, so is its monotonic subsequence, which much converge by Theorem 1.14.

1.2 Convergence of sequences and the Cauchy property

One slightly inconvenient thing about Definition 1.1 is that, to prove a sequence converges from first principles, we must first know what its limit is. In this section we will study a criterion which turns out to be *equivalent* to convergence, but which makes no mention of limits. This will turn out to be extremely useful later on, when we come to analyze the properties of functions defined by power series.

Definition 1.16 A real sequence (a_n) is **Cauchy** (or **has the Cauchy property**) if, for each $\varepsilon > 0$, there exists $N \in \mathbb{Z}^+$ such that, for all $n, m \geq N$, $|a_n - a_m| < \varepsilon$.

Remark This looks very similar to the definition of convergence. It says that, given any $\varepsilon > 0$ (no matter how small), there is a point in the sequence beyond which *all terms lie closer than distance ε from one another.* Note that it makes no mention of any limit.

We will show that real sequences converge if and only if they have the Cauchy property. Once again, this is revision if you studied MATH1026, but is probably new if not.

Lemma 1.17 *If (a_n) converges then (a_n) is Cauchy.*

[1] Weiertstraß wenn du Deutsch bist.

Proof: Assume $a_n \to L$. Let $\varepsilon \in (0, \infty)$ be given. Then $\varepsilon/2 > 0$, so there exists $N \in \mathbb{Z}^+$ such that, for all $n \geq N$,

$$|a_n - L| < \frac{\varepsilon}{2}.$$

But then, for all $n, m \geq N$,

$$\begin{aligned} |a_n - a_m| &= |(a_n - L) - (a_m - L)| \\ &\leq |a_n - L| + |a_m - L| \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon \end{aligned}$$

by the Triangle Inequality. Hence (a_n) is Cauchy. □

Lemma 1.18 *If (a_n) is Cauchy then (a_n) is bounded.*

Proof: Since (a_n) is Cauchy, there exists $N \in \mathbb{Z}^+$ such that, for all $n, m \geq N$, $|a_n - a_m| < 1$. In particular, for all $n \geq N$, $|a_n - a_N| < 1$, so $|a_n| < |a_N| + 1$. Consider the finite set

$$A = \{|a_1|, |a_2|, \ldots, |a_{N-1}|, |a_N| + 1\}.$$

Since A is finite, it has a maximum element, K say. Then, for all $n < N$, $|a_n| \leq K$ (since each of these numbers appears in the set A) and for all $n \geq N$, $|a_n| < |a_N| + 1 \leq K$. Hence, for all $n \in \mathbb{Z}^+$, $|a_n| \leq K$. □

Remark Since every convergent sequence is Cauchy, this gives a sneaky proof of Proposition 1.6 (every convergent sequence is bounded).

Recall that if a sequence converges to a limit L, every subsequence of that sequence also converges to L (Proposition 1.11). In general, knowing that a single subsequence converges to L tells us nothing about convergence of the original sequence, however. For example $a_n = (1 + (-1)^n)n$ has the subsequence $b_k = a_{2k+1} = (1 - 1)(2k + 1) = 0$ which converges to 0. The sequence (a_n), however, does not converge to 0: indeed, it isn't even bounded.

However, if we know that (a_n) is *Cauchy*, and it has a convergent subsequence, we *can* conclude that (a_n) converges:

Lemma 1.19 *Let (a_n) be Cauchy, and assume some subsequence of (a_n) converges to L. Then (a_n) converges to L.*

Proof: Let (a_{n_k}) be the subsequence which converges to L. Let $\varepsilon > 0$ be given. Since $a_{n_k} \to L$, there exists $N_1 \in \mathbb{Z}^+$ such that, for all $k \geq N_1$, $|a_{n_k} - L| < \varepsilon/2$. Further, since (a_n) is Cauchy, there exists $N_2 \in \mathbb{Z}^+$ such that, for all $n, m \geq N_2$,

$|a_n - a_m| < \varepsilon/2$. Let $N = \max\{N_1, N_2\}$. Then, for all $n \geq N$,

$$\begin{aligned}
|a_n - L| &= |(a_n - a_{n_N}) + (a_{n_N} - L)| \\
&\leq |a_n - a_{n_N}| + |a_{n_N} - L| && \text{(Triangle Inequality)} \\
&< \frac{\varepsilon}{2} + |a_{n_N} - L| && \text{(since } n \geq N_2\text{, and } n_N \geq N \geq N_2\text{)} \\
&< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} && \text{(since } N \geq N_1\text{).}
\end{aligned}$$

Hence, $a_n \to L$. □

Our target theorem (a sequence converges if and only if it is Cauchy) now follows immediately from the Bolzano-Weierstrass Theorem (Theorem 1.15).

Theorem 1.20 *A real sequence converges if and only if it is Cauchy.*

Proof: If (a_n) converges, it is Cauchy, by Lemma 1.17. Conversely, if (a_n) is Cauchy, it is bounded (Lemma 1.18), so has a convergent subsequence (by the Bolzano-Weierstrass Theorem) and so converges (Lemma 1.19). □

Let's put this theory to some use.

Example 1.21 Let (a_n) be the sequence defined so that $a_1 = 1$, $a_2 = 2$, and for all $n \geq 3$,

$$a_n = \frac{1}{2}(a_{n-1} + a_{n-2}),$$

so each subsequent term is generated by averaging the previous two.
Claim: (a_n) converges.

Proof: It's not at all obvious what the limit of this sequence is, so a direct ε—N proof would be a tall order (how do we estimate $|a_n - L|$ when we don't know L?). A nice idea would be try to use the Monotone Convergence Theorem, that is, show that (a_n) is monotone (increasing or decreasing) and bounded. Unfortunately, this sequence isn't monotone, as you can check by computing the first few terms, so that strategy fails too. Instead, we'll show that the sequence is Cauchy.

We first show (by induction) that, for all $n \in \mathbb{Z}^+$,

$$|a_n - a_{n+1}| = \frac{1}{2^{n-1}}. \tag{1.1}$$

This holds for $n = 1$:

$$|a_1 - a_2| = |1 - 2| = 1 = \frac{1}{2^{1-1}}$$

Assume it holds for some $n = k \in \mathbb{Z}^+$. Then, by the definition of (a_n)

$$\begin{aligned}
|a_{k+1} - a_{k+2}| &= \left| a_{k+1} - \frac{1}{2}(a_{k+1} + a_k) \right| \\
&= \frac{1}{2}|a_{k+1} - a_k| \\
&= \frac{1}{2}\frac{1}{2^{k-1}} = \frac{1}{2^{(k+1)-1}}
\end{aligned}$$

by the induction hypothesis. Hence if (1.1) holds for $n = k$, it holds for $n = k + 1$. Hence, by induction, (1.1) holds for all $n \in \mathbb{Z}^+$.

We now show that (1.1) implies that (a_n) is Cauchy. Let $\varepsilon \in (0, \infty)$ be given. We must show that there exists $N \in \mathbb{Z}^+$ such that, for all $n, m \geq N$, $|a_n - a_m| < \varepsilon$. Note that this inequality holds trivially if $n = m$, and is symmetric under interchange of n and m, so it's enough to show that there exists $N \in \mathbb{Z}^+$ such that, for all $m > n \geq N$, $|a_n - a_m| < \varepsilon$. Since the sequence 2^n is unbounded above, there exists $N \in \mathbb{Z}^+$ such that, $2^N > 4/\varepsilon$. I claim that this N is the one we want. Assume $m > n$. Then,

$$\begin{aligned}
|a_n - a_m| &= |(a_n - a_{n+1}) + (a_{n-1} - a_{n-2}) + \cdots + (a_{m-1} - a_m)| \\
&\leq |a_n - a_{n+1}| + |a_{n+1} - a_{n+2}| + \cdots + |a_{m-1} - a_m| \qquad \text{(by the triangle inequality)} \\
&= \frac{1}{2^{n-1}} + \frac{1}{2^n} + \cdots + \frac{1}{2^{m-2}} \qquad \text{(by (1.1)} \\
&= \frac{1}{2^{n-1}} \left(1 + \frac{1}{2} + \frac{1}{2^2} + \cdots + \frac{1}{2^{m-n-1}}\right) \\
&< \frac{1}{2^{n-2}}
\end{aligned}$$

since the geometric series $\sum_{n=0}^{\infty}(1/2)^n$ converges to 2 (if you've forgotten why, I'll remind you in a few weeks). Hence, for all $m > n \geq N$, with $2^N > 4/\varepsilon$,

$$|a_n - a_m| < \frac{1}{2^{n-2}} \leq \frac{1}{2^{N-2}} = \frac{4}{2^N} < \varepsilon.$$

Since (a_n) is Cauchy, it converges by Theorem 1.20. $\square$

Exercise Show that $a_n \to \frac{5}{3}$. (Hint: think about the subsequence (a_{2n-1}) of odd numbered terms.)

Remark We proved that (a_n) was Cauchy by exploiting the fact that the sequence of distances between consecutive terms, $|a_n - a_{n+1}|$, converges to 0 very fast (exponentially fast, in fact). It's tempting to think that the Cauchy property is equivalent to the property that $|a_n - a_{n+1}| \to 0$, but this is **false**!

Counterexample 1.22 (the Harmonic Series) Consider the sequence

$$a_n = 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}.$$

This is the sequence of *partial sums* of the harmonic series $\displaystyle\sum_{n=1}^{\infty} \frac{1}{n}$, which, as you proved in year 1, **diverges**. Hence, by Theorem 1.20 it is **not** Cauchy. Note that

$$|a_n - a_{n+1}| = \frac{1}{n+1} \to 0$$

however.

Summary

- A real sequence (a_n) **converges** to a **limit** L (written $a_n \to L$, or $\lim_{n\to\infty} a_n = L$) if, for each $\varepsilon > 0$, there exists $N \in \mathbb{Z}^+$ such that, for all $n \geq N$, $|a_n - L| < \varepsilon$.
- We can use this definition to prove many basic facts about convergent sequences:
  - If a sequence converges, its limit is unique.
  - Convergent sequences are bounded.
  - Limits preserve non-strict inequalities.
  - The **Squeeze Rule**: If $a_n \leq b_n \leq c_n$, $a_n \to L$ and $c_n \to L$, then $b_n \to L$.
  - The **Algebra of Limits**: If $a_n \to A$ and $b_n \to B$ then $a_n + b_n \to A + B$, $a_n b_n \to AB$ and $a_n/b_n \to A/B$ (provided both sides make sense).
  - Any subsequence of a convergent sequence converges (to the same limit).
  - The **Monotone Convergence Theorem**: If a_n is bounded and monotonic (increasing or decreasing) then it converges.
  - The **Bolzano-Weierstrass Theorem**: Every bounded sequence has a convergent subsequence.
- A real sequence (a_n) is **Cauchy** if, for each $\varepsilon > 0$, there exists $N \in \mathbb{Z}^+$ such that, for all $n, m \geq N$, $|a_n - a_m| < \varepsilon$.
- A real sequence converges if and only if it is Cauchy. This allows us to prove that some sequences converge even if we don't know their limit.

Chapter 2

Limits of functions and continuity

2.1 Limits at infinity

Note it's not essential in Definition 1.1 that the quantity N is a positive **integer**. We could just have well decided that

$\lim_{n\to\infty} a_n = L$ if, for each $\varepsilon > 0$, there exists $K \in \mathbb{R}$ such that, for all $n > K$, $|a_n - L| < \varepsilon$.

Both say that, given any positive number ε (no matter how small), there's a point in the sequence after which all terms lie within distance ε of L. We just use different means (a positive integer N or a real number K) to specify that point.

Having realized this, we see that the definition can be extended immediately to deal with *functions* of a real variable, not just sequences.

Definition 2.1 Let $D \subseteq \mathbb{R}$ be unbounded above and $f : D \to \mathbb{R}$. Then f has **limit** L **at infinity** if, for each $\varepsilon > 0$, there exists $K \in \mathbb{R}$ such that, for all $x \in D$ with $x > K$, $|f(x) - L| < \varepsilon$. The shorthand for this is $\lim_{x\to\infty} f(x) = L$.

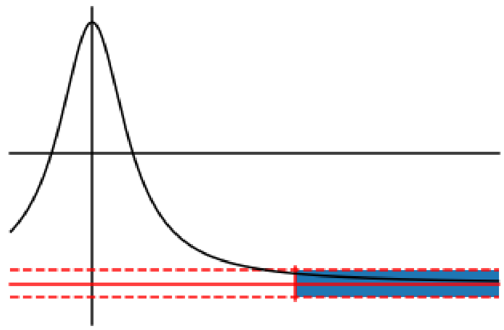
Example 2.2 Let $f : \mathbb{R} \to \mathbb{R}$, $f(x) = (1 - x^2)/(1 + x^2)$. Claim: $\lim_{x\to\infty} f(x) = -1$.

Proof: Let $\varepsilon > 0$ be given. Let $K = \sqrt{\dfrac{2}{\varepsilon}}$. Then, for all $x > K$,

$$\begin{aligned} |f(x) - (-1)| &= \left|\frac{1 - x^2}{1 + x^2} + 1\right| \\ &= \frac{2}{1 + x^2} \\ &< \frac{2}{x^2} \\ &< \frac{2}{K^2} = \varepsilon \end{aligned}$$

□

Again, it's useful to have a geometric picture of what this means:



Exercise 2.3 Let $f : [0, \infty) \to \mathbb{R}$ such that $f(x) = \dfrac{\sqrt{x}}{x+3}$. Prove that

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

The definition says that, given any positive number ε (no matter how small), there's a real number K such that, for all $x > K$, $f(x)$ is closer than distance ε from L. This makes precise the informal idea that $f(x)$ gets "arbitrarily close" to L as x gets "sufficiently large." Note we must demand that D, the domain of f, is unbounded above, or else there will exist some real number K such that there *are* no values of $x \in D$ with $x > K$. If we allowed this situation, then *any* real number L would satisfy the definition of $\lim_{x\to\infty} f(x)$!

Exercise 2.4 Write down a precise definition of the limit of $f : D \to \mathbb{R}$ at **minus infinity**. What property must D have in order for this definition to make sense? Use your definition to prove that

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

2.2 Cluster points

Limits at infinity are quite useful, but for the purposes of calculus, we really need to make proper sense of limits like

$$\lim_{x\to a} f(x)$$

where a is a real number. For example,

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

Our first task is to clearly define the kind of point $a \in \mathbb{R}$ at which it makes sense to define the limit of a given function f. Informally, we know that the limit "as x tends to a" concerns the behaviour of $f(x)$ "sufficiently close" to (but *not at*) a. Just as $\lim_{x\to\infty} f(x)$ only makes sense if the domain of f is unbounded above (so that $f(x)$ is well-defined for x arbitrarily large), $\lim_{x\to a} f(x)$ only makes sense if the domain of f contains points which are "arbitrarily close" to (but different from) a.

Definition 2.5 Let $D \subseteq \mathbb{R}$. Then $a \in \mathbb{R}$ is a **cluster point** of D if, for each $\varepsilon > 0$, there exists $x \in D$ with $0 < |x - a| < \varepsilon$. Equivalently: for each $\varepsilon > 0$, the set $(D \backslash \{a\}) \cap (a - \varepsilon, a + \varepsilon)$ is nonempty.

Example 2.6 For each of the following subsets of $\mathbb{R}$, write down the set of all cluster points:

$$A = (0, 1], \qquad B = \mathbb{Q}, \qquad C = \{1/n \, : \, n \in \mathbb{Z}^+\}, \qquad D = \mathbb{Z}.$$

Solution Let's denote the set of cluster points of a set X by $\mathrm{cl}(X)$. Then

$$\begin{aligned}
\mathrm{cl}(A) &= [0, 1] \\
\mathrm{cl}(B) &= \mathbb{R} \\
\mathrm{cl}(C) &= \{0\} \\
\mathrm{cl}(D) &= \emptyset
\end{aligned}$$

From this we see that it's possible for a number that is *not* in a set to be a cluster point of that set (e.g. 0 is a cluster point of $(0, 1]$ and $\sqrt{2}$ is a cluster point of $\mathbb{Q}$), and that an element of a set may *fail* to be a cluster point of the set (e.g. 5 is not a cluster point of $\mathbb{Z}$).

We can also characterize cluster points using sequences. This often turns out to be more convenient than the original definition.

Proposition 2.7 *$a \in \mathbb{R}$ is a cluster point of $D \subseteq \mathbb{R}$ if and only if there is a sequence in $D \backslash \{a\}$ that converges to a.*

Proof: Exercise (see Problem Set 1). □

Note that the sequence is in $D \backslash \{a\}$, that is, it never takes the value a.

2.3 Limits of functions

Cluster points of the domain of a function are exactly the points at which it makes sense to define a limit of the function.

Definition 2.8 Let $D \subseteq \mathbb{R}$, $f : D \to \mathbb{R}$ and $a \in \mathbb{R}$ be a cluster point of D. Then f has **limit** L **at** a if, for each $\varepsilon > 0$, there exists $\delta > 0$ such that, for all $x \in D$ with $0 < |x - a| < \delta$, $|f(x) - L| < \varepsilon$. The shorthand for this is $\lim\limits_{x \to a} f(x) = L$.

Example 2.9 Let $f : \mathbb{R} \to \mathbb{R}$, $f(x) = x^2$. Claim: $\lim\limits_{x \to 2} f(x) = 4$.

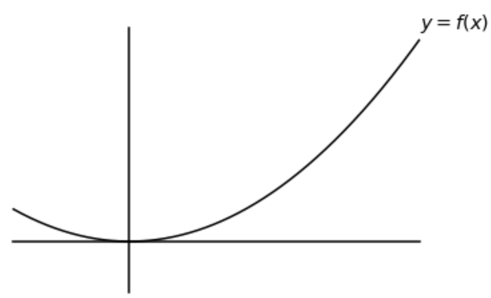
Proof: Let $\varepsilon \in (0, \infty)$ be given. Let $\delta = \boxed{\min\{1, \frac{\varepsilon}{5}\}}$.
Then, for all $x \in \mathbb{R}$ with $0 < |x - 2| < \delta$,

$$
\begin{aligned}
|f(x) - 4| &= |x^2 - 4| \\
&= |x + 2||x - 2| \\
&\leq 5|x - 2| \qquad (\text{since } \delta \leq 1 \text{ so } x \in (1, 3)) \\
&< 5\delta \\
&\leq \varepsilon
\end{aligned}
$$

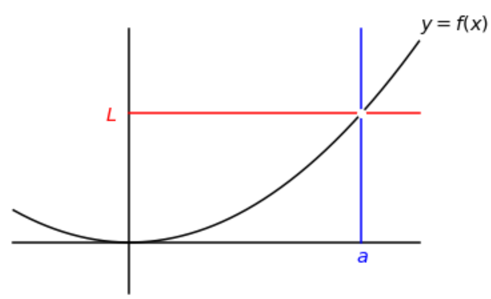
□

The definition says that, given any positive number ε (no matter how small), we can find another positive number δ which is so small that, whenever x is within distance δ of a, *and different from* a, $f(x)$ is within distance ε of L. This makes precise the informal idea that $f(x)$ is "arbitrarily close" to L for all x "sufficiently close" to, but different from, a.

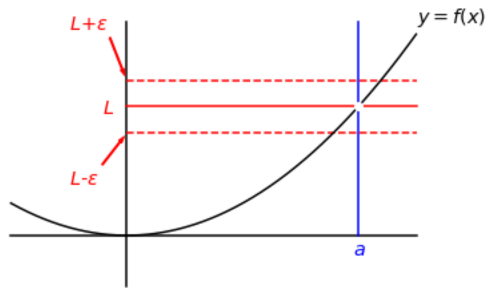
It's helpful to have a picture of what the definition means. Imagine the graph of the function f:



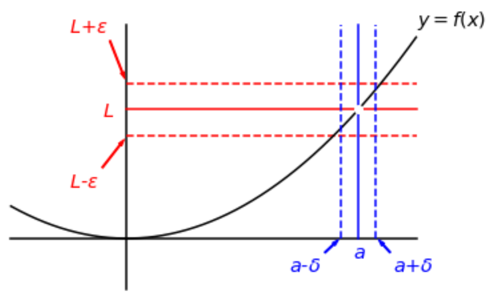
Delete from this the point $(a, f(a))$ (because the limit at a concerns the behaviour of f close to, but not at a) and mark on the lines $y = L$ and $x = a$, where L is the claimed limit:



Now, choose any $\varepsilon > 0$, and mark on the lines $y = L \pm \varepsilon$:



The claim is that, no matter how small we choose ε, there exists $\delta > 0$ such that the part of the graph between the vertical lines $x = a \pm \delta$ lies entirely between the horizontal lines $y = L \pm \varepsilon$:



A proof directly from the definition of limit (Definition 2.8), as in Example 2.9, is called a proof *from first principles*, or an ε—δ proof. It's very important that you get plenty of practice constructing such proofs, because this is the best way to properly understand the meaning of limits.

Example 2.10 Let $f : [0, \infty) \to \mathbb{R}$ be defined such that $f(x) = \sqrt{x}$. Claim: $\lim\limits_{x \to 1} f(x) = 1$.

Proof: Let $\varepsilon \in (0, \infty)$ be given. Then let $\delta = \boxed{\varepsilon}$. Then, for all $x \in [0, \infty)$ such that $0 < |x - 1| < \delta$,

$$
\begin{aligned}
|f(x) - 1| &= |\sqrt{x} - 1| \\
&= \left|(\sqrt{x} - 1)\frac{\sqrt{x}+1}{\sqrt{x}+1}\right| \\
&= \frac{|x-1|}{\sqrt{x}+1} \\
&\leq |x - 1| \\
&< \delta = \varepsilon.
\end{aligned}
$$

□

In Examples 2.9 and 2.10, the point a at which we defined the limit was actually in the domain of f (and the limit happened to be $f(a)$). Limits are more interesting when a is *not* in the domain of f.

Example 2.11 Let $f : \mathbb{R}\backslash\{-3, 3\} \to \mathbb{R}$, $f(x) = \dfrac{x-3}{x^2-9}$. Claim: $\lim_{x\to 3} f(x) = \dfrac{1}{6}$.

Proof: Let $\varepsilon > 0$ be given. Let $\delta = \min\{1, 30\varepsilon\}$. Then, for all $x \in \mathbb{R}\backslash\{-3, 3\}$ with $0 < |x - 3| < \delta$,

$$\begin{aligned} \left|f(x) - \frac{1}{6}\right| &= \left|\frac{x-3}{(x-3)(x+3)} - \frac{1}{6}\right| = \left|\frac{1}{x+3} - \frac{1}{6}\right| = \left|\frac{3-x}{6(x+3)}\right| \\ &\le \frac{1}{30}|x-3| \qquad \text{(since } \delta \le 1 \text{ so } x \in (2,4) \text{ whence } x+3 > 5\text{)} \\ &< \frac{1}{30}\delta \\ &\le \varepsilon \end{aligned}$$

□

You may be wondering where the heck the number

$$\delta = \min\{1, 30\varepsilon\}$$

came from in the above proof. The answer is that, before I started to write the proof down, I did the following rough work: I want to show that I can make $f(x)$ as close as I like to $1/6$ by making x sufficiently close to (but different from) 3. So I try to find an upper bound on $|f(x) - 1/6|$ in terms of $|x - 3|$:

$$\begin{aligned} \left|f(x) - \frac{1}{6}\right| &= \left|\frac{x-3}{(x-3)(x+3)} - \frac{1}{6}\right| = \left|\frac{1}{x+3} - \frac{1}{6}\right| \\ &= \left|\frac{3-x}{6(x+3)}\right| \\ &= \frac{1}{6}\frac{1}{|x+3|}|x-3|. \end{aligned}$$

I like the factor of $|x - 3|$: this is the thing I can make as small as I like. The factor of $1/6$ is no problem. The factor of $1/|x + 3|$ is not so congenial, however. It can get arbitrarily large – but only if x gets close to -3. So, if I demand that $|x - 3| < 1$, for example, then $x > 2$ so $|x + 3| \ge x + 3 > 5$. This is helpful, because it forces $1/|x + 3| < 1/5$. So, for all x with $0 < |x - 3| < 1$, I know that

$$\left|f(x) - \frac{1}{6}\right| \le \frac{1}{6\times 5}|x-3|.$$

To make this less than ε, it's enough to also demand that $|x - 3| < 30\varepsilon$. So, I need $\delta \le 1$ and $\delta < 30\varepsilon$, hence my choice of δ.

It's important to understand that there are infinitely many other correct choices of δ. This is the δ that I arrived at by reasoning as above. You may argue differently and come up with something different.

2.3.1 Some basic properties of limits

Since we have a precise definition of limits, we are in a position to prove some basic facts about them.

Theorem 2.12 (Uniqueness of limits) *If f has a limit at a, this limit is unique.*

Proof: Let $f : D \to \mathbb{R}$ and a be a cluster point of D. Assume, towards a contradiction, that both L_1 and $L_2 \neq L_1$ satisfy the definition of limit of f at a. Let $\varepsilon = |L_2 - L_1|/2 > 0$. Since $\lim\limits_{x\to a} f(x) = L_1$, there exists $\delta_1 > 0$ such that, for all $x \in D$ with $0 < |x - a| < \delta_1$, $|f(x) - L_1| < \varepsilon$. Similarly, since $\lim\limits_{x\to a} f(x) = L_2$, there exists $\delta_2 > 0$ such that, for all $x \in D$ with $0 < |x - a| < \delta_2$, $|f(x) - L_2| < \varepsilon$. Let $\delta = \min\{\delta_1, \delta_2\}$. **Since a is a cluster point of D**, there exists $x_* \in D$ such that $0 < |x_* - a| < \delta$. But then $0 < |x_* - a| < \delta_1$, so $|f(x_*) - L_1| < \varepsilon$, and, furthermore, $0 < |x_* - a| < \delta_2$, so $|f(x_*) - L_2| < \varepsilon$. Hence

$$2\varepsilon = |L_2 - L_1| = |f(x_*) - L_1 - (f(x_*) - L_2)| \leq |f(x_*) - L_1| + |f(x_*) - L_2| < \varepsilon + \varepsilon$$

a contradiction. □

So it makes sense to talk about *the* limit of a function at a. Note that this proof made essential use of the condition that a is a cluster point of D, the domain of f (see the text highlighted in bold above). Without this assumption, we can't be sure that the point $x_* \in D\backslash\{a\}$ distance less than δ from a actually exists.

We can also establish that limits of functions obey some basic arithmetic rules:

Theorem 2.13 (Algebra of Limits) *Let a be a cluster point of D, $f : D \to \mathbb{R}$, $g : D \to \mathbb{R}$, $\lim\limits_{x\to a} f(x) = L$ and $\lim\limits_{x\to a} g(x) = K$. Then*

(i) $\lim\limits_{x\to a}(f(x) + g(x)) = L + K$,

(ii) $\lim\limits_{x\to a} f(x)g(x) = LK$,

(iii) $\lim\limits_{x\to a} \dfrac{1}{f(x)} = \dfrac{1}{L}$ *if, in addition,* $0 \notin f(D)$ *and* $L \neq 0$.

We *could* give direct ε—δ proofs of each part of this theorem. They would be structurally very similar to the analogous proofs of Proposition 1.9 (the Algebra of Limits for sequences). This is actually quite a good exercise, and I encourage you to have a go. However, we're going to prove Theorem 2.13 by exploiting a sneaky and *very* powerful trick: we can reformulate limits of functions entirely in terms of convergence of sequences:

Theorem 2.14 *Let a be a cluster point of D and $f : D \to \mathbb{R}$. Then $\lim\limits_{x\to a} f(x) = L$ if and only if, for all sequences (x_n) in $D\backslash\{a\}$ such that $x_n \to a$, $f(x_n) \to L$.*

Proof: We must prove two things:

(i) If $\lim_{x \to a} f(x) = L$ then, for all sequences (x_n) in $D \backslash \{a\}$ such that $x_n \to a$, $f(x_n) \to L$, and

(ii) If for all sequences (x_n) in $D \backslash \{a\}$ such that $x_n \to a$, $f(x_n) \to L$, then $\lim_{x \to a} f(x) = L$.

(i) Assume $\lim_{x \to a} f(x) = L$ and let (x_n) be any sequence in $D \backslash \{a\}$ such that $x_n \to a$. We must show that $f(x_n) \to L$. So, let $\varepsilon > 0$ be given. By assumption, there exists $\delta > 0$ such that, for all $x \in D$ with $0 < |x - a| < \delta$, $|f(x) - L| < \varepsilon$. But $x_n \to a$, and δ is a positive number, so there exists $N \in \mathbb{Z}^+$ such that, for all $n \geq N$, $|x_n - a| < \delta$. Furthermore, $x_n \in D \backslash \{a\}$, so for all n, $|x_n - a| > 0$. Hence, for all $n \geq N$, $0 < |x_n - a| < \delta$. But then (by the definition of δ), for all $n \geq N$, $|f(x_n) - L| < \varepsilon$. Since $\varepsilon > 0$ was arbitrary, it follows that $f(x_n) \to L$.

(ii) We will prove the *contrapositive*: if f does **not** have limit L at a, then there exists a sequence (x_n) in $D \backslash \{a\}$ such that $x_n \to a$ but $f(x_n) \not\to L$. So, assume $\lim_{x \to a} f(x) = L$ is **false**. Our first job is to figure out precisely what this condition means. It is the **negation** of the statement $\lim_{x \to a} f(x) = L$, that is

$$\neg \left[\forall \varepsilon \in (0, \infty), \exists \delta \in (0, \infty), \forall x \in (D \backslash \{a\}) \cap (a - \varepsilon, a + \varepsilon), |f(x) - L| < \varepsilon\right]$$

$$\Leftrightarrow \quad \exists \varepsilon \in (0, \infty), \neg \left[\exists \delta \in (0, \infty), \forall x \in (D \backslash \{a\}) \cap (a - \varepsilon, a + \varepsilon), |f(x) - L| < \varepsilon\right]$$

$$\Leftrightarrow \quad \exists \varepsilon \in (0, \infty), \forall \delta \in (0, \infty), \neg \left[\forall x \in (D \backslash \{a\}) \cap (a - \varepsilon, a + \varepsilon), |f(x) - L| < \varepsilon\right]$$

$$\Leftrightarrow \quad \exists \varepsilon \in (0, \infty), \forall \delta \in (0, \infty), \exists x \in (D \backslash \{a\}) \cap (a - \varepsilon, a + \varepsilon), \neg \left[|f(x) - L| < \varepsilon\right]$$

$$\Leftrightarrow \quad \exists \varepsilon \in (0, \infty), \forall \delta \in (0, \infty), \exists x \in (D \backslash \{a\}) \cap (a - \varepsilon, a + \varepsilon), |f(x) - L| \geq \varepsilon.$$

So *there exists* $\varepsilon > 0$ such that, for all $\delta > 0$, there is some $x \in D$ with $0 < |x - a| < \delta$ such that $|f(x) - L| \geq \varepsilon$. Since this holds for each and every $\delta > 0$, it holds, in particular, if $\delta = 1/n$, where $n \in \mathbb{Z}^+$. That is, for each $n \in \mathbb{Z}^+$, there exists $x_n \in D$ with $0 < |x_n - a| < 1/n$ such that $|f(x_n) - L| \geq \varepsilon$. Now $x_n \to a$ (by the **Squeeze Rule**), and $x_n \in D \backslash \{a\}$ (since $0 < |x_n - a|$), but $|f(x_n) - L| \geq \varepsilon$ for all n so certainly $f(x_n) \not\to L$. $\square$

Armed with Theorem 2.14, we can now make short work of proving Theorem 2.13 – it follows almost immediately from Proposition 1.9:

Proof of Theorem 2.13: Let (x_n) be any sequence in $D \backslash \{a\}$ converging to a. Since $\lim_{x \to a} f(x) = L$, $f(x_n) \to L$ by Theorem 2.14. Since $\lim_{x \to a} g(x) = K$, $g(x_n) \to L$ by Theorem 2.14. Hence

(i) $f(x_n) + g(x_n) \to L + K$ by Proposition 1.9(i) and

(ii) $f(x_n)g(x_n) \to LK$ by Proposition 1.9(ii).

This holds for any such sequence (x_n) so $f(x) + g(x)$ has limit $L + K$ at a, and $f(x)g(x)$ has limit LK at a by Theorem 2.14.

If, in addition, $f(x) \neq 0$ for all $x \in D$ and $L \neq 0$, then $f(x_n) \neq 0$ for all n, so

(iii) $1/f(x_n) \to 1/L$ by Proposition 1.9(iii).

This holds for any such sequence (x_n) so $1/f(x)$ has limit $1/L$ at a by Theorem 2.14. □

Note that this proof used both directions of Theorem 2.14: since $\lim_{x\to a} f(x) = L$ we know that $f(x_n) \to L$ (the "only if" direction), and since $f(x_n) + g(x_n) \to L + K$ we know that $\lim_{x\to a}(f(x) + g(x)) = L + K$ (the "if" direction).

Exercise 2.15 Read through the proof of Theorem 2.13 and, at every place where Theorem 2.14 is cited, determine whether the "if" direction or the "only if" direction is being used.

2.3.2 Continuity and limits

You may have been told (by some shameless charlatan) that a function is continuous "if you can draw its graph without taking your pen off the paper." It should by now be clear that this is *not* an acceptable mathematical definition. What if you have no pen, or paper, or hands? What if you're simply not very good at drawing? Does the continuity of a function depend on your skill as a draughtsman? Clearly this is utter piffle. A mathematical definition should be formulated only in terms of the object being defined, not how we are able to interact with (or think about) it.

Luckily, we can use convergence of sequences to give a precise (and rather elegant) definition of continuity for functions:

Definition 2.16 Let $f : D \to \mathbb{R}$ and $a \in D$. Then f is **continuous at** a if, for all sequences (x_n) in D such that $x_n \to a$, $f(x_n) \to f(a)$. f is **continuous** if it is continuous at a for all $a \in D$. If f is not continuous, it is **discontinuous**.

For example, all polynomial functions

$$f : \mathbb{R} \to \mathbb{R}, \qquad f(x) = a_0 + a_1 x + \cdots + a_k x^k$$

are continuous. This follows from the Algebra of Limits for sequences and the rather obvious fact that $f(x) = x$ is continuous.[1]

Example 2.17 A simple example of a discontinuous function is a "step function":

$$f : \mathbb{R} \to \mathbb{R}, \qquad f(x) = \begin{cases} 0 & x < 0 \\ 1 & x \geq 0 \end{cases}$$

To see this, note that the sequence $x_n = -1/n \to 0$, but $f(x_n) = f(-1/n) = 0$, which does **not** converge to $f(0) = 1$.

[1]Proof: If $x_n \to a$ then $f(x_n) = x_n \to a = f(a)$.

A step function is discontinuous at a single point. Somewhat counterintuitively, it's possible for a function $f : \mathbb{R} \to \mathbb{R}$ to be *continuous* at only a single point:

Example 2.18 Let $f : \mathbb{R} \to \mathbb{R}$ be such that $f(x) = x$ if $x \in \mathbb{Q}$ and $f(x) = 0$ if $x \notin \mathbb{Q}$. Claim: f is continuous as 0 and discontinuous everywhere else.

Proof: Assume that f is continuous at a. Let (r_n) be any rational sequence converging to a. Then $f(r_n) = r_n \to a$. Let (i_n) be any irrational sequence converging to a. Then $f(i_n) = 0 \to 0$. But f is continuous at a, so $f(r_n) \to f(a)$ and $f(i_n) \to f(a)$. Hence $f(a) = a$ and $f(a) = 0$, that is $a = 0$. So *if* f is continuous at a *then* $a = 0$. That is, f is discontinuous at every $a \neq 0$.

It remains to show that f is continuous at 0 (careful: we have **not** proved this yet!). Let (x_n) be any sequence converging to 0. Then

$$|f(x_n)| = \left\{ \begin{array}{cc} |x_n|, & \text{if } x_n \in \mathbb{Q}, \\ 0, & \text{if } x_n \notin \mathbb{Q}, \end{array} \right\} \leq |x_n| \to 0$$

so $f(x_n) \to 0 = f(0)$ by the Squeeze Rule. Hence f is continuous at 0. $\square$

Here's an even more counterintuitive example: a function $f : \mathbb{R} \to \mathbb{R}$ which is continuous at every irrational number, but discontinuous at every rational number! It was invented by Carl Johannes Thomae but, in the anglosphere, is usually referred to by the (somewhat mysterious) name given to it by John Horton Conway:

Example 2.19 (Stars Over Babylon) Let $f : \mathbb{R} \to \mathbb{R}$ be the function which maps every irrational number to 0, and maps each rational number p/q, expressed in lowest terms (that is, p and q have greatest common divisor 1), to $1/q$. So

$$\begin{aligned} f(\sqrt{2}) &= 0 \\ f(3/12) &= = f(1/4) = 1/4 \\ f(5\frac{1}{2}) &= f(11/2) = \frac{1}{2} \\ f(0) &= f(0/1) = 1, \end{aligned}$$

for example. Note that $f(x) \geq 0$ for all $x \in \mathbb{R}$, and $f(x) = 0$ if **and only if** x is irrational.

Claim 1: f is discontinuous at every $a \in \mathbb{Q}$.
Proof: Assume, towards a contradiction, that $a \in \mathbb{Q}$ and f is continuous at a. Then, by the density of the irrationals in the reals, for each $n \in \mathbb{Z}^+$, there exists $i_n \in \mathbb{R}\backslash\mathbb{Q}$ such that $a < i_n < a + \frac{1}{n}$. By the Squeeze Rule, $i_n \to a$. Since f is continuous at a, $f(i_n) \to f(a)$ and $f(a) > 0$ (since a is rational). But $f(i_n)$ is the constant sequence 0, so $f(i_n) \to 0$, a contradiction. $\square$

Claim 2: f is continuous at every $a \in \mathbb{R}\backslash\mathbb{Q}$.

Proof: Again, assume towards a contradiction, that $a \in \mathbb{R}\backslash\mathbb{Q}$ and f is *discontinuous* at a. Then there exists a sequence x_n such that $x_n \to a$ but $f(x_n) \nrightarrow f(a) = 0$. Hence, there exists $\varepsilon > 0$ such that, for each $k \in \mathbb{Z}^+$, there exists $n_k \geq k$ such that $f(x_{n_k}) = |f(x_{n_k}) - 0| \geq \varepsilon$. That is, there is a subsequence $y_k = x_{n_k}$ of (x_n) such that $f(y_k) \geq \varepsilon > 0$. Clearly, (y_k) is a sequence of *rational numbers* (since $f(x) = 0$ if x is *irrational*).

Consider the rational sequence (y_k). Since it's a subsequence of a sequence converging to a, it converges to a. Since it's rational, we can write it in the form

$$y_k = \frac{p_k}{q_k}$$

where $p_k \in \mathbb{Z}$, $q_k \in \mathbb{Z}^+$ and $\gcd(p_k, q_k) = 1$. Then $f(y_k) = 1/q_k$. But $f(y_k) \geq \varepsilon$, so

$$0 < q_k \leq \frac{1}{\varepsilon},$$

that is, the sequence (q_k) is bounded. Hence, by the Bolzano-Weierstrass Theorem, it has a convergent subsequence, (q_{k_l}) say. But any convergent sequence of integers is eventually constant, that is, there exists $L \in \mathbb{Z}^+$ such that, for all $l \geq L$, $q_{k_l} = q_{k_L} = q$ say. (To see this, note that (q_{k_l}) is Cauchy , so there exists $L \in \mathbb{Z}^+$ such that, for all $m \geq L$, $|q_{k_L} - q_{k_m}| < 1/2$, and the only integer within distance $1/2$ of q_{k_L} is q_{k_L} itself.) Consider now the subsequence

$$y_{k_l} = \frac{p_{k_l}}{q_{k_l}} = \frac{p_{k_l}}{q},$$

for all $l \geq L$. This converges to a, so $p_{k_l} \to qa$, which is irrational (note $q > 0$). But (p_{k_l}) is also a sequence of integers, so if it converges, it is eventually constant, and hence its limit is an integer. This is a contradiction. $\square$

It's not hard to show that every function $f : D \to \mathbb{R}$ is automatically continuous at a if a is *not* a cluster point of D.

Proposition 2.20 *Let $f : D \to \mathbb{R}$ and assume $a \in D$ is not a cluster point of D. Then f is continuous at a.*

Proof: Let (x_n) be any sequence in D converging to a. Since a is *not* a cluster point of D, there exists $\varepsilon > 0$ such that no element of D has $0 < |x - a| < \varepsilon$. Hence, if $|x - a| < \varepsilon$ and $x \in D$, then $x = a$. Now $x_n \to a$, so there exists $N \in \mathbb{Z}^+$ such that, for all $n \geq N$, $|x_n - a| < \varepsilon$ and hence $x_n = a$. Hence, for all $n \geq N$, $f(x_n) = f(a)$, that is, the sequence $(f(x_n))$ is constant and equals $f(a)$, for all $n \geq N$. Hence $f(x_n) \to f(a)$. $\square$

So in considering whether a function is continuous, we may restrict attention only to cluster points of (and in) its domain. For such points we have two useful ways to reformulate continuity:

Theorem 2.21 *Let $f : D \to \mathbb{R}$, $a \in D$ and a be a cluster point of D. Then the following are equivalent:*

(i) f is continuous at a;

(ii) $\lim_{x\to a} f(x) = f(a)$;

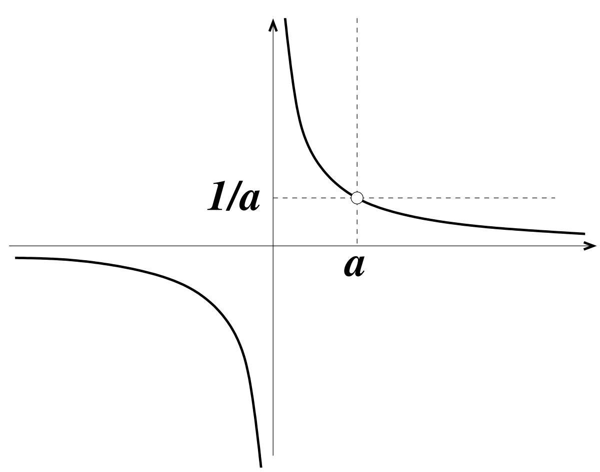
(iii) For each $\varepsilon > 0$ there exists $\delta > 0$ such that, for all $x \in D$ with $|x - a| < \delta$, $|f(x) - f(a)| < \varepsilon$.

Proof: (i) $\Rightarrow$ (ii) follows immediately from Theorem 2.14.
(ii) $\Rightarrow$ (iii): assume $\lim_{x\to a} f(x) = f(a)$ and let $\varepsilon > 0$ be given. Then (by Definition 2.8) there exists $\delta > 0$ such that, for all $x \in D$ with $0 < |x - a| < \delta$, $|f(x) - f(a)| < \varepsilon$. Clearly, if $x = a$ then $|f(x) - f(a)| = 0 < \varepsilon$. Hence, for all $x \in D$ with $|x - a| < \delta$, $|f(x) - f(a)| < \varepsilon$.
(iii) $\Rightarrow$ (i): assume (iii) holds, and let (x_n) be any sequence in D converging to a. We must show that $f(x_n) \to L$. So, let $\varepsilon > 0$ be given. By assumption, there exists $\delta > 0$ such that, for all $x \in D$ with $|x - a| < \delta$, $|f(x) - f(a)| < \varepsilon$. Now $x_n \to a$ and $\delta > 0$, so there exists $N \in \mathbb{Z}^+$ such that, for all $n \geq N$, $|x_n - a| < \delta$. But then, for all $n \geq N$, $|f(x_n) - f(a)| < \varepsilon$, by definition of δ. Hence $f(x_n) \to f(a)$. Since this holds for any sequence (x_n) in D converging to a, f is continuous at a. □

Many mathematicians take condition (iii) in Theorem 2.21 as the *definition* of continuity at a, instead of the sequential definition, Definition 2.16. It's important to be able to use both. Here's an example which demonstrates that the whole "continuous if you can draw its graph without taking your pen off the paper" pseudo-definition isn't just embarrassingly jejeune, it's also plain wrong.

Example 2.22 Let $f : \mathbb{R}\backslash\{0\} \to \mathbb{R}$, $f(x) = \dfrac{1}{x}$. Claim: f is continuous.[2]

Proof: We must show that f is continuous at a for every $a \neq 0$. To do this , we will use Theorem 2.21, that is, show that f, a satisfy the ε—δ criterion of part (iii).



So, choose and fix $a \neq 0$ and let some positive number ε be given. We must show that there exists $\delta > 0$ such that, for all $x \in \mathbb{R}\backslash\{0\}$ with $|x - a| < \delta$, $|f(x) - f(a)| < \varepsilon$.

[2]Try drawing the graph of that without taking your pen off the paper!

Now

$$\begin{aligned}
|f(x) - f(a)| &= \left|\frac{1}{x} - \frac{1}{a}\right| && (2.1)\\
& && (2.2)\\
&= \frac{|x-a|}{|a||x|} && (2.3)
\end{aligned}$$

So first let's insist that $\delta \leq |a|/2$. Then, for all $x \in (a-\delta, a+\delta)$, $|x| \geq |a|/2$, and hence

$$\frac{1}{|x|} \leq \frac{2}{|a|}$$

Given this, we see from (2.3) that, to ensure $|f(x) - f(a)| < \varepsilon$, it suffices to require that $|x-a| < |a|^2\varepsilon/2$. We may now write out the argument explicitly.

Given any $\varepsilon > 0$, let $\delta = \min\{|a|/2, |a|^2\varepsilon/2\} > 0$. Then for all $x \in \mathbb{R}\backslash\{0\}$ with $|x-a| < \delta$,

$$\begin{aligned}
|f(x) - f(a)| &= \frac{|x-a|}{|a||x|}\\
&\leq \frac{2}{|a|^2}|x-a| \qquad \text{(since } |x| > |a| - \delta \geq |a|/2\text{)}\\
&< \frac{2}{a^2}\delta\\
&\leq \varepsilon.
\end{aligned}$$

Hence, by Theorem 2.21, f is continuous at a. Since $a \in \mathbb{R}\backslash\{0\}$ was arbitrary, f is continuous. □

This kind of argument is called an ε—δ proof of continuity. Often an argument making direct use of Definition 2.16 is simpler.

Example 2.22 revisited Prove that $f : \mathbb{R}\backslash\{0\} \to \mathbb{R}$, $f(x) = \dfrac{1}{x}$ is continuous directly from Definition 2.16.

Proof: Let $a \in \mathbb{R}\backslash\{0\}$ and (x_n) be any sequence in $\mathbb{R}\backslash\{0\}$ such that $x_n \to a$. Then

$$f(x_n) = \frac{1}{x_n} \to \frac{1}{a} = f(a)$$

by the Algebra of Limits. Hence f is continuous at a. □

Here's another result which follows very quickly from the sequential definition of continuity (the composition of two continuous functions is continuous):

Theorem 2.23 *Let D and E be subsets of $\mathbb{R}$, $f : D \to E$, $g : E \to \mathbb{R}$, f be continuous at a and g be continuous at $f(a)$. Then $g \circ f : D \to \mathbb{R}$ is continuous at a.*

Proof: We will show directly that $g \circ f : D \to \mathbb{R}$, $(g \circ f)(x) = g(f(x))$ satisfies Definition 2.16. So, let (x_n) be any sequence in D converging to a. Since f is continuous at a, $y_n := f(x_n) \to f(a)$. Hence, since g is continuous at $f(a)$, $g(y_n) \to g(f(a))$. Hence $g(f(x_n)) \to g(f(a))$. $\square$

Exercise 2.24 Give an ε—δ proof of Theorem 2.23.

We finish this chapter by recalling two famous and important theorems about continuous functions on a closed, bounded *interval*,

$$[a, b] = \{x \in \mathbb{R} \,:\, a \leq x \leq b\}.$$

Theorem 2.25 (Intermediate Value Theorem) *Let $f : [a, b] \to \mathbb{R}$ be continuous and y be any number between $f(a)$ and $f(b)$. Then there exists $c \in [a, b]$ such that $f(c) = y$.*

Theorem 2.26 (Extreme Value Theorem) *Let $f : [a, b] \to \mathbb{R}$ be continuous. Then f is bounded (above and below) and attains both a minimum and a maximum value.*

These two Theorems may look obvious, but it's important to realize that we can (and, in MATH1026, *did*) prove them rigorously from Definitions 2.16 and 1.1. The whole point of Real Analysis is to go beyond intuitive, hand waving reasoning, to back up all our assertions with precise, rigorous reasoning. After all, as we shall see, not everything that is "obvious" is true!

Summary

- A real number a is a **cluster point** of a set $D \subseteq \mathbb{R}$ if, for each $\delta > 0$, there exists $x \in D$ with $0 < |x - a| < \delta$.
- A function $f : D \to \mathbb{R}$ has **limit** L at a, a cluster point of D, if, for each $\varepsilon > 0$, there exists $\delta > 0$ such that, for all $x \in D$ with $0 < |x - a| < \delta$, $|f(x) - L| < \varepsilon$. The shorthand for this is $\lim_{x \to a} f(x) = L$.
- We can reformulate limits in terms of convergence of sequences:
  Theorem $\lim_{x \to a} f(x) = L$ if and only if, for all sequences (x_n) in $D \backslash \{a\}$ converging to a, $(f(x_n))$ converges to L.
- $f : D \to \mathbb{R}$ is **continuous** at $a \in D$ if, for all sequences (x_n) in D converging to a, $f(x_n)$ converges to $f(a)$.
- **Theorem** If a is a cluster point of D (the domain of f), the following are equivalent:
  (i) f is continuous at a
  (ii) $\lim_{x \to a} f(x) = f(a)$
  (iii) For each $\varepsilon > 0$ there exists $\delta > 0$ such that, for all $x \in D$ with $|x - a| < \delta$, $|f(x) - f(a)| < \varepsilon$.

Chapter 3

Differentiable functions

3.1 The main definition

Now that we have a rigorous definition of the limit of a function, it is straightforward to define derivatives.

Definition 3.1 Let $f : D \to \mathbb{R}$, where D is some subset of $\mathbb{R}$, and $a \in D$ be a cluster point of D. Then f is **differentiable at** a if the limit

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

exists. In this case, we denote the limit $f'(a)$ and call it the **derivative of** f **at** a. We say that f is **differentiable** if it is differentiable at a for all $a \in D$.

Remarks

- In general, in order to define $\lim_{x \to a} g(x)$, we only need a to be a *cluster point* of the domain of g: it isn't necessary in general for a to be in the domain of g, so $g(a)$ may or may not exist. For example

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

  exists, although the function is undefined at $x = 1$. Note, however, that the limit defining $f'(a)$ contains the number $f(a)$, so to be differentiable at a, the point a must be both an *element* and a *cluster point* of the domain of f. For example, a function $f : (0, \infty) \to \mathbb{R}$ cannot be differentiable at 0 (since $0 \notin (0, \infty)$), and a function $f : \mathbb{Z} \to \mathbb{R}$ cannot be differentiable anywhere (since $\mathbb{Z}$ has no cluster points).

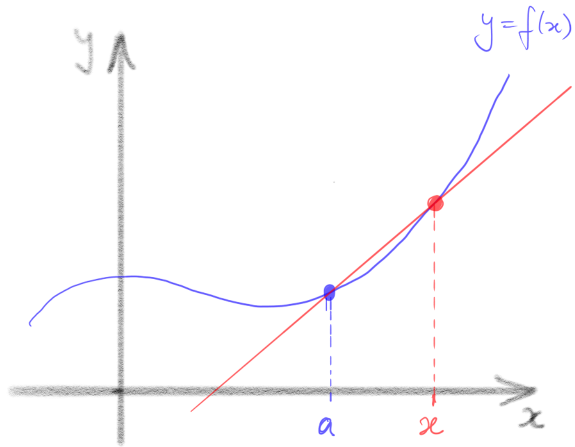
- The limit in Definition 3.1 is defined as in Definition 2.8. That is, $f : D \to \mathbb{R}$ is differentiable at a if there exists a real number, denoted $f'(a)$, such that, for each $\varepsilon > 0$, there exists $\delta > 0$ such that, for all $x \in D$ with $0 < |x - a| < \delta$,

  $$\left| \frac{f(x) - f(a)}{x - a} - f'(a) \right| < \varepsilon.$$

- Geometrically, the quantity

$$s(x) = \frac{f(x) - f(a)}{x - a}$$

is the slope of the straight line passing through the points $(a, f(a))$ and $(x, f(x))$. Such a straight line is called a **chord** on the graph $y = f(x)$. As x approaches a, the points defining the chord get "arbitrarily close" to one another, and the chords approach a straight line through $(a, f(a))$ with slope $f'(a)$, the so-called **tangent line**. This is why the derivative $f'(a)$ is often interpreted as the slope of the graph $y = f(x)$ at the point $(a, f(a))$. Note, however, that this is just an *interpretation*. The *definition* of $f'(a)$ is in terms of limits which are, in turn, defined precisely in Definition 2.8.



- If $f : D \to \mathbb{R}$ is differentiable, then its derivative is also a function $f' : D \to \mathbb{R}$, mapping D to $\mathbb{R}$. There is a very popular alternative notation for the derivative in this case, namely

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

This notation is convenient in some circumstances, but it tends to blur the distinction between a function (in this case f') and the *value* of the function at a particular point (in this case $f'(x)$), so we will tend to avoid it.

Let's verify that some simple, familiar functions have the derivatives we expect. In each case, we will give a direct ε—δ proof that the limit

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

exists and has the expected value.

Example 3.2 Let $f : \mathbb{R} \to \mathbb{R}$ such that $f(x) = c$, a constant. Then f is differentiable (everywhere) and $f'(a) = 0$ for all $a \in \mathbb{R}$.

Proof: We must show that, for each $\varepsilon > 0$ there is some $\delta > 0$ such that

$$\left|\frac{f(x) - f(a)}{x - a} - 0\right| < \varepsilon$$

for all x satisfying $0 < |x - a| < \delta$. So, let $\varepsilon > 0$ be given. Let $\delta = 1 > 0$. Then for all $x \in \mathbb{R}$ with $0 < |x - a| < \delta$,

$$\left|\frac{f(x) - f(a)}{x - a} - 0\right| \quad = \quad \left|\frac{c - c}{x - a}\right| = 0 < \varepsilon.$$

□

So, for a constant function, the same δ will work for every $\varepsilon > 0$. (We chose $\delta = 1$ but any other $\delta > 0$ works equally well.) Here's another example that's equally obliging:

Example 3.3 Let $f : \mathbb{R} \to \mathbb{R}$ such that $f(x) = x$. Then f is differentiable (everywhere) and $f'(a) = 1$ for all $a \in \mathbb{R}$.

Proof: We must show that, for each $\varepsilon > 0$ there is some $\delta > 0$ such that

$$\left|\frac{f(x) - f(a)}{x - a} - 1\right| < \varepsilon$$

for all x satisfying $0 < |x - a| < \delta$. So, let $\varepsilon > 0$ be given. Let $\delta = 1 > 0$. Then for all $x \in \mathbb{R}$ with $0 < |x - a| < \delta$,

$$\left|\frac{f(x) - f(a)}{x - a} - 1\right| = \left|\frac{x - a}{x - a} - 1\right| = 0 < \varepsilon.$$

□

The next examples are not quite so straightforward.

Example 3.4 Let $f : \mathbb{R} \to \mathbb{R}$ such that $f(x) = x^2$. Then f is differentiable and $f'(a) = 2a$ for all $a \in \mathbb{R}$.

Proof: Choose and fix $a \in \mathbb{R}$. For any given $\varepsilon > 0$, let $\delta = \varepsilon$. Then for all $x \in \mathbb{R}$ with $0 < |x - a| < \delta$,

$$\left|\frac{f(x) - f(a)}{x - a} - 2a\right| \quad = \quad \left|\frac{x^2 - a^2}{x - a} - 2a\right| = |x + a - 2a| = |x - a| < \delta = \varepsilon.$$

□

Example 3.5 Let $f : \mathbb{R}\backslash\{0\} \to \mathbb{R}$ such that $f(x) = 1/x$. Then f is differentiable and $f'(a) = -1/a^2$ for all $a \in \mathbb{R}\backslash\{0\}$.

Proof: Choose and fix $a \neq 0$. For any given $\varepsilon > 0$, let

$$\delta = \min\left\{\frac{|a|}{2}, \frac{|a|^3\varepsilon}{2}\right\} > 0.$$

Then, for all $x \in \mathbb{R}\backslash\{0\}$ with $0 < |x-a| < \delta$,

$$\begin{aligned}
\left|\frac{f(x)-f(a)}{x-a} + \frac{1}{a^2}\right| &= \left|\frac{(1/x)-(1/a)}{x-a} + \frac{1}{a^2}\right| \\
&= \left|\frac{a-x}{ax(x-a)} + \frac{1}{a^2}\right| \\
&= \left|-\frac{1}{ax} + \frac{1}{a^2}\right| \\
&= \frac{1}{|a|}\left|\frac{1}{a} - \frac{1}{x}\right| \\
&= \frac{1}{|a|^2|x|}\,|x-a| \\
&\leq \frac{2}{|a|^3}\,|x-a| \qquad \text{(since } |x-a| < |a|/2 \text{ so } |x| > |a|/2\text{)} \\
&< \varepsilon \qquad \text{(since } |x-a| < |a|^3\varepsilon/2\text{).}
\end{aligned}$$

□

Exercise 3.6 Give a direct ε—δ proof that $f : [0,\infty) \to \mathbb{R}$, $f(x) = \sqrt{x}$ is differentiable at every $a \in (0,\infty)$, and that

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

OK, all is as we expected. What about functions which are *not* differentiable? How do we *prove* they aren't? A function $f : D \to \mathbb{R}$ is *not* differentiable at a precisely if the limit

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

does not exist. We can prove this by thinking carefully about the negation of the statement defining a limit (Definition 2.8), but it's usually easier to exploit Theorem 2.14. Recall that this says that a function, g say, has limit L at a if and only if, for all sequences (x_n) in $D\backslash\{a\}$ converging to a, $g(x_n)$ converges to L. So to prove that $\lim_{x\to a} g(x)$ doesn't exist, it's enough to find just one sequence (x_n) in $D\backslash\{a\}$ converging to a whose image sequence $(g(x_n))$ does not converge.

Example 3.7 Let $f : \mathbb{R} \to \mathbb{R}$ such that $f(x) = |x|$. Then f is not differentiable at 0.

Proof: Assume, towards a contradiction, that f is differentiable at 0 with derivative $f'(0)$. Then

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

Let $x_n=(-1)^n/n$. Note that this is a sequence in $\mathbb{R}\backslash\{0\}$ converging to 0. Hence, by Theorem 2.14,

$$\frac{f(x_n)-f(0)}{x_n-0}\to f'(0).$$

But

$$\frac{f(x_n)-f(0)}{x_n-0}=\frac{|(-1)^n/n|}{(-1)^n/n}=\frac{1}{(-1)^n}=(-1)^n$$

which does not converge, a contradiction. Hence f is *not* differentiable at 0. $\square$

Exercise 3.6 revisited Use a sequential argument to prove that $f:[0,\infty)\to\mathbb{R}$, $f(x)=\sqrt{x}$ is **not** differentiable at 0.

There is a link between differentiablity and continuity, as we now show.

Proposition 3.8 *Let $f:D\to\mathbb{R}$ be differentiable at $a\in D$. Then f is continuous at a.*

Proof: By Theorem 2.21, it suffices to show that $\lim_{x\to a} f(x)=f(a)$. Define

$$s:D\backslash\{a\}\to\mathbb{R},\qquad s(x)=\frac{f(x)-f(a)}{x-a}.$$

Then, by assumption, s has a limit (denoted $f'(a)$) at a. But for all $x\in D\backslash\{a\}$,

$$f(x)=f(a)+(x-a)s(x),$$

and hence, by the Algebra of Limits (Theorem 2.13),

$$\lim_{x\to a} f(x)=f(a)+0\times f'(a)=f(a).$$

$\square$

Remark. The converse of Proposition 3.8 is false: if f is continuous at a, a cluster point of its domain, it does not follow that f is differentiable there. We've already seen a counteraxample: $f(x)=|x|$ is continuous at 0 but is not differentiable at 0 (see Example 3.7). In this case, the function fails to be differentiable only at a single isolated point. It is straightforward to construct functions which are *differentiable* only at a single isolated point.

Example 3.9 Let $f:\mathbb{R}\to\mathbb{R}$ be defined such that

$$f(x)=\begin{cases} x & \text{if } x\in\mathbb{Q},\\ x-x^2 & \text{if } x\notin\mathbb{Q}.\end{cases}$$

I claim that f is differentiable at 0, with $f'(0)=1$, but is not differentiable anywhere else.

Proof: Let $a \in \mathbb{R}$ and assume that f is differentiable at a. Then, by Proposition 3.8, f is continuous at a. For each $n \in \mathbb{Z}^+$ there exist $r_n \in \mathbb{Q}$ and $i_n \in \mathbb{R}\backslash\mathbb{Q}$ such that $a < r_n < a + 1/n$ and $a < i_n < a + 1/n$. Clearly $r_n \to a$ and $i_n \to a$ by the Squeeze Rule. Hence, by the definition of continuity, $f(r_n) \to f(a)$ and $f(i_n) \to f(a)$. But r_n is rational, so $f(r_n) = r_n \to a$, and i_n is irrational, so $f(i_n) = i_n - i_n^2 \to a - a^2$. Both these limits equal $f(a)$, so $a = a - a^2$, whence $a = 0$. So *if* f is differentiable at a, then $a = 0$. Equivalently, f is *not* differentiable at any $a \neq 0$.

It remains to show that f is differentiable at 0 with $f'(0) = 1$, that is,

$$\lim_{x \to 0} \frac{f(x) - f(0)}{x - 0} = 1.$$

In this case, it's easier to use Theorem 2.14 than to give a direct ε—δ argument. So, let (x_n) be any sequence in $\mathbb{R}\backslash\{0\}$ converging to 0, and consider

$$s_n = \frac{f(x_n) - f(0)}{x_n - 0} = \frac{f(x_n)}{x_n}.$$

We must show that $s_n \to 1$. If x_n is rational, then $|s_n - 1| = |x_n/x_n - 1| = 0$, whereas if x_n is irrational then

$$|s_n - 1| = \left| \frac{x_n - x_n^2}{x_n} - 1 \right| = |x_n|.$$

Hence, for all n, $0 \leq |s_n - 1| \leq |x_n|$, so $|s_n - 1| \to 0$ by the Squeeze Rule. Hence $s_n \to 1$. It follows that $f'(0) = 1$, as claimed. □

So this (admittedly rather bizarre) function is differentiable at 0, and has a positive derivative there. Naive intuition would suggest, therefore, that the function should be increasing, at least for values of x sufficiently close to 0. In fact, this is **false!**

Example 3.9 continued Claim: the function f is not increasing on any neighbourhood of 0. That is, there does not exist $\varepsilon > 0$ such that f is increasing on $(-\varepsilon, \varepsilon)$.
Proof: Let $\varepsilon > 0$ be given. Then there exists an irrational number x such that $0 < x < \varepsilon$. By definition, $f(x) = x - x^2 < x$. Similarly, there exists a *rational* number y such that $x - x^2 < y < x$. But then $y < x$ and $f(y) = y > x - x^2 = f(x)$. Hence, f is not increasing on the interval $(-\varepsilon, \varepsilon)$. This is true no matter which positive number ε we choose, which establishes the claim. □

Your reaction to this may well be "so what, that's a really crazy function – after all, it's only differentiable at the single point 0." We will see later an example of a function $f : \mathbb{R} \to \mathbb{R}$ which is differentiable *everywhere* and has $f'(0) = 1$, but still is not increasing on any neighbourhood of 0.

3.2 The rules of differentiation

In principle, if we are given an explicit function, like

$$f : \mathbb{R} \to \mathbb{R}, \qquad f(x) = (x^3 + 1)^2,$$

we can show it is differentiable and compute its derivative by directly applying Definition 3.1. This quickly becomes complicated and tedious, however. So we next develop some "derivative theorems" which will allow us to reduce differentiation of many functions to an algorithmic process. You are (I would hope) already familiar with these and quite fluent in applying them. The point of this section is to show that, now we have a mathematically precise definition of the derivative, we can rigorously *prove* them, thus putting differential calculus on a solid foundation.

Proposition 3.10 (Linearity) *Let $f : D \to \mathbb{R}$, $g : D \to \mathbb{R}$ be differentiable at $a \in D$ with derivatives $f'(a)$ and $g'(a)$, respectively, and c be a real constant. Then*

(i) $cf : D \to \mathbb{R}$ is differentiable at a with derivative $cf'(a)$.

(ii) $f + g : D \to \mathbb{R}$ is differentiable at a with derivative $f'(a) + g'(a)$.

Proof: These both follow immediately from the Algebra of Limits (Theorem 2.13):

$$\begin{aligned}
\lim_{x\to a} \frac{cf(x) - cf(a)}{x - a} &= \lim_{x\to a} c\left(\frac{f(x) - f(a)}{x - a}\right) = cf'(a) \\
\lim_{x\to a} \frac{f(x) + g(x) - (f(a) + g(a))}{x - a} &= \lim_{x\to a} \left(\frac{f(x) - f(a)}{x - a} + \frac{g(x) - g(a)}{x - a}\right) = f'(a) + g'(a)
\end{aligned}$$

□

Remark Actually the first part of this proof assumes that $\lim_{x\to a} c = c$. Have we actually proved this? Not directly. (You should compose a direct ε–δ proof in your head now. For a given $\varepsilon > 0$, how small must we take δ?) We *have* proved it indirectly however: constant functions are differentiable (Example 3.2), hence continuous (Proposition 3.8), so equal their limit at each point (Theorem 2.21).
Proposition 3.10 followed easily from the Algebra of Limits. Our next differentiation rule, and its proof, are considerably more subtle. Recall that the *composition* of two functions $f : A \to B$ and $g : B \to C$, denoted $g \circ f : A \to C$ is the function which first "does" f, then feeds the result into g, that is

$$(g \circ f)(x) = g(f(x)).$$

This only makes sense if the range of f is a subset of the domain of g. We've already proved that the composition of two continuous functions is continuous (Theorem 2.23). There's a very famous rule which tells us how to compute the derivative of a composition of two functions, if we know the derivatives of its component pieces.

Theorem 3.11 (Chain Rule) *Let $f : D \to E$ be differentiable at $a \in D$ and $g : E \to \mathbb{R}$ be differentiable at $f(a) \in E$. Then $g \circ f : D \to \mathbb{R}$ is differentiable at a, and*

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

To prove this, we will need to slightly alter the way we think of the derivative. The idea is that, if a function $g : \mathbb{R}\backslash\{a\} \to \mathbb{R}$ has limit L at a, we can extend the function to the domain $\mathbb{R}$ by defining $g(a) = L$, and this extended function is continuous at

a by Theorem 2.21. Conversely, if the function $g : \mathbb{R}\backslash\{a\} \to \mathbb{R}$ has a continuous extension to $\mathbb{R}$, it has a limit at a and that limit is $g(a)$. So we can reinterpret "g has a limit at a" as saying that "g has an extension which is continuous at a." Applying this idea to the difference quotient of a function f,

$$g(x) := \frac{f(x) - f(a)}{x - a},$$

we obtain Carathéodory's criterion for differentiablity:

Proposition 3.12 (Carathéodory's Criterion) *Let $f : D \to \mathbb{R}$ and $a \in D$ be a cluster point of D. Then f is differentiable at $a \in D$ if and only if there exists a function $\phi : D \to \mathbb{R}$ that is continuous at a and satisfies*

$$f(x) - f(a) = \phi(x)(x - a) \qquad (\clubsuit)$$

for all $x \in D$. In this case, $f'(a) = \phi(a)$.

Proof: ($\Rightarrow$): Assume f is differentiable at a with derivative $f'(a)$. Define the function

$$\phi : D \to \mathbb{R}, \qquad \phi(x) = \begin{cases} \frac{f(x)-f(a)}{x-a} & \text{if } x \neq a \\ f'(a) & \text{if } x = a. \end{cases}$$

Clearly ϕ satisifes ($\clubsuit$) for all $x \in D\backslash\{a\}$, and ($\clubsuit$) holds automatically at $x = a$ since both sides equal 0. Furthermore,

$$\lim_{x\to a} \phi(x) = \lim_{x\to a} \frac{f(x) - f(a)}{x - a} = f'(a) = \phi(a),$$

so ϕ is continuous at a by Theorem 2.21.
($\Leftarrow$): Assume that a function continuous at a satisfying ($\clubsuit$) exists. Then, dividing ($\clubsuit$) by $(x - a)$,

$$\lim_{x\to a} \frac{f(x) - f(a)}{x - a} = \lim_{x\to a} \phi(x) = \phi(a)$$

by Theorem 2.21, since ϕ is continuous at a. Hence, f is differentiable at a, with derivative $\phi(a)$. □

Proof of the Chain Rule: By Carathéodory's Criterion, there exist functions $\phi : D \to \mathbb{R}$ and $\psi : E \to \mathbb{R}$ such that ϕ is continuous at a, ψ is continuous at $f(a)$ and, for all $x \in D$, and $y \in E$,

$$\phi(x)(x - a) = f(x) - f(a), \qquad \psi(y)(y - f(a)) = g(y) - g(f(a)).$$

Define the function

$$\Phi : D \to \mathbb{R}, \qquad \Phi(x) = \psi(f(x))\phi(x).$$

By Theorem 2.23 (and the Algebra of Limits), Φ is continuous at a. Furthermore, for all $x \in D$,

$$\begin{aligned} \Phi(x)(x - a) &= \psi(f(x))\phi(x)(x - a) \\ &= \psi(f(x))(f(x) - f(a)) \\ &= g(f(x)) - g(f(a)). \end{aligned}$$

Hence, by Carathéodory's Criterion, $g \circ f$ is differentiable at a, and its derivative is

$$\Phi(a) = \psi(f(a))\phi(a) = g'(f(a))f'(a).$$

$\square$

Example 3.13 Let $f : \mathbb{R} \to \mathbb{R}$ such that $f(x) = (x^2 + 1)^2$. Then f is differentiable and $f'(x) = 4(x^2 + 1)x$.

Proof: $f = h \circ g$ where $h, g : \mathbb{R} \to \mathbb{R}$

$$\begin{aligned} g(x) &= x^2 + 1 \\ h(x) &= x^2. \end{aligned}$$

Now, by Examples 3.4 and 3.2 and the Linearity Property (Proposition 3.10) g, h are differentiable, with derivatives

$$\begin{aligned} g'(x) &= 2x + 0 \\ h'(x) &= 2x. \end{aligned}$$

Hence, by the Chain Rule, f is differentiable, and

$$f'(x) = h'(g(x))g'(x) = 2g(x)2x = 4(x^2 + 1)x.$$

$\square$

Remark You are probably more familiar with the Chain Rule expressed something like the following:

$$\frac{dy}{dx} = \frac{dy}{du}\frac{du}{dx}. \qquad (*)$$

Certainly, this has a look of plausibility about it (don't the "du's" just cancel?) and is arguably more memorable than

$$(g \circ f)'(a) = g'(f(a))f'(a). \qquad (\dagger)$$

Nonetheless, it is vastly inferior. Ask yourself what, exactly, does $(*)$ mean? The derivative of a function y (where? At x?) equals the derivative (where? At u?) of the same (?) function – but with respect to a different variable, whatever that means — times the derivative of some other function u (but isn't u a variable? Can we use the same symbol to denote both a function and a variable?) evaluated somewhere (x? Maybe?). Note that "the function y thought of as a function of u instead of x," where "u is a function of x," is actually a *different* function! And to interpret $(*)$ we're forced to talk about how we *think* about the function y, instead of what it actually *is*. This is the hallmark of imprecise, badly formulated mathematics.

By contrast $(\dagger)$ is a completely unambiguous, precise statement standing on its own. We don't need to explain separately "how we're thinking of" the terms it contains in order to make sense of it. Be warned: if ever I ask for a statement of the Chain Rule, it is Theorem 3.11 I want, not wooly, ambiguous stuff in the style of $(*)$. □

Example 3.14 Let $g : \mathbb{R} \to \mathbb{R}$ be $g(x) = f(\sqrt{2}f(\sqrt{2}x))$, where $f : \mathbb{R} \to \mathbb{R}$ is the bizarre function defined in Example 3.9. Prove that g is differentiable at 0 and compute $g'(0)$.

Proof: Recall that f is differentiable at 0, and $f'(0) = 1$. Now, let $h : \mathbb{R} \to \mathbb{R}$, $h(x) = \sqrt{2}x$. This is differentiable (everywhere, by Linearity) and $h'(x) = \sqrt{2}$. Further, let $k : \mathbb{R} \to \mathbb{R}$, $k = f \circ h$. Then f is differentiable at $h(0) = 0$, and h is differentiable at 0, so, by the Chain Rule, k is differentiable at 0 and

$$k'(0) = f'(h(0))h'(0) = 1 \times \sqrt{2}$$

Now note that $g = k \circ k$:

$$\begin{aligned} k \circ k(x) &= k(k(x)) = k(f(\sqrt{2}x)) \\ &= f(\sqrt{2}f(\sqrt{2}x)) = g(x) \end{aligned}$$

k is differentiable at $k(0) = 0$, so, by the Chain Rule, g is differentiable at 0 and

$$g'(0) = k'(k(0))k'(0) = \sqrt{2} \times \sqrt{2} = 2$$

□

Remark Note that we managed to compute $g'(0)$ without ever writing down a formula for $g(x)$! I invite you to calculate a formula for $g(x)$; it's surprisingly difficult.

You may have expected us to prove the Product Rule before the Chain Rule: it is, after all, considerably simpler. The reason we proved the Chain Rule first is that we can deduce the Product Rule from it.

Proposition 3.15 (Product Rule) *Let $f : D \to \mathbb{R}$, $g : D \to \mathbb{R}$ be differentiable at $a \in D$ with derivatives $f'(a)$ and $g'(a)$, respectively. Then $fg : D \to \mathbb{R}$ is differentiable at a with derivative $f'(a)g(a) + f(a)g'(a)$.*

Proof: Let $s : \mathbb{R} \to \mathbb{R}$ such that $s(x) = x^2$ and recall that this is differentiable with derivative $s'(x) = 2x$ by Example 3.4. Hence, by the Chain Rule and the Linearity Property, $s \circ f$, $s \circ g$ and $s \circ (f + g)$ are all differentiable at a, with derivatives

$$\begin{aligned} (s \circ f)'(a) &= 2f(a)f'(a) \\ (s \circ g)'(a) &= 2g(a)g'(a) \\ (s \circ (f + g))'(a) &= 2(f(a) + g(a))(f'(a) + g'(a)). \end{aligned}$$

Hence, by Proposition 3.10,

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

is differentiable at a with derivative

$$\begin{aligned}(fg)'(a) &= \frac{1}{2}\{2(f(a)+g(a))(f'(a)+g'(a)) - 2f(a)f'(a) - 2g(a)g'(a)\}\\ &= f'(a)g(a) + f(a)g'(a).\end{aligned}$$

□

Exercise 3.16 Give an alternative proof of the Product Rule using the Algebra of Limits for functions (Theorem 2.13).

Having established the Linearity Property and the Product Rule, it's not hard to prove that all polynomial functions are differentiable, with the derivative we expect.

Proposition 3.17 *Every polynomial function $p:\mathbb{R}\to\mathbb{R}$, $p(x) = a_0 + a_1x + \cdots + a_mx^m$ is differentiable, and its derivative is another polynomial function $p':\mathbb{R}\to\mathbb{R}$, namely*

$$p'(x) = a_1 + 2a_2x + \cdots + ma_mx^{m-1}.$$

Proof: Exercise. Try proof by induction (on the degree of p). □

Remark Proposition 3.17 retrospectively justifies our assertion that all polynomials are continuous: since they're differentiable, they're certainly continuous (by Proposition 3.8). □

Example 3.18 Let m be a positive integer and $f:\mathbb{R}\backslash\{0\}\to\mathbb{R}$ be the function $f(x) = 1/x^m$. Then f is differentiable, and $f'(x) = -m/x^{m+1}$.

Proof: $f = h\circ g$ where $g:\mathbb{R}\backslash\{0\}\to\mathbb{R}$, $g(x) = 1/x$ and $h:\mathbb{R}\to\mathbb{R}$, $h(x) = x^m$. By Examples 3.5 and 3.17, both g and h are differentiable with derivatives

$$\begin{aligned}g'(x) &= -\frac{1}{x^2}\\ h'(x) &= mx^{m-1}.\end{aligned}$$

Hence, by the Chain Rule, f is differentiable, and

$$\begin{aligned}f'(x) &= h'(g(x))g'(x)\\ &= mg(x)^{m-1}\left(-\frac{1}{x^2}\right)\\ &= -m\left(\frac{1}{x}\right)^{m-1}\left(\frac{1}{x}\right)^2\\ &= -\frac{m}{x^{m+1}}\end{aligned}$$

by Example 3.5 and Proposition 3.17. □

Generalizing the trick used in this proof, we can obtain another useful rule of differentiation:

Proposition 3.19 (Quotient Rule) *Let $f : D \to \mathbb{R}$ and $g : D \to \mathbb{R}\backslash\{0\}$, be differentiable at $a \in D$. Then f/g is differentiable at a and*

$$\left(\frac{f}{g}\right)'(a) = \frac{f'(a)g(a) - f(a)g'(a)}{g(a)^2}.$$

Proof: Exercise. □

3.3 Open sets and the Localization Lemma

We will often have to deal with functions $f : \mathbb{R} \to \mathbb{R}$ that are defined *piecewise*, by combining different formulae for $f(x)$ valid on different subsets of the domain. A simple example is

$$f : \mathbb{R} \to \mathbb{R}, \qquad f(x) = |x| = \begin{cases} x, & x \in [0, \infty), \\ -x, & x \in (-\infty, 0). \end{cases}$$

It would be handy if we could deduce information about the differentiablity of such functions directly from their constituent pieces. Luckily, we can, because limits, and hence derivatives, are inherently *local* objects. That is, if two functions f, g coincide for all x "close to" a point a, they have the same derivative at a. To formulate this statement precisely (and hence be able to prove it) we need to introduce the concept of *open subsets* of $\mathbb{R}$.

Definition 3.20 $U \subseteq \mathbb{R}$ is **open** if, for each $a \in U$ there exists $\delta > 0$ such that $(a - \delta, a + \delta) \subseteq U$.

Example 3.21 (i) $\mathbb{R}$ itself is open: given any $a \in \mathbb{R}$, $(a - 52, a + 52) \subseteq \mathbb{R}$, for example.

(ii) $(0, 1)$ is open: given any $a \in (0, 1)$ we can take $\delta = \min\{|a|, |1 - a|\} > 0$. Then $(a - \delta, a + \delta) \subseteq (0, 1)$.

(iii) $[0, 1)$ is not open: $0 \in [0, 1)$ but, for all $\delta > 0$, $(-\delta, \delta) \not\subseteq [0, 1)$.

(iv) $\emptyset$ is open (careful!).

Example 3.22 Determine whether the following sets are open.

$$\mathbb{Q}, \qquad \mathbb{R}\backslash\mathbb{Q}, \qquad \mathbb{R}\backslash\mathbb{Z}, \qquad \mathbb{R}\backslash\{1/n \,:\, n \in \mathbb{Z}^+\}.$$

- $\mathbb{Q}$

  This is **not** open: it contains 0 (for example), but for all $\delta > 0$ there is an irrational number x between $-\delta$ and δ, so $(0-\delta, 0+\delta)$ is not a subset of $\mathbb{Q}$.

- $\mathbb{R}\backslash\mathbb{Q}$

  This is **not** open, by almost identical reasoning: it contains $\sqrt{2}$ (for example), but for all $\delta > 0$ there is a rational number x between $\sqrt{2}-\delta$ and $\sqrt{2}+\delta$, so $(\sqrt{2}-\delta, \sqrt{2}+\delta)$ is not a subset of $\mathbb{R}\backslash\mathbb{Q}$.

- $\mathbb{R}\backslash\mathbb{Z}$

  This is open: Given any $a \notin \mathbb{Z}$, its distance to the nearest integer $\delta > 0$, and $(a-\delta, a+\delta)$ contains no integers. Hence $(a-\delta, a+\delta) \subset \mathbb{R}\backslash\mathbb{Z}$.

- $\mathbb{R}\backslash\{1/n : n \in \mathbb{Z}^+\}$

  This is **not** open: it contains 0, but given any $\delta > 0$ there exists $n \in \mathbb{Z}^+$ such that $n > 1/\delta$, and hence $0 < 1/n < \delta$. It follows that $(-\delta, \delta)$ is not a subset of $\mathbb{R}\backslash\{1/n : n \in \mathbb{Z}^+\}$.

We can now give a precise formulation of the idea that derivatives are local.

Lemma 3.23 (Localization Lemma) *Let $f : D \to \mathbb{R}$ coincide with a differentiable function $g : U \to \mathbb{R}$ on some open set $U \subseteq D$. Then f is differentiable, with derivative g', on U.*

Proof: Let $a \in U$. We must show that

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

Let $\varepsilon > 0$ be given. Since U is open, there exists $\delta_1 > 0$ such that $(a-\delta_1, a+\delta_1) \subseteq U \subseteq D$. Since g is differentiable at a, there exists $\delta_2 > 0$ such that for all $x \in U$ with $0 < |x-a| < \delta_2$,

$$\left|\frac{g(x)-g(a)}{x-a} - g'(a)\right| < \varepsilon.$$

Let $\delta = \min\{\delta_1, \delta_2\} > 0$. Then for all $x \in D$ with $0 < |x-a| < \delta$, $x \in U$ so $f(x) = g(x)$ (and $f(a) = g(a)$), and $0 < |x-a| < \delta_2$, so

$$\left|\frac{f(x)-f(a)}{x-a} - g'(a)\right| = \left|\frac{g(x)-g(a)}{x-a} - g'(a)\right| < \varepsilon.$$

$\square$

Example 3.24 The function $f : \mathbb{R} \to \mathbb{R}$, $f(x) = |x|$ coincides with the differentiable function $g(x) = x$ on the open set $(0, \infty)$. Hence, by the Localization Lemma, f is differentiable on $(0, \infty)$ with derivative $f'(x) = 1$. Similarly, f coincides with the differentiable function $h(x) = -x$ on the open set $(-\infty, 0)$. Hence, by the Localization Lemma, f is differentiable on $(-\infty, 0)$ with derivative $f'(x) = -1$.

Warning! The condition that U is an *open* set is crucial for the Localization Lemma. For example, $f(x) = |x|$ coincides with the differentiable function $g(x) = x$ on $U = [0, \infty)$, but f is *not* differentiable on U (it fails to be differentiable at 0).

Summary

- A function $f : D \to \mathbb{R}$ is **differentiable at** $a \in D$ if
$$\lim_{x \to a} \frac{f(x) - f(a)}{x - a}$$
exists. In this case, we denote the limit $f'(a)$ and call it the **derivative** of f at a. We say that f is **differentiable** if it is differentiable at a for all $a \in D$.
- The limit in $f'(a)$ is defined as in Definition 2.8.
- If a function is differentiable at a, it is continuous at a. The converse is false.
- Using this definition of $f'(a)$, we can prove that derivatives obey the usual rules of differential calculus, to wit:
$$\begin{array}{lrcl} \text{Linearity:} & (\alpha f + \beta g)'(a) & = & \alpha f'(a) + \beta g'(a), \\ \text{Product Rule:} & (fg)'(a) & = & f'(a)g(a) + f(a)g'(a), \\ \text{Quotient Rule:} & (f/g)'(a) & = & \dfrac{f'(a)g(a) - f(a)g'(a)}{g(a)^2}, \\ \text{Chain Rule:} & (g \circ f)'(a) & = & g'(f(a))f'(a). \end{array}$$
- $U \subseteq \mathbb{R}$ is **open** if for each $a \in U$ there exists $\delta > 0$ such that $(a - \delta, a + \delta) \subseteq U$.
- Derivatives are **local**: if $f : D \to \mathbb{R}$ coincides with a differentiable function $g : U \to \mathbb{R}$ on some open set $U \subseteq D$, then f is differentiable, with derivative g', on U.

Chapter 4

Functions differentiable on an interval

Two applications of differential calculus are particularly pervasive in the natural sciences. In one, we use the derivative of a function to deduce its maximum (or minimum) value and the input which produces that value. In the second, we are given information, not about the function itself, but about its derivative (which is often interpreted as a "rate of change" with respect to time). We then seek to reconstruct the function itself from this information. These two applications rely fundamentally on the Interior Extremum Theorem, and the Mean Value Theorem, respectively, and the main purpose of this section is to state and prove these. Both are statements concerning functions which are differentiable (at least) on an open interval (a, b). The functions in question may be differentiable elsewhere too (perhaps on the whole of $\mathbb{R}$, in fact), so it's important when you read "f is differentiable on (a, b)" not to add the unwarranted assumption "and only on (a, b)".

4.1 The Interior Extremum Theorem

Recall that $f : D \to \mathbb{R}$ attains a **maximum** at $a \in D$ if $f(x) \leq f(a)$ for all $x \in D$. Similarly, f attains a **minimum** at $a \in D$ if $f(x) \geq f(a)$ for all $x \in D$. Note that these definitions have *absolutely nothing to do with calculus!* In fact, they make perfectly good sense whatever set D, the domain of f, is: it need not be a subset of $\mathbb{R}$!

We say that f attains an **extremum** at $a \in D$ if it attains either a maximum or minimum at a. (The word "extremum" means "maximum or minimum", in much the same way that "monotonic" means "increasing or decreasing". Its plural is "extrema".)

Theorem 4.1 (The Interior Extremum Theorem) *Let $f : (a, b) \to \mathbb{R}$ be differentiable and f attain an extremum at $c \in (a, b)$. Then $f'(c) = 0$.*

Proof: Consider first the case that f attains a maximum at $c \in (a, b)$. By assumption, $f'(c)$ exists. Hence, by Theorem 2.14, given any sequence (x_n) in $(a, b)\backslash\{c\}$

which converges to c,

$$s(x_n) := \frac{f(x_n) - f(c)}{x_n - c} \to f'(c). \tag{4.1}$$

This holds, in particular, for the sequence $x_n = c + (b-c)/(n+1) \in (c, b)$. Note that $x_n > c$ for all c, so $x_n - c > 0$, and $f(c)$ is the maximum value of f, so $f(x_n) - f(c) \leq 0$. Hence $s(x_n) \leq 0$ for all n so, by Proposition 1.7, its limit is non-positive, i.e. $f'(c) \leq 0$.

But (4.1) also holds in the case where $x_n = c-(c-a)/(n+1) \in (a, c)$. Now $x_n - c < 0$ for all n, and $f(x_n) - f(c) \leq 0$ (since, as before, f attains its maximum at c), so $s(x_n) \geq 0$ for all n. Hence, by Proposition 1.7, its limit is non-negative, i.e. $f'(c) \geq 0$.

Hence, $f'(c) = 0$, as was to be shown.

Consider now the case that f attains a minimum at $c \in (a, b)$. Let $g = -f$. Then g is differentiable on (a, b) and attains a maximum at c, so, as we just proved, $g'(c) = 0$. Hence $f'(c) = -g'(c) = 0$. □

Definition 4.2 Let $f : D \to \mathbb{R}$ be differentiable. Then $c \in D$ is a **critical point** of f if $f'(c) = 0$.

Given a differentiable function $f : [a, b] \to \mathbb{R}$, we know it attains both a maximum and a minimum by the Extreme Value Theorem. Either or both of these extrema might occur at an endpoint of the interval $[a, b]$. But if they don't, they occur at some interior point $c \in (a, b)$, and this point must be a *criticial point* of f by the Interior Extremum Theorem. So to find the maximum and minimum values attained by f, we just need to evaluate it at the endpoints a, b and any interior critical points, and extract the extreme values.

Example 4.3 Find the maximum and minimum values attained by the function

$$f : [0, 3] \to \mathbb{R}, \qquad f(x) = x^2 - 2x + 7.$$

Solution: First note that, since f is continuous, it certainly attains both a maximum and a minimum by the Extreme Value Theorem (Theorem 2.26). Since f is differentiable on $(0, 3)$, it follows from the Interior Extremum Theorem that each extremum occurs either at an endpoint, that is, 0 or 3, or at an interior critical point of f. Now

$$f'(x) = 2x - 2 = 0$$

if and only if $x = 1$, so 1 is the only critical point of f. Since $f(1) = 6$, $f(0) = 7$, and $f(3) = 10$, we deduce that f attains a maximum value of 10 at the right endpoint 3, and a minimum value of 6 at the interior critical point 1.

Remark In this case, using calculus was really overkill, since we could have deduced the same information by simply completing the square:

$$f(x) = (x - 1)^2 + 6.$$

We can also use information about the derivative to show, in some circumstances, that the extrema of a differentiable function on $[a, b]$ do *not* occur at its endpoints.

Proposition 4.4 *Let $f : [a, b] \to \mathbb{R}$ be differentiable.*

(i) If $f'(a) > 0$, then $f(a)$ is not the maximum value attained by f.

(ii) If $f'(a) < 0$, then $f(a)$ is not the minimum value attained by f.

(iii) If $f'(b) > 0$, then $f(b)$ is not the minimum value attained by f.

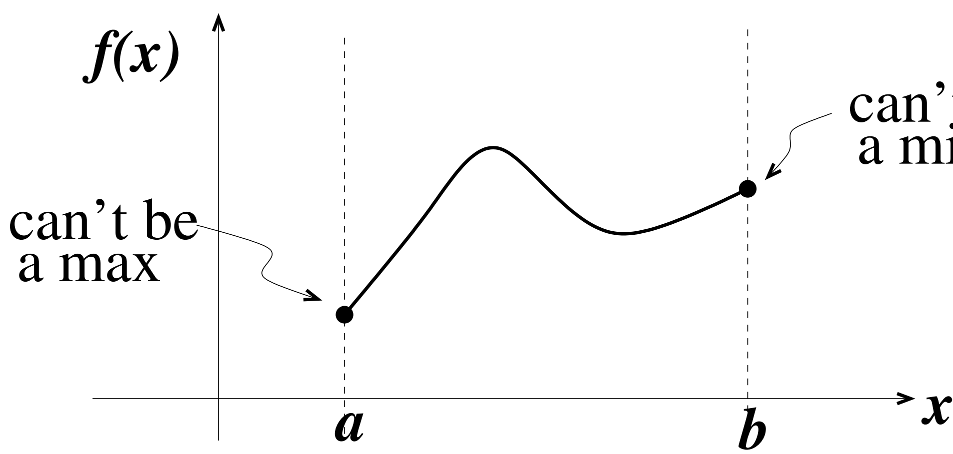
(iv) If $f'(b) < 0$, then $f(b)$ is not the maximum value attained by f.

Proof: The proofs of each part are very similar, so we will just prove (i).

Assume $f'(a) > 0$. Assume further, towards a contradiction, that f does attain its maximum at a. Since f is differentiable at a,

$$s_n = \frac{f(x_n) - f(a)}{x_n - a} \to f'(a) > 0,$$

for any sequence (x_n) in $(a, b]$ converging to a (by Theorem 2.14). But $f(x_n) \leq f(a)$ by assumption, and $x_n > a$, so $s_n \leq 0$ for all n. Hence $f'(a) \leq 0$ by Proposition 1.7, a contradiction. $\square$



Exercise 4.5 Write out the proof of (at least) one of the other parts of Proposition 4.4.

4.2 The Mean Value Theorem

Our next theorem has hypotheses (assumptions on f) which look very restrictive – so restrictive that one might wonder whether the theorem is of any practical use at all. In fact, as we shall see, it has very powerful and useful consequences.

Theorem 4.6 (Rolle's Theorem) *Let f be continuous on $[a, b]$ and differentiable on (a, b). Assume $f(a) = f(b)$. Then there exists $c \in (a, b)$ such that $f'(c) = 0$.*

Proof: By the Extreme Value Theorem, f attains both a maximum and a minimum value on $[a, b]$. If both of these occur at the endpoints of $[a, b]$ then, since $f(a) = f(b)$, the maximum value coincides with the minimum value, whence it follows that

$f : [a, b] \to \mathbb{R}$ is constant. But then $f'(x) = 0$ for all $x \in (a, b)$, so $c = (b + a)/2$, for example, has $f'(c) = 0$.

Hence, we may assume that either the maximum or the minimum value does not occur at an endpoint. But then f attains an extremum at some interior point $c \in (a, b)$, so $f'(c) = 0$ by the Interior Extremum Theorem. □

Theorem 4.7 (The Mean Value Theorem) *Let f be continuous on $[a, b]$ and differentiable on (a, b). Then there exists $c \in (a, b)$ such that*

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

Proof: Let $\alpha = (f(b) - f(a))/(b - a)$ and define $g : [a, b] \to \mathbb{R}$ such that

$$g(x) = f(x) - \alpha(x - a).$$

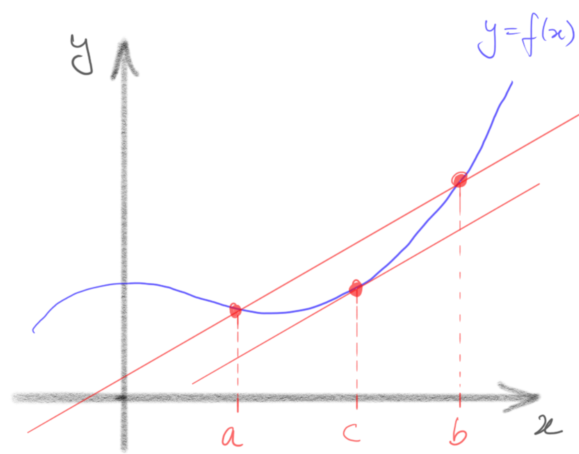
Then g is continuous, and is differentiable on (a, b). Furthermore $g(a) = f(a)$ and $g(b) = f(a)$. So g satisfies the hypotheses of Rolle's Theorem, and we deduce that there exists $c \in (a, b)$ such that $g'(c) = 0$. But

$$g'(c) = f'(c) - \alpha$$

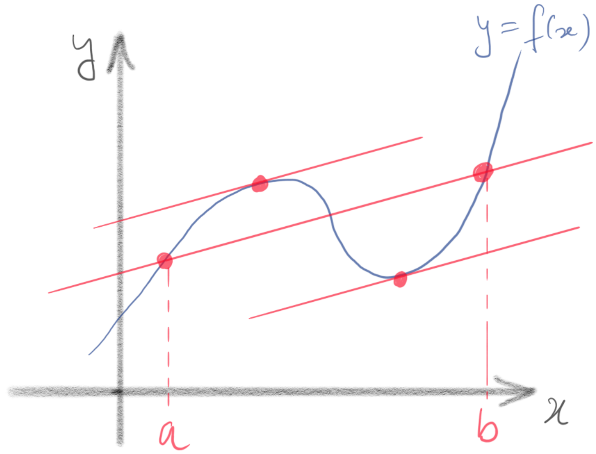
which completes the proof. □

Remarks

- The Mean Value Theorem has a nice geometric interpretation. Consider the graph $y = f(x)$ of a differentiable function $f : \mathbb{R} \to \mathbb{R}$. For each pair of distinct numbers $a < b$, we can construct the chord (straight line) passing through the points $(a, f(a))$ and $(b, f(b))$ on the graph. Its slope is $(f(b)-f(a))/(b-a)$. The Mean Value Theorem asserts that, at some point $(c, f(c))$ on the graph between these two points, the tangent line to the graph is parallel to the chord.



- The Mean Value Theorem guarantees the existence of a point c with the required $f'(c)$. It says nothing about uniqueness: there could be more than one such point.



- Clearly Rolle's Theorem is a special case of the Mean Value Theorem (geometrically, the case where the chord is horizontal). It's slightly surprising at first sight, therefore, that Rolle's Theorem actually implies the Mean Value Theorem.

You perhaps don't realize it, but you have probably been using the Mean Value Theorem rather a lot. We have seen (Example 3.2) that if a function is constant then its derivative is zero everywhere. When one solves a differential equation, one often uses the converse fact: if a function has zero derivative everywhere, then it must be constant. But how do we *know* this? It's a consequence of the Mean Value Theorem!

Example 4.8 Suppose we know that $f : \mathbb{R} \to \mathbb{R}$ is differentiable and satisfies $f'(x) = x^2$ for all x, and $f(0) = 2$. Then surely

$$f(x) = \frac{x^3}{3} + 2,$$

right? But how do we *know* this? Certainly, $F(x) := \frac{x^3}{3} + 2$ is an example of a function whose derivative at each x is x^2, and whose value at 0 is 2, so it's *possible* that $f = F$. But, perhaps there's more than one function satisfying those properties?

In fact, we can *prove* that $f = F$, as follows. Let $g : \mathbb{R} \to \mathbb{R}$ be

$$g(x) = f(x) - F(x) = f(x) - \frac{x^3}{3} - 2.$$

Then, for all x,

$$g'(x) = f'(x) - x^2 = 0.$$

Now assume, towards a contradiction, that g is not constant. Then there exist $a, b \in \mathbb{R}$ with $a < b$ such that $g(a) \neq g(b)$. But then, by the Mean Value Theorem applied to g on $[a, b]$, there exists $c \in (a, b)$ such that

$$g'(c) = \frac{g(b) - g(a)}{b - a} \neq 0$$

a contradiction since $g'(x) = 0$ everywhere. Hence, g *is* constant, so for all $x \in \mathbb{R}$ $g(x) = g(0) = 0$, that is, $f = F$. $\square$

When we first learn calculus, we are taught to think of $f'(x)$ as the "rate of change of f with respect to the variable x". So if, at a particular point $a \in D$, $f'(a) > 0$, we expect that f should be strictly increasing, at least sufficiently close to a. You've already seen (Example 3.9) that this expectation can be false: the relationship between f' and the increasing/decreasing behaviour of f can be subtle. Before proceeding further, let's make absolutely precise what we mean by *increasing* and *decreasing* functions.

Definition 4.9 $f : D \to \mathbb{R}$ is **increasing** if for all $x, y \in D$ with $x < y$, $f(x) \leq f(y)$. It is **strictly increasing** if for all $x, y \in D$ with $x < y$, $f(x) < f(y)$. Similarly, f is **decreasing** if for all $x, y \in D$ with $x < y$, $f(x) \geq f(y)$, and **strictly decreasing** if for all $x, y \in D$ with $x < y$, $f(x) > f(y)$.

Example 4.10 $f : \mathbb{R} \to \mathbb{R}$, $f(x) = x^3$. This is strictly increasing: if $x < y$ then

$$\begin{aligned} f(x) - f(y) &= (x-y)(x^2 + xy + y^2) \\ &= (x-y)((x+y/2)^2 + 3y^2/4) \\ &< 0 \end{aligned}$$

since $((x + y/2)^2 + 3y^2/4) > 0$ unless $x = y = 0$.

Note that Definition 4.9 makes no mention of the derivative of f and may, in principle, apply when f is not differentiable, or even continuous. For example, the step function defined in Example 2.17, is an increasing function. Our next result uses the Mean Value Theorem to establish some useful links between f' and the increasing/decreasing behaviour of f.

Proposition 4.11 *Let $I \subseteq \mathbb{R}$ be an interval and $f : I \to \mathbb{R}$ be differentiable. Then*

(i) f is increasing if and only if $f'(x) \geq 0$ for all $x \in I$.

(ii) f is decreasing if and only if $f'(x) \leq 0$ for all $x \in I$.

(iii) f is constant if and only if $f'(x) = 0$ for all $x \in I$.

(iv) If $f'(x) > 0$ for all $x \in I$ then f is strictly increasing.

(v) If $f'(x) < 0$ for all $x \in I$ then f is strictly decreasing.

Proof:

(i) Assume $f'(x) \geq 0$ for all x in I, but that f is *not* increasing. Then there exist $a, b \in I$ with $a < b$ such that $f(a) > f(b)$. But then, by the MVT, at some $c \in (a, b)$,

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

contradicting the assumption that $f' \geq 0$. Hence f is increasing.

Conversely, assume that f is increasing. If I has a left endpoint $a \in I$, then f attains a minimum at a, so $f'(a) \geq 0$ by Proposition 4.4. Similarly, if I has a right endpoint $b \in I$ then f attains a maximum at b so $f'(b) \geq 0$ by Proposition 4.4. Let x be an interior point of I, so $x \in (a, b) \subseteq I$. Consider the sequence

$$x_n = x + \frac{b - x}{n + 1} \in (x, b) \subset I.$$

This converges to x, and f is differentiable at x, so

$$s_n := \frac{f(x_n) - f(x)}{x_n - x} \to f'(x).$$

But $s_n \geq 0$ for all n, since f is increasing and $x_n > x$. Hence it's limit $f'(x) \geq 0$ by Proposition 1.7.

(ii) Exercise (just modify the proof of part (i) in the obvious way).

(iii) If $f(x) = c$ for all x then $f'(x) = 0$ (Example 3.2). Conversely, if $f'(x) = 0$ for all x, then f is increasing by part (i) and decreasing by part (ii). So for all $x, y \in I$ with $x < y$, $f(x) \leq f(y)$ and $f(x) \geq f(y)$, whence $f(x) = f(y)$. Hence f is constant.

(iv) Assume $f'(x) > 0$ for all x in I, but that f is *not* strictly increasing. Then there exist $a, b \in I$ with $a < b$ such that $f(a) \geq f(b)$. But then, by the MVT, at some $c \in (a, b)$,

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

contradicting the assumption that $f' > 0$. Hence f is strictly increasing. (Note this is almost exactly the same as the proof of the "if" direction of part (i).)

(v) Exercise (just modify the proof of part (iv) in the obvious way).

□

Remarks

- Note that the converses of parts (iv) and (v) of Proposition 4.11 are **false**! Example 4.10 is a counterexample: $f(x) = x^3$ is strictly increasing, but $f'(0) = 0$, so its derivative is not positive everywhere.

- The condition that I is an **interval** is important for Proposition 4.11. For example,

$$f : \mathbb{R}\backslash\{0\} \to \mathbb{R}, \qquad f(x) = \frac{1}{x}$$

has $f'(x) = -1/x^2 < 0$ everywhere, but $f(-1) = -1 < 1 = f(1)$, so f is not a decreasing function. This doesn't contradict Proposition 4.11 since $\mathbb{R}\backslash\{0\}$ is not an interval.

- Since strictly increasing and strictly decreasing functions are automatically injective, Proposition 4.11 gives us a sneaky way of showing that some functions are injective.

Example 4.12 Let $f : \mathbb{R} \to \mathbb{R}$ such that $f(x) = 2x + \sin x$. Then f is strictly increasing, and hence injective.

Proof: The function f is differentiable, with derivative[1]

$$f'(x) = 2 + \cos x \geq 1 > 0.$$

Hence, by Proposition 4.11, f is strictly increasing, and hence injective. $\square$

Looking back at Example 4.8, we can see that we really used part (iii) of Proposition 4.11: since $g(x) = f(x) - F(x)$ has derivative 0 everywhere on the interval $\mathbb{R}$, it must be constant, and $g(0) = 2 - 2 = 0$, so $f(x) = F(x)$.

We finish this section by studying a function $f : \mathbb{R} \to \mathbb{R}$ with some very counter-intuitive properties.

Example 4.13 Consider the function $f : \mathbb{R} \to \mathbb{R}$ defined so that

$$f(x) = \begin{cases} x + 2x^2 \sin(1/x) & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}$$

Claim: f is differentiable everywhere, has $f'(0) = 1 > 0$, but is not increasing on any neighbourhood of 0.

Proof: f coincides with the function $g : U \to \mathbb{R}$, $g(x) = x + 2x^2 \sin(1/x)$ on the open set $U = \mathbb{R}\backslash\{0\}$. g is differentiable by the Product and Chain Rules, so f is differentiable on U by the Localization Lemma, and for all $a \neq 0$,

$$f'(a) = g'(a) = 1 + 4a \sin(1/a) - 2\cos(1/a). \tag{4.2}$$

Consider now the case $a = 0$. Let (x_n) be any sequence in $\mathbb{R}\backslash\{0\}$ converging to 0. Then

$$s_n := \frac{f(x_n) - f(0)}{x_n - 0} = 1 + 2x_n \sin(1/x_n).$$

Now $-|x_n| \leq x_n \sin(1/x_n) \leq |x_n|$, so $x_n \sin(1/x_n) \to 0$ by the Squeeze Rule, and hence $s_n \to 1$ (by the Algebra of Limits). Hence, f is differentiable at 0 and $f'(0) = 1$.

For the final part, assume, to the contrary, that there exists $\varepsilon > 0$ such that $f : (-\varepsilon, \varepsilon) \to \mathbb{R}$ is increasing. Then, by Proposition 4.11, $f'(x) \geq 0$ for all $x \in (-\varepsilon, \varepsilon)$. But there exists $m \in \mathbb{Z}^+$ such that $m > 1/(2\pi\varepsilon)$. Then $x = 1/(2\pi m) \in (0, \varepsilon)$ and, by equation (4.2),

$$f'(x) = 1 + 4x \times 0 - 2 \times 1 = -1 < 0,$$

a contradiction. $\square$

Note that we proved f can't be increasing on any interval containing 0 by showing that its derivative takes negative values on every such interval. So we are again using Proposition 4.11.

[1] Strictly speaking, you haven't yet proved that sin is differentiable, with derivative cos, so this argument requires you to suspend disbelief for the time being.

4.3 The Extended Mean Value Theorem and L'Hospital's Rule

Sometimes a more general version of the Mean Value Theorem is useful.

Theorem 4.14 (Extended Mean Value Theorem) *Let f and g be real functions which are continuous on $[a, b]$ and differentiable on (a, b), and assume that, for all $x \in (a, b)$, $g'(x) \neq 0$. Then there exists $c \in (a, b)$ such that*

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

Proof: Exercise. Here are some hints:

(i) Define the constant $\alpha = \dfrac{f(b) - f(a)}{g(b) - g(a)}$. (Question: how do we know $g(b) \neq g(a)$?)

(ii) Apply Rolle's Theorem to $h(x) = f(x) - \alpha g(x)$.

First note that the conditions on g imply that $g(a) \neq g(b)$ (since if $g(b) = g(a)$ then $g'(x) = 0$ for some $x \in (a, b)$ by Rolle's Theorem). Hence,

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

is a well-defined real constant. Let $h : [a, b] \to \mathbb{R}$ such that $h(x) = f(x) - \alpha g(x)$. Then h is continuous on $[a, b]$ and differentiable on (a, b), and, as one may easily verify,

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

Hence h satisfies the hypotheses of Rolle's Theorem, and we conclude that there exists $c \in (a, b)$ such that $h'(c) = 0$. But then $f'(c) = \alpha g'(c)$, whence the result immediately follows.

$\square$

If we apply Theorem 4.14 in the case where $g(x) = x$, we immediately deduce the usual Mean Value Theorem[2]. We will see that other choices of g can be useful. For

[2] Or, even better, apply it in the case $g(x) = x - a$. Then the *proof* of Theorem 4.14 reduces to the proof of Theorem 4.7.

example, we can use Theorem 4.14 to rigorously justify a popular trick for computing limits called "L'Hospital's Rule".

Theorem 4.15 (L'Hospital's Rule) *Let $I \subseteq \mathbb{R}$ be an open interval, $f, g : I \to \mathbb{R}$ be differentiable functions, $a \in I$, and $f(a) = g(a) = 0$. Assume that, for all $x \in I \backslash \{a\}$, $g(x) \neq 0$ and $g'(x) \neq 0$. Then*

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

provided the limit on the right exists.

Proof: Let (x_n) be any sequence in $I \backslash \{a\}$ converging to a, and for each n, consider f, g restricted to the closed interval with endpoints a and x_n. These functions satisfy the hypotheses of Theorem 4.14, so there exists c_n, between a and x_n, such that

$$\frac{f'(c_n)}{g'(c_n)} = \frac{f(x_n) - f(a)}{g(x_n) - g(a)} = \frac{f(x_n)}{g(x_n)}.$$

By the Squeeze Rule, $c_n \to a$, and so, assuming the limit

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

exists, call it $L \in \mathbb{R}$, say, $f'(c_n)/g'(c_n) \to L$ (by the definition of limit). Hence $f(x_n)/g(x_n) \to L$, as was to be shown. $\square$

Example 4.16 Claim: $\displaystyle\lim_{x \to 0} \frac{1 - \cos x}{\sin x} = 0.$

Proof: Let $I = (-\pi/2, \pi/2)$, $f(x) = 1 - \cos x$, $g(x) = \sin x$, and $a = 0 \in I$. Then f, g satisfy the hypotheses of L'Hospital's Rule on I.

$$\lim_{x \to 0} \frac{f'(x)}{g'(x)} = \lim_{x \to 0} \frac{\sin x}{\cos x} = \frac{0}{1} = 0.$$

Hence, by L'Hospital's Rule,

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

also. $\square$

Example 4.17 Claim: $\displaystyle\lim_{x \to \pi} \frac{\sin x}{e^x - e^\pi} = -e^{-\pi}.$

Proof: To fit this problem into the template of Theorem 4.15, we define

$$\begin{aligned} f(x) &= \sin x \\ g(x) &= e^x - e^\pi \\ I &= \mathbb{R} \\ a &= \pi \end{aligned}$$

These satisfy the hypotheses of the theorem. Then

$$\begin{aligned} f'(x) &= \cos x \\ g'(x) &= e^x \\ \Rightarrow \lim_{x\to\pi} \frac{f'(x)}{g'(x)} &= \lim_{x\to\pi} \frac{\cos x}{e^x} \\ &= \frac{\cos \pi}{e^\pi} \qquad \text{(since both cos and exp are continuous functions)} \\ &= -e^{-\pi} \end{aligned}$$

Hence, by Theorem 4.15,

$$\lim_{x\to\pi} \frac{\sin x}{e^x - e^\pi} = -e^{-\pi}$$

$\square$

4.4 Higher derivatives and Taylor's Theorem

If a function $f : D \to \mathbb{R}$ is differentiable (everywhere on D) then its derivative defines another function $f' : D \to \mathbb{R}$, and we can ask whether this function, in turn, is differentiable (recall that, in general, it might not even be continuous). If f' is differentiable, then *its* derivative is denoted $f'' : D \to \mathbb{R}$ and called the *second derivative* of f. Similarly, if $f'' : D \to \mathbb{R}$ is differentiable, its derivative $f''' : D \to \mathbb{R}$ is the *third derivative* of f. Proceeding inductively, we can define the n^{th} derivative of f, denoted $f^{(n)}$ to be the derivative (if it exists), of $f^{(n-1)}$, where $f^{(0)}$ is, by definition, just the function f itself (so $f^{(1)} = f'$, $f^{(2)} = f''$ etc.). If $f^{(n)} : D \to \mathbb{R}$ exists, we say that f is n times differentiable, and if f is n times differentiable for all $n \in \mathbb{Z}^+$, we say that f is *smooth*. It is convenient to develop some terminology and notation for describing the differentiability of functions:

Definition 4.18 A function $f : D \to \mathbb{R}$ is **continuously differentiable** if it is differentiable and $f' : D \to \mathbb{R}$ is continuous. The function is n **times continuously differentiable** if all its derivatives up to the n^{th} exist everywhere on D, and are continuous. We denote the set of such functions by $C^n(D)$ and say that "f is C^n" if $f \in C^n(D)$. The function f is **smooth** if it is C^n for all $n \in \mathbb{Z}^+$. The set of smooth functions on D is denoted $C^\infty(D)$. For completeness, we denote the set of **continuous** functions $D \to \mathbb{R}$ by $C^0(D)$.

Example 4.19 (i) Every polynomial function $p : \mathbb{R} \to \mathbb{R}$, $p(x) = a_0 + a_1 x + \cdots + a_m x^m$, is smooth. This follows immediately from Proposition 3.17.

(ii) The function $f : \mathbb{R} \to \mathbb{R}$, $f(x) = |x|^3$ is C^2 but not C^3. (Exercise: prove it.)

(iii) The function $f : \mathbb{R} \to \mathbb{R}$ defined in Example 4.13 is differentiable, but not C^1 (because f' is discontinuous).

An interesting interpretation of the Mean Value Theorem can be given as follows. Let $I \subseteq \mathbb{R}$ be an open interval, $f : I \to \mathbb{R}$ be differentiable, and choose and fix some $a \in I$. Then for any other $x \in I$, we may apply the Mean Value Theorem to $f : [a, x] \to \mathbb{R}$ (if $x > a$) or $f : [x, a] \to \mathbb{R}$ (if $x < a$) to deduce that there exists c between a and x such that

$$f(x) = f(a) + f'(c)(x - a).$$

Note that the same equation holds trivially in the case where $x = a$ (with $c = x = a$). So, for x close to a (meaning that $|x - a|$ is small) we can approximate f by the constant $f(a)$, and the error in this approximation is of size $|f'(c)||x - a|$, where c is somewhere between a and x (and, of course, depends on x).

The moral is that differentiability (once) of f on an open interval allows one to approximate f by a degree 0 polynomial function (i.e. a constant), with an error controlled by f'. It turns out that if f is n times differentiable one can do better: one can approximate f by a degree $(n - 1)$ polynomial function, with an error controlled by $f^{(n)}$. To prove this, we will again use the Extended Mean Value Theorem (Theorem 4.14).

Theorem 4.20 (Taylor's Theorem) *Let $I \subseteq \mathbb{R}$ be an interval, $f : I \to \mathbb{R}$ be $(n+1)$ times differentiable, and $a, x \in I$. Then there exists c between a and x such that*

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

Proof: First note that the claimed equation holds trivially (with $c = a$) in the case where $x = a$. Choose and fix $a, x \in I$, $x \neq a$, and consider the function $F : I \to \mathbb{R}$,

$$F(t) = f(x) - f(t) - f'(t)(x - t) - \frac{f''(t)}{2!}(x - t)^2 - \cdots - \frac{f^{(n)}(t)}{n!}(x - t)^n.$$

Then the claim to be proved is that there exists c between a and x such that

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

Now, by the Product Rule, F is differentiable, and, as is easily verified (exercise!),

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

Let $G(t) = (x-t)^{n+1}$. Then F, G satisfy the hypotheses of Theorem 4.14 on the interval with endpoints a, x (note that $G'(t) = 0$ only if $t = x$), so there exists c between a and x such that

$$\begin{aligned} \frac{F'(c)}{G'(c)} &= \frac{F(x)-F(a)}{G(x)-G(a)} = \frac{F(a)}{G(a)} \\ \text{i.e.} \qquad \frac{-f^{(n+1)}(c)(x-c)^n}{n![-(n+1)(x-c)^n]} &= \frac{F(a)}{(x-a)^{n+1}}, \end{aligned}$$

whence the claimed expression for $F(a)$ immediately follows. □

The polynomial

$$p_n(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n$$

is often called the n^{th} **Taylor approximant** of the function f about a, and

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

is called the **remainder**.

Example 4.21 Let $f : [0, \infty) \to \mathbb{R}$ be the function $f(x) = \sqrt{x}$. Construct the third Taylor approximant for f about 4, and hence find an approximation to $f(5) = \sqrt{5}$. Use Taylor's Theorem to find upper and lower bounds on $\sqrt{5}$.

Solution The first four derivatives of f are

$$\begin{aligned} f'(x) &= \frac{x^{-1/2}}{2}, \\ f''(x) &= -\frac{x^{-3/2}}{4}, \\ f'''(x) &= \frac{3x^{-5/2}}{8}, \\ f^{(4)}(x) &= -\frac{15x^{-7/2}}{16}, \end{aligned}$$

so $f(4) = 2$, $f'(4) = \frac{1}{4}$, $f''(4) = -2^{-5}$, and $f'''(4) = 3/2^8$. It follows that the third Taylor approximant for f about 4 is

$$\begin{aligned} p_3(x) &= f(4) + f'(4)(x-4) + \frac{f''(4)}{2!}(x-4)^2 + \frac{f'''(4)}{3!}(x-4)^3 \\ &= 2 + \frac{1}{4}(x-4) - \frac{1}{64}(x-4)^2 + \frac{1}{512}(x-4)^3. \end{aligned}$$

Using the approximation $f(x) \approx p_3(x)$ gives

$$\sqrt{5} = f(5) \approx p_3(5) = 2 + \frac{1}{4} - \frac{1}{64} + \frac{1}{512} = 2.236328125.$$

By Taylor's Theorem, the true value of $f(5)$ is

$$f(5) = p_3(5) + \frac{f^{(4)}(c)}{4!}(5-4)^4 = p_3(5) - \frac{5}{2^7 c^{7/2}}$$

for some $c \in (4, 5)$. It follows that the true value of $\sqrt{5}$ is strictly less than $p_3(5)$, but strictly greater than $p_3(5) - 5/(2^7 4^{7/2})$. That is, we have established that

$$2.236022949 < \sqrt{5} < 2.236328125.$$

For comparison, my pocket calculator tells me that $\sqrt{5} \approx 2.236067977$. $\square$

Summary

- The **Interior Extremum Theorem**: If $f : (a, b) \to \mathbb{R}$ is differentiable, and has an extremum (maximum or minimum) at $c \in (a, b)$, then $f'(c) = 0$.
- The **Mean Value Theorem**: If $f : [a, b] \to \mathbb{R}$ is continuous, and differentiable on (a, b), then there exists $c \in (a, b)$ such that
$$f'(c) = \frac{f(b) - f(a)}{b - a}.$$
- Let $f : I \to \mathbb{R}$ be differentiable, where $I \subseteq \mathbb{R}$ is an interval. We can use the Mean Value Theorem to prove that
  - If $f'(x) = 0$ for all $x \in I$ then f is constant.
  - If $f'(x) > 0$ for all $x \in I$ then f is strictly increasing.
  - If $f'(x) < 0$ for all $x \in I$ then f is strictly decreasing.

  We often use the first of these when we solve ordinary differential equations.
- The **Extended Mean Value Theorem**: If $f : [a, b] \to \mathbb{R}$ and $g : [a, b] \to \mathbb{R}$ are continuous, and differentiable on (a, b), and $g'(x) \neq 0$ for all $x \in (a, b)$, then there exists $c \in (a, b)$ such that
$$\frac{f'(c)}{g'(c)} = \frac{f(b) - f(a)}{g(b) - g(a)}.$$
- We can use the Extended Mean Value Theorem to prove Taylor's Theorem.
- **Taylor's Theorem**: If f is $(n + 1)$ times differentiable on some open interval I and $a, x \in I$, then there exists c between a and x such that
$$f(x) = f(a) + \frac{f'(a)}{1!}(x - a) + \cdots + \frac{f^{(n)}(a)}{n!}(x - a)^n + \frac{f^{(n+1)}(c)}{n!}(x - a)^{n+1}.$$
- Taylor's Theorem allows us to approximate $(n+1)$ times differentiable functions by a polynomial of degree n, with an error which is controlled by $f^{(n+1)}$.

Chapter 5

Integration

5.1 Suprema and infima

In this section we will develop a mathematcally rigorous theory of integration. This will make frequent use of the idea of the *supremum* and *infimum* of a bounded set $A \subset \mathbb{R}$. Recall that A is **bounded above** if there exists $K \in \mathbb{R}$ such that, for all $x \in A$, $x \leq K$. Any such K is called an **upper bound** on A. “Bounded below” and “lower bound” are defined analogously: $A \subset \mathbb{R}$ is **bounded below** if there exists $L \in \mathbb{R}$ such that, for all $x \in A$, $x \geq L$, and any such L is called a **lower bound** on A. A set is **bounded** if it is bounded above and below.

Definition 5.1 Let A be a nonempty subset of $\mathbb{R}$. The **supremum** of A, denoted $\sup A$, is the least upper bound on A, if this exists. That is, $\sup A$ is a real number with two properties:

(i) $\sup A$ is an upper bound on A (i.e. for all $x \in A$, $x \leq \sup A$), and

(ii) no number less than $\sup A$ is an upper bound on A.

Similarly, the **infimum** of A, denoted $\inf A$, is the greatest lower bound on A, if this exists. That is, $\inf A$ is a real number with two properties:

(i) $\inf A$ is a lower bound on A (i.e. for all $x \in A$, $x \geq \inf A$), and

(ii) no number greater than $\inf A$ is a lower bound on A.

One of the defining properties of $\mathbb{R}$ is:

The Axiom of Completeness Every nonempty subset of $\mathbb{R}$ which is bounded above has a supremum in $\mathbb{R}$.

Example 5.2 Let $A = (-\infty, 1)$. This is nonempty and bounded above, by 1, for example. Hence, by the Axiom of Completeness, it must have a supremum. In fact, $\sup A = 1$.

Proof: We've already observed that 1 is an upper bound on A. Let $K < 1$. Then $x = (K+1)/2 < (1+1)/2 = 1$ so $x \in A$. But $x > (K+K)/2 = K$, so K is not an upper bound on A. Hence 1 is the *least* upper bound on A. $\square$

Exercise 5.3 Let $B = \{1/n \,:\, n \in \mathbb{Z}^+\}$. Find $\sup B$ and $\inf B$. Rigorously justify your answers (i.e. *prove* they satisfy Definition 5.1).

Warning! Do not confuse "supremum" with "maximum" or "infimum" with "minimum". Note that, in the Example 5.2, $A = (-\infty, 1)$ does *not* have a maximum element! It does have a supremum, however, $1 \notin A$. This set is unbounded below, so has no infimum. In Example 5.3, B has an infimum, but does *not* have a minimum element: given any $1/n \in B$ there is always some other element $1/m \in B$ which is smaller (e.g. $1/m = 1/(n+1)$). Do we know that every (nonempty) set which is bounded below has an infimum (greatest lower bound)? Yes!

Theorem 5.4 *Let $A \subset \mathbb{R}$ be nonempty and bounded below. Then A has an infimum in $\mathbb{R}$.*

Proof: Let B be the set of lower bounds on A. Then B is nonempty (since A is bounded below) and is bounded above (by any element of A – such an element exists since A is nonempty). Hence B has a supremum, L say. I claim that $L = \inf A$. First note that L is an upper bound on B, so if $K > L$ it isn't in B, so is *not* a lower bound on A. Second, note that every element x of A is an upper bound on B, and L is the *least* upper bound on B, so $L \leq x$. Hence L is a lower bound on A. $\Box$

Although "supremum" and "maximum" are not the same thing, they behave similarly in many respects. Here's an example. If sets A and B both have maxima, and A is a subset of B, then clearly $\max A \leq \max B$. The same holds for suprema. If B is bounded above and $A \subseteq B$, then any upper bound on B is also an upper bound on A. In particular, $\sup B$ is an upper bound on A. Since $\sup A$ is the *least* upper bound on A, it follows that $\sup A \leq \sup B$. That is

$$A \subseteq B \qquad \Rightarrow \qquad \sup A \leq \sup B.$$

By similar reasoning,

$$A \subseteq B \qquad \Rightarrow \qquad \inf A \geq \inf B.$$

5.2 Dissections and Riemann sums

How do we define the area of a bounded subset of the plane $\mathbb{R}^2$? If the subset is a rectangle, the answer is easy: its area is its length times its width. Similarly, if the subset is a union of non-overlapping rectangles it's easy: we just add up the areas of all the constituent rectangles. But what if the subset is more complicated: the region bounded by the x-axis, the vertical lines $x = a$ and $x = b > a$ and the graph $y = f(x)$ of some non-constant function, for example? One approach is to define the area of such a region to be the unique real number (if it exists) which is no bigger than the total area of any collection of rectangles which covers the region, and no smaller than the total area of any collection of rectangles which is covered by the region. This is the underlying idea that leads to the *Riemann integral*.

We begin by identifying the collections of rectangles we will use. These are determined by *dissecting* the interval $[a, b]$ into a finite collection of subintervals.

Definition 5.5 A **dissection** of a closed bounded interval $[a, b]$ is a finite subset $\mathscr{D}$ of $[a, b]$ containing both a and b. By convention, if $\mathscr{D}$ has $n + 1$ elements, we label these $a_0, a_1, \ldots, a_n$, so that

$$a = a_0 < a_1 < a_2 < \cdots < a_n = b,$$

and say that $\mathscr{D}$ is a dissection of **size** n. We say that $\mathscr{D}$ is a **regular** dissection if $a_j - a_{j-1} = (b - a)/n$ for all j, that is, if the points in the dissection are regularly spaced.

Remark The size of a dissection is one *less* than the number of elements it contains. This is the number of *subintervals* $[a_{j-1}, a_j]$ into which the dissection divides the interval $[a, b]$.

Definition 5.6 Let $f : [a, b] \to \mathbb{R}$ be a bounded function and $\mathscr{D}$ be a dissection of size n of $[a, b]$. For each $j \in \{1, 2, \ldots, n\}$, let

$$\begin{aligned} m_j &= \inf\{f(x) \,:\, a_{j-1} \le x \le a_j\}, \\ \text{and} \quad M_j &= \sup\{f(x) \,:\, a_{j-1} \le x \le a_j\}. \end{aligned}$$

Note that these numbers exist, since f is bounded. The **lower Riemann sum** of f with respect to $\mathscr{D}$ is

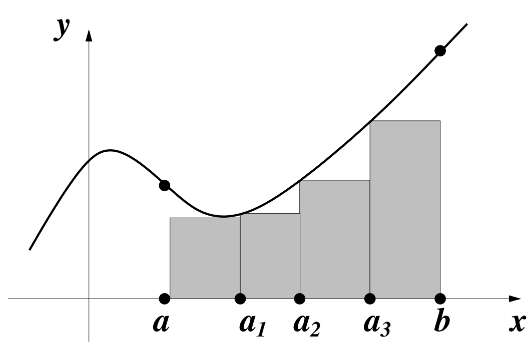
$$l_{\mathscr{D}}(f) = \sum_{j=1}^{n} m_j (a_j - a_{j-1}),$$

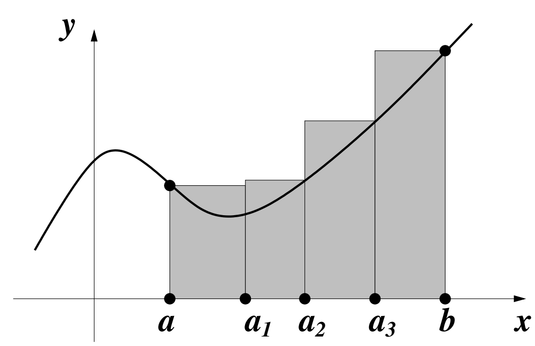
and the **upper Riemann sum** of f with respect to $\mathscr{D}$ is

$$u_{\mathscr{D}}(f) = \sum_{j=1}^{n} M_j (a_j - a_{j-1}).$$

$\square$

The idea is that a dissection of size n divides $[a, b]$ into n subintervals, $[a_{j-1}, a_j]$ for $j = 1, 2, \ldots, n$. If $f(x) \ge 0$ for all x, then the lower Riemann sum can be visualized as the total area of the tallest rectangles with bases $[a_{j-1}, a_j]$ which fit under the graph $y = f(x)$ between $x = a$ and $x = b$. So $l_{\mathscr{D}}(f)$ is an underestimate of the area under the graph. Similarly, the upper Riemann sum can be visualized as the total area of the shortest rectangles with bases $[a_{j-1}, a_j]$ which the graph $y = f(x)$ between $x = a$ and $x = b$ fits under and is thus an overestimate of the area under the graph. This is illustrated below





Example 5.7 $\mathscr{D}_1 = \{0, 1\}$, $\mathscr{D}_2 = \{0, \frac{1}{2}, 1\}$, and $\mathscr{D}_3 = \{0, \frac{1}{2}, \frac{3}{4}, 1\}$ are dissections of $[0, 1]$. The function $f : [0, 1] \to \mathbb{R}$, $f(x) = x^2$ is bounded, and its lower and upper Riemann sums with respect to these dissections are:

$$\begin{aligned}
l_{\mathscr{D}_1}(f) &= 0(1-0) = 0 \\
u_{\mathscr{D}_1}(f) &= 1(1-0) = 1 \\
l_{\mathscr{D}_2}(f) &= 0(\frac{1}{2}-0) + \frac{1}{4}(1-\frac{1}{2}) = \frac{1}{8} \\
u_{\mathscr{D}_2}(f) &= \frac{1}{4}(\frac{1}{2}-0) + 1(1-\frac{1}{2}) = \frac{5}{8} \\
l_{\mathscr{D}_3}(f) &= 0(\frac{1}{2}-0) + \frac{1}{4}(\frac{3}{4}-\frac{1}{2}) + \frac{9}{16}(1-\frac{3}{4}) = \frac{13}{64} \\
u_{\mathscr{D}_3}(f) &= \frac{1}{4}(\frac{1}{2}-0) + \frac{9}{16}(\frac{3}{4}-\frac{1}{2}) + 1(1-\frac{3}{4}) = \frac{33}{64}.
\end{aligned}$$

Note that

$$l_{\mathscr{D}_1} \le l_{\mathscr{D}_2} \le l_{\mathscr{D}_3} \le u_{\mathscr{D}_3} \le u_{\mathscr{D}_2} \le u_{\mathscr{D}_1}.$$

We will see shortly that this is no accident.

Proposition 5.8 *Let $f : [a, b] \to \mathbb{R}$ be bounded above by M and below by m, and let $\mathscr{D}$ be any dissection of $[a, b]$. Then*

$$m(b-a) \le l_{\mathscr{D}}(f) \le u_{\mathscr{D}}(f) \le M(b-a).$$

Proof: Let $\mathscr{D} = \{a_0, a_1, \ldots, a_n\}$ and m_j, M_j be defined as in Definition 5.6. Then, for all j, $m \le m_j \le M_j \le M$, so

$$l_{\mathscr{D}}(f) = \sum_{j=1}^{n} m_j(a_j - a_{j-1}) \le \sum_{j=1}^{n} M_j(a_j - a_{j-1}) = u_{\mathscr{D}}(f)$$

and

$$\begin{aligned}
l_{\mathscr{D}}(f) &\ge \sum_{j=1}^{n} m(a_j - a_{j-1}) \\
&= m\left\{(a_1 - a_0) + (a_2 - a_1) + (a_3 - a_2) + \cdots + (a_n - a_{n-1})\right\} = m(a_n - a_0) \\
&= m(b-a)
\end{aligned}$$

and

$$\begin{aligned}
u_{\mathscr{D}}(f) &\le \sum_{j=1}^{n} M(a_j - a_{j-1}) \\
&= M\left\{(a_1 - a_0) + (a_2 - a_1) + (a_3 - a_2) + \cdots + (a_n - a_{n-1})\right\} = M(a_n - a_0) \\
&= M(b-a)
\end{aligned}$$

since the sums "telescope." $\square$

5.3 Definition of the Riemann integral

It follows from Proposition 5.8 that the set of all lower Riemann sums

$$\{l_{\mathscr{D}}(f) \,:\, \mathscr{D} \text{ any dissection of } [a,b]\}$$

is bounded above, by $M(b-a)$, and the set of all upper Riemann sums

$$\{u_{\mathscr{D}}(f) \,:\, \mathscr{D} \text{ any dissection of } [a,b]\}$$

is bounded below, by $m(b-a)$, where

$$m = \inf\{f(x) \,:\, x \in [a,b]\}, \qquad M = \sup\{f(x) \,:\, x \in [a,b]\}.$$

Hence, by the Axiom of Completeness and Theorem 5.4, the following definition makes sense.

Definition 5.9 Let $f : [a,b] \to \mathbb{R}$ be a bounded function. Then its **lower Riemann integral** is

$$l(f) = \sup\{l_{\mathscr{D}}(f) \,:\, \mathscr{D} \text{ any dissection of } [a,b]\},$$

and its **upper Riemann integral** is

$$u(f) = \inf\{u_{\mathscr{D}}(f) \,:\, \mathscr{D} \text{ any dissection of } [a,b]\}.$$

We say that f is **Riemann integrable** (on $[a,b]$) if $l(f) = u(f)$. In that case, we denote this commmon value by

$$\int_a^b f \quad \text{or} \quad \int_a^b f(x)dx,$$

and call this number the **Riemann integral** of f (over $[a,b]$). $\square$

Geometrically, we can think of $l(f)$ as the least upper bound on the collection of all underestimates of the area under the curve $y = f(x)$, and $u(f)$ as the greatest lower bound on the collection of overestimates of the area under the curve. Then $\int_a^b f$ is, loosely, the unique number (where this exists) which is smaller than every overestimate and larger than every underestimate. As it stands, to compute this number, or even show that it exists, we need to consider the collection of all possible dissections of $[a,b]$. This is a very large and complicated set, so, before proceeding further, we need to develop some tools for handling $l(f)$ and $u(f)$ which allow us to avoid considering all possible dissections. The key idea is that of *refinement* of a dissection.

Definition 5.10 Given dissections $\mathscr{D}, \mathscr{D}'$ of $[a,b]$, we say that $\mathscr{D}'$ is a **refinement** of $\mathscr{D}$ if $\mathscr{D} \subseteq \mathscr{D}'$. If $\mathscr{D}' \backslash \mathscr{D}$ contains k points, we say that $\mathscr{D}'$ is a **k-point refinement** of $\mathscr{D}$. Note that $\mathscr{D}$ is the unique 0-point refinement of itself.

The idea is that a refinement $\mathscr{D}'$ of $\mathscr{D}$ contains all the points in $\mathscr{D}$ and (unless $\mathscr{D}' = \mathscr{D}$) some more points too, so it splits $[a,b]$ up into more subintervals, at least some of which are narrower. Intuitively, one expects that passing from $\mathscr{D}$ to a refinement of $\mathscr{D}$ can only improve, that is, increase, the underestimate $l_{\mathscr{D}}(f)$. Similarly, passing to a refinement of $\mathscr{D}$, one expects, can only reduce the overestimate $u_{\mathscr{D}}(f)$. This expectation turns out to be essentially correct and is fundamental.

Lemma 5.11 (Refinement Lemma) *Let $f : [a,b] \to \mathbb{R}$ be bounded, $\mathscr{D}, \mathscr{D}'$ be dissections of $[a,b]$, and $\mathscr{D}'$ be a refinement of $\mathscr{D}$. Then*

$$l_{\mathscr{D}}(f) \leq l_{\mathscr{D}'}(f) \leq u_{\mathscr{D}'}(f) \leq u_{\mathscr{D}}(f).$$

Proof: We first prove this in the special case where $\mathscr{D}'$ is a 1-point refinement of $\mathscr{D}$. So, let $\mathscr{D} = \{a = a_0, a_1, \ldots, a_n = b\}$ and $\mathscr{D}' = \mathscr{D} \cup \{z\}$ where $z \notin \mathscr{D}$. Then there exists $k \in \{1, \ldots, n\}$ such that $z \in (a_{k-1}, a_k)$. Let

$$\begin{aligned}
m_j &= \inf\{f(x) \,:\, a_j \leq x \leq a_{j-1}\} \\
m' &= \inf\{f(x) \,:\, a_{k-1} \leq x \leq z\} \\
m'' &= \inf\{f(x) \,:\, z \leq x \leq a_k\} \\
M_j &= \sup\{f(x) \,:\, a_j \leq x \leq a_{j-1}\} \\
M' &= \sup\{f(x) \,:\, a_{k-1} \leq x \leq z\} \\
M'' &= \sup\{f(x) \,:\, z \leq x \leq a_k\}
\end{aligned}$$

and note that, since $[a_{k-1}, z]$ and $[z, a_k]$ are subsets of $[a_{k-1}, a_k]$, we know immediately that $m', m'' \geq m_k$, and $M', M'' \leq M_k$. Now

$$\begin{aligned}
l_{\mathscr{D}'}(f) &= \sum_{j \in \{1,2,\ldots,n\} \setminus \{k\}} m_j(a_j - a_{j-1}) + m'(z - a_{k-1}) + m''(a_k - z) \\
&= l_{\mathscr{D}}(f) - m_k(a_k - a_{k-1}) + m'(z - a_{k-1}) + m''(a_k - z) \\
&\geq l_{\mathscr{D}}(f) - m_k(a_k - a_{k-1}) + m_k(z - a_{k-1}) + m_k(a_k - z) \\
&= l_{\mathscr{D}}(f),
\end{aligned}$$

and

$$\begin{aligned}
u_{\mathscr{D}'}(f) &= \sum_{j \in \{1,2,\ldots,n\} \setminus \{k\}} M_j(a_j - a_{j-1}) + M'(z - a_{k-1}) + M''(a_k - z) \\
&= u_{\mathscr{D}}(f) - M_k(a_k - a_{k-1}) + M'(z - a_{k-1}) + M''(a_k - z) \\
&\leq u_{\mathscr{D}}(f) - M_k(a_k - a_{k-1}) + M_k(z - a_{k-1}) + M_k(a_k - z) \\
&= u_{\mathscr{D}}(f).
\end{aligned}$$

Hence, by Proposition 5.8,

$$l_{\mathscr{D}}(f) \leq l_{\mathscr{D}'}(f) \leq u_{\mathscr{D}'}(f) \leq u_{\mathscr{D}}(f),$$

that is, the Refinement Lemma holds for every 1-point refinement of every dissection.

Consider now the case where $\mathscr{D}' = \mathscr{D} \cup \{z_1, z_2, \ldots, z_p\}$ is a p-point refinement of $\mathscr{D}$. Then we can define a chain of dissections $\mathscr{D}_i$, $i = 0, \ldots, p$, by $\mathscr{D}_0 = \mathscr{D}$ and

$\mathscr{D}_i = \mathscr{D}_{i-1} \cup \{z_i\}$ for each $i = 1, \ldots, p$. Note that $\mathscr{D}_p = \mathscr{D}'$, and that each $\mathscr{D}_i$ is a 1-point refinement of $\mathscr{D}_{i-1}$. Hence, by the result just proved,

$$l_{\mathscr{D}}(f) = l_{\mathscr{D}_0}(f) \leq l_{\mathscr{D}_1}(f) \leq l_{\mathscr{D}_{2}(f)} \leq \cdots \leq l_{\mathscr{D}_p}(f) = l_{\mathscr{D}'}(f)$$

and

$$u_{\mathscr{D}}(f) = u_{\mathscr{D}_0}(f) \geq u_{\mathscr{D}_1}(f) \geq u_{\mathscr{D}_{2}(f)} \geq \cdots \geq u_{\mathscr{D}_p}(f) = u_{\mathscr{D}'}(f).$$

By Proposition 5.8, $l_{\mathscr{D}'}(f) \leq u_{\mathscr{D}'}(f)$, and so

$$l_{\mathscr{D}}(f) \leq l_{\mathscr{D}'}(f) \leq u_{\mathscr{D}'}(f) \leq u_{\mathscr{D}}(f),$$

as was to be proved. □

It follows immediately from the Refinement Lemma that *every* upper Riemann sum is at least as large as *every* lower Riemann sum, whatever (possibly different) dissections we use to compute them:

Lemma 5.12 *Let $\mathscr{D}, \widehat{\mathscr{D}}$ be two dissections of $[a, b]$ and $f : [a, b] \to \mathbb{R}$ be bounded. Then $l_{\mathscr{D}}(f) \leq u_{\widehat{\mathscr{D}}}(f)$.*

Proof: $\mathscr{D}' = \mathscr{D} \cup \widehat{\mathscr{D}}$ is a refinement of both $\mathscr{D}$ and $\widehat{\mathscr{D}}$, so by the Refinement Lemma,

$$l_{\mathscr{D}}(f) \leq l_{\mathscr{D}'}(f) \leq u_{\mathscr{D}'}(f) \leq u_{\widehat{\mathscr{D}}}(f).$$

□

From this, it follows that the upper Riemann integral is no less than the lower Riemann integral.

Lemma 5.13 *Let $f : [a, b] \to \mathbb{R}$ be bounded. Then $l(f) \leq u(f)$.*

Proof: Assume, towards a contradiction, that $l(f) > u(f)$. Then $l(f)$ is not a lower bound on the set of upper Riemann sums of f (since $u(f)$ is, by definition, the *greatest* lower bound on this set). Hence, there exists a dissection $\mathscr{D}$ such that $u_{\mathscr{D}}(f) < l(f)$. Hence, $u_{\mathscr{D}}(f)$ is not an upper bound on the set of lower Riemann sums of f (since $l(f)$ is, by definition, the *least* upper bound on this set), so there exists $\mathscr{D}'$ such that $l_{\mathscr{D}'}(f) > u_{\mathscr{D}}(f)$. But this contradicts Lemma 5.12. □

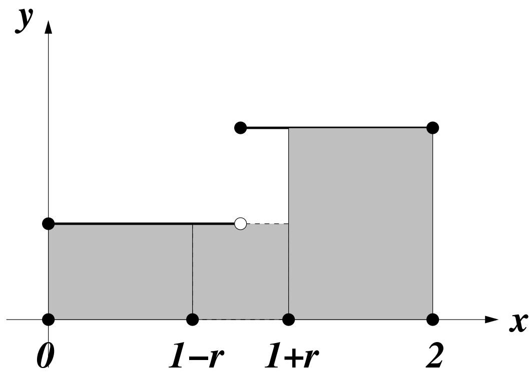
We are now, finally, in a position to compute some nontrivial Riemann integrals.

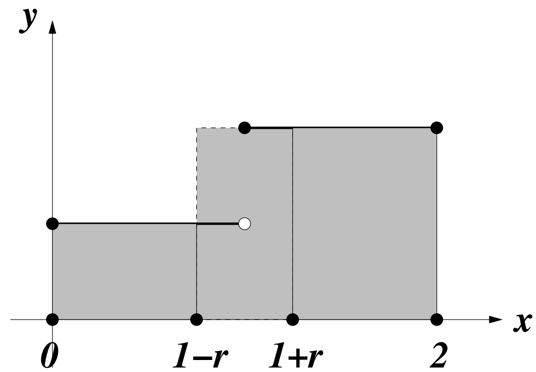
Example 5.14 Let $f : [0, 2] \to \mathbb{R}$ be the "step" function

$$f(x) = \begin{cases} 1 & \text{if } 0 \leq x < 1, \\ 2 & \text{if } 1 \leq x \leq 2. \end{cases}$$

Show that f is Riemann integrable on $[0, 2]$ and compute $\int_0^2 f$.

Solution For each $r \in (0, 1)$ define the dissection $\mathscr{D}_r = \{0, 1 - r, 1 + r, 2\}$ of $\mathscr{D}_r$





Then the lower and upper Riemann sums of f with respect to this dissection are

$$\begin{aligned} l_{\mathscr{D}_r}(f) &= 1(1-r-0)+1((1+r)-(1-r))+2(2-(1+r)) = 3-r \\ u_{\mathscr{D}_r}(f) &= 1(1-r-0)+2((1+r)-(1-r))+2(2-(1+r)) = 3+r. \end{aligned}$$

Hence, $l(f)$ is the supremum of a set (the set of all lower sums) which contains $\{3-r\ :\ 0<r<1\} = (2,3)$, so $l(f) \geq 3$. Similarly, $u(f)$ is the infimum of a set (the set of all upper sums) which contains $\{3+r\ :\ 0<r<1\} = (3,4)$, so $u(f) \leq 3$. Hence $l(f) \geq u(f)$. But by Lemma 5.13, $l(f) \leq u(f)$, so $l(f) = u(f)$, that is, f is Riemann integrable. Furthermore

$$\int_0^2 f = l(f) \geq 3 \quad \text{and} \quad \int_0^2 f = u(f) \leq 3$$

so we conclude that $\displaystyle\int_0^2 f = 3.$ □

Example 5.15 Let $f : [0,1] \to \mathbb{R}$ be the function $f(x) = x^2$. Show that f is Riemann integrable on $[0,1]$ and compute $\int_0^1 f$.

Solution For each integer $n \geq 1$ let $\mathscr{D}_n$ be the *regular* dissection of $[0,1]$ of size n, that is, the dissection that divides $[0,1]$ into n subintervals of equal width,

$$\mathscr{D}_n = \left\{0, \frac{1}{n}, \frac{2}{n}, \ldots, \frac{n-1}{n}, 1\right\}.$$

Each subinterval $[a_{j-1}, a_j] = [(j-1)/n, j/n]$ has width $1/n$, and, since the function f is increasing,

$$\begin{aligned} m_j &= \inf\left\{f(x)\ :\ \frac{j-1}{n} \leq x \leq \frac{j}{n}\right\} \\ &= f\left(\frac{j-1}{n}\right) = \frac{(j-1)^2}{n^2}, \\ M_j &= \sup\left\{f(x)\ :\ \frac{j-1}{n} \leq x \leq \frac{j}{n}\right\} \\ &= f\left(\frac{j}{n}\right) = \frac{j^2}{n^2}, \end{aligned}$$

so the lower and upper Riemann sums with respect to $\mathscr{D}_n$ are

$$\begin{aligned} l_{\mathscr{D}_n}(f) &= \frac{1}{n}\sum_{j=1}^{n}\frac{(j-1)^2}{n^2} = \frac{1}{n^3}\sum_{k=1}^{n-1}k^2, \\ u_{\mathscr{D}_n}(f) &= \frac{1}{n}\sum_{j=1}^{n}\frac{j^2}{n^2} = \frac{1}{n^3}\sum_{j=1}^{n}j^2. \end{aligned}$$

I claim that, for all $n \in \mathbb{Z}^+$,

$$\sum_{j=1}^{n} j^2 = \frac{1}{6}n(n+1)(2n+1),$$

and leave the proof of this as an exercise (hint: use induction!). Hence, for each $n \in \mathbb{Z}^+$,

$$\begin{aligned} l_{\mathscr{D}_n}(f) &= \frac{1}{6}\left(1-\frac{1}{n}\right)\left(2-\frac{1}{n}\right), \\ u_{\mathscr{D}_n}(f) &= \frac{1}{6}\left(1+\frac{1}{n}\right)\left(2+\frac{1}{n}\right). \end{aligned}$$

By definition, $l(f) \geq l_{\mathscr{D}_n}(f)$ for all n (it is an upper bound on a set containing every $l_{\mathscr{D}_n}(f)$), and $l_{\mathscr{D}_n}(f) \to \frac{1}{3}$. Hence $l(f) \geq \frac{1}{3}$ (Theorem 1.7).

Similarly, $u(f) \leq u_{\mathscr{D}_n}(f)$ for all n (it is a lower bound on a set containing every $u_{\mathscr{D}_n}(f)$), and $u_{\mathscr{D}_n}(f) \to \frac{1}{3}$. Hence $u(f) \leq \frac{1}{3}$ (Proposition 1.7).

Hence $u(f) \leq \frac{1}{3} \leq l(f)$. But, by Lemma 5.13, $l(f) \leq u(f)$, so we conclude that $l(f) = u(f)$, that is, f is Riemann integrable. Furthermore,

$$\int_0^1 f = l(f) \geq \frac{1}{3} \quad \text{and} \quad \int_0^1 f = u(f) \leq \frac{1}{3},$$

whence $\int_0^1 f = \frac{1}{3}$. $\square$

It is convenient to extend the definition of dissection to include the case where the interval is $[a,a] = \{a\}$. The only dissection of $[a,a]$ is the singleton set $\mathscr{D} = \{a\}$ (a dissection of size 0). Every function $f : [a,a] \to \mathbb{R}$ is bounded, above and below by $f(a)$, so the one and only lower Riemann sum is

$$l_{\{a\}}(f) = f(a)(a-a) = 0,$$

and the one and only upper Riemann sum is

$$u_{\{a\}}(f) = f(a)(a-a) = 0.$$

It follows that every function is integrable on $[a,a]$, and

$$\int_a^a f = 0.$$

Given a Riemann integrable function f on $[a,b]$ where $a \leq b$, it is also convenient to define

$$\int_b^a f = -\int_a^b f.$$

I leave it as an exercise to verify that all the results we prove about $\int_a^b f$ trivially extend to the case $a \geq b$ with these conventions.

5.4 A sequential characterization of integrability

We can generalize the argument used in Example 5.15 to give a rather elegant (and useful) characterization of Riemann integrability in terms of *sequences* of dissections.

Theorem 5.16 *Let $f : [a,b] \to \mathbb{R}$ be a bounded function. Then f is Riemann integrable if and only if there exists a sequence $(\mathscr{D}_n)$ of dissections of $[a,b]$ such that*

$$u_{\mathscr{D}_n}(f) - l_{\mathscr{D}_n}(f) \to 0.$$

In this case,

$$\int_a^b f = \lim l_{\mathscr{D}_n}(f) = \lim u_{\mathscr{D}_n}(f).$$

Proof: Let $\mathcal{D}$ be the set of all dissections of $[a,b]$, $\mathcal{L} = \{l_{\mathscr{D}}(f) : \mathscr{D} \in \mathcal{D}\}$ the set of all lower Riemann sums of f, and $\mathcal{U} = \{u_{\mathscr{D}}(f) : \mathscr{D} \in \mathcal{D}\}$ the set of all upper Riemann sums of f, so $l(f) = \sup \mathcal{L}$ and $u(f) = \inf \mathcal{U}$.

We first prove the "only if" ($\Rightarrow$) direction. So assume that $u(f) = l(f)$ (that is, f is Riemann integrable). For each $n \in \mathbb{Z}^+$, $u(f) + 1/n > u(f)$ the infimum of $\mathcal{U}$, so is *not* a lower bound on $\mathcal{U}$. Hence there exists $\mathscr{D}'_n \in \mathcal{D}$ such that $u_{\mathscr{D}'_n}(f) < u(f) + 1/n$. Similarly, $l(f) - 1/n < l(f) = \sup \mathcal{L}$, so is not an upper bound on $\mathcal{L}$. Hence there exists $\mathscr{D}''_n \in \mathscr{D}$ such that $l_{\mathscr{D}''_n}(f) > l(f) - 1/n$. Let $\mathscr{D}_n = \mathscr{D}'_n \cup \mathscr{D}''_n$ and note this is a refinement of both $\mathscr{D}'_n$ and $\mathscr{D}''_n$. Hence, by the Refinement Lemma,

$$\begin{aligned} u(f) \leq \quad & u_{\mathscr{D}_n}(f) & \leq u_{\mathscr{D}'_n}(f) < u(f) + \frac{1}{n} \\ l(f) - \frac{1}{n} < l_{\mathscr{D}''_n}(f) \leq \quad & l_{\mathscr{D}_n}(f) & \leq l(f). \end{aligned}$$

Hence, by the Squeeze Rule, $u_{\mathscr{D}_n}(f) \to u(f)$ and $l_{\mathscr{D}_n}(f) \to l(f)$. But $u(f) = l(f)$, so $u_{\mathscr{D}_n}(f) - l_{\mathscr{D}_n}(f) \to 0$ by the Algebra of Limits. Furthermore

$$\int_a^b f = u(f) = \lim u_{\mathscr{D}_n}(f) = \lim l_{\mathscr{D}_n}(f).$$

We now prove the "if" ($\Leftarrow$) direction. So assume that a sequence of dissections $\mathscr{D}_n$ exists such that $u_{\mathscr{D}_n}(f) - l_{\mathscr{D}_n}(f) \to 0$. For all n, $u_{\mathscr{D}_n}(f) \geq u(f)$ and $l_{\mathscr{D}_n}(f) \leq l(f)$, so $u_{\mathscr{D}_n}(f) - l_{\mathscr{D}_n}(f) \geq u(f) - l(f)$. Hence, by Proposition 1.7

$$0 = \lim(u_{\mathscr{D}_n}(f) - l_{\mathscr{D}_n}(f)) \geq u(f) - l(f),$$

that is, $l(f) \geq u(f)$. But $l(f) \leq u(f)$ by Proposition 5.12, so $l(f) = u(f)$, that is, f is Riemann integrable. It remains to show that $u(f) = \lim u_{\mathscr{D}_n}(f)$ (from which it follows, by the Algebra of Limits that $l(f) = \lim l_{\mathscr{D}_n}(f)$). Now

$$0 \leq u_{\mathscr{D}_n}(f) - u(f) \leq u_{\mathscr{D}_n}(f) - u(f) + l(f) - l_{\mathscr{D}_n}(f) = u_{\mathscr{D}_n}(f) - l_{\mathscr{D}_n}(f) \to 0$$

so, by the Squeeze Rule, $u_{\mathscr{D}_n}(f) - u(f) \to 0$. Hence $u_{\mathscr{D}_n}(f) \to u(f)$.

□

To illustrate the power of Theorem 5.16, let's use it to prove that *every* monotonic (i.e. increasing or decreasing) function is Riemann integrable.

Theorem 5.17 *Let $f : [a, b] \to \mathbb{R}$ be monotonic (increasing or decreasing). Then f is Riemann integrable.*

Proof: I will prove the theorem in the case where f is increasing, and leave the case where f is decreasing as an exercise(just modify the argument below in the obvious way).

So, assume f is increasing. The f is bounded (below by $f(a)$ and above by $f(b)$), so $l(f)$ and $u(f)$ exist. For each $n \in \mathbb{Z}^+$ let $\mathscr{D}_n$ be the regular dissection of $[a, b]$ of size n, that is, the dissection which divides $[a, b]$ into n subintervals $[a_{j-1}, a_j]$ each of width $(b - a)/n$.

Then, for each $j = 1, \dots, n$,

$$\begin{aligned} m_j &= \inf\{f(x) : a_{j-1} \leq x \leq a_j\} = f(a_{j-1}) \\ M_j &= \sup\{f(x) : a_{j-1} \leq x \leq a_j\} = f(a_j), \end{aligned}$$

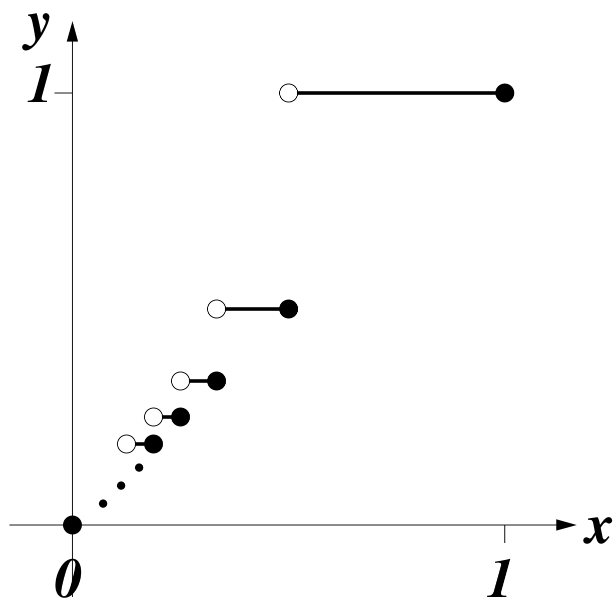
and so

$$\begin{aligned} u_{\mathscr{D}_n}(f) - l_{\mathscr{D}_n}(f) &= \frac{b-a}{n} \sum_{j=1}^{n} (f(a_j) - f(a_{j-1})) \\ &= \frac{b-a}{n}(f(b) - f(a)) \end{aligned}$$

since the sum telescopes. Hence $u_{\mathscr{D}_n}(f) - l_{\mathscr{D}_n}(f) \to 0$, so f is Riemann integrable on $[a, b]$ by Theorem 5.16. □

Note that Theorem 5.17 partially covers both Examples 5.14 and 5.15: in both these examples the function in question is increasing so Theorem 5.17 implies that they're both Riemann integrable. It does not, however, tell us anything about the *value* of $\int_a^b f$, only that it exists. Note also that the function in Example 5.14 is discontinuous, at a single point ($x = 1$), which illustrates that while a discontinuous function cannot be differentiable (Proposition 3.8), it certainly can be integrable. In fact, we can construct examples of functions which are discontinuous at infinitely many points in $[a, b]$ and yet are still Riemann integrable on $[a, b]$.

Example 5.18 Consider the function $f : [0, 1] \to \mathbb{R}$ defined so that, $f(0) = 0$ and, for all $x \in (\frac{1}{n+1}, \frac{1}{n}]$, where $n \in \mathbb{Z}^+$, $f(x) = \frac{1}{n}$. By construction, f is increasing and so is Riemann integrable on $[0,1]$ by Theorem 5.17. Note, however, that f is discontinuous at every point $\frac{1}{n}$ for $n \geq 2$.



If even such an extreme example as this is Riemann integrable, you might wonder whether there are any bounded functions which *fail* to be integrable. Fear not: such functions certainly do exist. Here is an example.

Example 5.19 Let $f : [0,1] \to \mathbb{R}$ such that $f(x) = 0$ if $x \in \mathbb{Q}$ and $f(x) = 1$ if $x \notin \mathbb{Q}$. I claim that f is not Riemann integrable.

Proof: Let $\mathscr{D} = \{a_0, \ldots, a_n\}$ be any dissection of $[0,1]$. Then, every subinterval $[a_{j-1}, a_j]$ contains both rational and irrational members, so $m_j = 0$ and $M_j = 1$ for all j. Hence

$$\begin{aligned} l_{\mathscr{D}}(f) &= \sum_{j=1}^{n} 0(a_j - a_{j-1}) = 0 \\ u_{\mathscr{D}}(f) &= \sum_{j=1}^{n} 1(a_j - a_{j-1}) = a_n - a_0 = 1. \end{aligned}$$

Since this is true for *all* dissections of $[0,1]$, $l(f) = \sup\{0\} = 0$ and $u(f) = \inf\{1\} = 1$, so $l(f) \neq u(f)$, that is, f is not Riemann integrable. □

Theorem 5.17 gives us one interesting class of functions $f : [a,b] \to \mathbb{R}$ that are Riemann integrable: those that are monotonic. Our next theorem gives us another rather more useful class: continuous functions.

Theorem 5.20 *Let $f : [a,b] \to \mathbb{R}$ be continuous. Then f is Riemann integrable.*

Proof: First note that, since f is continuous, it is certainly bounded (by the Extreme Value Theorem), so $l(f)$ and $u(f)$ exist.

Consider the sequence $(\mathscr{D}_n)$ of regular dissections of $[a,b]$ of size 2^n. I claim that $u_{\mathscr{D}_n}(f) - l_{\mathscr{D}_n}(f) \to 0$, so f is Riemann integrable by Theorem 5.16. Note that $\mathscr{D}_n \subset \mathscr{D}_{n+1}$ for all n, so by the Refinement Lemma, $l_{\mathscr{D}_n}(f)$ is an increasing sequence and $u_{\mathscr{D}_n}(f)$ is a decreasing sequence. Hence $l_{\mathscr{D}_n}(f) \to K$ and $u_{\mathscr{D}_n}(f) \to L$ for some numbers K and L by the Monotone Convergence Theorem, so $u_{\mathscr{D}_n}(f) - l_{\mathscr{D}_n}(f) \to K - L \geq 0$ by the Algebra of Limits and Proposition 1.7. We must show that $K - L = 0$.

Assume, towards a contradiction, that $\varepsilon = K - L > 0$. As usual, let

$$\begin{aligned} m_j &= \inf\{f(x) : a_{j-1} \leq x \leq a_j\} \\ \text{and} \quad M_j &= \sup\{f(x) : a_{j-1} \leq x \leq a_j\}. \end{aligned}$$

Then

$$u_{\mathscr{D}_n}(f) - l_{\mathscr{D}_n}(f) = \sum_{j=1}^{2^n} \frac{b-a}{2^n}(M_j - m_j) \geq \varepsilon.$$

This sum consists of 2^n terms, each non-negative, so at least one term must be greater than or equal to $\varepsilon/2^n$. That is, there must exist $j \in \{1, 2, 3, \dots, 2^n\}$ such that

$$M_j - m_j \geq \frac{\varepsilon}{b-a}.$$

But $f : [a_{j-1}, a_j] \to \mathbb{R}$ is continuous so, by the Extreme Value Theorem, f attains both a maximum and minimum value on $[a_{j-1}, a_j]$, that is, there are points, x_n and y_n say, in $[a_{j-1}, a_j]$ such that $f(x_n) = M_j$ and $f(y_n) = m_j$. Hence, for each $n \in \mathbb{Z}^+$ there exist $x_n, y_n \in [a,b]$ such that

$$|x_n - y_n| \leq \frac{b-a}{2^n}, \tag{5.1}$$

since both lie in an interval of this width, and

$$f(x_n) - f(y_n) \geq \frac{\varepsilon}{b-a}. \tag{5.2}$$

Consider the sequence (x_n). It is bounded, so, by the Bolzano–Weierstrass Theorem (Theorem 1.15), it has a convergent subsequence $x_{n_k} \to c \in [a,b]$. By (5.1),

$$x_{n_k} - \frac{b-a}{2^{n_k}} \leq y_{n_k} \leq x_{n_k} + \frac{b-a}{2^{n_k}}$$

so $y_{n_k} \to c$ also, by the Squeeze Rule. Now f is continuous, so $f(x_{n_k}) \to f(c)$ and $f(y_{n_k}) \to f(c)$, and hence $f(x_{n_k}) - f(y_{n_k}) \to f(c) - f(c) = 0$. But this contradicts (5.2) and Proposition 1.7. □

5.5 Elementary properties of the Riemann integral

Theorem 5.16 is a convenient tool for establishing many useful properties of the Riemann integral.

We will first prove that the Riemann integral is *linear*, that is,

$$\int_a^b (\alpha f + \beta g) = \alpha \int_a^b f + \beta \int_a^b g,$$

for any constants α, β and any integrable functions f, g. To do this, we need the following:

Lemma 5.21 *Let $f, g : [a, b] \to \mathbb{R}$ be bounded functions and $\mathscr{D}$ be a dissection of $[a, b]$. Then*

$$l_{\mathscr{D}}(f + g) \geq l_{\mathscr{D}}(f) + l_{\mathscr{D}}(g) \quad \textit{and} \quad u_{\mathscr{D}}(f + g) \leq u_{\mathscr{D}}(f) + u_{\mathscr{D}}(g).$$

Proof: I will prove the first inequality, and leave the second as an exercise.
Let $\mathscr{D} = \{a_0, a_1, \ldots, a_n\}$ and $I_j = [a_{j-1}, a_j]$. Let

$$\begin{aligned} m_j(f) &= \inf\{f(x) : x \in I_j\}, \\ M_j(f) &= \sup\{f(x) : x \in I_j\}, \end{aligned}$$

and $m_j(g), M_j(g), m_j(g + f)$, and $M_j(g + f)$ be defined similarly. Then, for all $x \in I_j$, $f(x) + g(x) \geq m_j(f) + m_j(g)$, so $m_j(f) + m_j(g)$ is certainly a lower bound on $\{f(x) + g(x) : x \in I_j\}$. Since $m_j(f + g)$ is the *greatest* lower bound on this set, it follows that $m_j(f + g) \geq m_j(f) + m_j(g)$. Hence

$$\begin{aligned} l_{\mathscr{D}}(f + g) &= \sum_{j=1}^{n} m_j(f + g)(a_j - a_{j-1}) \\ &\geq \sum_{j=1}^{n} (m_j(f) + m_j(g))(a_j - a_{j-1}) = l_{\mathscr{D}}(f) + l_{\mathscr{D}}(g). \end{aligned}$$

□

Theorem 5.22 (Linearity of the Riemann Integral) *Let f, g be Riemann integrable on $[a, b]$, and $\alpha \in \mathbb{R}$ be a constant. Then*

(i) αf is Riemann integrable on $[a, b]$, and $\int_a^b \alpha f = \alpha \int_a^b f$,

(ii) $f + g$ is Riemann integrable on $[a, b]$, and $\int_a^b (f + g) = \int_a^b f + \int_a^b g$.

Proof: (i) By Theorem 5.16, there is a sequence of dissections $(\mathscr{D}_n)$ such that $l_{\mathscr{D}_n}(f) \to \int_a^b f$ and $u_{\mathscr{D}_n}(f) \to \int_a^b f$. It follows directly from Definition 5.6 that $l_{\mathscr{D}_n}(\alpha f) = \alpha l_{\mathscr{D}_n}(f)$ and $u_{\mathscr{D}_n}(\alpha f) = \alpha u_{\mathscr{D}_n}(f)$ if $\alpha \geq 0$, and $l_{\mathscr{D}_n}(\alpha f) = \alpha u_{\mathscr{D}_n}(f)$ and $u_{\mathscr{D}_n}(\alpha f) = \alpha l_{\mathscr{D}_n}(f)$ if $\alpha < 0$. In either case, $l_{\mathscr{D}_n}(\alpha f) \to \alpha \int_a^b f$ and $u_{\mathscr{D}_n}(\alpha f) \to \alpha \int_a^b f$ by the Algebra of Limits, and the claim follows from Theorem 5.16.
(ii) Let $(\mathscr{D}_n)$ be as defined above. By Theorem 5.16, there is also a sequence of dissections $(\mathscr{D}'_n)$ such that $l_{\mathscr{D}'_n}(g) \to \int_a^b g$ and $u_{\mathscr{D}'_n}(g) \to \int_a^b g$. Let $\mathscr{D}''_n = \mathscr{D}_n \cup \mathscr{D}'_n$.

This is a refinement of both $\mathscr{D}_n$ and $\mathscr{D}'_n$ and so, by the Refinement Lemma

$$\begin{aligned}
&l_{\mathscr{D}_n}(f) \leq l_{\mathscr{D}''_n}(f) \leq \int_a^b f, \\
&l_{\mathscr{D}'_n}(g) \leq l_{\mathscr{D}''_n}(g) \leq \int_a^b g, \\
&\int_a^b f \leq u_{\mathscr{D}''_n}(f) \leq u_{\mathscr{D}_n}(f), \\
&\int_a^b f \leq u_{\mathscr{D}''_n}(g) \leq u_{\mathscr{D}'_n}(g).
\end{aligned}$$

Hence, by the Squeeze Rule, $l_{\mathscr{D}''_n}(f) \to \int_a^b f$, $l_{\mathscr{D}''_n}(g) \to \int_a^b g$, $u_{\mathscr{D}''_n}(f) \to \int_a^b f$ and $u_{\mathscr{D}''_n}(f) \to \int_a^b g$. Now, by Lemma 5.21,

$$l_{\mathscr{D}''_n}(f) + l_{\mathscr{D}''_n}(g) \leq l_{\mathscr{D}''_n}(f+g) \leq l(f+g),$$

and

$$u(f+g) \leq u_{\mathscr{D}''_n}(f+g) \leq u_{\mathscr{D}''_n}(f) + u_{\mathscr{D}''_n}(g).$$

But both $l_{\mathscr{D}''_n}(f) + l_{\mathscr{D}''_n}(g)$ and $u_{\mathscr{D}''_n}(f) + u_{\mathscr{D}''_n}(g)$ converge to $\int_a^b f + \int_a^b g$ so, by Proposition 1.7, $\int_a^b f + \int_a^b g \leq l(f+g)$ and $u(f+g) \leq \int_a^b f + \int_a^b g$. It follows that $l(f+g) \geq u(f+g)$ and hence, by Lemma 5.13, that $l(f+g) = u(f+g)$, that is, $f+g$ is Riemann integrable. Furthermore

$$\begin{aligned}
&\int_a^b (f+g) = l(f+g) \geq \int_a^b f + \int_a^b g \\
\text{and} \qquad &\int_a^b (f+g) = u(f+g) \leq \int_a^b f + \int_a^b g,
\end{aligned}$$

so $\int_a^b (f+g) = \int_a^b f + \int_a^b g$, as claimed. □

Let's denote by $L([a,b])$ the set of Riemann integrable functions on $[a,b]$. Theorem 5.22 establishes that $L([a,b])$ is in fact a **vector space** over $\mathbb{R}$: it is closed under the obvious (vector) addition and scalar multiplication operations. Better still, it shows that the map

$$L([a,b]) \to \mathbb{R}, \qquad f \mapsto \int_a^b f$$

is a **linear map** between vector spaces.

We next prove that Riemann integration preserves inequalities.

Proposition 5.23 *Let $f, g : [a,b] \to \mathbb{R}$ be Riemann integrable and assume $f(x) \leq g(x)$ for all $x \in [a,b]$. Then*

$$\int_a^b f \leq \int_a^b g.$$

Proof: By Theorem 5.16 there exist sequences of dissections $\mathscr{D}'_n$, $\mathscr{D}''_n$ such that $l_{\mathscr{D}'_n}(f) \to \int_a^b f$ and $l_{\mathscr{D}''_n} \to \int_a^b g$. Let $\mathscr{D}_n = \mathscr{D}'_n \cup \mathscr{D}''_n$. Then, by the Refinement Lemma (and the Squeeze Rule) $l_{\mathscr{D}_n}(f) \to \int_a^b f$ and $l_{\mathscr{D}_n} \to \int_a^b g$ also. But, for each n, $l_{\mathscr{D}_n}(g) \geq l_{\mathscr{D}_n}(f)$ since $\inf g \geq \inf f$ on any subset of $[a, b]$. Hence, $\int_a^b g - \int_a^b f$, being the limit of a convergent non-negative sequence, is ≥ 0 (Proposition 1.7). $\square$

From this we can deduce a kind of "triangle inequality" for Riemann integrals.

Proposition 5.24 (Integral Triangle Inequality) *Let $f : [a, b] \to \mathbb{R}$ be Riemann integrable.*

$$\left| \int_a^b f \right| \leq \int_a^b |f|.$$

Before proving this, we should first establish that $|f|$ is Riemann integrable (so we can be sure the number on the right of the claimed inequality exists).

Lemma 5.25 *Let $f : [a, b] \to \mathbb{R}$ be Riemann integrable. Then $|f| : [a, b] \to \mathbb{R}$ is Riemann integrable.*

Proof: Let $f_+ : [a, b] \to \mathbb{R}$ denote the "positive part" of the function f, that is

$$f_+(x) = \begin{cases} f(x), & \text{if } f(x) \geq 0 \\ 0, \text{if } f(x) < 0. \end{cases}$$

We will show that f_+ is Riemann integrable.

Let $\mathscr{D} = \{a = a_0, a, \ldots, a_n = b\}$ be any dissection of $[a, b]$ and consider, for a particular subinterval $I_j = [a_{j-1}, a_j]$ the numbers $m_j(f), m_j(f_+), M_j(f)$ and $M_j(f_+)$. There are 3 possible cases:

- If $f \geq 0$ on I_j, then $f_+ = f$ (on I_j), so
$$M_j(f_+) - m_j(f_+) = M_j(f) - m_j(f).$$
- If $f < 0$ on I_j then $f_+ = 0$ (on I_j), so
$$M_j(f_+) - m_j(f_+) = 0 - 0 = 0 \leq M_j(f) - m_j(f).$$
- If f takes both negative and non-negative values on I_j, then $m_j(f_+) = 0 > m_j(f)$ while $M_j(f_+) = M_j(f)$, so
$$M_j(f_+) - m_j(f_+) < M_j(f) - m_j(f).$$

In all cases, we see that, for all j,

$$M_j(f_+) - m_j(f_+) \leq M_j(f) - m_j(f).$$

Hence, for any dissection $\mathscr{D}$,

$$\begin{aligned} u_{\mathscr{D}}(f_+) - l_{\mathscr{D}}(f_+) &= \sum_{j=1}^{n} (M_j(f_+) - m_j(f_+))(a_j - a_{j-1}) \\ &\leq \sum_{j=1}^{n} (M_j(f) - m_j(f))(a_j - a_{j-1}) = u_{\mathscr{D}}(f) - l_{\mathscr{D}}(f). \quad (5.3) \end{aligned}$$

Since f is Riemann integrable, there exists, by Theorem 5.16, a sequence $(\mathscr{D}_n)$ of dissections of $[a, b]$ such that

$$u_{\mathscr{D}_n}(f) - l_{\mathscr{D}_n}(f) \to 0.$$

By inequality (5.3) and the Squeeze Rule, $u_{\mathscr{D}_n}(f_+) - l_{\mathscr{D}_n}(f_+) \to 0$ also, so, again by Theorem 5.16, f_+ is also Riemann integrable.

But $|f| = 2f_+ - f$,

$$2f_+(x) - f(x) = \begin{cases} 2f(x) - f(x) = f(x), & \text{if } f(x) \geq 0 \\ 0 - f(x) = -f(x), & \text{if } f(x) < 0, \end{cases}$$

so $|f|$ is also Riemann integrable by Theorem 5.22 . □

Proof of Proposition 5.24: Note that $-|f(x)| \leq f(x) \leq |f(x)|$ for all $x \in [a, b]$, so

$$-\int_a^b |f| \leq \int_a^b f \leq \int_a^b |f|$$

by Proposition 5.23. The claimed inequality immediately follows. □

We conclude this section by proving the "join rule" for integration:

Proposition 5.26 (Join Rule) *Let f be Riemann integrable on $[a, b]$ and on $[b, c]$. Then f is Riemann integrable on $[a, c]$ and*

$$\int_a^c f = \int_a^b f + \int_b^c f.$$

Proof: By Theorem 5.16, there exist sequences of dissections $(\mathscr{D}'_n)$ of $[a, b]$ and $(\mathscr{D}''_n)$ of $[b, c]$ such that $l_{\mathscr{D}'_n}(f) \to \int_a^b f$, $u_{\mathscr{D}'_n}(f) \to \int_a^b f$, $l_{\mathscr{D}''_n}(f) \to \int_b^c f$, and $u_{\mathscr{D}''_n}(f) \to \int_b^c f$. Let $\mathscr{D}_n = \mathscr{D}'_n \cup \mathscr{D}''_n$. Then $\mathscr{D}_n$ is a dissection of $[a, c]$ and it follows directly from Definition 5.6 that $l_{\mathscr{D}_n}(f) = l_{\mathscr{D}'_n}(f) + l_{\mathscr{D}''_n}(f)$ and $u_{\mathscr{D}_n}(f) = u_{\mathscr{D}'_n}(f) + u_{\mathscr{D}''_n}(f)$. Hence, by the Algebra of Limits, $l_{\mathscr{D}_n}(f) \to \int_a^b f + \int_b^c f$ and $u_{\mathscr{D}_n}(f) \to \int_a^b f + \int_b^c f$, so the claim follows from Theorem 5.16. □

Summary

- A **dissection** of $[a, b]$ is a finite subset $\mathscr{D} = \{a_0, a_1, \ldots, a_n\}$ of $[a, b]$ such that
$$a = a_0 < a_1 < a_2 < \cdots < a_n = b.$$
- Given a bounded function $f : [a, b] \to \mathbb{R}$, and a dissection $\mathscr{D}$, the **lower Riemann sum** is
$$l_{\mathscr{D}}(f) = \sum_{j=1}^{n} \inf\{f(x) \ : \ a_{j-1} \leq x \leq a_j\}(a_j - a_{j-1}),$$
and the **upper Riemann sum** is
$$u_{\mathscr{D}}(f) = \sum_{j=1}^{n} \sup\{f(x) \ : \ a_{j-1} \leq x \leq a_j\}(a_j - a_{j-1}).$$
- The **lower Riemann integral** of f is
$$l(f) = \sup\{l_{\mathscr{D}}(f) \ : \ \mathscr{D} \text{ any dissection of } [a, b]\},$$
and the **upper Riemann integral** of f is
$$u(f) = \inf\{u_{\mathscr{D}}(f) \ : \ \mathscr{D} \text{ any dissection of } [a, b]\}.$$
- f is **Riemann integrable** if $l(f) = u(f)$, and in this case we denote their common value by
$$\int_a^b f$$
and call it the **Riemann integral** of f (on, or over, $[a, b]$).
- A bounded function $f : [a, b] \to \mathbb{R}$ is Riemann integrable if and only if there exists some sequence $(\mathscr{D}_n)$ of dissections of $[a, b]$ such that $l_{\mathscr{D}_n}(f) - u_{\mathscr{D}_n}(f) \to 0$ (and, in this case, $\int_a^b f = \lim l_{\mathscr{D}_n}(f) = \lim u_{\mathscr{D}_n}(f)$).
- We used this theorem to prove that all **continuous** functions, and all **monotonic** functions are Riemann integrable. We also used it to prove that
$$\int_a^b f + \int_b^c f = \int_a^c f, \qquad \int_a^b (\alpha f + \beta g) = \alpha \int_a^b f + \beta \int_a^b g, \qquad \text{and} \qquad \left| \int_a^b f \right| \leq \int_a^b |f|$$
where α, β are constants.

Chapter 6

The Fundamental Theorem of the Calculus

So far we have developed some powerful theoretical tools to show that a given function is integrable and to establish relationships between the integrals of related integrable functions, but we don't have any really convenient techniques for actually computing Riemann integrals. To compute

$$\int_0^1 x^2\,dx,$$

for example, we had to resort to exhibiting a sequence of dissections whose upper and lower sums converge to a common limit, in this case, $\frac{1}{3}$ (see Example 5.15). Theorem 5.16 tells us that, in principle, we can always compute Riemann integrals like this, but in practice this method of computing integrals is very onerous and, unless the integrand (the function to be integrated) is fairly simple, is likely to be intractable. Consider, for example, attempting to compute

$$\int_0^\pi \sin x\,dx$$

using this method. We know that this integral exists (Theorem 5.20) because we know that sin is continuous, but we have no hope of computing it directly using Riemann sums. In this chapter, we establish a fundamental connexion between Riemann integration and differentiation which will give us a convenient means of computing $\int_a^b f$ whenever we can dream up a function whose *derivative* equals f. Once we have done this we will rarely have to resort to computing sequences of Riemann sums to compute integrals.

6.1 The first form

It's important to realize that the Riemann integral of a function $f : [a, b] \to \mathbb{R}$ is a single *number* $\int_a^b f$, not a *function*. If $f : \mathbb{R} \to \mathbb{R}$ is continuous, it is Riemann integrable on any interval $[a, b] \subset \mathbb{R}$. If we allow (say) the right endpoint of the interval $[a, b]$ to vary while keeping a fixed, we can think of $\int_a^b f$ as a function of b. It's natural to ask what analytic properties this new function has. The (first form of) the Fundamental Theorem of the Calculus says that $b \mapsto \int_a^b f$ is differentiable, and its derivative is f.

Theorem 6.1 (Fundamental Theorem of the Calculus version 1) *Let $f : I \to \mathbb{R}$ be continuous, where $I \subseteq \mathbb{R}$ is an interval. Choose $a \in I$ and define $F : I \to \mathbb{R}$ by*

$$F(x) = \int_a^x f.$$

Then F is differentiable, and $F' = f$.

Proof: First note that f is continuous on $[a, x] \subseteq I$ (or $[x, a]$ if $x < a$) for all $x \in [a, b]$, so F is well-defined by Theorem 5.20. We wish to compute

$$\lim_{y \to x} \frac{F(y) - F(x)}{y - x}.$$

We will do this using the sequential criterion for limits (Theorem 2.14). So, let (y_n) be any sequence in $I \backslash \{x\}$ such that $y_n \to x$, and

$$s_n = \frac{F(y_n) - F(x)}{y_n - x}.$$

We must show that $s_n \to f(x)$.

For each n, either $y_n > x$ or $y_n < x$. If $y_n > x$ then, by Theorem 5.26,

$$s_n = \frac{1}{y_n - x} \int_x^{y_n} f,$$

whereas if $y_n < x$,

$$s_n = \frac{1}{x - y_n} \int_{y_n}^{x} f.$$

In either case, by the Extreme Value Theorem, there exist w_n and z_n between x and y_n such that $f(w_n)$ is the minimum value of f on the closed interval with endpoints x and y_n, and $f(z_n)$ is the maximum value of f on this interval. Hence, by Proposition 5.8,

$$\begin{array}{lrcll} & \frac{1}{y_n - x} f(w_n)(y_n - x) \leq & s_n & \leq \frac{1}{y_n - x} f(z_n)(y_n - x) & \text{if } y_n > x, \\ \text{and} & \frac{1}{x - y_n} f(w_n)(x - y_n) \leq & s_n & \leq \frac{1}{x - y_n} f(z_n)(x - y_n) & \text{if } y_n < x. \end{array}$$

In either case, we conclude that

$$f(w_n) \leq s_n \leq f(z_n).$$

Now, $y_n \to x$ so $w_n \to x$ and $z_n \to x$ also, by the Squeeze Rule, and f is continuous, so $f(w_n) \to f(x)$ and $f(z_n) \to f(x)$. Hence, by the Squeeze Rule again, $s_n \to f(x)$, which completes the proof. $\square$

This theorem tells us something interesting and far from obvious: any function which is continuous on an interval is the derivative of some differentiable function on that interval. Note that the converse of this is **false**: the derivative of a function may fail to be a continuous function (see Example 4.13 for a counterexample).

Theorem 6.1 says that (provided f is continuous!) $\int_a^b f$ is a differentiable function of the upper integration limit b, and its derivative is f. What if we fix b and think of it as a function of a? Is this function also differentiable? If so, what is *its* derivative?

Corollary 6.2 *Let $f : I \to \mathbb{R}$ be continuous, where $I \subseteq \mathbb{R}$ is an interval. Choose $b \in I$ and define $F : I \to \mathbb{R}$ by*

$$F(x) = \int_x^b f.$$

Then F is differentiable, and $F' = -f$.

Proof: Choose any $a \in I$ and define $G : I \to \mathbb{R}$, $G(x) = \int_a^x f$. Then by Proposition 5.26

$$G(x) + F(x) = \int_a^x f + \int_x^b f = \int_a^b f = C$$

a constant. G is differentiable and $G' = f$ by Theorem 6.1, so $F(x) = C - G(x)$ is differentiable and $F'(x) = 0 - F'(x) = -f(x)$. □

Of course, we can apply all the usual rules of differentiation to functions defined in this way (as integrals of other functions).

Exercise 6.3 Let $f : \mathbb{R} \to \mathbb{R}$,

$$f(x) = \int_x^{x^2+1} \frac{t}{t^2+1} dt.$$

Compute $f'(x)$.

Define $F : \mathbb{R} \to \mathbb{R}$, $G : \mathbb{R} \to \mathbb{R}$, $g : \mathbb{R} \to \mathbb{R}$ and $h : \mathbb{R} \to \mathbb{R}$ such that

$$\begin{aligned} g(x) &= \frac{x}{x^2+1}, \\ F(x) &= \int_0^x g, \\ G(x) &= \int_x^0 g, \\ h(x) &= x^2+1. \end{aligned}$$

Then

$$f(x) = F(h(x)) + g(x).$$

Hence, by the Chain Rule,

$$f'(x) = F'(h(x))h'(x) + G'(x).$$

But, Theorem 6.1 and by Corollary 6.2,

$$\begin{aligned} F' &= g \\ G' &= -g \end{aligned}$$

so

$$\begin{aligned} f'(x) &= g(h(x))h'(x) - g(x) \\ &= \frac{h(x)}{h(x)^2+1}h'(x) - \frac{x}{x^2+1} \\ &= \frac{2x(x^2+1)}{(x^2+1)^2+1} - \frac{x}{x^2+1}. \end{aligned}$$

□

6.2 The second form and a practical method for computing integrals

Theorem 6.1 immediately implies a second theorem, also called the Fundamental Theorem of the Calculus, which renders the job of computing many Riemann integrals almost trivial:

Theorem 6.4 (Fundamental Theorem of the Calculus version 2) *Let $f : [a, b] \to \mathbb{R}$ be continuous and $F : [a, b] \to \mathbb{R}$ be any differentiable function such that $F' = f$. Then*

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

Proof: Define the function $g : [a, b] \to \mathbb{R}$,

$$g(x) = \left(\int_a^x f\right) - F(x).$$

Then, by Theorem 6.1 and the definition of F,

$$g'(x) = f(x) - f(x) = 0$$

for all $x \in [a, b]$. Hence, by Proposition 4.11, g is constant, so $g(b) = g(a)$, that is

$$\left(\int_a^b f\right) - F(b) = \left(\int_a^a f\right) - F(a) = 0 - F(a),$$

and the result immediately follows. □

Example 6.5 Compute the Riemann integrals

(i) $\displaystyle\int_{-1}^{1} x^2\,dx$,

(ii) $\displaystyle\int_{0}^{\pi} \sin x\,dx$,

Solution (i) Let $f(x) = x^2$ and $F(x) = x^3/3$. Then f is continuous and $F' = f$ on $[-1, 1]$ so, by Theorem 6.4,

$$\int_{-1}^{1} f = F(1) - F(-1) = \frac{2}{3}.$$

(ii) Let $f(x) = \sin x$ and $F(x) = -\cos x$. Then f is continuous and $F' = f$ on $[0, \pi]$ so, by Theorem 6.4,

$$\int_{0}^{\pi} f = F(\pi) - F(0) = -\cos\pi + \cos 0 = 2.$$

Having established Theorem 6.4, we see that we can explicitly compute

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

if we can think of an **antiderivative** of f, that is, any function whose *derivative* is $f(x)$. This trick is so pervasive in integral calculus that it leads the unwary to identify *integration* of f (that is, the process of computing a Riemann integral $\int_a^b f$) with the process of writing down an antiderivative of f. Indeed, it is common practice to call an antiderivative of f an *indefinite integral* of f and to denote any such function by the symbol

$$\int f(x)\,dx.$$

This notation has some practical advantages, but it also, unfortunately, generates a huge amount of confusion. Note that

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

is a single number and that its *definition* has nothing whatsoever to do with antiderivatives of f: it is the unique real number which is no greater than any upper Riemann sum of f on $[a, b]$ and no less than any lower Riemann sum of f on $[a, b]$. In particular, it is *not* a function of the "variable" x. Indeed, we could equally well have written it $\int_a^b f(y)\,dy$ or $\int_a^b f(\Gamma_{\widehat{\aleph}})\,d\Gamma_{\widehat{\aleph}}$, which is one reason to prefer the simpler notation $\int_a^b f$. It *is* a function of a and b, however.

By contrast, $\int f(x)\,dx$ is just a (slightly ambiguous) symbol denoting any function whose derivative (at x) is $f(x)$. It *is* a function of x, not a single number, and its *definition* has nothing to do with Riemann sums. That these two things turn out to be closely related is (as its name suggests) a very important theorem: the Fundamental Theorem of the Calculus. To understand calculus properly it is important to maintain a clear conceptual distinction between the Riemann integral $\int_a^b f$ and any antiderivative that one might use to compute it.

6.3 The natural logarithm

Definition 6.6 The **(natural) logarithm** is the function

$$\ln : (0,\infty) \to \mathbb{R}, \qquad \ln x = \int_1^x \frac{1}{t} dt.$$

Remarks

- The function $f(t) = 1/t$ is continuous on $(0,\infty)$, and hence Riemann integrable on $[1,x]$ (if $x \geq 1$) or $[x,1]$ (if $0 < x < 1$), so the function ln is well-defined (Theorem 5.20).
- ln is differentiable, by the Fundamental Theorem of the Calculus version 1 (Theorem 6.1), and
$$\ln'(x) = \frac{1}{x}.$$
- It follows immediately from the definition that
$$\ln 1 = \int_1^1 \frac{1}{t} dt = 0.$$

The logarithm function obeys a very useful identity.

Proposition 6.7 *For all* $x, y \in (0,\infty)$*,* $\ln(xy) = \ln x + \ln y$*.*

Proof: Choose and fix $y \in (0,\infty)$ and consider the function

$$f : (0,\infty) \to \mathbb{R}, \qquad f(x) = \ln(xy) - \ln x - \ln y.$$

By the Chain Rule,

$$f'(x) = \frac{y}{xy} - \frac{1}{x} = 0$$

so f is constant (Proposition 4.11). Hence, for all x, $f(x) = f(1) = \ln y - 0 - \ln y = 0$. □

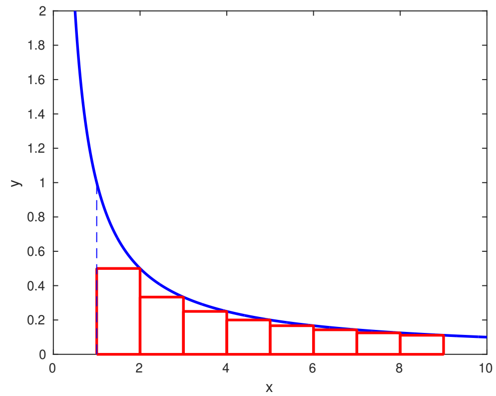
A second useful identity quickly follows from this.

Proposition 6.8 *For all* $x \in (0,\infty)$ *and* $n \in \mathbb{Z}$*,* $\ln x^n = n \ln x$*.*

Proof: Exercise. (*Hint*: proof by induction.) □

Proposition 6.9 *The function* $\ln : (0,\infty) \to \mathbb{R}$ *is smooth, strictly increasing, and bijective.*

Proof: Let $f : (0,\infty) \to \mathbb{R}$, $f(x) = 1/x$. We have noted that ln is differentiable with derivative f. But f is smooth, so ln is smooth. Also, for all $x \in (0,\infty)$, $f(x) > 0$, so ln is strictly increasing, and hence injective (Proposition 4.11). It remains to show that ln is surjective.



For each $n \in \mathbb{Z}^+$, $n \geq 2$, let $\mathscr{D}_n$ be the regular dissection of $[1, n]$ of size $(n-1)$, that is

$$\mathscr{D}_n = \{1, 2, \ldots, n\}.$$

Then

$$\ln n = \int_1^n f \quad \geq \quad l_{\mathscr{D}_n}(f) = \sum_{j=1}^{n-1} (1) f(j+1) = \sum_{k=2}^{n} \frac{1}{k} =: s_n,$$

where we have used the fact that f is monotonically decreasing (the case $n = 8$ is depicted above). The sequence (s_n) is unbounded above (Example 1.22), so the sequence $\ln n$ is also unbounded above. Hence, given any $K \geq 0$, there exists $n \in \mathbb{Z}^+$ such that $\ln n > K$. But $\ln 1 = 0$, and ln is continuous (since it is differentiable) so, by the Intermediate Value Theorem, there exists $x \in [1, n]$ such that $\ln x = K$. Hence, ln takes all non-negative values. Let $L < 0$. As we just showed, there exists $x \in [1, \infty)$ such that $\ln x = -L$. But then, by Proposition 6.8, $\ln(1/x) = -\ln x = L$. So ln also takes all negative values, and we conclude that $\ln : (0, \infty) \to \mathbb{R}$ is surjective, hence bijective. $\square$

It is common to define *natural logarithm* to mean the inverse function to the exponential function. Definition 6.6 has no obvious connexion with the exponential function (which as we will see, is defined using a convergent power series), so this coincidence of terminology will need to be justified.

Summary

- There is an important link between Riemann integrals and derivatives, given by the **Fundamental Theorem of the Calculus**:
  - Version 1: if f is continuous and $F(x) = \int_a^x f$, then $F' = f$.
  - Version 2: if f is continuous and F is some antiderivative of f (that is, a function satisfying $F' = f$), then $\int_a^b f = F(b) - F(a)$.
- Version 2 of the Fundamental Theorem of the Calculus provides a flexible and convenient method for computing many Riemann integrals.
- The **natural logarithm** function is
$$\ln : (0, \infty) \to \mathbb{R}, \qquad \ln x = \int_1^x \frac{1}{t} dt.$$
- ln is smooth, increasing, and bijective.
- For all $x, y \in (0, \infty)$,
$$\ln(xy) = \ln x + \ln y.$$
  We can prove this using version 1 of the FTC.

Chapter 7

Uniform convergence

7.1 Pointwise convergence versus uniform convergence

It is often convenient (or even necessary) to define a function $f : \mathbb{R} \to \mathbb{R}$ as a limit of a sequence of other functions $f_n : \mathbb{R} \to \mathbb{R}$, for example, using a power series,

$$f(x) = \sum_{k=0}^{\infty} \frac{x^k}{k!}$$

or a Fourier series,

$$g(x) = \sum_{k=1}^{\infty} \frac{4}{(2k+1)\pi} \sin(2k+1)x.$$

These are limits of the sequences of functions

$$f_n(x) = \sum_{k=0}^{n} \frac{x^k}{k!} \qquad \text{and} \qquad g_n(x) = \sum_{k=0}^{n} \frac{4}{(2k+1)\pi} \sin kx$$

respectively. We would like to be able to tell which properties of the terms of the sequence are shared by the limit function. For example, all of the functions f_n, g_n defined above are smooth (that is infinitely differentiable) on the whole of $\mathbb{R}$. Can we conclude that f and g are also smooth? The answer is no, not immediately: it turns out that f *is* smooth (it's the exponential function) but g isn't even continuous. To address this kind of question we will need to develop some new tools.

First, let's ask, given a sequence of functions, $f_n : D \to \mathbb{R}$, where $D \subseteq \mathbb{R}$ is some fixed set, what does it *mean* to say that (f_n) converges to $f : D \to \mathbb{R}$? The obvious answer (implicitly used abvove) is:

Definition 7.1 A sequence of functions $f_n : D \to \mathbb{R}$ **converges pointwise** to a function $f : D \to \mathbb{R}$ if, for each fixed $x \in D$, the real sequence $(f_n(x))$ converges (in the sense of Definition 1.1) to the real number $f(x)$.

Here's an instructive example:

Example 7.2 Consider the sequence $f_n : [0,1] \to \mathbb{R}$, $f_n(x) = x^n$. If $0 \le x < 1$ then

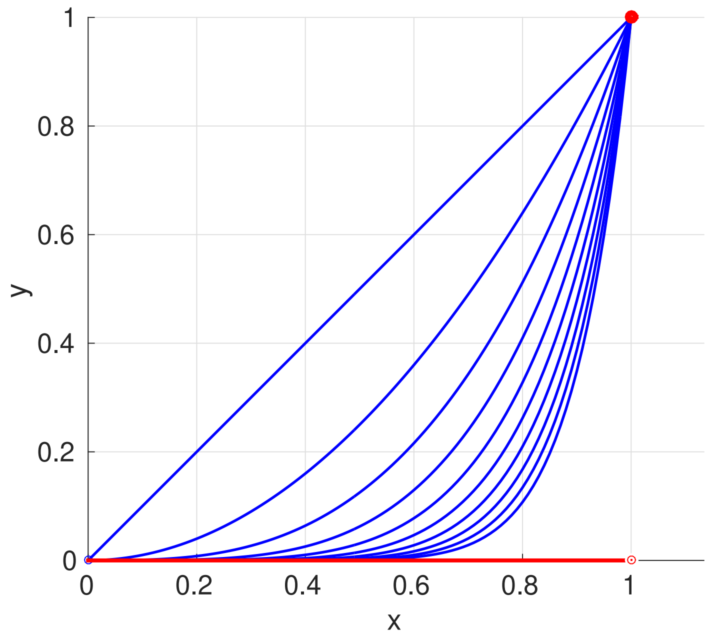
$$f_n(x) = x^n \to 0.$$

If $x = 1$, however,

$$f_n(1) = 1^n = 1 \to 1.$$

Hence (f_n) converges pointwise to the discontinuous function

$$f(x) = \begin{cases} 0, & 0 \le x < 1, \\ 1, & x = 1. \end{cases} \tag{7.1}$$



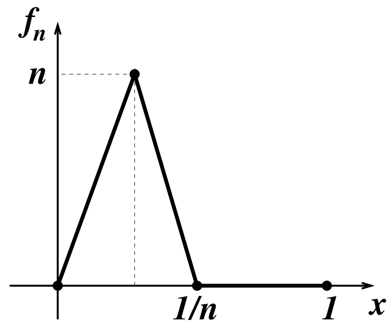
This illustrates a key weakness of pointwise convergence: it doesn't necessarily preserve continuity. Another weakness is that pointwise convergence doesn't interact well with integration. That is, a sequence of Riemann integrable functions $f_n : [a, b] \to \mathbb{R}$ may converge pointwise to a function $f : [a, b] \to \mathbb{R}$ which is not Riemann integrable or, even if it is, one may find that

$$\lim_{n\to\infty} \int_a^b f \ne \int_a^b f.$$

Example 7.3 For each $n \in \mathbb{Z}^+$, let $f_n : [0, 1] \to \mathbb{R}$ be the function

$$f_n(x) = \begin{cases} 2n^2 x, & 0 \le x \le \frac{1}{2n}, \\ 2n - 2n^2 x, & \frac{1}{2n} \le x \le \frac{1}{n}, \\ 0, & \frac{1}{n} \le x \le 1 \end{cases}$$

whose graph is depicted below:



Choose and fix any $x \in (0,1]$. Then, for all $n > 1/x$, $x > 1/n$ so $f_n(x) = 0$. Hence, for each fixed $x \in (0,1]$, the real sequence $f_n(x)$ is eventually 0. Hence $f_n(x) \to 0$. Furthermore $f_n(0) = 0$ for all n, so this sequence also converges to 0. So, if we define $f : [0,1] \to \mathbb{R}$ to be the constant function $f(x) = 0$, then for all $x \in [0,1]$, $\lim_{n\to\infty} f_n(x) = f(x)$.

Note that each f_n is continuous, hence Riemann integrable, and clearly

$$\int_0^1 f_n = \frac{1}{2} \times \frac{1}{n} \times n = \frac{1}{2},$$

a (constant) sequence that converges to $1/2$. But

$$\int_0^1 f = 0.$$

□

Nonetheless, pointwise limits do have some nice properties.

Proposition 7.4 *If (f_n) converges pointwise, its limit is unique.*

Proof: Assume, towards a contradiction, that (f_n) converges to both f and g, where $f \neq g$. Then, by assumption, there exists $c \in D$ such that $f(c) \neq g(c)$. But $f_n(c) \to f(c)$ and $f_n(c) \to g(c)$, which contradicts the uniqueness of the limit of a convergent real sequence. □

Similarly, we can show that pointwise limits have an Algebra of Limits property.

Proposition 7.5 *Let (f_n) converge to f pointwise and (g_n) to g pointwise on D. Then*

(i) $(f_n + g_n)$ converges to $f + g$ pointwise on D and

(ii) $(f_n g_n)$ converges to fg pointwise on D.

Proof: Exercise. Just apply the usual Algebra of Limits for sequences at each fixed $x \in D$. □

A much stronger, and more useful, notion of convergence is that of *uniform* convergence. To define this, it's convenient to make a preliminary definition.

Definition 7.6 The **sup norm** of a bounded function $f : D \to \mathbb{R}$ is

$$\|f\| := \sup\{|f(x)| \ : \ x \in D\}.$$

Example 7.7 Compute the sup norm of $f : \mathbb{R} \to \mathbb{R}$, $f(x) = \dfrac{x^2}{1+x^2}$.
Solution: For all $x \in \mathbb{R}$,

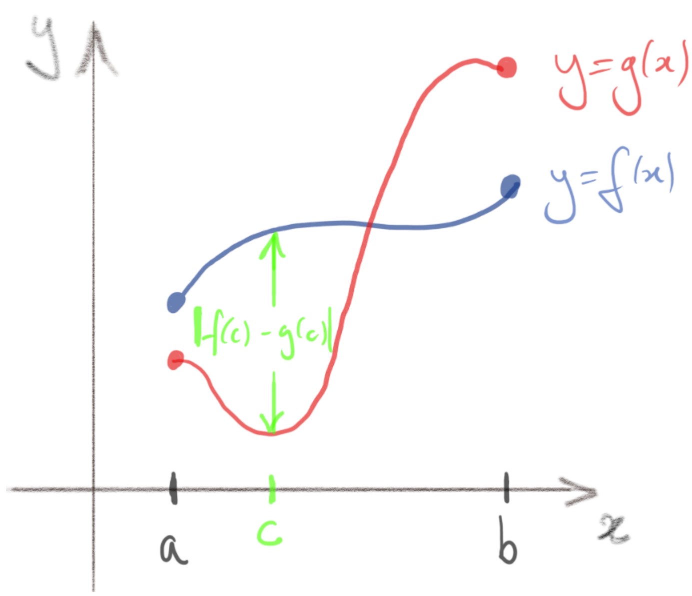
$$0 \leq f(x) = \frac{x^2}{1+x^2} = 1 - \frac{1}{1+x^2} < 1$$

and we see that the range of f is $[0, 1)$. Hence,

$$\|f\| = \sup\{|f(x)| : x \in \mathbb{R}\} = 1.$$

$\square$

In the case where $f, g : [a, b] \to \mathbb{R}$ are continuous, we can give $\|f - g\|$ a geometric interpretation: it's the maximum separation between $(x, f(x))$ and $(x, g(x))$ as x ranges over $[a, b]$. We may think of this as a "distance" between the functions f and g.



So $\|f - g\|$ defines a sort of "distance" between bounded functions, in the same way that $|a - b|$ defines a distance between real numbers a and b. This line of thought leads to the main definition of this chapter:

Definition 7.8 A sequence of bounded functions $f_n : D \to \mathbb{R}$ **converges uniformly** to a function $f : D \to \mathbb{R}$ if the real sequence $\|f_n - f\| \to 0$.

Example 7.2 revisited Claim: The sequence $f_n : [0, 1] \to \mathbb{R}$, $f_n(x) = x^n$ does *not* converge uniformly to f, the function defined in (7.1).

Proof: For each fixed n,

$$|f_n(x) - f(x)| = |x^n - f(x)| = \begin{cases} x^n, & 0 \leq x < 1, \\ 0, & x = 1. \end{cases}$$

Hence

$$\{|f_n(x) - f(x)| \,:\, x \in [0,1]\} = \{x^n \,:\, 0 \leq x < 1\} \cup \{0\}$$
$$= [0,1)$$

and so

$$\|f_n - f\| = \sup\{|f_n(x) - f(x)| \,:\, x \in [0,1]\}$$
$$= \sup[0,1) = 1$$

for all n. This sequence clearly doesn't converge to 0. □

So pointwise convergence certainly does *not* imply uniform convergence. However, the converse holds:

Theorem 7.9 *If a sequence $f_n : D \to \mathbb{R}$ converges uniformly to $f : D \to \mathbb{R}$, it converges pointwise to f.*

Proof: Assume the sequence of functions $f_n : D \to \mathbb{R}$ converges uniformly to $f : D \to \mathbb{R}$. Then, for each fixed $x \in D$,

$$0 \leq |f_n(x) - f(x)| \leq \sup\{|f_n(y) - f(y)| \,:\, y \in D\} = \|f_n - f\| \to 0,$$

so $|f_n(x) - f(x)| \to 0$ by the Squeeze Rule. Hence $f_n(x) \to f(x)$, Since this holds for each $x \in D$, f_n converges pointwise to f. □

Corollary 7.10 *If (f_n) converges uniformly, its limit is unique.*

Proof: Assume f_n converges uniformly to both f and g. Then f_n converges pointwise to f and f_n converges pointwise to g (Theorem 7.9). Hence $f = g$ (Proposition 7.4). □

We have seen that pointwise convergence doesn't always preserve continuity. That is, a sequence of continuous functions can converge pointwise to a discontinuous function (Example 7.2). A key advantage of *uniform* convergence is that it *does* preserve continuity.

Theorem 7.11 *Let $f_n : D \to \mathbb{R}$ be a sequence of continuous functions converging uniformly to $f : D \to \mathbb{R}$. Then f is continuous.*

Before diving into the proof of this, let's outline the strategy. We wish to show that the limit function f is continuous at a for each $a \in D$. Using the ε—δ criterion for continuity (Theorem 2.21), this amounts (roughly speaking) to showing that $|f(x) - f(a)|$ can be made as small as we like by taking $|x - a|$ sufficiently small. The key observation is that, for any fixed $n \in \mathbb{Z}^+$,

$$\begin{aligned} |f(x) - f(a)| &= |f(x) - f_n(x) + f_n(x) - f_n(a) + f_n(a) - f(a)| \\ &\leq |f_n(x) - f(x)| + |f_n(x) - f_n(a)| + |f_n(a) - f(a)| \\ &\leq \|f_n - f\| + |f_n(x) - f_n(a)| + \|f_n - f\|. \end{aligned}$$

Now $\|f_n - f\|$ can be made as small as we like by choosing n suitably large (since f_n converges to f uniformly, meaning $\|f_n - f\| \to 0$). Having chosen n suitably large, f_n is a fixed continuous function, so $|f_n(x) - f_n(a)|$ can be made as small as we like by taking $|x - a|$ sufficiently small. OK, let's write it up:

Proof of Theorem 7.11: We must prove that f is continuous at a for all $a \in D$. By Theorem 2.21, it suffices to show that, for each $\varepsilon > 0$, there exists $\delta > 0$ such that, for all $x \in D$ with $|x - a| < \delta$, $|f(x) - f(a)| < \varepsilon$.

So let a fixed $\varepsilon > 0$ be given. Since f_n converges to f uniformly, $\|f_n - f\| \to 0$. Hence, there exists $N \in \mathbb{Z}^+$ such that, for all $n \geq N$, $\|f_n - f\| < \varepsilon/3$. In particular,

$$\|f_N - f\| = \sup\{|f_N(x) - f(x)| \,:\, x \in D\} < \frac{\varepsilon}{3},$$

and hence

$$|f_N(x) - f(x)| < \frac{\varepsilon}{3} \qquad \text{for all } x \in D. \tag{7.2}$$

Furthermore, f_N is, by assumption, continuous, so, by Theorem 2.21, there exists $\delta > 0$ such that, for all $x \in D$ with $|x - a| < \delta$,

$$|f_N(x) - f_N(a)| < \frac{\varepsilon}{3}.$$

Hence, for all $x \in D$ with $|x - a| < \delta$,

$$\begin{aligned} |f(x) - f(a)| &= |f(x) - f_N(x) + f_N(x) - f_N(a) + f_N(a) - f(a)| \\ &\leq |f_N(x) - f(x)| + |f_N(x) - f_N(a)| + |f_N(a) - f(a)| \\ &< \frac{\varepsilon}{3} + |f_N(x) - f_N(a)| + \frac{\varepsilon}{3} \qquad \text{by (7.2)} \\ &< \varepsilon. \end{aligned}$$

□

This gives an indirect way of showing that $f_n(x) = x^n$ does not converge uniformly on $[0, 1]$. Assume, towards a contradiction, that f_n converges uniformly to some function f. Then f_n converges pointwise to f (Theorem 7.9) and hence f must be the discontinuous function defined in equation (7.1) (by Proposition 7.4). But this contradicts Theorem 7.11 (because each f_n is continuous on $[0, 1]$).

Exercise 7.12 Determine whether the following sequences of functions converge pointwise. For those that do, determine whether they converge uniformly.

(i) $f_n : [0,1] \to \mathbb{R}$, $f_n(x) = x/n$.

(ii) $f_n : [0,\pi] \to \mathbb{R}$, $f_n(x) = \sin(x + x^2/n)$.

Solution:

(i) For each fixed $x \in [0,1]$, $x/n \to 0$, so f_n converges pointwise to the constant function $f(x) = 0$. Furthermore

$$\|f_n - f\| = \sup\{|x/n - 0| : x \in [0,1]\} = \sup\{\frac{x}{n} : x \in [0,1]\} = \frac{1}{n} \to 0.$$

Hence f_n converges to 0 uniformly

(ii) For each fixed $x \in [0,\pi]$, $x + x^2/n \to x$ so, assuming sin is continuous (we haven't actually proved this yet, it's on our to-do list), $f_n(x) \to \sin x$. Hence $f_n \to \sin$ pointwise. Further, given any $x \in [0,\pi]$ and $n \in \mathbb{Z}^+$ we may apply the Mean Value Theorem to the function sin on the interval $[x, x + x^2/n]$ to deduce that there exists $c \in (x, x + x^2/n)$ such that

$$\frac{\sin(x + x^2/n) - \sin x}{(x + x^2/n) - x} = \cos c,$$

so

$$f_n(x) - \sin x = \frac{x^2}{n}\cos c$$

and hence

$$|f_n(x) - \sin x| \leq \frac{x^2}{n}.$$

Since this holds for all $x \in [0,\pi]$,

$$\|f_n - \sin\| = \sup\{|f_n(x) - \sin x| : x \in [0,\pi]\} \leq \frac{\pi^2}{n}$$

so $\|f_n - \sin\| \to 0$ by the Squeeze Rule. Hence f_n converges to sin uniformly on $[0,\pi]$

7.2 Uniform convergence and calculus

We saw in Example 7.3 that a sequence of continuous, and hence Riemann integrable, functions $f_n : [0,1] \to \mathbb{R}$ which converge pointwise to a continuous function f, may have integrals which do *not* converge to the integral of f. So pointwise convergence does not justify a manoeuvre like

$$\lim_{n\to\infty} \int_a^b f_n = \int_a^b \lim_{n\to\infty} f_n,$$

even if both sides of the equation exist. Things are much nicer if we know the convergence $f_n \to f$ is uniform, however:

Theorem 7.13 *Let $f_n : [a,b] \to \mathbb{R}$ be a sequence of continuous functions converging uniformly to $f : [a,b] \to \mathbb{R}$. Then*

$$\int_a^b f_n \to \int_a^b f.$$

Proof: First note that f is necessarily continuous (Theorem 7.11) and hence Riemann integrable (Theorem 5.20), so the proposed limit $\int_a^b f$ certainly exists. Now, for all n,

$$\begin{aligned} 0 \leq \left| \int_a^b f_n - \int_a^b f \right| &= \left| \int_a^b (f_n - f) \right| && \text{(Theorem 5.22)} \\ &\leq \int_a^b |f_n - f| && \text{(Theorem 5.24)} \\ &\leq (b-a)\|f_n - f\| && \text{(Theorem 5.23).} \end{aligned}$$

Hence, $\int_a^b f_n \to \int_a^b f$ by the Squeeze Rule. □

We deduce from this that the sequence $f_n : [0,1] \to \mathbb{R}$, of "triangle" functions defined in Example 7.3 does *not* converge uniformly. Note that its pointwise limit (the constant function 0) is continuous, so we couldn't deduce this from Theorem 7.11.

Example 7.14 Claim: $\lim_{n\to\infty}\int_0^\pi \sin(x+\frac{x^2}{n})dx = 2.$

Proof: We saw in Exercise 7.12 that $f_n : [0,\pi]\to\mathbb{R}$, $f_n(x) = \sin(x+x^2/n)$, converges uniformly to sin. Hence, by Theorem 7.13,

$$\begin{aligned}\int_0^\pi f_n &\to \int_0^\pi \sin x\, dx \\ &= [-\cos x]_0^\pi = 2\end{aligned}$$

by the Fundamental Theorem of the Calculus (version 2). □

Note we have no hope of computing the integrals $\int_0^\pi f_n$ exactly (I invite you to try to write down an antiderivative for f_n), so this limit can't be computed directly.

Exercise 7.15 Compute

$$\lim_{n\to\infty}\int_0^1 \frac{n}{n+x^n}dx,$$

rigorously justifying your answer.

Solution: Let $f_n : [0,1]\to\mathbb{R}$, $f_n(x) = \frac{n}{n+x^n}$. Note that $f_n(1) = \frac{n}{n+1}\to 1$ and, for all $x\in[0,1)$, $x^n\to 0$ so

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

by the Algebra of Limits. Hence f_n converges pointwise to the constant function $f(x) = 1$. Now, for all $x\in[0,1]$ and $n\in\mathbb{Z}^+$,

$$|f_n(x)-f(x)| = 1-\frac{n}{n+x^n} = \frac{x^n}{n+x^n}\le\frac{1}{n+x^n}\le\frac{1}{n}$$

and so

$$\|f_n - f\|\le\frac{1}{n}$$

whence $\|f_n - f\|\to 0$ by the Squeeze Rule. So f_n converegs to the constant function 1 uniformly. Hence, by Theorem 7.13,

$$\lim_{n\to\infty}\int_0^1 f_n = \int_0^1 f = 1.$$

Our next goal is to prove an analogue of Theorem 7.13 for *derivatives*. Under what circumstances can we be sure that a sequence f_n of differentiable functions converges to a differentiable function? The functions f_n in the following theorem are assumed to be *continuously differentiable*, meaning that each function $f_n : [a, b] \to \mathbb{R}$ is differentiable and that its derivative $f_n' : [a, b] \to \mathbb{R}$ is continuous. (Recall that every differentiable function is continuous, but its *derivative* isn't necessarily continuous: see Example 4.13 for a counterexample.)

Theorem 7.16 *Let $f_n : [a, b] \to \mathbb{R}$ be a sequence of continuously differentiable functions which converge pointwise to $f : [a, b] \to \mathbb{R}$, and whose derivatives $f_n' : [a, b] \to \mathbb{R}$ converge uniformly to $g : [a, b] \to \mathbb{R}$. Then f is continuously differentiable and $f' = g$.*

Proof: For each $n \in \mathbb{Z}^+$ let $F_n : [a, b] \to \mathbb{R}$ be defined by

$$F_n(x) = \int_a^x f_n'.$$

Note that this exists since f_n' is, by assumption, continuous. By version 2 of the Fundamental Theorem of the Calculus,

$$F_n(x) = f_n(x) - f_n(a).$$

By Theorem 7.13, for each fixed x,

$$f_n(x) - f_n(a) = F_n(x) \to \int_a^x g,$$

since f_n' converges uniformly to g on $[a, x]$. But f_n converges pointwise to f, so $f_n(x) \to f(x)$ and $f_n(a) \to f(a)$. Hence, $f_n(x) - f_n(a) \to f(x) - f(a)$ (by the Algebra of Limits). But limits are unique, so

$$f(x) - f(a) = \int_a^x g.$$

Hence, by version 1 of the Fundamental Theorem of the Calculus, f is differentiable and $f'(x) = g(x)$. □

At the moment, it's hard to see why Theorem 7.16 is useful, because it's hard to imagine a situation where we can check that f_n' converges uniformly to a continuous function g without already knowing that the limit f of f_n is differentiable. For example, if we apply this Theorem to the sequence of functions in Exercise 7.15 we can deduce (at some effort) that the constant function $f(x) = 1$ is differentiable and has derivative 0 – hardly a startling revelation!

We will see in the next chapter that Theorem 7.16 can be extremely powerful when applied to the sequence of partial sums of a convergent power series. But before we can explore that application, we need to develop a way to prove that a sequence f_n of functions converges uniformly to *something* f even when we can't really say explicitly what f is (except tautologically: $f(x) = \lim_{n\to\infty} f_n(x)$).

7.3 Completeness of the set of bounded functions

Recall that a real sequence (a_n) is **Cauchy** if, for each $\varepsilon > 0$, there exists $N \in \mathbb{Z}^+$ such that, for all $n, m \geq N$, $|a_n - a_m| < \varepsilon$ (Definition 1.16), and that this condition turns out to be *equivalent* to being convergent, that is, (a_n) converges if and only if it is Cauchy (Theorem 1.20). This Theorem allows us to prove that a real sequence converges to *something* $A \in \mathbb{R}$ even when we have no idea what A actually is. The aim of this section is to repeat this trick for uniform convergence of sequences of functions.

For a sequence of *functions* $f_n : D \to \mathbb{R}$, where D is some fixed subset of $\mathbb{R}$, we can define what it means to be *uniformly* Cauchy, in much the same way that we defined uniform convergence (Definition 7.8):

Definition 7.17 A sequence of bounded functions $f_n : D \to \mathbb{R}$ is **uniformly Cauchy** if, for each $\varepsilon > 0$, there exists $N \in \mathbb{Z}^+$ such that, for all $n, m \geq N$, $\|f_n - f_m\| < \varepsilon$.

As usual, $\|\cdot\|$ denotes the sup norm here. Note that this is only well defined for *bounded* functions, which is why Definition 7.17 specifies that each $f_n : D \to \mathbb{R}$ is bounded.

Exercise 7.18 Show, directly from Definition 7.17, that the sequence

$$f_n : [-1, 1] \to \mathbb{R}, \qquad f_n(x) = \frac{x^3}{n}$$

is uniformly Cauchy.

Solution: For all $x \in [-1, 1]$ and $n, m \in \mathbb{Z}^+$,

$$|f_n(x) - f_m(x)| = \left|\frac{x^3}{n} - \frac{x^3}{m}\right| = |x|^3 \left|\frac{1}{n} - \frac{1}{m}\right|$$

and so

$$\|f_n - f_m\| \leq \left|\frac{1}{n} - \frac{1}{m}\right| \leq \frac{1}{n} + \frac{1}{m}$$

by the Triangle Inequality. Hence, given any $\varepsilon > 0$ we may choose any $N \in \mathbb{Z}^+$ with $N > 2/\varepsilon$. Then, for all $n, m \geq N$,

$$\|f_n - f_m\| \leq \frac{1}{N} + \frac{1}{N} < \varepsilon.$$

Our goal in this section is to prove that a sequence of bounded functions is uniformly convergent if and only if it is uniformly Cauchy. The "only if" direction of this statement is much easier to prove than the "if" direction, so let's start with that.

Theorem 7.19 *Let $f_n : D \to \mathbb{R}$ be a sequence of bounded functions converging uniformly to a bounded function $f : D \to \mathbb{R}$. Then (f_n) is uniformly Cauchy.*

The proof of this follows the proof of Lemma 1.17 (the analogous statement for real sequences) almost word for word. To make this work, we need a "triangle inequality" for the sup norm.

Lemma 7.20 *For all bounded functions $f, g : D \to \mathbb{R}$,*

$$\|f + g\| \leq \|f\| + \|g\|.$$

Proof: For all $x \in D$, $|f(x) + g(x)| \leq |f(x)| + |g(x)| \leq \|f\| + \|g\|$, since $\|f\|$ is an upper bound on $\{|f(x)| : x \in D\}$ and $\|g\|$ is an upper bound on $\{|g(x)| : x \in D\}$. Hence, $\|f\| + \|g\|$ is an upper bound on $\{|f(x) + g(x)| : x \in D\}$. But $\|f + g\|$ is, by definition, the *least* upper bound on this set. Hence $\|f + g\| \leq \|f\| + \|g\|$. □

Proof of Theorem 7.19: Let $\varepsilon > 0$ be given. Since (f_n) converges to f uniformly, there exists $N \in \mathbb{Z}^+$ such that, for all $n \geq N$, $\|f_n - f\| < \varepsilon/2$. Hence, for all $n, m \geq N$,

$$\|f_n - f_m\| = \|f_n - f - (f_m - f)\| \leq \|f_n - f\| - \|f_m - f\| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2}$$

by Lemma 7.20. □

OK, now for the hard(er) part. We first note that, if (f_n) is *uniformly* Cauchy, it is certainly *pointwise* Cauchy, hence pointwise convergent. That is:

Lemma 7.21 *Let $f_n : D \to \mathbb{R}$ be a uniformly Cauchy sequence of bounded functions. Then (f_n) converges pointwise to some function $f : D \to \mathbb{R}$.*

Proof: Choose and fix $x \in D$. I claim that the real sequence $(f_n(x))$ is Cauchy. To see this, let $\varepsilon > 0$ by given. Since (f_n) is uniformly Cauchy, there exists $N \in \mathbb{Z}^+$ such that, for all $n, m \geq N$, $\|f_n - f_m\| < \varepsilon$. Hence, for all $n, m \geq N$,

$$|f_n(x) - f_m(x)| \leq \|f_n - f_m\| < \varepsilon.$$

Since the real sequence $(f_n(x))$ is Cauchy, it converges, by Theorem 1.20, to some limit $f(x) \in \mathbb{R}$. Allowing x to vary in D, we obtain a function $f : D \to \mathbb{R}$ to which (f_n) converges pointwise. □

Note that we don't (yet) know that the pointwise limit $f : D \to \mathbb{R}$ of our uniformly Cauchy sequence of bounded functions is itself a bounded function! We need to prove this.

Lemma 7.22 *Let a uniformly Cauchy sequence of bounded functions $f_n : D \to \mathbb{R}$ converge pointwise to $f : D \to \mathbb{R}$. Then f is bounded.*

Proof: We must show that there exists $M > 0$ such that, for all $x \in D$, $|f(x)| \leq M$.

Since (f_n) is uniformly Cauchy, there exists $N \in \mathbb{Z}^+$ such that, for all $n, m \geq N$, $\|f_n - f_m\| < 1$. Furthermore, $f_N : D \to \mathbb{R}$ is a bounded function, so there exists $K > 0$ such that, for all $x \in D$, $|f_N(x)| \leq K$. I claim that $|f|$ is bounded above by $K + 1$.

To see this, choose and fix $x \in D$. Then, for all $n \in \mathbb{Z}^+$,

$$\begin{aligned} |f(x)| &= |f(x) - f_n(x) + f_n(x) - f_N(x) + f_N(x)| \\ &\leq |f(x) - f_n(x)| + |f_n(x) - f_N(x)| + |f_N(x)| \\ &\leq |f(x) - f_n(x)| + \|f_n - f_N\| + K \end{aligned}$$

Hence, for all $n \geq N$,

$$|f(x)| \leq |f(x) - f_n(x)| + 1 + K. \tag{7.3}$$

But, by assumption, $f_n(x) \to f(x)$, so the sequence on the right hand side of (7.3) converges to $1 + K$. Since this sequence is bounded below by $|f(x)|$, it follows that $1 + K \geq |f(x)|$ (Proposition 1.7). This is true whatever $x \in D$ we choose, so $|f|$ is bounded above by $M = K + 1$. □

So we now know that every uniformly Cauchy sequence of bounded functions converges *pointwise* to a *bounded* function. The last job is to show that this convergence is actually *uniform* (recall uniform convergence implies pointwise convergence, but not vice versa, so this is not automatic).

Theorem 7.23 *A sequence $f_n : D \to \mathbb{R}$ of bounded functions converges uniformly if and only if it is uniformly Cauchy.*

Proof: We have already proved the "only if" part (Theorem 7.19).

Assume that (f_n) is uniformly Cauchy. Then (f_n) converges pointwise to some bounded function $f : D \to \mathbb{R}$ by Lemmas 7.21 and 7.22. We must show that $\|f_n - f\| \to 0$.

Let $\varepsilon > 0$ be given. Since (f_n) is uniformly Cauchy, there exists $N \in \mathbb{Z}^+$ such that, for all $n, m \geq N$, $\|f_n - f_m\| < \varepsilon/4$. I claim that, for all $n \geq N$,

$$\|f_n - f\| < \varepsilon.$$

To see this, choose and fix $x \in D$. Then, for all $n, m \geq N$,

$$\begin{aligned} |f_n(x) - f(x)| &= |f_n(x) - f_m(x) + f_m(x) - f(x)| \\ &\leq |f_n(x) - f_m(x)| + |f_m(x) - f(x)| \\ &\leq \|f_n - f_m\| + |f_m(x) - f(x)| \\ &< \frac{\varepsilon}{4} + |f_m(x) - f(x)|. \end{aligned} \tag{7.4}$$

Note this inequality holds for *all* $m \geq N$. Now (f_n) converges pointwise to f, so $f_n(x) \to f(x)$. Hence, there exists $N_1 \in \mathbb{Z}^+$ (depending on x) such that, for all $m \geq N_1$, $|f_m(x) - f(x)| < \varepsilon/4$. Applying (7.4) in the case $m = \max\{N, N_1\}$, we see that, for all $n \geq N$,

$$|f_n(x) - f(x)| < \frac{\varepsilon}{4} + \frac{\varepsilon}{4} = \frac{\varepsilon}{2}.$$

This inequality holds for all $x \in D$, and N is independent of x, so, for all $n \geq N$,

$$\|f_n - f\| = \sup\{|f_n(x) - f(x)| \,:\, x \in D\} \leq \frac{\varepsilon}{2} < \varepsilon.$$

Since $\varepsilon > 0$ was arbitrary, $\|f_n - f\| \to 0$. □

Let us denote the set of bounded functions on D

$$B(D) := \{f : D \to \mathbb{R} \,:\, f \text{ is bounded}\}$$

and the set of continuous functions on D

$$C(D) := \{f : D \to \mathbb{R} \,:\, f \text{ is continuous}\}$$

For general sets D, neither of these is a subset of the other. For example $f(x) = x^2$ is in $C(\mathbb{R})$ but not $B(\mathbb{R})$, and $g(x) = 1$ for $x \geq 0$, $g(x) = 0$ for $x < 0$ is in $B(\mathbb{R})$ but not $C(\mathbb{R})$. In the case where $D = [a, b]$, a closed bounded interval, however, every continuous function is bounded (by the Extreme Value Theorem), so $C([a, b]) \subset B([a, b])$.

We have just shown that the set $B(D)$ enjoys (with respect to the sup norm) a fundamental property of the set $\mathbb{R}$ (with respect to absolute value): every Cauchy sequence in the set converges to a limit in the set. Mathematicians call this property **completeness**. So what we have just shown is that $B(D)$ is *complete* with respect to the sup norm. In the case where $D = [a, b]$, it's not hard to deduce that $C(D)$ is also complete with respect to the sup norm.

Theorem 7.24 *If a sequence of continuous functions $f_n : [a, b] \to \mathbb{R}$ is uniformly Cauchy, then it converges uniformly to some continuous function $f : [a, b] \to \mathbb{R}$.*

Proof: Each f_n is a bounded function (by the Extreme Value Theorem), so (f_n) converges uniformly to some bounded function $f : [a, b] \to \mathbb{R}$ by Theorem 7.23. But f is continuous by Theorem 7.11. □

The neat thing about Theorems 7.23 and 7.24 is that they allow us to prove that a sequence of functions converges *uniformly* without knowing what its limit is. As we will see, this is extremely useful when we come to consider functions defined by power series.

Summary

- A sequence of functions $f_n : D \to \mathbb{R}$ **converges pointwise** to $f : D \to \mathbb{R}$ if, for each $x \in D$, $f_n(x) \to f(x)$.
- The **sup norm** of a function $f : D \to \mathbb{R}$ is $\|f\| = \sup\{|f(x)| \,:\, x \in D\}$.
- A sequence of functions $f_n : D \to \mathbb{R}$ **converges uniformly** to $f : D \to \mathbb{R}$ if the real sequence $\|f_n - f\|$ converges to 0.
- If f_n converges uniformly to f, it converges pointwise to f. The converse is **false**.
- If a sequence of continuous functions $f_n : D \to \mathbb{R}$ converges uniformly to f then f is continuous.
- If a sequence of continuous functions $f_n : [a, b] \to \mathbb{R}$ converges uniformly to f then $\displaystyle\int_a^b f_n \to \int_a^b f$.
- If a sequence of continuously differentiable functions $f_n : [a, b] \to \mathbb{R}$ converges pointwise to f, and its sequence of derivatives $f_n' : [a, b] \to \mathbb{R}$ converges uniformly to g, then f is differentiable and $f' = g$.
- A sequence $f_n : D \to \mathbb{R}$ of bounded functions is **uniformly Cauchy** if, for each $\varepsilon > 0$, there exists $N \in \mathbb{Z}^+$ such that, for all $n, m \geq N$, $\|f_n - f_m\| < \varepsilon$.
- A sequence of bounded functions is uniformly convergent if and only if it is uniformly Cauchy.

Chapter 8

Power series

8.1 Convergence tests for series

We begin with a brisk review of the basics concerning convergence of series. All this material was covered in detail in MATH1026. (The main ideas were covered in MATH1055, albeit with some of the proofs omitted.)

Recall that a **series** is a formal infinite sum

$$\sum_{n=0}^{\infty} a_n = a_0 + a_1 + a_2 + \cdots ,$$

that its k^{th} **partial sum** is the real number

$$s_k := \sum_{n=0}^{k} a_n = a_0 + a_1 + \cdots + a_k,$$

and that the series **converges** precisely if the sequence (s_k) converges (in the sense of Definition 1.1). In this case, we also use $\sum_{n=0}^{\infty} a_n$ to denote its limit.

The series **converges absolutely** if

$$\sum_{n=0}^{\infty} |a_n|$$

converges. Absolute convergence implies convergence, but convergence does *not* imply absolute convergence. For example, the **alternating harmonic series**

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

converges, but it does not converge absolutely since the **harmonic series**

$$\sum_{n=1}^{\infty} \frac{1}{n}$$

diverges.

If a series $\sum_{n=0}^{\infty} a_n$ converges then its sequence of terms (a_n) converges to 0. This fact is often called the **Divergence Test** since we can apply its contrapositive: if $a_n \nrightarrow 0$ then $\sum a_n$ diverges. Note that the *converse is false*, as illustrated by the harmonic series: the sequence of terms $a_n = 1/n \to 0$, but the corresponding *series* $\sum_{n=1}^{\infty} a_n$ diverges.

If all the terms of a series are positive, $a_n > 0$, then the sequence of partial sums $s_k = \sum_{n=0}^{k} a_n$ is an increasing sequence,

$$s_{k+1} = s_k + a_{k+1} > s_k,$$

so, by the Monotone Convergence Theorem, the series converges if (and only if) s_k is bounded above. This observation allows one to establish several useful convergence tests applicable to the case $a_n > 0$. The first allows us to prove convergence (or divergence) of a series $\sum a_n$ by comparing it with a (simpler) series $\sum b_n$ that we already know converges (or diverges).

Theorem 8.1 (The Comparison Test) *Let $a_n > 0$ and $b_n > 0$.*

(i) If a_n/b_n is bounded above and $\sum b_n$ converges, then $\sum a_n$ converges.

(ii) If b_n/a_n is bounded above and $\sum b_n$ diverges, then $\sum a_n$ diverges.

To use the comparison test effectively, you need a stock of simple example series whose convergence/divergence you've already established. Here's a useful and simple family of examples:

Example 8.2 Claim: $\displaystyle\sum_{n=1}^{\infty} \frac{1}{n^p}$, where $p \in (0, \infty)$, converges if $p > 1$ and diverges if $0 < p \leq 1$.

Proof: Exercise. The idea is to bound

$$s_k = \sum_{n=1}^{k} \frac{1}{n^p}$$

in terms of

$$\int_1^k \frac{1}{x^p} dx.$$

$\square$

Example 8.3 Claim: $\displaystyle\sum_{n=1}^{\infty} \frac{n}{n^2+1}$ diverges.

Proof: Let $a_n = n/(n^2+1)$ and $b_n = 1/n$. Then

$$\frac{b_n}{a_n} = \frac{n^2+1}{n^2} = 1 + \frac{1}{n^2} \leq 2,$$

so the sequence (b_n/a_n) is bounded above. Since $\displaystyle\sum_{n=1}^{\infty} b_n$ diverges, we conclude from part (ii) of the Comparison Test that $\sum_{n=1}^{\infty} a_n$ also diverges. $\square$

Exercise 8.4 Determine whether $\displaystyle\sum_{n=1}^{\infty} \frac{n \sin n}{n^3+2}$ converges.

Our next convergence test is perhaps the most useful – as well as being the easiest to use!

Theorem 8.5 (The Ratio Test) *Let $a_n > 0$ for all n and $a_{n+1}/a_n \to L$. Then*

(i) if $L < 1$, the series $\sum a_n$ converges,

(ii) if $L > 1$, the series $\sum a_n$ diverges.

Example 8.6 Claim: $\displaystyle\sum_{n=1}^{\infty} n^{50}\left(\frac{49}{50}\right)^n$ converges.

Proof: Let $a_n = n^{50}(49/50)^n > 0$, and note that

$$\frac{a_{n+1}}{a_n} = \frac{(n+1)^{50}}{n^{50}}\frac{49}{50} = \left(1+\frac{1}{n}\right)^{50}\frac{49}{50} \to \frac{49}{50} < 1.$$

Hence, by the Ratio Test, $\sum_{n=1}^{\infty} a_n$ converges. $\square$

Note that if the terms a_n of your series aren't all positive, you can always try applying the Comparison or Ratio Test to $\sum |a_n|$ to determine whether $\sum a_n$ converge absolutely. If it does, it certainly converges. Discarding all information about the sign of a_n does lose us potentially vital knowledge, however. If the terms *alternate* in sign, we have a particularly neat convergence test:

Theorem 8.7 (The Alternating Series Test) *Let (a_n) be a positive, decreasing sequence that converges to 0. Then*

$$\sum_{n=1}^{\infty} (-1)^{n+1} a_n$$

converges.

Example 8.8 The series $\displaystyle\sum_{n=1}^{\infty} (-1)^{n+1}\frac{1}{\sqrt{n}}$ converges, by the Alternating Series Test, since $a_n = 1/\sqrt{n}$ is positive, decreasing and converges to 0. As usual with convergence tests, we've proved that the series converges but have no idea what its limit is.

8.2 Power series and their radius of convergence

A **power series** is a series of the form

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

where x is interpreted as a real *variable*. The series depends on the choice of value for $x \in \mathbb{R}$. The terms of the series are $a_n x^n$, and the k^{th} partial sum is the polynomial function

$$f_k(x) = \sum_{n=0}^{k} a_n x^n = a_0 + a_1 x + a_2 x^2 + \cdots + a_k x^k,$$

just as for series in general. Note that, if $x = 0$, then every partial sum is $f_k(0) = a_0$, so the series certainly converges (to a_0) in that case. In general, a power series may converge for some values of x but diverge for others. If the series converges for a particular choice of x, its limit will, in general, depend on x. So a power series defines a real-valued function on some subset (perhaps all) of $\mathbb{R}$. The following example is fundamental (and, hopefully, familiar).

Example 8.9 (Geometric series) Claim: The series

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

converges to $1/(1-x)$ if $|x| < 1$, and diverges otherwise.

Proof: If $x = 1$, the k-th partial sum is $s_k = k+1$, and this sequence clearly diverges. If $x \neq 1$ then

$$\begin{aligned} s_k &= 1 + x + x^2 + \cdots x^k \\ \Rightarrow \quad x s_k &= x + x^2 + x^3 + \cdots + x^{k+1} \\ \Rightarrow \quad (1-x) s_k &= 1 - x^{k+1} \\ \Rightarrow \quad s_k &= \frac{1 - x^{k+1}}{1-x}. \end{aligned}$$

The sequence x^k converges to 0 if $|x| < 1$ and diverges if $x = -1$. Hence, $s_k \to 1/(1-x)$ if $|x| < 1$ and diverges otherwise. $\square$

So if we use the power series to define a function

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

then $f : D \to \mathbb{R}$ where $D = (-1, 1)$ and for all $x \in D$,

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

Definition 8.10 The **radius of convergence** of a power series $\sum_{n=0}^{\infty} a_n x^n$ is

$$R := \sup\{|x| \,:\, \textstyle\sum_{n=0}^{\infty} |a_n x^n| \text{ converges}\}.$$

If this set is unbounded above, we say that $R = \infty$.

The radius of convergence of a power series tells us almost everything about the subset of $\mathbb{R}$ on which it converges:

Theorem 8.11 *Let a power series have radius of convergence $R > 0$. Then the series* **converges absolutely** *for $|x| < R$ and* **diverges** *for $|x| > R$.*

So the series converges on $(-R, R)$ and diverges on $(-\infty, -R) \cup (R, \infty)$. The only information not revealed by Theorem 8.11 is whether the series converges at $x = \pm R$. So, it's important to be able to compute radii of convergence. Luckily this can usually be achieved with a simple application of the Ratio Test.

Example 8.12 Claim: $\displaystyle\sum_{n=0}^{\infty} \frac{nx^{3n+1}}{n^2+1}$ has radius of convergence $R = 1$.

Proof: Let $b_n = |nx^{3n+1}/(n^2+1)| > 0$ and note that

$$\begin{aligned}
\frac{b_{n+1}}{b_n} &= \frac{|(n+1)x^{3n+4}|}{n^2+2n+3} \times \frac{n^2+1}{|nx^{3n+1}|} \\
&= \frac{(n+1)(n^2+1)|x|^3}{(n^2+2n+3)n} \\
&= \frac{(1+1/n)(1+1/n^2)}{1+2/n+3/n^2}|x|^3 \\
&\to |x|^3.
\end{aligned}$$

Hence, by the Ratio Test, the power series converges absolutely if $|x| < 1$, but not if $|x| > 1$. Comparing with Definition 8.10, we see that $R = 1$. $\square$

It follows that this series converges absolutely on $(-1, 1)$ and diverges on $\mathbb{R}\backslash[-1, 1]$ (Theorem 8.11).

Question What precisely is the set of values of x in $\mathbb{R}$ for which the power series in Example 8.12 converges? $[-1, 1)$

Many functions are most conveniently *defined* by power series:

Definition 8.13

$$\begin{aligned}
\exp : \mathbb{R} \to \mathbb{R}, \qquad \exp(x) &= \sum_{n=0}^{\infty} \frac{x^n}{n!}, \\
\sin : \mathbb{R} \to \mathbb{R}, \qquad \sin(x) &= \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{(2n+1)!}, \\
\cos : \mathbb{R} \to \mathbb{R}, \qquad \cos(x) &= \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n}}{(2n)!}.
\end{aligned}$$

Exercise 8.14 Show that these functions are well defined (in fact, the series defining them converge absolutely) for all $x \in \mathbb{R}$. This amounts to showing that each of them has radius of convergence $R = \infty$.

So we can use power series to define functions like exp, sin, and cos. But what do we know about such functions? Are they continuous? Differentiable? Smooth? To answer these questions we'll need to apply the theory of uniform convergence developed in the previous chapter.

8.3 Uniform convergence of series of functions

In the language of chapter 7, Theorem 8.11 establishes that the sequence of polynomial functions

$$f_k(x) = \sum_{n=0}^{k} a_n x^n$$

converges **pointwise** on $(-R, R)$ to some function $f(x)$, which we denote

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

As we learned, pointwise convergence is a rather weak property, from which we can't usually deduce much about the limit function (e.g. it may be discontinuous, even if each function in the sequence is continuous). We can deduce much more about the limit function if we can establish that the convergence is actually **uniform**. Recall also that we can show that a sequence of functions converges uniformly even if we have no information about its limit function by showing that the sequence is **uniformly Cauchy**.

In this section we will develop a test which allows us to establish quickly and easily that a *series* of functions is uniformly Cauchy, and hence uniformly convergent. We will state the test for general series of functions, not just polynomials, because there are many other useful choices, for example, **Fourier series**

$$\sum_{n=1}^{\infty} a_n \sin nx.$$

Theorem 8.15 (The Weierstrass M Test) *Let $g_n : D \to \mathbb{R}$ be a sequence of functions and, M_n a sequence of non-negative real numbers such that*

(i) for all $x \in D$, $|g_n(x)| \leq M_n$, and

(ii) the series $\sum_{n=0}^{\infty} M_n$ converges.

Then the sequence of functions

$$f_k : D \to \mathbb{R}, \qquad f_k(x) = \sum_{n=0}^{k} g_n(x)$$

converges uniformly.

Proof: We will prove that (f_k) is uniformly Cauchy. Let $\varepsilon > 0$ be given. By assumption, the sequence

$$s_k = \sum_{n=0}^{k} M_n$$

converges, so is Cauchy (Theorem 1.20). Hence, there exists $N \in \mathbb{Z}^+$ such that for all $k > l \geq N$, $|s_k - s_l| < \varepsilon/2$. Since $M_n \geq 0$ for all n,

$$|s_k - s_l| = s_k - s_l = \sum_{n=l+1}^{k} M_n.$$

Hence, for all $k > l \geq N$,

$$\sum_{n=l+1}^{k} M_n < \frac{\varepsilon}{2}.$$

But, for all $x \in D$, and all n, $|g_n(x)| \leq M_n$. Hence, for all $x \in D$, and all $k > l \geq N$

$$\begin{aligned} |f_k(x) - f_l(x)| &= \left| \sum_{n=0}^{k} g_n(x) - \sum_{n=0}^{l} g_n(x) \right| \\ &= \left| \sum_{n=l+1}^{k} g_n(x) \right| \\ &\leq \sum_{n=l+1}^{k} |g_n(x)| \\ &\leq \sum_{n=l+1}^{k} M_n \\ &< \frac{\varepsilon}{2}. \end{aligned}$$

Hence, for all $k > l \geq N$,

$$\|f_k - f_l\| = \sup\{|f_k(x) - f_l(x)| : x \in D\} \leq \frac{\varepsilon}{2} < \varepsilon.$$

So (f_k) is uniformly Cauchy and hence uniformly convergent by Theorem 7.23. □

Exercise 8.16 Let $\sum_{n=1}^{\infty} a_n$ be an absolutely convergent series. Prove that the Fourier series

$$\sum_{n=1}^{\infty} a_n \sin nx$$

converges uniformly to a continuous function $f : \mathbb{R} \to \mathbb{R}$.

Solution: Let $g_n(x) = a_n \sin nx$ and $M_n = |a_n|$. Then, by assumption,

$$\sum_{n=1}^{\infty} M_n$$

converges (that's what it means to say that $\sum_{n=1}^{\infty} a_n$ converges absolutely) and, for all $x \in \mathbb{R}$,

$$|g_n(x)| = |a_n||\sin nx| \leq |a_n| = M_n.$$

Hence, by the Weierstrass M Test, the series with terms g_n converges uniformly to some function $f : \mathbb{R} \to \mathbb{R}$. Since each of the partial sums

$$f_k(x) = \sum_{n=1}^{k} a_n \sin nx$$

is continuous, f is continuous by Theorem 7.11.

Returning to our main focus of study, it would be nice if we could use the M test to prove that power series converge uniformly on $(-R, R)$, where R is the radius of convergence. In fact, that's a bit too much to hope for, as we can deduce by thinking about the geometric series

$$\sum_{n=1}^{\infty} x^n.$$

This has $R = 1$ (as we showed directly in Example 8.9). If it converged *uniformly* its limit function $f : (-1, 1) \to \mathbb{R}$ would perforce be bounded. But we showed that

$$f(x) = \frac{1}{1-x}$$

which grows unbounded above as x approaches 1. So we deduce that this series does *not* converge uniformly on $(-1, 1)$. Nil desperandum! We can prove something almost as good as uniform convergence on $(-R, R)$:

Theorem 8.17 *Let the power series* $\sum_{n=1}^{\infty} a_n x^n$ *have radius of convergence* $R > 0$ *and* ρ *be any constant in* $(0, R)$. *Then the series converges uniformly on* $[-\rho, \rho]$.

Proof: We apply the Weierstrass M Test with $g_n(x) = a_n x^n$, $D = [-\rho, \rho]$ and $M_n = |a_n|\rho^n$. Certainly, for all $x \in [-\rho, \rho]$

$$|g_n(x)| = |a_n||x|^n \leq |a_n|\rho^n = M_n$$

so condition (i) is satisfied. Since $|\rho| < R$, the series converges absolutely at $x = \rho$, that is

$$\sum_{n=0}^{\infty} M_n = \sum_{n=0}^{\infty} |a_n|\rho^n$$

converges, by Theorem 8.11. Hence condition (ii) is satisfied also. So the series converges uniformly on $[-\rho, \rho]$ by the Weierstrass M Test (Theorem 8.15). $\square$

Each of the partial sums of a power series is a polynomial, and hence is continuous. It follows from Theorems 7.11 and 8.17 that the limit function

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

is continuous on $[-\rho, \rho]$ for every $\rho \in (0, R)$. But that implies that f is continuous on $(-R, R)$ since, at every point $a \in (-R, R)$ we can choose $\rho = (a + R)/2$, say (or any other $\rho \in (a, R)$). So we immediately deduce:

Corollary 8.18 *Let the power series* $\sum_{n=1}^{\infty} a_n x^n$ *have radius of convergence* $R > 0$. *Then this series converges to a continuous function* $f : (-R, R) \to \mathbb{R}$.

Excellent! But we want (much) more: we're actually going to prove that the limit function is *smooth* (that is, infinitely differentiable) on $(-R, R)$.

8.4 Differentiability of power series

Given a power series $f(x) = \sum_{n=0}^{\infty} a_n x^n$ of radius of convergence R, we know that f is a continuous function on $(-R, R)$. Is f differentiable on $(-R, R)$? If it is, what is f'? Your first guess is probably

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

the series obtained by differentiating f term by term. Actually, this turns out to be absolutely correct, but *proving* it is not entirely trivial. Note that, in writing down our formula for f' we have swapped the order of two limits (the limit defining the derivative, and the limit defining the series). We need to justify this. Also, how do we know that the proposed power series for $f'(x)$ even converges on $(-R, R)$? Maybe its radius of convergence is less than R? Given the growth given by the extra factor of n, this seems like a genuine worry.

So, our first task will be to prove that a power series and the obvious candidate for its derivative always have the same radius of convergence. Since we will use this "obvious candidate" series a lot, it's helpful to develop some notation and a name for it.

Definition 8.19 Given a power series $f(x) = \sum_{n=0}^{\infty} a_n x^n$, its **termwise derivative** is the power series

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

Careful! At the moment, termwise **derivative** is just a name: we haven't (yet) established that $\widehat{f}$ is the derivative of f.

Lemma 8.20 *The power series $f(x) = \sum_{n=0}^{\infty} a_n x^n$ and its termwise derivative $\widehat{f}(x) = \sum_{n=1}^{\infty} n a_n x^{n-1}$ have the same radius of convergence.*

Proof: We define the following subsets of $[0, \infty)$,

$$A = \{|x| \ : \ \sum |a_n x^n| \text{ converges}\}, \qquad B = \{|x| \ : \ \sum |n a_n x^{n-1}| \text{ converges}\},$$

and recall that $\sup A$ and $\sup B$ are the radii of convergence of f and $\widehat{f}$, respectively. We will prove that

(i) for all x with $|x| < \sup B$, $f(x)$ converges absolutely (so $|x| \in A$, and hence $\sup A \geq \sup B$), and

(ii) for all x with $|x| < \sup A$, $\widehat{f}(x)$ converges absolutely (so $|x| \in B$, and hence $\sup B \geq \sup A$).

It follows immediately that $\sup A = \sup B$, which is what we seek to prove.
(i) Let $x \in \mathbb{R}$ with $|x| < \sup B$. Then $s_k = \sum_{n=1}^{k} |n a_n x^{n-1}|$ converges (Theorem 8.11), and hence is bounded above, and

$$t_k = \sum_{n=0}^{k} |a_n x^n| = |a_0| + |x| \sum_{n=1}^{k} |a_n x^{n-1}| \leq |a_0| + |x| s_k.$$

So (t_k) is increasing and bounded above, and hence converges, by the Monotone Convergence Theorem. Hence, $|x| \in A$.
(ii) Let $x \in \mathbb{R}$ with $|x| < \sup A$. If $x = 0$ then $\widehat{f}(x)$ certainly converges, so we may assume $|x| > 0$. Choose $\rho \in (|x|, \sup A)$. Then $t_k = \sum_{n=0}^{k} |a_n| \rho^n$ converges (Theorem 8.11), and hence is bounded above. Now

$$s_k = \sum_{n=1}^{k} |n a_n x^{n-1}| = |x|^{-1} \sum_{n=1}^{k} n \left(\frac{|x|}{\rho} \right)^n |a_n| \rho^n.$$

We know that the sequence $n(|x|/\rho)^n \to 0$, since $(|x|/\rho) < 1$ (exponentials beat polynomials), so it must be bounded, that is, there exists $K > 0$ such that, for all n, $n(|x|/\rho)^n \leq K$. But then

$$s_k \leq |x|^{-1} \sum_{n=0}^{k} K |a_n| \rho^n = \frac{K}{|x|} t_k.$$

Hence, s_k is increasing and bounded above, so converges, by the Monotone Convergence Theorem. Hence, $|x| \in B$. $\square$

We now show that the termwise derivative $\widehat{f}(x)$ really is the derivative of $f(x)$. We do this by applying Theorem 7.16 to the sequence of partial sums of $f(x)$. It's a good idea to read the statement of Theorem 7.16 carefully before attempting to understand the next proof.

Theorem 8.21 *Suppose that* $f(x) = \displaystyle\sum_{n=0}^{\infty} a_n x^n$ *has radius of convergence* $R > 0$. *Then* f *is differentiable on* $(-R, R)$, *and*

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

Proof: Choose and fix $\rho \in (0, R)$, and consider the sequence of functions

$$f_k : [-\rho, \rho] \to \mathbb{R}, \qquad f_k(x) = \sum_{n=0}^{k} a_n x^n.$$

This is a sequence of continuously differentiable functions which converges pointwise to f (Theorem 8.11), and its sequence of derivatives

$$f_k' : [-\rho, \rho] \to \mathbb{R}, \qquad f_k'(x) = \sum_{n=1}^{k} n a_n x^{n-1}$$

converges uniformly to $\widehat{f}$ (Lemma 8.20 and Theorem 8.17). Hence, by Theorem 7.16, f is differentiable and $f' = \widehat{f}$. This holds for all $x \in [-\rho, \rho]$, but $\rho \in (0, R)$ was arbitrary, so holds for all $x \in (-R, R)$. $\square$

This is the most important theorem in this chapter. It says that the derivative of a power series exists on its open interval of convergence and is just the power series obtained by termwise differentiation.

Since $f'(x)$ is also a power series with radius of convergence R, we can apply Theorem 8.21 to $f'(x)$ and deduce that $f : (-R, R) \to \mathbb{R}$ is actually *twice* differentiable. Further, $f''(x)$ also has radius of convergence R, so is differentiable on $(-R, R)$, that is, f is *three times* differentiable. In fact, we can keep applying Theorem 8.21 as often as we like, and we conclude that f is *smooth*:

Corollary 8.22 *Let* $f(x) = \sum_{n=0}^{\infty} a_n x^n$ *have radius convergence* $R > 0$. *Then* $f : (-R, R) \to \mathbb{R}$ *is a smooth function, and its* k^{th} *derivative is given by the power series*

$$f^{(k)}(x) = \sum_{n=k}^{\infty} \frac{n!}{(n-k)!} a_n x^{n-k},$$

which also has radius of convergence R. *In particular,*

$$a_k = \frac{f^{(k)}(0)}{k!}$$

for all $k \geq 0$.

Proof: For each integer $k \geq 0$, define

$$g_k(x) = \sum_{n=0}^{\infty} \frac{(n+k)!}{n!} a_{n+k} x^n.$$

I claim that each power series $g_k(x)$ has radius of convergence R and that, for all $x \in (-R, R)$, $f^{(k)}(x) = g_k(x)$. We prove this by induction on k.

Certainly the claim holds for $k = 0$, since $g_0 = f$. So, assume that the claim holds for some value $k \geq 0$, and consider g_{k+1}. Defining coefficients $b_n = \frac{(n+k)!}{n!} a_{n+k}$ so that $g_k(x) = \sum_{n=0}^{\infty} b_n x^n$, then

$$\begin{aligned} g_{k+1}(x) &= \sum_{n=0}^{\infty} \frac{(n+k+1)!}{n!} a_{n+k+1} x^n = \sum_{m=1}^{\infty} \frac{(m+k)!}{(m-1)!} a_{m+k} x^{m-1} \\ &= \sum_{m=1}^{\infty} m b_m x^{m-1}. \end{aligned}$$

Hence, by Lemma 8.20, g_{k+1} has the same radius of convergence as g_k, and by Theorem 8.21, $g_{k+1}(x)$ coincides with the derivative of g_k. But, by our induction hypothesis, g_k has radius of convergence R and coincides with $f^{(k)}$, so g_{k+1} has radius of convergence R and coincides with $f^{(k+1)}$. Hence, if the claim holds for some $k \geq 0$, it also holds for $k + 1$. Hence, by induction, the claim holds for all integers $k \geq 0$.

It follows that $f^{(k)}(0) = g_k(0) = k! a_k$, which completes the proof. □

Corollary 8.22 applies to any power series with non-zero radius of convergence so, in particular, it applies to the series defining exp, sin, and cos.

Proposition 8.23 *The functions* $\exp : \mathbb{R} \to \mathbb{R}$*,* $\sin : \mathbb{R} \to \mathbb{R}$*,* $\cos : \mathbb{R} \to \mathbb{R}$ *defined in Definition 8.13 are smooth, and their derivatives are*

$$\exp' = \exp, \qquad \sin' = \cos, \qquad \cos' = -\sin.$$

Proof: That these functions are smooth follows immediately from Corollary 8.22. Furthermore, by Theorem 8.21

$$\exp'(x) = \sum_{n=1}^{\infty} n \left(\frac{1}{n!}\right) x^{n-1} = \sum_{n=1}^{\infty} \frac{x^{n-1}}{(n-1)!} = \sum_{m=0}^{\infty} \frac{x^m}{m!} = \exp(x)$$

for all $x \in \mathbb{R}$, so $\exp' = \exp$. The formulae for $\sin'$ and $\cos'$ follow from similar arguments. □

With a bit of ingenuity, we can often use power series to exactly sum (that is, exactly compute the limit of) a given series. Here's a simple example.

Example 8.24 Claim: $\displaystyle\sum_{n=1}^{\infty} \frac{n}{7^n} = \frac{7}{36}$.

Proof: The geometric series $f(x) = \sum_{n=0}^{\infty} x^n$ has radius of convergence $R = 1$ and, for all $x \in (-1, 1)$,

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

Hence, by Theorem 8.21, the power series

$$\sum_{n=1}^{\infty} nx^{n-1}$$

also has radius of convergence $R = 1$ and, for all $x \in (-1, 1)$ converges to

$$f'(x) = \frac{1}{(1-x)^2}.$$

Evaluating this at $x = 1/7$ yields

$$\sum_{n=1}^{\infty} \frac{n}{7^{n-1}} = f'(1/7) = \left(\frac{7}{6}\right)^2.$$

Hence

$$\sum_{n=1}^{\infty} \frac{n}{7^n} = \frac{1}{7} \sum_{n=1}^{\infty} \frac{n}{7^{n-1}} = \frac{7}{36}.$$

□

Exercise 8.25 Exactly sum the series $\displaystyle\sum_{n=1}^{\infty} \frac{n^2}{2^n}$. Rigorously justify your answer (of course!).

8.5 Properties of the exponential function

In this section we will prove that the function $\exp : \mathbb{R} \to \mathbb{R}$, defined as a power series, has all the properties you're familiar with. We begin with the following fundamental identity.

Lemma 8.26 *For all* $x, y \in \mathbb{R}$, $\exp(x + y) = \exp(x)\exp(y)$.

Proof: For each fixed number $b \in \mathbb{R}$, define the function $f : \mathbb{R} \to \mathbb{R}$,

$$f(x) = \exp(x)\exp(b - x).$$

This is a product of differentiable functions, so is differentiable, and by the Product and Chain Rules, and Proposition 8.23,

$$f'(x) = \exp'(x)\exp(b - x) - \exp(x)\exp'(b - x) = 0.$$

Since $f'(x) = 0$ for all x in the interval $\mathbb{R}$, it follows that f is constant (Proposition 4.11). Hence, for all $x \in \mathbb{R}$ and any $b \in \mathbb{R}$, $f(x) = f(b)$, that is,

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

Applying this in the case where $b = x + y$ establishes the claim. □

Lemma 8.26 explains why $\exp(x)$ is often denoted e^x and thought of as the constant $e = \exp(1) = \sum_{n=0}^{\infty} \frac{1}{n!}$ "raised to the power x". The point is that, written in this alternative notation, Lemma 8.26 looks like one of the standard algebraic rules of integer exponents:

$$e^{x+y} = e^x \times e^y \qquad \text{in analogy with} \qquad a^{n+m} = a^n \times a^m.$$

This analogy, and the associated notation, are a useful mnemonic, but it's important to realize that they are just that, an analogy. We don't literally *define* $\exp(x)$ to be the irrational number e "raised to the power x", for obvious reasons: what on earth does it mean to raise e to the power $\sqrt{2}$, for example? In fact, the logic works in the opposite direction. Having proved that the function exp, defined as a power series, satisfies a rule analogous to the behaviour of integer (and rational) exponents, we *define* e^x, for any exponent x (rational or irrational) to be $\exp(x)$.

Proposition 8.27 *For all $x \in \mathbb{R}$, $\exp(x) > 0$.*

Proof: We first note that, for all $x \in \mathbb{R}$, $\exp(x) \neq 0$ since, if there exists $x \in \mathbb{R}$ such that $\exp(x) = 0$, then Lemma 8.26 in the case $y = -x$ implies that

$$1 = \exp(0) = 0 \times \exp(-x) = 0.$$

Now $\exp(0) = 1 > 0$, and exp is differentiable, hence continuous, so it follows from the Intermediate Value Theorem that $\exp(x) > 0$ for all $x \in \mathbb{R}$ (if $\exp(x) < 0$ then there exists c between x and 0 such that $\exp(c) = 0$). □

So $\exp : \mathbb{R} \to (0, \infty)$. I claim that exp is bijective and that $\ln : (0, \infty) \to \mathbb{R}$ is its inverse function. Recall (Definition 6.6) that

$$\ln : (0, \infty) \to \mathbb{R}, \qquad \ln x := \int_1^x \frac{1}{t} dt.$$

Proposition 8.28 $\exp : \mathbb{R} \to (0, \infty)$ *is bijective and* $\ln : (0, \infty) \to \mathbb{R}$ *is its inverse function.*

Proof: We must show that, for all $x \in \mathbb{R}$, $\ln(\exp(x)) = x$, and for all $y \in (0, \infty)$, $\exp(\ln y) = y$. Consider the function

$$f : \mathbb{R} \to \mathbb{R}, \qquad f(x) = \ln(\exp(x)) - x.$$

By the Chain Rule,

$$f'(x) = \frac{\exp(x)}{\exp(x)} - 1 = 0$$

for all x, so f is constant (Corollary 4.11). Hence, for all x, $f(x) = f(0) = \ln 1 - 0 = 0$, that is, $\ln(\exp(x)) = x$. Now, for all $y \in (0, \infty)$,

$$\ln(\exp(\ln y)) = \ln y,$$

as we have just shown. But ln is injective, so if $\ln z = \ln y$ then $z = y$. Hence $\exp(\ln y) = y$, as was to be shown.

It follows immediately that exp is surjective:

$$\forall y \in (0, \infty), \quad \exp(\ln(y)) = y,$$

and injective:

$$\exp(x) = \exp(x') \quad \Rightarrow \quad \ln(\exp(x)) = \ln(\exp(x')) \quad \Rightarrow \quad x = x'.$$

$\square$

So, for each positive number y, $\ln y$ is the real number whose exponential is y. Recall that *Euler's number* is

$$e = \exp(1) = \sum_{n=0}^{\infty} \frac{1}{n!}.$$

It follows that $\ln e = 1$. There is an alternative way to define e, as the limit of the sequence $(1 + \frac{1}{n})^n$. You are now in a position to prove this.

Exercise 8.29 Prove that the sequence $x_n = \left(1 + \dfrac{1}{n}\right)^n$ converges to e.

Solution: Let

$$y_n = \ln x_n = n \ln\left(1 + \frac{1}{n}\right) = \frac{\ln\left(1 + \frac{1}{n}\right) - \ln 1}{\left(1 + \frac{1}{n}\right) - 1}.$$

Now ln is differentiable at 1 with derivative $\ln'(1) = 1/1 = 1$, that is,

$$\lim_{s \to 1} \frac{\ln s - \ln 1}{s - 1} = 1.$$

Hence, by the sequential criterion for limits (Theorem 2.14), given any sequence s_n in $(0, \infty) \backslash \{1\}$ which converges to 1,

$$\frac{\ln s_n - \ln 1}{s_n - 1} \to 1.$$

Applying this in the case $s_n = 1 + 1/n$, we see that $y_n \to 1$. But

$$x_n = \exp y_n$$

by Proposition 8.28, and exp is differentiable, hence continuous, so $x_n \to \exp(1)$, as was to be proved.

Proposition 8.28 implies that, for all $a > 0$, $\ln a$ is the unique real number y such that $\exp(y) = a$. This motivates the following definition:

Definition 8.30 Let $a > 0$ and $x \in \mathbb{R}$. Then a **to the power** x is

$$a^x = \exp(x \ln a).$$

It follows from Proposition 6.8 that this definition of a^x coincides with the more obvious definition of a^x in the case where x is an integer.

So now we can make sense of the function $f(x) = x^r$ for all $x > 0$ and any constant $r \in \mathbb{R}$ (even if r is irrational!). It is not hard to prove that $f : (0, \infty) \to \mathbb{R}$ is differentiable, and has the derivative we expect.

Proposition 8.31 *Let $f : (0, \infty) \to \mathbb{R}$ be the function $f(x) = x^r$, where r is any real constant. Then f is differentiable, and for all $x \in (0, \infty)$,*

$$f'(x) = rx^{r-1}.$$

Proof: Exercise. (Use the Chain Rule on $f(x) = \exp(r \ln x)$.) □

8.6 Analyticity versus smoothness

So far we have only considered power series based at 0, but the whole theory extends immediately to power series with some other base point $x_0 \in \mathbb{R}$, that is, series of the form

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

The point is that $f(x) = g(x - x_0)$, where $g(x) = \sum_{n=0}^{\infty} a_n x^n$, so (by the Chain Rule) f is smooth on $(x_0 - R, x_0 + R)$ where R is the radius of convergence of g, and the derivative of f is again its termwise derivative. We will refer to R as the radius of convergence of f.

A function f is **analytic** if it locally coincides with some convergent power series, everywhere in its domain. More precisely:

Definition 8.32 Let $U \subseteq \mathbb{R}$ be open. Then $f : U \to \mathbb{R}$ is **analytic** if, for each $x_0 \in U$, there exists $\varepsilon > 0$ and a real sequnce (a_n) such that, for all $x \in (x_0 - \varepsilon, x_0 + \varepsilon)$,

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

Example 8.33 $\exp : \mathbb{R} \to \mathbb{R}$ is analytic since

$$\exp(x) = \exp(x_0) \exp(x - x_0) = \sum_{n=0}^{\infty} \frac{\exp(x_0)}{n!} (x - x_0)^n$$

for all $x, x_0 \in \mathbb{R}$ (so for any x_0 we can take any $\varepsilon > 0$ in the definition above).

Exercise 8.34 Prove that $f : \mathbb{R}\backslash\{0\} \to \mathbb{R}$, $f(x) = 1/x$, is analytic.

If $f : U \to \mathbb{R}$ is analytic, it is certainly smooth, by Corollary 8.22 and the Localization Lemma, since it coincides on an open set $(x_0 - \varepsilon, x_0 + \varepsilon)$ with a convergent power series

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

So all analytic functions are smooth. Furthermore, by Corollary 8.22, we know that the coefficients of this series are

$$a_n = \frac{f^{(n)}(x_0)}{n!}.$$

That is, on $(x_0 - \varepsilon, x_0 + \varepsilon)$, the function f coincides with its **Taylor series**

$$f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n.$$

Now, given *any* smooth function on an open set U and a point x_0, we can compute all the derivative $f^{(n)}(x_0)$, and hence construct its Taylor series. Does it follow that all smooth functions are analytic? Rather remarkably, the answer is **no!** The rest of this section is devoted to constructing a class of functions $f : \mathbb{R} \to \mathbb{R}$ which are smooth but **not** analytic.

Definition 8.35 Given a polynomial $p(x) = a_0 + a_1 x + \cdots + a_k x^k$, we define the function $f_p : \mathbb{R} \to \mathbb{R}$ such that

$$f_p(x) = \begin{cases} 0, & x \leq 0 \\ p(1/x)\exp(-1/x), & x > 0. \end{cases}$$

We will prove that the function f_p is smooth but not (except in the trivial case $p(x) = 0$) analytic. This follows fairly quickly once we establish that $\lim_{x \to 0} f_p(x) = 0$.

Lemma 8.36 (Exponentials beat powers – sequences) *For any $k \in \mathbb{N}$ the sequence*

$$a_n = n^k \exp(-n) \to 0.$$

Proof:

$$\frac{a_{n+1}}{a_n} = \left(1 + \frac{1}{n}\right)^k \exp(-1) \to \frac{1}{e} < 1$$

so $\sum_{n=1}^{\infty} a_n$ converges by the Ratio Test. Hence $a_n \to 0$ by the Divergence Test. □

Lemma 8.37 (Exponentials beat powers – functions) *For any $k \in \mathbb{N}$,*

$$\lim_{x \to \infty} x^k \exp(-x) = 0.$$

Proof: Let $g(x) = x^k \exp(-x)$. We must show that, for each $\varepsilon > 0$, there exists $K \in \mathbb{R}$ such that, for all $x > K$, $|g(x) - 0| < \varepsilon$.

So, let $\varepsilon > 0$ be given. Since the sequence $g(n) = n^k \exp(-n) \to 0$ (Lemma 8.36), there exists $N \in \mathbb{Z}^+$ such that, for all $n \geq N$, $|g(n) - 0| < \varepsilon$, and hence $0 < g(n) < \varepsilon$. Let $K = \max\{k, N\}$. Then for all $x > K$,

$$g'(x) = (kx^{k-1} - x^k)\exp(-x) = -x^{k-1}(x-k)\exp(-x) < 0,$$

(by the Chain and Product Rules), so g is strictly decreasing (by Proposition 4.11). Hence, for all $x > K$, $0 < g(x) < g(K) < \varepsilon$ (since $K \geq N$). □

Lemma 8.38 *Let $f_p : \mathbb{R} \to \mathbb{R}$ be the function associated to a polynomial p, as in Definition 8.35. Then*

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

Proof: Let $p(x) = a_0 + a_1 x + \cdots + a_k x^k$, and for each $j \in \mathbb{N}$, denote by m_j the monomial of degree j (so $m_j(x) = x^j$). Then

$$f_p = a_0 f_{m_0} + a_1 f_{m_1} + \cdots + a_k f_{m_k}$$

so, by the Algebra of Limits, it suffices to prove the claim in the case where $p(x) = m_k(x) = x^k$, a general monomial.

Let $\varepsilon > 0$ be given. By Lemma 8.37, there exists $K > 0$ such that, for all $x > K$, $|x^k \exp(-x) - 0| < \varepsilon$. Let $\delta = 1/K > 0$. Then, for all $x \in (0, \delta)$,

$$0 \leq f_p(x) = (1/x)^k \exp(-1/x) < \varepsilon,$$

since $1/x > 1/\delta = K$, and for all $x \in (-\delta, 0)$,

$$0 = f_p(x) < \varepsilon.$$

Hence, for all $x \in \mathbb{R}$ with $0 < |x - 0| < \delta$, $|f_p(x) - 0| < \varepsilon$. □

Theorem 8.39 *Let $f_p : \mathbb{R} \to \mathbb{R}$ be the function associated to a polynomial p, as in Definition 8.35. Then f_p is differentiable and $f_p' = f_q$ where $q(x) = x^2(p(x) - p'(x))$.*

Proof: f_p coincides with the differentiable function $p(1/x)\exp(-1/x)$ on the open set $(0, \infty)$, so is differentiable on $(0, \infty)$ with derivative

$$f'(x) = p'(x^{-1})(-x^{-2})\exp(-x^{-1}) + p(x^{-1})\exp(-x^{-1})(x^{-2}) = q(x^{-1})\exp(-x^{-1})$$

by the Chain and Product Rules, and the Localization Lemma (Lemma 3.23). Similarly, f_p coincides with the differentiable function 0 on the open set $(-\infty, 0)$, so is differentiable on $(-\infty, 0)$ with derivative 0 by the Localization Lemma. It remains to show that f_p is differentiable at 0 and $f_p'(0) = 0$. That is, we must prove that

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

But $f_p(x)/x = f_P(x)$ where $P(x) = xp(x)$, so this follows immediately from Lemma 8.38. □

Corollary 8.40 *The function f_p defined in Definition 8.35 is smooth and has $f_p^{(n)}(0) = 0$ for all $n \geq 0$.*

Proof: Denote by X the set of functions $\{f_p : p \text{ a polynomial}\}$. Then for all $f_p \in X$, $f_p(0) = 0$ (by definition), f_p is differentiable, and $f_p' \in X$ (Theorem 8.39). Hence every element of X is infinitely differentiable and has all its derivatives zero at 0. □

It follows that almost every function f_p is smooth, but not analytic. (The only exception is f_0, where we take $p(x)$ to be the trivial polynomial $p = 0$. Clearly $f_0 \equiv 0$ which is trivially analytic.) Let's consider the simplest example.

Example 8.41 Consider the special case $p(x) = 1$. The corresponding function is

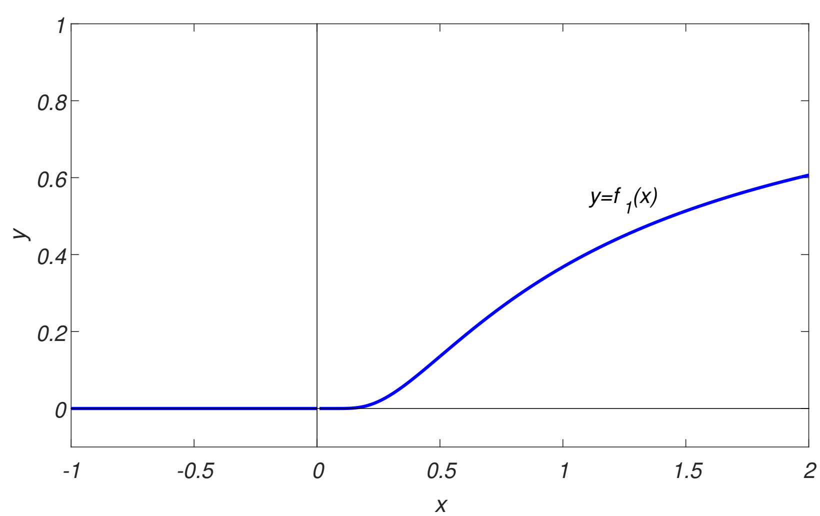
$$f(x) = \begin{cases} \exp(-1/x), & x > 0 \\ 0, & x \leq 0. \end{cases}$$

By Corollary 8.40, this is smooth and all its derivatives at 0 are 0. I claim that f is **not** analytic.

Assume, towards a contradiction, that f *is* analytic. Then there exist $\varepsilon > 0$ and $\{a_n : n \in \mathbb{N}\}$ such that

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

for all $x \in (-\varepsilon, \varepsilon)$. But then, by Corollary 8.22, $a_n = f^{(n)}(0)/n! = 0$. Hence $f(x) = 0$ for all $x \in (-\varepsilon, \varepsilon)$. But $f(\varepsilon/2) = \exp(-2/\varepsilon) > 0$, a contradiction.



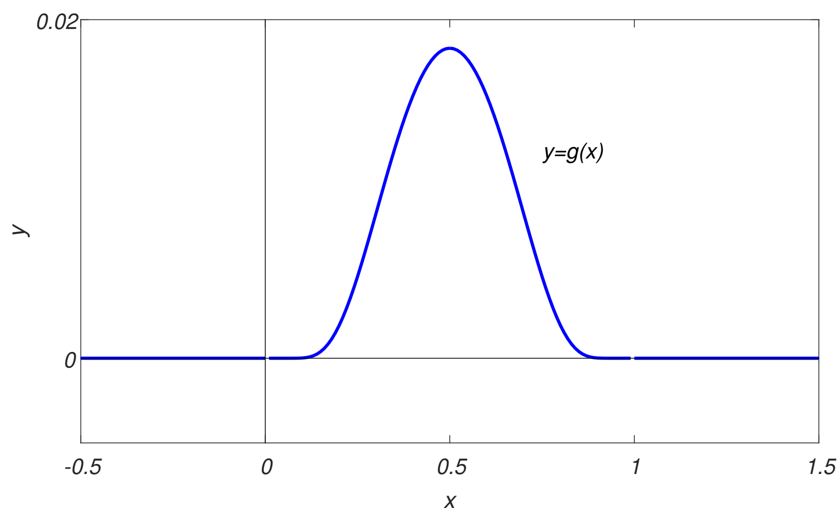
□

It's not hard to construct from Example 8.41 smooth functions which vanish exactly outside some bounded interval (a, b) but are strictly positive on (a, b). Such functions are often called *bump functions*.

Example 8.42 Given an interval $[a,b]$ define $g:\mathbb{R}\to\mathbb{R}$ by

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

where f is the smooth function of Example 8.41. Then g is smooth by the product and chain rules. If $x\le a$, $x-a\le 0$ so $f(x-a)=0$. If $x\ge b$, $b-x\le 0$ so $f(b-x)=0$. Hence, $g(x)=0$ for all $x\notin(a,b)$. If $x\in(a,b)$, both $x-a$ and $b-x$ are positive, so $f(x-a)>0$ and $f(b-x)>0$, whence $g(x)>0$. So g is strictly positive on (a,b). Here's a plot of g in the case $[a,b]=[0,1]$:



Summary

- A **power series** is a series of the form $\sum_{n=0}^{\infty} a_n x^n$ where x is a real variable. Its **radius of convergence** is

$$R = \sup\{|x| \ : \ \textstyle\sum_{n=0}^{\infty} a_n x^n \text{ converges absolutely}\}.$$

- If a power series has radius of convergence R, it converges absolutely for all $|x| < R$ and diverges for all $|x| > R$.

- The **Weierstrass M Test**: if $|g_n(x)| \leq M_n$ for all $x \in D$ and the series $\sum_{n=0}^{\infty} M_n$ converges, then the series $\sum_{n=0}^{\infty} g_n(x)$ converges uniformly on D.

- Using the M test, we can prove that a power series with radius of convergence $R > 0$ converges **uniformly** on every interval $[-\rho, \rho]$, where $\rho \in (0, R)$.

- A power series $f(x) = \sum_{n=0}^{\infty} a_n x^n$ with radius of convergence $R > 0$ is **differentiable** on $(-R, R)$ and its derivative is

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

  This power series also has radius of convergence R.

- Convergent power series are, in fact, **smooth**.

- A function $f : U \to \mathbb{R}$, where $U \subseteq \mathbb{R}$ is an open set, is **analytic** if for each $x_0 \in U$ there is $\varepsilon > 0$ and a power series $\sum_{n=0}^{\infty} a_n (x - x_0)^n$ which converges to $f(x)$ for all $x \in (x_0 - \varepsilon, x_0 + \varepsilon)$.

- All analytic functions are smooth, but **not all smooth functions are analytic**. An example of a smooth but non-analytic function is

$$f : \mathbb{R} \to \mathbb{R}, \qquad f(x) = \begin{cases} \exp(-1/x), & x > 0 \\ 0, & x \leq 0. \end{cases}$$