Math40002: Analysis I

We will pick up where MATH40002 left off in the Autumn term, rigorously developing the basic tools of calculus: continuity, differentiation, and integration.

We will continue to have infinite fun.

Syllabus

Continuity: Review of continuity. Sequential continuity. Uniform continuity. Intermediate and extreme value theorems. Inverse function theorem for monotonic functions.

Differentiation: Definitions, examples, and properties. Mean value theorem. Higher derivatives and convexity. Differentiation of series.

Integration: Definition, examples, and properties of Riemann–Darboux integral. Fundamental theorem of calculus. Techniques: integration by parts, substitution.

Books

Martin Liebeck, *A Concise Introduction to Pure Mathematics.*
Mary Hart, *Guide to Analysis.*

KG Binmore, *Mathematical Analysis, A Straightforward Approach.*
David Brannan, *A first course in mathematical analysis.*
Steven Lay, *Analysis: with an introduction to proof.*
Stephen Abbott, *Understanding analysis.*

I will post gappy notes on Blackboard, and a complete version at the end of term. The course discussion site (`http://tinyurl.com/DiscussMATH40002`) will remain active, and I encourage you to use it, though participation will not count for marks.

Assessment

20% Autumn term assessments, including the January test
4% Mini Blackboard tests every few weeks (due 27 Jan, 10 Feb, 02 Mar, 16 Mar)
5% Midterm test – Thursday, 13 February, 2pm
1% Formal write-up of a midterm problem, due 21 February
70% May exam

Contents

1 Continuity

We begin by recalling what it means for a function to be continuous.

Definition. Given a function $f : \mathbb{R} \to \mathbb{R}$, we say that f is *continuous* at $a \in \mathbb{R}$ if and only if

$$\forall \epsilon > 0 \ \exists \delta > 0 \text{ such that } |x - a| < \delta \implies |f(x) - f(a)| < \epsilon$$

We say that f is continuous on $\mathbb{R}$ (or just "continuous") if it is continuous at all $a \in \mathbb{R}$.

We also saw the equivalent notion of *sequential continuity*, which is sometimes more convenient.

Theorem 1.1

$f : \mathbb{R} \to \mathbb{R}$ is continuous at $a \in \mathbb{R} \iff f(x_n) \to f(a)$ $\forall$ sequences $x_n \to a$.

Let's warm up by proving some basic properties of continuity.

Proposition 1.2. *Let $f, g : \mathbb{R} \to \mathbb{R}$ be functions which are both continuous at $a \in \mathbb{R}$. Then the functions $f + g$, $f - g$, and $f \cdot g$ are all continuous at a, and if $g(a) \neq 0$ then $\frac{f}{g}$ is also continuous at a.*

Proof. We'll make use of sequential continuity here. For any sequence $x_n \to a$, we have $f(x_n) \to f(a)$ and $g(x_n) \to g(a)$ by hypothesis. The algebra of limits tells us that

$$\lim_{n\to\infty} \big(f(x_n) + g(x_n)\big) = \lim_{n\to\infty} f(x_n) + \lim_{n\to\infty} g(x_n) = f(a) + g(a).$$

Since this works for *every* sequence $x_n \to a$, we conclude that $f + g$ is continuous.

The same argument works for $f - g$ and $f \cdot g$, and also for $\frac{f}{g}$ if $g(a) \neq 0$. □

Now we can start to assemble a big collection of continuous functions almost for free – no need for epsilons and deltas! We know that linear functions $f(x) = mx + c$ are continuous, so we get

- any function $f(x) = x^n$ is continuous, where $n \geq 0$ is an integer;
- any **polynomial** $p(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$ is continuous, where the a_i are real numbers;
- any **rational function** $\frac{p(x)}{q(x)}$, where p and q are polynomials, is continuous at all $a \in \mathbb{R}$ where $q(a) \neq 0$.

The first two can be proved by induction on n (try it!), and the third follows from the second.

Similarly, we saw last term that $E : \mathbb{C} \to \mathbb{C}$ defined by $E(z) = \sum_{n=0}^{\infty} \frac{z^n}{n!}$ is continuous on $\mathbb{C}$. Since $x \mapsto \sin(x)$ and $x \mapsto \cos(x)$ can be defined in terms of E (how?), this tells us where the **trigonometric functions** are continuous:

- $\sin(x)$ and $\cos(x)$ are continuous on $\mathbb{R}$;
- $\tan(x) = \frac{\sin(x)}{\cos(x)}$ is continuous wherever $\cos(x) \neq 0$, meaning at all x *except* $x = \frac{(2k+1)\pi}{2}$, $k \in \mathbb{Z}$;
- likewise, $\sec(x)$, $\csc(x)$, $\cot(x)$ are continuous at any x where $\cos(x) \neq 0$, $\sin(x) \neq 0$, and $\sin(x) \neq 0$ respectively.

Proposition 1.3. *If $f : \mathbb{R} \to \mathbb{R}$ is continuous at a, and $g : \mathbb{R} \to \mathbb{R}$ is continuous at $f(a)$, then the composition $g \circ f$ defined by $x \mapsto g(f(x))$ is continuous at a.*

Proof. Let x_n be any sequence such that $x_n \to a$. Since f is continuous at a, we have $f(x_n) \to f(a)$. Now we can plug the convergent sequence $y_n = f(x_n)$, with limit $y = f(a)$, into g: since g is continuous at y, we have $g(y_n)) \to g(y)$. But this is just another way of saying that $g(f(x_n)) \to g(f(a))$. Again, this works for any sequence $x_n \to a$, so $g(f(x))$ must be continuous at a. $\square$

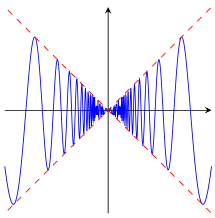
Example 1.4. We know that $E : \mathbb{C} \to \mathbb{C}$ defined by $E(z) = \sum_{n=0}^{\infty} \frac{z^n}{n!}$ is continuous on $\mathbb{C}$, and we saw that this implies that $\sin(x)$ is continuous on $\mathbb{R}$. So now

$$f(z) = \sin\left(\frac{z^2+1}{z-2}\right)$$

is continuous at all $a \in \mathbb{R}$ where $\frac{z^2+1}{z-2}$ is continuous, meaning at all $a \neq 2$.

Example 1.5. Consider the function $f : \mathbb{R} \to \mathbb{R}$ defined by

$$f(x) = \begin{cases} x\sin(1/x), & x \neq 0 \\ 0, & x = 0. \end{cases}$$



Since $\frac{1}{x}$ is a rational function, it is continuous at all $x \neq 0$. Thus

- $\sin(\frac{1}{x})$ is continuous at all $x \neq 0$, since $x \mapsto \sin(x)$ is continuous;
- $x\sin(\frac{1}{x})$ is continuous at all $x \neq 0$, since both x and $\sin(\frac{1}{x})$ are;

and finally we saw last term that f is continuous at $x = 0$ as well, so f is continuous on all of $\mathbb{R}$.

Mentimeter question 1. Let $f, g : \mathbb{R} \to \mathbb{R}$ be functions, and let $h(x) = g(f(x))$. Which of the following is always true?

1. If f and g are not continuous, then h is not continuous.
2. If f is continuous and g is not, then h is not.
3. If f is not continuous but g is, then h is.
4. More than one of choices 1, 2, 3.
5. None of choices 1, 2, 3. ✓

The easiest thing to try would be functions that are constant, or maybe piecewise constant, to see what happens. For item 1, we might try something like

$$f(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0 \end{cases} \qquad \text{and} \qquad g(x) = \begin{cases} 0, & -2 \leq x \leq 2 \\ 1, & |x| > 2 \end{cases}$$

so that $g(f(x)) = 0$ is constant. For 2, we take $f = 0$ and any discontinuous g, and then $h = f(0)$ is constant. For 3, we take $g(x) = x$ and any discontinuous f, and then $h = f$ is discontinuous.

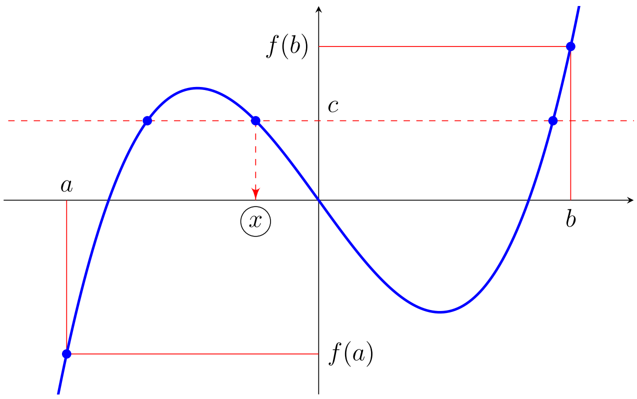
1.1 The intermediate value theorem

We now come to the first major application of continuity.

Theorem 1.6: Intermediate value theorem

Let $f : [a, b] \to \mathbb{R}$ be a continuous function, and pick a value c between $f(a)$ and $f(b)$. Then there is some $x \in [a, b]$ such that $f(x) = c$.

Here's a picture of what this might look like:



Note that there might be several possible values of x; in this picture there are three.

This theorem certainly sounds obvious – a continuous function can't jump, so it can't skip any values – but that doesn't mean it's easy to prove! We wouldn't even know where to begin if we hadn't already built up the notions of limits and continuity from scratch.

Proof. We can assume without loss of generality that $f(a) < c < f(b)$. Indeed:

- if $f(a) = c$ we take $x = a$, and if $f(b) = c$ then we take $x = b$;
- if $f(a) = f(b)$ then $c = f(a)$ and we are already done; and
- if $f(a) > f(b)$ then we can take $g(x) = -f(x)$, so that $g(a) < g(b)$, and ask for x with $g(x) = -c$ instead.

So now we consider the set

$$S_c = \{y \in [a, b] \mid f(y) < c\}.$$

This set is nonempty, because $a \in S_c$, and it is bounded above by b. Thus if we let $x = \sup S_c$ then we have $a \leq x \leq b$.

Claim 1: $f(x) \geq c$.

Let's suppose this claim is false, so $f(x) < c$, and take $\epsilon = c - f(x) > 0$. Note that if $f(x) < c$ then $x \neq b$, so we must have $x < b$ here. Since f is continuous at x, we have

$$\exists \delta > 0 \text{ such that } |f(y) - f(x)| < \epsilon \ \forall y \in (x - \delta, x + \delta) \cap [a, b].$$

In particular, this means that

$$f(y) < f(x) + \epsilon = c \ \forall y \in (x - \delta, x + \delta) \cap [a, b].$$

So all of these y belong to S_c, and if we choose $y \in (x, x + \delta) \cap [a, b]$, say

$$y = x + \frac{1}{2}\min(\delta, b - x),$$

then we have $y > x$, so x isn't actually an upper bound for S_c, contradiction.

Claim 2: $f(x) \leq c$.

Supposing again that this claim is false, so that $f(x) > c$ (and hence $x \neq a$), we take $\epsilon = f(x) - c > 0$ and use the continuity of f at x to find $\delta > 0$ such that

$$|f(y) - f(x)| < \epsilon \ \forall y \in (x - \delta, x + \delta) \cap [a, b].$$

For any such y we have $f(y) > f(x) - \epsilon = c$, so none of these y are in S_c. But then

$$m = \max\left(x - \frac{\delta}{2}, a\right)$$

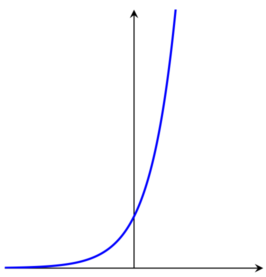
is strictly less than x, and it is also an upper bound for S_c, since x is an upper bound and no $y \in (m, x)$ belongs to S_c. This contradicts the fact that x is the *least* upper bound of S_c.

We have now shown that $f(x) \geq c$ and $f(x) \leq c$, so together these tell us that $f(x) = c$ and we are done. $\square$

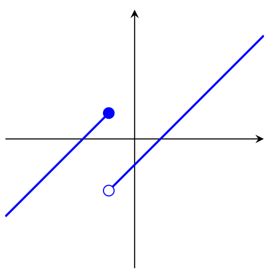
Mentimeter question 2. Let $f : \mathbb{R} \to \mathbb{R}$. Which of the following is true?

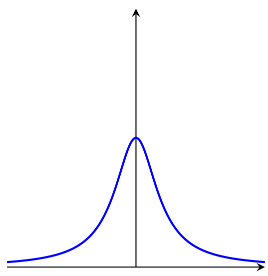
1. If f is continuous, then $\forall c \in \mathbb{R}\ \exists x$ such that $f(x) = c$.
2. If f is not continuous, then $\exists c \in \mathbb{R}$ such that $\forall x\ f(x) \neq c$.
3. If f is continuous and $f(-1) = -1$, $f(0) = 2$, and $f(3) = -5$, then there is more than one $x \in \mathbb{R}$ such that $f(x) = 0$. $\checkmark$
4. If f is continuous and $\lim_{x\to\infty} f(x) = 0$, then $\exists x \in \mathbb{R}$ such that $f(x) = 0$.

Here are pictures of some counterexamples to choices 1, 2, and 4:



$f(x) = \exp(x)$





$f(x) = 1/(x^2 + 1)$

As an application, the **fundamental theorem of algebra** says that every non-constant polynomial

$$p(x) = a_d x^d + a_{d-1}x^{d-1} + \cdots + a_1 x + a_0, \qquad d \geq 1,\ a_d \neq 0$$

has a root: there is some x such that $p(x) = 0$. We won't prove this here (it's one of the highlights of complex analysis!), but we can prove a special case.

Proposition 1.7. *If $p(x)$ is a polynomial of* odd *degree, then $p(x)$ has a root.*

Proof. We can divide both sides of $p(x) = 0$ by a_d, so we might as well assume that $a_d = 1$ and then try to solve

$$p(x) = x^d + a_{d-1}x^{d-1} + \cdots + a_1 x + a_0 = 0.$$

Polynomials are continuous, so we just need to find some $a, b \in \mathbb{R}$ such that $p(a) < 0$ and $p(b) > 0$, and then we can apply the intermediate value theorem to the interval $[a, b]$, taking $c = 0$, to find x such that $p(x) = c = 0$.

We'll look for b first. The key idea is that for x very large, the x^d term above is much bigger than the others, so that $p(x)$ behaves like x^d in some sense. We can factor

$$p(x) = x^d \left(1 + \frac{a_{d-1}}{x} + \cdots + \frac{a_1}{x^{d-1}} + \frac{a_0}{x^d}\right),$$

and if we call the part in parentheses $q(x)$, then $\lim_{x\to\infty} q(x) = 1$. This means that there's some $N > 0$ such that $q(x) > \frac{1}{2}$ for all $x \geq N$, and then

$$p(x) = x^d q(x) > \frac{x^d}{2} > 0 \quad \forall x \geq N.$$

So we can just take $b = N$, and then $p(b) > 0$.

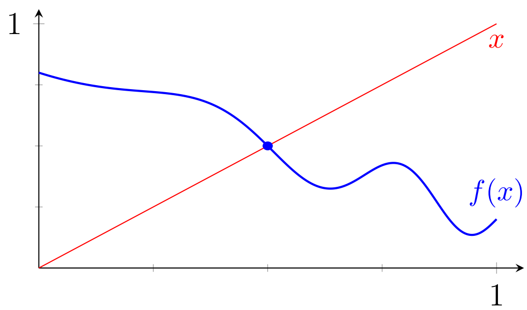
We use the same idea to look for a. We also have $\lim_{x\to-\infty} q(x) = 1$, so there's some $M < 0$ such that $q(x) > \frac{1}{2}$ for all $x \leq M$, and then since $x^d < 0$ for $x \leq M$ we have

$$p(x) = x^d q(x) < \frac{x^d}{2} < 0 \quad \forall x \leq M.$$

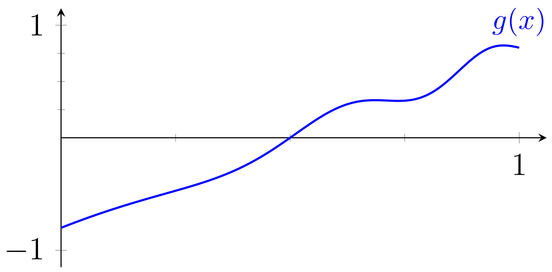
So if we take $a = M$ then we have $p(a) < 0$, and we're done. $\square$

Here's another application of the intermediate value theorem.

Proposition 1.8. *If $f : [0, 1] \to [0, 1]$ is continuous, then it has a* ***fixed point****: there is an $x \in [0, 1]$ such that $f(x) = x$.*



Proof. Rearranging the conclusion we want a little bit, we get $x - f(x) = 0$. So we'll define $g : [0,1] \to \mathbb{R}$ by $g(x) = x - f(x)$, which is also continuous, and try to prove that there's some $x \in [0,1]$ for which $g(x) = 0$.



If we have $f(0) = 0$ or $f(1) = 1$ then there's nothing left to prove, so we can assume from now on that $f(0) > 0$ and $f(1) < 1$. In this case we have

$$g(0) = 0 - f(0) < 0 \qquad \text{and} \qquad g(1) = 1 - f(1) > 0.$$

Since 0 lies between $g(0)$ and $g(1)$, the intermediate value theorem tells us that there's some $x \in [0,1]$ for which $g(x) = 0$, and this means that $f(x) = x$. $\square$

1.2 The extreme value theorem

We wish to study the maxima and minima of continuous functions, so we'll begin with some terminology.

Definition. Let S be a subset of $\mathbb{R}$, and let $f : S \to \mathbb{R}$ be a function. We say that f is *bounded above* if

$$\exists M \in \mathbb{R} \text{ such that } f(x) \le M \ \forall x \in S.$$

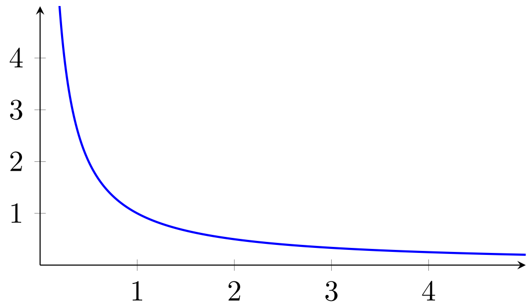
Similarly, f is *bounded below* if

$$\exists m \in \mathbb{R} \text{ such that } f(x) \ge m \ \forall x \in S.$$

If f is both bounded above and bounded below, we say that f is *bounded.*

A function can be continuous but not bounded, and vice versa; we have to pay close attention to the domain where f is defined. Here are some examples:

Example 1.9. The function $f : \{x > 0\} \to \mathbb{R}$ given by $f(x) = \frac{1}{x}$ is bounded below by 0, but not bounded above.



Supposing f were bounded above by some $M \in \mathbb{R}$, we could take

$$x = \frac{1}{\max(2|M|, 1)} \qquad \Longrightarrow \qquad f(x) = \max(2|M|, 1) > M,$$

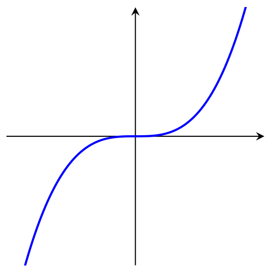
a contradiction. (We have $2|M| > M$ unless $M = 0$, so the "$\max(\cdot, 1)$" takes care of that case.)

Example 1.10. The function $f : \mathbb{R} \setminus \{0\} \to \mathbb{R}$ given by $f(x) = \frac{1}{x}$ is neither bounded above nor bounded below. If it were bounded below by some $m \in \mathbb{R}$, we could take $x = -1/\max(2|m|, 1)$ and get the same contradiction as before.

Example 1.11. Define $f : \mathbb{R} \to \mathbb{R}$ by $f(x) = \begin{cases} 0, & x \in \mathbb{Q} \\ 1, & x \notin \mathbb{Q}. \end{cases}$ Then f isn't continuous anywhere, but it's bounded below by $m = 0$ and above by $M = 1$.

The above examples all seem to involve discontinuities, if not on their domains then right at the boundary. But even continuous functions on all of $\mathbb{R}$ can be unbounded:

Example 1.12. The function $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^3$ is neither bounded above nor bounded below. If it were bounded above by M, then we could take $x = \max(\sqrt[3]{M}, 0) + 1$ and we'd get $f(x) > M$, and the bounded below case is similar.



On the other hand, things look a little better for continuous functions on closed intervals of finite length.

Theorem 1.13

If $f : [a,b] \to \mathbb{R}$ is continuous, then it is bounded.

Proof. Let's prove that f is bounded above. If not, then we can define a sequence

$$x_1, x_2, x_3, \dots$$

by taking $x_n \in [a,b]$ which satisfies $f(x_n) > n$ for all n. The sequence $\big(x_n\big)$ is bounded below by a and bounded above by b, so the Bolzano–Weierstrass theorem says that it has a convergent subsequence

$$x_{n_1}, x_{n_2}, x_{n_3}, \dots$$

for some integers $1 \le n_1 < n_2 < n_3 < \dots$ which go to infinity. We have $a \le x_{n_i} \le b$ for all i, so if we call the limit of this subsequence x, then $a \le x \le b$ as well. (Why?)

Now we use the (sequential) continuity of f and the fact that $x \in [a,b]$ to say that

$$f(x_{n_i}) \to f(x).$$

But by construction we have $f(x_{n_i}) \ge n_i$, and $n_i \to \infty$, so $f(x_{n_i}) \to \infty$ and this is a contradiction.

The proof that f is bounded below is essentially the same. Or we could reduce it to the case we've already done: the function $g(x) = -f(x)$ is continuous on $[a,b]$, hence bounded above by some M, and since $g(x) \le M$ for all $x \in [a,b]$ we must have $f(x) \ge -M$ for all $x \in [a,b]$. □

Theorem 1.14: Extreme value theorem

Let $f : [a,b] \to \mathbb{R}$ be a continuous function. Then f is bounded, and it attains its lower and upper bounds: there are $c, d \in [a,b]$ such that

$$f(c) \le f(x) \le f(d) \qquad \forall x \in [a,b].$$

Equivalently, we have $f(c) = \inf\limits_{x\in[a,b]} f(x)$ and $f(d) = \sup\limits_{x\in[a,b]} f(x)$.

Proof 1. We just proved that f is bounded, so let $S = \sup\limits_{x\in[a,b]} f(x)$. Since S is the *least* upper bound of $\{f(x) \mid x \in [a,b]\}$, we know that $S - \frac{1}{n}$ is *not* an upper bound for any $n \geq 1$, so for each n we know that

$$\exists x_n \in [a,b] \quad \text{such that} \quad S - \frac{1}{n} < f(x_n) \leq S.$$

Again by Bolzano–Weierstrass, the x_n have a convergent subsequence x_{n_i}, with some limit $d \in [a,b]$. Since f is continuous we have

$$f(x_{n_i}) \to f(d),$$

and the sequence $f(x_{n_i})$ is squeezed between $S - \frac{1}{n_i} \to S$ and S, so its limit $f(d)$ must be S.

Likewise, if we let $I = \inf\limits_{x \in [a,b]} f(x)$, then for all n we can find elements $x_n \in [a,b]$ such that $I \leq f(x_n) < I + \frac{1}{n}$. We take c to be the limit of a convergent subsequence, and it follows just as before that $f(c) = I$. ☐

Proof 2. Once again f is bounded above, so we let $S = \sup\limits_{x \in [a,b]} f(x)$. If we assume that $f(x) < S$ for all $x \in [a,b]$, then we have a well-defined function

$$g : [a,b] \to \mathbb{R}, \qquad g(x) = \frac{1}{S - f(x)}$$

which takes only positive values since $S - f(x) > 0$ for all x. We know that g is continuous, so g is bounded as well.

Let $M > 0$ be an upper bound for g. Then we do some rearranging to get

$$\begin{aligned} g(x) = \frac{1}{S - f(x)} \leq M \qquad &\Longleftrightarrow \qquad \frac{1}{M} \leq S - f(x) \\ &\Longleftrightarrow \qquad f(x) \leq S - \frac{1}{M}. \end{aligned}$$

So $S - \frac{1}{M}$ is also an upper bound for $\{f(x) \mid a \leq x \leq b\}$, contradicting the assumption that S was the least upper bound. It must have been true that $f(x) = S$ for some $x \in [a,b]$ after all, and we take d to be that value of x. (And again, the proof that f attains its infimum is basically the same.) ☐

We can combine the intermediate and extreme value theorems to understand the images of continuous functions.

Proposition 1.15. *Let $f : [a,b] \to \mathbb{R}$ be a continuous function. Then there are $c, d \in [a,b]$ such that the image $f([a,b])$ is the closed interval $[f(c), f(d)]$.*

Proof. By the extreme value theorem, there are $c, d \in [a, b]$ such that

$$f(c) = \inf_{x\in[a,b]} f(x), \qquad\qquad f(d) = \sup_{x\in[a,b]} f(x).$$

So the image $f([a, b])$ is at least a subset of the interval $[f(c), f(d)]$. In fact, for any $y \in [f(c), f(d)]$, the intermediate value theorem says that we can find $e \in [c, d]$ (or $e \in [d, c]$, in case $d < c$) such that $f(e) = y$. Such an e must belong to the interval $[a, b]$, since it lies between c and d, so then $y = f(e) \in f([a, b])$ and we conclude that the image is the whole interval. □

So the image of a closed, bounded interval under a continuous map is closed and bounded. This map need not be one-to-one, and in fact we can tell exactly when this happens.

Proposition 1.16. *If $f : [a, b] \to \mathbb{R}$ or $f : \mathbb{R} \to \mathbb{R}$ is continuous, then f is injective if and only if it is strictly monotonic.*

Proof. The direction $\Leftarrow$ is easy: if f is strictly monotonic, then for any distinct points $x < y$ in the domain we have $f(x) < f(y)$ if f is monotone increasing, or $f(x) > f(y)$ if f is monotone decreasing. In any case, $f(x) \neq f(y)$.

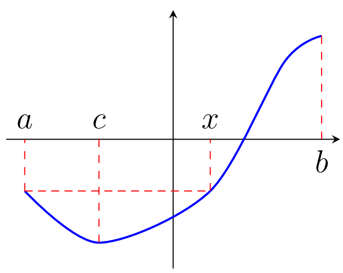
In order to prove $\Rightarrow$, let's first consider the case where f is defined on $[a, b]$. If f is injective then $f(a) \neq f(b)$, so we can assume without loss of generality that $f(a) < f(b)$.

<u>Claim 1</u>: $f(a) = \inf_{x\in[a,b]} f(x)$ and $f(b) = \sup_{x\in[a,b]} f(x)$.

The extreme value theorem says that there are some $c, d \in [a, b]$ such that

$$f(c) = \inf_{x\in[a,b]} f(x) \qquad \text{and} \qquad f(d) = \sup_{x\in[a,b]} f(x).$$

If $f(c) < f(a)$ then we have $f(c) < f(a) < f(b)$, so by the intermediate value theorem we can find some $x \in [c, b]$ such that $f(x) = f(a)$. We have $a \notin [c, b]$, so this says that $a \neq x$, contradicting the assumption that f is injective. Therefore $f(a) = f(c)$. Likewise, if $f(b) < f(d)$ then we can find $y \in [a, d]$ with $f(y) = f(b)$, and again this is impossible, so $f(b) = f(d)$.

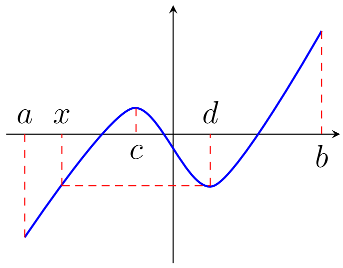


<u>Claim 2</u>: f is strictly monotone increasing on $[a, b]$.

Supposing otherwise, there are $c, d \in [a,b]$ such that $c < d$ but $f(c) \geq f(d)$; since f is injective we can sharpen this to $f(c) > f(d)$. Since

$$f(a) = \inf_{x \in [a,b]} f(x) \leq f(d) < f(c),$$

we can use the intermediate value theorem to find $x \in [a,c]$ such that $f(x) = f(d)$. But then $x \leq c < d$, so this contradicts the fact that f is injective.



These claims prove that an injective, continuous function $f : [a,b] \to \mathbb{R}$ must be strictly monotone. If instead we have an injective, continuous function $f : \mathbb{R} \to \mathbb{R}$ which is *not* strictly monotonic, then there are

- $a, b \in \mathbb{R}$ such that $a < b$ and $f(a) \geq f(b)$, since f is not strictly increasing;
- $c, d \in \mathbb{R}$ such that $c < d$ and $f(c) \leq f(d)$, since f is not strictly decreasing.

We pick some constant $N > \max(|a|, |b|, |c|, |d|)$ and consider the restriction

$$g = f|_{[-N,N]} : [-N, N] \to \mathbb{R}.$$

On the one hand, g is injective since f is. On the other hand, its domain contains the intervals $[a,b]$ and $[c,d]$, and it is neither strictly monotone increasing on the first nor strictly monotone decreasing on the second, so g cannot be strictly monotonic. But this contradicts the case of the proposition that we already proved, so we conclude that f must have been strictly monotonic as well. $\square$

Thus for a continuous function $f : [a,b] \to \mathbb{R}$, we have

$$\begin{aligned} f \text{ injective} &\iff f \text{ strictly monotonic} \\ &\iff \boxed{f \text{ is a bijection } [a,b] \to [f(a), f(b)].} \end{aligned}$$

This means that we have an inverse function $f^{-1} : [f(a), f(b)] \to \mathbb{R}$, where $f^{-1}(x)$ is defined to be the unique value of y such that $f(y) = x$.

Similarly, any continuous injective function $f : \mathbb{R} \to \mathbb{R}$ also has an inverse. Here we have to be a little more careful about the domain of f^{-1}, because there are several possibilities for the image of f.

Mentimeter question 3. Let $f : \mathbb{R} \to \mathbb{R}$ be continuous and injective. Which of the following can *not* be the image of f?

1. $\mathbb{R}$
2. (a, ∞) for some $a \in \mathbb{R}$

3. $(-\infty, b]$ for some $b \in \mathbb{R}$ ✓
4. (a, b) for some $a, b \in \mathbb{R}$

In fact, $f(\mathbb{R})$ has to be an open interval. If it contained an endpoint, say if $f(\mathbb{R}) = (-\infty, b]$, then we'd have $f(x) = b$ for some $x \in \mathbb{R}$. But f is strictly monotonic, so either

$$f(x-1) > f(x) = b \quad \text{or} \quad f(x+1) > f(x) = b,$$

and this contradicts the assertion that b is an upper bound for $f(\mathbb{R})$.

Theorem 1.17

If $f : \mathbb{R} \to \mathbb{R}$ is continuous and injective, then $f^{-1} : f(\mathbb{R}) \to \mathbb{R}$ is continuous.

Proof. Fix a point $y \in f(\mathbb{R})$, say with $y = f(x)$. We need to show that $g = f^{-1}$ is continuous at y, meaning that

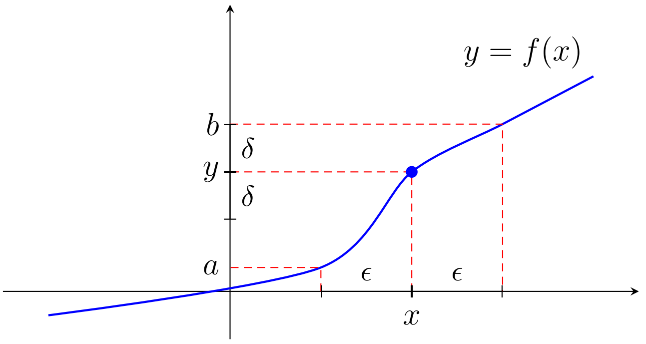
$$\forall \epsilon > 0 \ \exists \delta > 0 \text{ such that } |z - y| < \delta \implies |g(z) - g(y)| < \epsilon.$$

By definition we have $x = g(y)$.

Since f is continuous and injective, it is strictly monotonic; let's say without loss of generality that it's increasing. Then g is also strictly monotonically increasing. Indeed, for any $y_1 < y_2$ in $f(\mathbb{R})$, we have

$$f(g(y_1)) = y_1 < y_2 = f(g(y_2)),$$

and since f is strictly increasing this must mean that $g(y_1) < g(y_2)$.



Fix $\epsilon > 0$ and let $a = f(x - \epsilon)$ and $b = f(x + \epsilon)$. Then $a < y < b$, since $y = f(x)$ and f is strictly increasing. We also have $g(a) = x - \epsilon$ and $g(b) = x + \epsilon$, and since g is strictly increasing we have

$$x - \epsilon < g(z) < x + \epsilon \qquad \forall z \in (a, b).$$

Since $g(y) = x$, we can subtract it from each side and rewrite this as

$$-\epsilon < g(z) - g(y) < \epsilon \qquad \forall z \in (a, b).$$

We take $\delta = \min(b-y, y-a)$. Then $y+\delta \leq y+(b-y) = b$, and $y-\delta \geq y-(y-a) = a$, so

$$|z - y| < \delta \implies a < z < b \implies |g(z) - g(y)| < \epsilon.$$

So this value of δ suffices to prove that g is continuous at y. □

Example 1.18. We saw last term that the function $E(x) = \sum_{n=0}^{\infty} \frac{x^n}{n!}$ is continuous and strictly increasing, so it must be injective. We also recall that it satisfies $E(x) \neq 0$ and $E(-x) = \frac{1}{E(x)}$ for all $x \in \mathbb{R}$.

For $x \geq 0$ every term in the series is nonnegative, so we throw them out for $n \geq 2$ and get $E(x) \geq 1 + x$ for all $x \geq 0$. Thus $E(x)$ is *not* bounded above.

On the other hand, $E(x)$ is bounded below by 0: we already know that $E(0) = 1$ is positive, so supposing there were some a such that $E(a) < 0$, then the intermediate value theorem would say that $E(c) = 0$ for some c between a and 0. But 0 must be the infimum of $E(\mathbb{R})$, because for every $\epsilon > 0$ we have

$$E\left(-\frac{1}{\epsilon}\right) = \frac{1}{E\left(\frac{1}{\epsilon}\right)} \leq \frac{1}{1 + \frac{1}{\epsilon}} = \frac{\epsilon}{\epsilon + 1} < \epsilon.$$

So the intermediate value theorem says that every positive real number is in the image of E, hence $E(\mathbb{R}) = (0, \infty)$. In other words, E gives a continuous bijection

$$E : \mathbb{R} \xrightarrow{\sim} (0, \infty).$$

The inverse function E^{-1}, which we write as

$$\log : (0, \infty) \to \mathbb{R},$$

is therefore also a continuous bijection.

1.3 Open, closed, and compact sets

We'd like to state some upcoming theorems about continuity in fairly general terms. In order to do this, we'll first need to take a detour and introduce some language for subsets of $\mathbb{R}$. You'll get to see much more general versions of this if you take a topology module, which of course you should.

Definition. A set $S \subset \mathbb{R}$ is *open* if and only if

$$\forall x \in S\ \exists \delta > 0 \text{ such that } (x-\delta, x+\delta) \subset S.$$

In other words, if x is a point of an open set S, then S has to contain every other point within a small *neighborhood* of x.

Example 1.19. An "open interval" (a, b) is open.

Proof: for any $x \in (a, b)$, we take $\delta = \min(b-x, x-a)$. Then

$$x < y < x+\delta \implies a < y < x + (b-x) = b \implies y \in (a, b),$$

and similarly if $x - \delta < y < x$ then

$$a = x - (x-a) \leq x - \delta < y < b.$$

Essentially the same proof works for unbounded open intervals like (a, ∞), $(-\infty, b)$, or $\mathbb{R}$ itself.

Proposition 1.20. *Given a collection $\{S_\alpha\}$ of open subsets of $\mathbb{R}$, which may or may not be finite, the union $S = \bigcup_\alpha S_\alpha$ is open.*

Proof. For any $x \in S$ we must have $x \in S_\alpha$ for some α, and since S_α is open,

$$\exists \delta > 0 \text{ such that } (x-\delta, x+\delta) \subset S_\alpha.$$

But then this entire neighborhood of x belongs to $S = \bigcup_\alpha S_\alpha$ as well. ☐

Proposition 1.21. *Given* finitely many *open sets $S_1, \ldots, S_n \subset \mathbb{R}$, the intersection $S = \bigcap_{i=1}^{n} S_i$ is open.*

Proof. For any $x \in S$ we must have $x \in S_i$ for all i, and since S_i is open,

$$\forall i \in \{1, \ldots, n\}\ \exists \delta_i > 0 \text{ such that } (x-\delta_i, x+\delta_i) \subset S_i.$$

Let $\delta = \min(\delta_1, \ldots, \delta_n)$. Then

$$\forall i \in \{1, \ldots, n\}\ (x-\delta, x+\delta) \subset (x-\delta_i, x+\delta_i) \subset S_i,$$

so $(x-\delta, x+\delta)$ is also a subset of $\bigcap S_i = S$. ☐

Example 1.22. To see that finiteness is necessary in this proposition, we define open sets $S_n \subset \mathbb{R}$ for $n \in \mathbb{N}$ by

$$S_n = \left(-\frac{1}{n}, \frac{1}{n}\right).$$

To compute the intersection $S = \bigcap_{n=1}^{\infty} S_n$, we check that

- we have $0 \in S_n$ for all n, so $0 \in S$;
- if $x \neq 0$, then $x \notin S_n$ for n large enough, so $x \notin S$.

For example, we could take $n = \left\lceil \left|\frac{1}{x}\right| \right\rceil + 1$, so that $n > |\frac{1}{x}|$, and then

$$|x| > \frac{1}{n} \Longrightarrow x \notin S_n = \left(-\frac{1}{n}, \frac{1}{n}\right).$$

It follows that $S = \{0\}$, which is not open because it does not contain any neighborhood $(-\delta, \delta)$ of 0 where $\delta > 0$.

Definition. A set $S \subset \mathbb{R}$ is *closed* if and only if

$$\forall \text{sequence } \big(x_n\big) \subset S,\ x_n \to x \in \mathbb{R} \Longrightarrow x \in S.$$

In other words, the limit of any convergent subsequence of S must also be in S.

A set $S \subset \mathbb{R}$ is *compact* if and only if it is closed and bounded.

Example 1.23. A "closed interval" $[a, b]$ is compact.

Proof: It is bounded, so we just need to see that it is closed.

Given a convergent sequence $x_n \to x$, with $a \leq x_n \leq b$ for all n, we must have

$$a \leq \inf_n x_n \leq x \leq \sup_n x_n \leq b,$$

so $x \in [a, b]$ as well.

Warning. "Open" and "closed" are opposites in English, but not in maths! A subset of $\mathbb{R}$ which is not open is not necessarily closed either. Convince yourself of the following:

- that the half-open interval $(0, 1]$ is neither open nor closed;
- that $\mathbb{R}$ is both open and closed;
- that the empty set $\emptyset$ is both open and closed.

Mentimeter question 4. Which of the following sets are closed?

1. $\{\frac{1}{n} \mid n \in \mathbb{N}\}$
2. $[3, \infty)$ ✓
3. $\{40002\}$ ✓
4. $\mathbb{Q}$
5. $\mathbb{R}$ ✓

Proposition 1.24. *A set $S \subset \mathbb{R}$ is open if and only if its complement $T = \mathbb{R} \setminus S$ is closed.*

Proof. $\Longrightarrow$: Suppose that S is open and let $(x_n) \subset T$ converge to some $x \in \mathbb{R}$. This means that

$$\forall \epsilon > 0\ \exists N \in \mathbb{N} \text{ such that } n \geq N \Longrightarrow |x - x_n| < \epsilon.$$

This means that any open neighborhood $(x - \epsilon, x + \epsilon)$ contains some (in fact, infinitely many) of the x_n, which belong to T. But then no such neighborhood of x lies entirely in S, and since S is open we must have $x \notin S$, i.e., $x \in T$. So T is closed.

$\Longleftarrow$: Suppose that T is closed and fix $x \in S$. If S does *not* contain any δ-neighborhood of x, then

$$\forall \delta > 0\ \exists y \notin S \text{ such that } |x - y| < \delta.$$

So for each $n \geq 1$ we can find $x_n \in T$ such that $|x - x_n| < \frac{1}{n}$. Then $x_n \to x$, and since T is closed we must have $x \in T$, contradiction. So S must be open. □

Proposition 1.25. *A union of* finitely many *closed sets is closed. An intersection of arbitrarily many closed sets is also closed.*

Proof. Let $S_1, \ldots, S_n$ be finitely many closed sets. Their complements $T_i = \mathbb{R} \setminus S_i$ are open, and we have

$$\mathbb{R} \setminus \bigcup_{i=1}^{n} S_i = \bigcap_{i=1}^{n} \big(\mathbb{R} \setminus S_i\big) = \bigcap_{i=1}^{n} T_i,$$

which as an intersection of finitely many open sets is also open. The union $\bigcup_{i=1}^{n} S_i$ must therefore be closed, since its complement is open.

Exercise: prove the second part of this proposition. $\square$

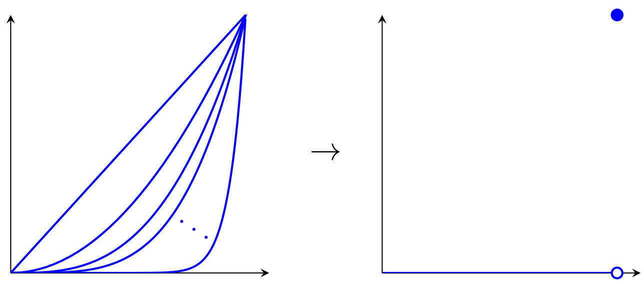
We can revisit some of our earlier theorems with these definitions in mind. The extreme value theorem actually holds for any continuous $f : S \to \mathbb{R}$ whose domain S is compact, because the proof only used the claims that the domain (originally $[a, b]$) was closed and bounded. Why isn't this true of the intermediate value theorem as well?

1.4 Uniform continuity and convergence

Suppose we have a sequence of functions defined on a set $S \subset \mathbb{R}$,

$$f_1, f_2, \cdots : S \to \mathbb{R}.$$

In what sense can we say these converge? If $\lim_{n\to\infty} f_n(x)$ exists for all $x \in S$, then we could define $f : S \to \mathbb{R}$ so that $f(x)$ is this limit. But it's possible that f is not continuous, even if all of the f_n are.

> **Example 1.26.** Define $f_n : [0, 1] \to \mathbb{R}$ by $f_n(x) = x^n$ for all $n \geq 1$. Let
>
> $$f(x) = \lim_{n\to\infty} f_n(x) = \lim_{n\to\infty} x^n.$$
>
> 
>
> Then $f(x) = 0$ for $0 \leq x < 1$, but $f(1) = 1$, so f is not continuous at 1.

We would like to have a stronger notion of continuity which does guarantee convergence to a continuous function.

Recall that given a function $f : S \to \mathbb{R}$, where $S \subset \mathbb{R}$, we say that f is *continuous* if and only if

$$\forall a \in S\ \forall \epsilon > 0\ \exists \delta > 0 \text{ such that } |x - a| < \delta \implies |f(x) - f(a)| < \epsilon.$$

Reading this carefully, the value of δ is allowed to depend on our choice of a. But sometimes this isn't necessary:

Definition. A function $f : S \to \mathbb{R}$ is said to be *uniformly continuous* if and only if

$$\forall \epsilon > 0\ \exists \delta > 0 \text{ such that } \forall x, y \in S,\ |x - y| < \delta \Longrightarrow |f(x) - f(y)| < \epsilon.$$

Note that this definition has its quantifiers in a different order. Before, for each $a \in S$ individually we examined whether f was continuous at a by checking over all $x \in S$. Now, for uniform continuity, we don't fix a; instead, we fix ϵ and choose δ, and then check the same condition for all pairs of points x, y simultaneously. This is certainly a stronger condition! In particular, it doesn't make sense to talk about being uniformly continuous at a point; rather, it's a property of the whole domain.

Proposition 1.27. *If $f : S \to \mathbb{R}$ is uniformly continuous, then it is continuous.*

Proof. We need to check that f is continuous at each point $a \in S$. By uniform continuity,

$$\boxed{\forall \epsilon > 0\ \exists \delta > 0 \text{ such that } \forall x, y \in S,\ |x - y| < \delta \Longrightarrow |f(x) - f(y)| < \epsilon.}$$

If we specialize to $y = a$, then we have

$$\boxed{\forall \epsilon > 0\ \exists \delta > 0 \text{ such that } \forall x \in S,\ |x - a| < \delta \Longrightarrow |f(x) - f(a)| < \epsilon.}$$

But this is exactly what it means to be continuous at a. ☐

Example 1.28. Let $f(x) = ax + b$. Then for any $x, y \in \mathbb{R}$ we have

$$|f(x) - f(y)| = |a| \cdot |x - y| \leq (|a| + 1)|x - y|,$$

so if $|x - y| < \delta$ then $|f(x) - f(y)| < (|a| + 1)\delta$. Thus if we are given any $\epsilon > 0$ and we set $\delta = \frac{\epsilon}{|a|+1}$, then it follows that

$$|x - y| < \delta \quad \Longrightarrow \quad |f(x) - f(y)| < (|a| + 1)\left(\frac{\epsilon}{|a| + 1}\right) = \epsilon.$$

This didn't depend on x and y, so f is uniformly continuous.

Example 1.29. Define $f : \mathbb{R} \to \mathbb{R}$ by $f(x) = x^2$, and fix $\epsilon > 0$. For any $\delta > 0$, if we set $y = x + \frac{\delta}{2}$ then we have

$$|f(y) - f(x)| = \left| \left(x + \frac{\delta}{2}\right)^2 - x^2 \right| = \frac{\delta}{2} \cdot \left| 2x + \frac{\delta}{2} \right|.$$

We can then choose $x = \frac{\epsilon}{\delta}$, and we have $|y - x| = \frac{\delta}{2} < \delta$ but $|f(y) - f(x)| > \frac{\delta}{2} \cdot 2x = \epsilon$. This works for any δ, so f cannot be uniformly continuous.

Example 1.30. Define $f : (0, 1] \to \mathbb{R}$ by $f(x) = \frac{1}{x}$. If f were uniformly continuous, then we could set $\epsilon = 1$ and know that

$$\exists \delta > 0 \text{ such that } |x - y| < \delta \Longrightarrow |f(x) - f(y)| < 1.$$

We should expect this to be false, because $f(x)$ grows much too quickly as $x \downarrow 0$.

In order to disprove it, let's take $x = \frac{1}{n}$ and $y = \frac{1}{n+1}$ for some really large $n \geq 1$. We've chosen these because they satisfy

$$|f(x) - f(y)| = |n - (n+1)| = 1,$$

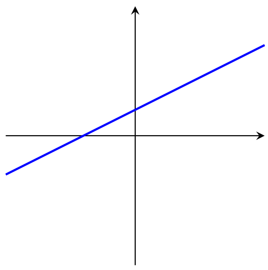
while still being as close together as we could ask for (for large enough n). In particular, we have

$$|x - y| = \left| \frac{1}{n} - \frac{1}{n+1} \right| = \frac{1}{n^2 + n} < \frac{1}{n^2},$$

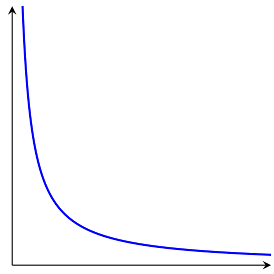
so no matter what value of $\delta > 0$ we were given, we could let $n = \frac{1}{\sqrt{\delta}}$ and we'd have $|x - y| < \delta$. In other words, we have

$$(x, y) = \left(\sqrt{\delta}, \frac{\sqrt{\delta}}{\sqrt{\delta} + 1} \right) \Longrightarrow |x - y| = \frac{\delta}{1 + \sqrt{\delta}} < \delta, \text{ but } |f(x) - f(y)| = 1.$$

So f is *not* uniformly continuous.



Uniformly continuous



Not uniformly continuous

In some sense, the reason for the lack of uniform continuity in this last example is that $f(x)$ approaches infinity in finite time, say if we walk backwards from $x = 1$ to $x = 0$. This can't happen if f satisfies the extreme value theorem, because then it's bounded. So this suggests that in this situation we should get uniform continuity for free.

Proposition 1.31. *If S is compact and $f : S \to \mathbb{R}$ is continuous, then f is uniformly continuous.*

Proof. Suppose that f is not uniformly continuous. Then there must be an $\epsilon > 0$ such that

$$\forall \delta > 0\ \exists x, y \in S \text{ such that } |x - y| < \delta \text{ and } |f(x) - f(y)| \geq \epsilon.$$

We'll take a sequence of such points x_i, y_i with $|x_i - y_i| < \frac{1}{i}$ for all i. By Bolzano–Weierstrass, there is a subsequence $\left(x_{i_j}\right)$ of the x_i which converges (since S is bounded) to some limit $x \in S$ (since S is closed). Then for all j we have

$$|x - y_{i_j}| \leq |x - x_{i_j}| + |x_{i_j} - y_{i_j}|$$

by the triangle inequality, and both terms on the right go to 0 as $j \to \infty$, so $y_{i_j} \to x$ as well.

Since f is sequentially continuous at x, we know that

$$\lim_{j\to\infty} f(x_{i_j}) = f(x) = \lim_{j\to\infty} f(y_{i_j}).$$

So on the one hand we have $|f(x_{i_j}) - f(y_{i_j})| \geq \epsilon$ for all j, but on the other hand we can find N such that for all $j \geq N$

$$|f(x_{i_j}) - f(x)| < \frac{\epsilon}{2} \quad \text{and} \quad |f(x) - f(y_{i_j})| < \frac{\epsilon}{2}.$$

We combine these and use the triangle inequality to get

$$|f(x_{i_j}) - f(y_{i_j})| \leq |f(x_{i_j}) - f(x)| + |f(x) - f(y_{i_j})| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon,$$

and this is a contradiction. □

We saw that the function $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = x^2$ is not uniformly continuous, and neither is $f : (0,1] \to \mathbb{R}$ given by $f(x) = \frac{1}{x}$. But this last proposition says that when restricted to $[1,2]$, for example, both of these become uniformly continuous. In other words, unlike continuity at a point, uniform continuity can depend very much on the domain of the function.

Mentimeter question 5. Which of the following is uniformly continuous?

1. $f : (-\infty, 0) \to \mathbb{R}$ given by $f(x) = e^x$
2. $f : (0,1) \to \mathbb{R}$ given by $f(x) = e^x$
3. $f : [1,\infty) \to \mathbb{R}$ given by $f(x) = e^x$
4. 1 and 2 ✓
5. 2 and 3
6. 1, 2, and 3
7. None of these

For any real numbers $y \le x \le 1$ with $x - y < \delta$, we have

$$e^x - e^y = e^y(e^{x-y} - 1) < e^1(e^\delta - 1).$$

So given $\epsilon > 0$ and any $x, y \le 1$ with $|x - y| < \delta$, we have $|e^x - e^y| < \epsilon$ as long as $e(e^\delta - 1) \le \epsilon$, so we set

$$\delta = \log\left(1 + \frac{\epsilon}{e}\right).$$

Then $|x - y| < \delta$ implies $|e^x - e^y| < \epsilon$ for any $x, y \le 1$, and this proves uniform continuity on both $(-\infty, 0)$ and $(0,1)$.

On the other hand, f can't be uniformly continuous on $[1,\infty)$. Suppose that for a given $\epsilon > 0$ we have $\delta > 0$ such that $|x - y| < \delta$ implies $|e^x - e^y| < \epsilon$ for all $x, y \ge 1$. Assuming $y \le x$, we have

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

So we take $y = \max(\log(2\epsilon/\delta), 1)$ and $x = y + \frac{\delta}{2}$, and we have

$$|x - y| = \frac{\delta}{2} < \delta \quad \text{but} \quad |e^x - e^y| \ge e^y(x - y) \ge \frac{2\epsilon}{\delta} \cdot \frac{\delta}{2} = \epsilon,$$

a contradiction.

We now discuss what it means for a sequence of functions to converge, generalizing the earlier notions of convergence we had for sequences of real numbers.

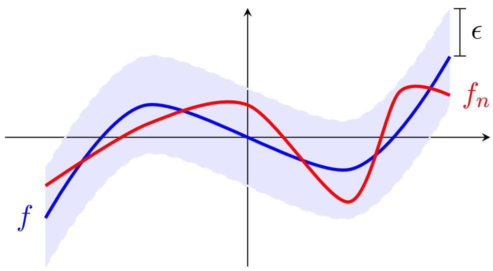
Definition. Let $f_1, f_2, \cdots : S \to \mathbb{R}$ be a sequence of functions defined on $S \subset \mathbb{R}$. We say that f_n converges *pointwise* to $f : S \to \mathbb{R}$ if

$$\forall x \in S \; \forall \epsilon > 0 \; \exists N \in \mathbb{N} \text{ such that } n \geq N \implies |f_n(x) - f(x)| < \epsilon.$$

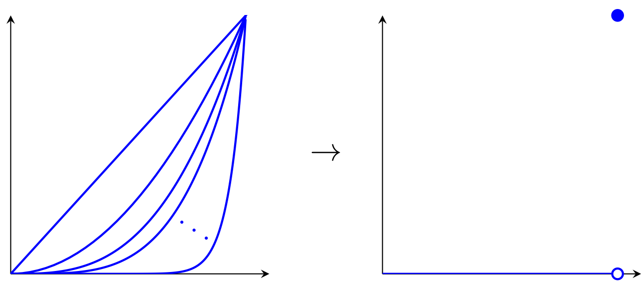
We say that f_n converges *uniformly* to f if

$$\forall \epsilon > 0 \; \exists N \in \mathbb{N} \text{ such that } \forall x \in S, \; n \geq N \implies |f_n(x) - f(x)| < \epsilon.$$

Note that uniform convergence of functions is almost the same as pointwise convergence, but we've rearranged the quantifiers to make it a stronger condition: for each $\epsilon > 0$, there has to be an N such that for all $n \geq N$, the supremum of the *function* $x \mapsto |f_n(x) - f(x)|$ is at most ϵ.



Example 1.32. Recall the sequence of functions $f_n(x) = x^n$ on $[0, 1]$:



These converge pointwise to the discontinuous function

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

but not uniformly. The problem is that for all $n \in \mathbb{N}$, we have

$$\lim_{x \uparrow 1} |f_n(x) - f(x)| = \lim_{x \uparrow 1} |x^n - 0| = 1$$

and so the supremum of $|f_n(x) - f(x)|$ is at least 1 for all n.

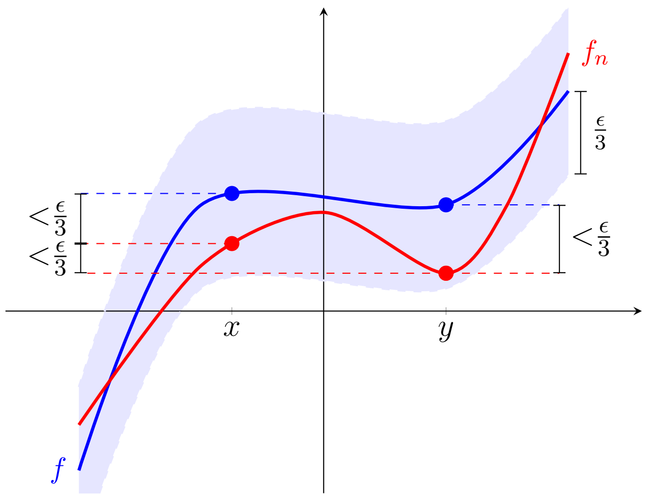
Theorem 1.33

If a sequence of uniformly continuous functions $f_n : S \to \mathbb{R}$ converges uniformly to $f : S \to \mathbb{R}$, then f is uniformly continuous.

Proof. Fixing $\epsilon > 0$, we want to find $\delta > 0$ such that

$$|x - y| < \delta \implies |f(x) - f(y)| < \epsilon.$$

We achieve this by trying to make sense of the following picture, which is an attempt to estimate $|f(y) - f(x)|$ by breaking it into several steps.



Since f_n converges uniformly, we can find $N \in \mathbb{N}$ such that

$$\forall x \in S, \ n \geq N \implies |f_n(x) - f(x)| < \frac{\epsilon}{3}.$$

For a fixed $n \geq N$, the uniform continuity of f_n says that we can also find $\delta > 0$ such that

$$|x - y| < \delta \implies |f_n(x) - f_n(y)| < \frac{\epsilon}{3}.$$

We apply the triangle inequality to see that if $|x - y| < \delta$, then

$$|f(x) - f(y)| \leq |f(x) - f_n(x)| + |f_n(x) - f_n(y)| + |f_n(y) - f(x)|$$

and each of the terms on the right is less than $\frac{\epsilon}{3}$, so $|f(x) - f(y)| < \epsilon$. $\square$

Often we just want to know that a sequence of functions converges to something continuous, rather than uniformly continuous, and then we can get away with weaker hypotheses.

Proposition 1.34. *Let S be any subset of $\mathbb{R}$. If a sequence of continuous functions $f_n : S \to \mathbb{R}$ converges uniformly to $f : S \to \mathbb{R}$, then f is continuous.*

Proof. We mostly repeat the same argument. Fix any $x \in S$ and $\epsilon > 0$, and then by uniform convergence,

$$\exists N > 0 \text{ such that } \forall y \in S\ \forall n \geq N, |f_n(y) - f(y)| < \frac{\epsilon}{3}.$$

Fixing some $n \geq N$, we know that f_n is continuous at x and so

$$\exists \delta > 0 \text{ such that } |y - x| < \delta \Longrightarrow |f_n(y) - f_n(x)| < \frac{\epsilon}{3}.$$

Now we put these together using the triangle inequality, and for all $y \in S$ with $|y - x| < \delta$, we get

$$|f(y) - f(x)| \leq |f(y) - f_n(y)| + |f_n(y) - f_n(x)| + |f_n(x) - f(x)|$$

which is strictly less than $\frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon$. □

This proposition gives another proof that the sequence $f_n : [0, 1] \to \mathbb{R}$ given by $f_n(x) = x^n$ does not converge uniformly, since if it did then its limit would have been continuous.

Finally, we might also be interested in whether a *series* of functions converges. Just as with real numbers, we say that a series

$$\sum_{i=1}^{\infty} f_i(x)$$

converges if and only if the sequence of partial sums

$$S_n(x) = \sum_{i=1}^{n} f_i(x)$$

converges, and it converges uniformly if and only if the sequence $S_n(x)$ converges uniformly. The following is a useful criterion.

Theorem 1.35: Weierstrass M-test

Let $f_1, f_2, \cdots : S \to \mathbb{R}$ be a sequence of continuous functions, and suppose there are constants $M_1, M_2, \ldots$ such that

$$\forall i\ \forall x \in S,\ |f_i(x)| \leq M_i.$$

If $\sum_{i=1}^{\infty} M_i$ converges, then the series $\sum_{i=1}^{\infty} f_i(x)$ converges uniformly to a continuous function $g : S \to \mathbb{R}$.

Proof. For each $n \in \mathbb{N}$ we define the nth partial sum of this series by

$$S_n(x) = \sum_{i=1}^{n} f_i(x).$$

The S_n are all continuous. For any $x \in S$, the comparison test for series tells us that

$$0 \leq |f_i(x)| \leq M_i \Longrightarrow \sum_{i=1}^{\infty} f_i(x) \text{ converges absolutely,}$$

so we can define $g : S \to \mathbb{R}$ by $g(x) = \sum_{i=1}^{\infty} f_i(x) = \lim_{n\to\infty} S_n(x)$.

The sequence $\left(\sum_{i=1}^{n} M_i\right)$ of partial sums is Cauchy because it converges, so given any $\epsilon > 0$:

$$\exists N \in \mathbb{N} \text{ such that } N \leq m \leq n \Longrightarrow \left|\sum_{i=m+1}^{n} M_i\right| < \frac{\epsilon}{2}.$$

For the same ϵ and N, we use the triangle inequality to show that $\forall x \in S$,

$$|S_n(x) - S_m(x)| = \left|\sum_{i=m+1}^{n} f_i(x)\right| \leq \sum_{i=m+1}^{n} |f_i(x)| \leq \sum_{i=m+1}^{n} M_i < \frac{\epsilon}{2},$$

and this bound is independent of $x \in S$. Taking limits as $n \to \infty$ gives

$$\forall m \geq N \ \forall x \in S, \ |g(x) - S_m(x)| \leq \frac{\epsilon}{2} < \epsilon.$$

Since we can find such an $N \in \mathbb{N}$ for any $\epsilon > 0$, the sequence $S_n(x)$ converges uniformly, and this means that the series $\sum_{i=1}^{\infty} f_i(x)$ does as well. □

Example 1.36. Suppose for some $r > 0$ that the series $\sum_{i=0}^{\infty} a_i r^i$ converges absolutely, where the a_i are a sequence of real numbers. For all $i \geq 0$, we take

$$f_i(x) = a_i x^i, \ M_i = |a_i| r^i \Longrightarrow \forall x \in [-r, r], \ |f_i(x)| \leq M_i.$$

Since $\sum_i M_i$ converges, the Weierstrass M-test then tells us that the *power series*

$$\sum_{i=0}^{\infty} a_i x^i = \sum_{i=0}^{\infty} f_i(x)$$

converges uniformly to a continuous function on the interval $[-r, r]$.

Example 1.37. The series $f(x) = \displaystyle\sum_{i=0}^{\infty} \frac{\cos(13^i \pi x)}{2^i}$ converges uniformly on all of $\mathbb{R}$, since if we take $M_i = \frac{1}{2^i}$ for all i then

$$\left| \frac{\cos(13^i \pi x)}{2^i} \right| \leq M_i \text{ and } \sum_{i=0}^{\infty} M_i \text{ converges.}$$

This last example is one of a family of functions constructed by Weierstrass which are famously continuous on all of $\mathbb{R}$ but not differentiable anywhere.

2 Differentiation

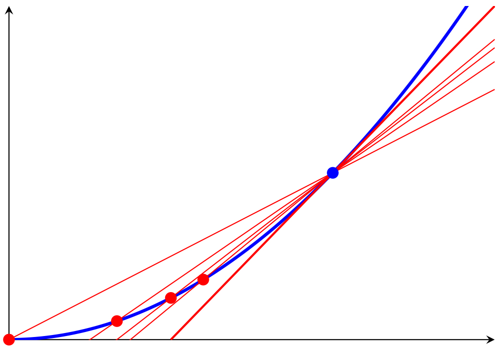
Definition. A function $f : \mathbb{R} \to \mathbb{R}$ is *differentiable* at $a \in \mathbb{R}$, with derivative $f'(a) \in \mathbb{R}$, iff

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

exists and is equal to $f'(a)$. This is equivalent to

$$\forall \epsilon > 0\ \exists \delta > 0 \text{ such that } 0 < |x - a| < \delta \ \Rightarrow \left| \frac{f(x) - f(a)}{x - a} - f'(a) \right| < \epsilon.$$

The quotient $\frac{f(x)-f(a)}{x-a}$ is the slope of the line segment through $(x, f(x))$ and $(a, f(a))$, so being differentiable at a means that these slopes get arbitrarily close to $f'(a)$ as x gets close to a.



Mentimeter question 6. Which of the following is equivalent to the definition of $f'(a)$?

1. $\lim\limits_{h \to 0} \frac{f(a+h) - f(a)}{h}$ ✓
2. $\lim\limits_{h \to 0} \frac{f(a+h) - f(a-h)}{2h}$
3. $\lim\limits_{x \downarrow a} \frac{f(x) - f(a)}{x - a}$
4. More than one of these.

The second option is not equivalent because this limit may exist when $\lim\limits_{x \to a} \frac{f(x) - f(a)}{x - a}$ does not. Consider the function

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

at $a = 0$; then $\lim_{h\to 0} \frac{f(h) - f(-h)}{2h} = 0$, but $\lim_{x\to 0} \frac{f(x) - f(0)}{x} = \lim_{x\to 0} \frac{-1}{x} = \infty$. For the third option, we note that the definition of derivative requires the limit to exist as $x \to a$ from either side, and so

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

at $a = 0$ satisfies $\lim_{x\downarrow 0} \frac{f(x) - f(a)}{x - a} = 0$ even though the same limit as $x \to 0$ does not exist.

In both cases, we constructed a counterexample out of a function that wasn't continuous or differentiable at $x = a$. We'll see soon that if f was differentiable, then it would have been continuous as well.

Example 2.1. Let $f(x) = x^2$. Then for any $a \in \mathbb{R}$,

$$\lim_{x\to a} \frac{f(x) - f(a)}{x - a} = \lim_{x\to a} \frac{x^2 - a^2}{x - a} = \lim_{x\to a} x + a = 2a.$$

So $f(x)$ is differentiable at a with derivative $f'(a) = 2a$.

We will also write $f'(x)$ to denote the function whose value at $x = a$ is $f'(a)$, assuming one exists.

Example 2.2. Let $f(x) = |x|$, and fix $a \in \mathbb{R}$. If $a > 0$ then

$$\lim_{x\to a} \frac{|x| - |a|}{x - a} = \lim_{x\to a} \frac{x - a}{x - a} = 1,$$

because we have $|x| = x$ for all x sufficiently close to a (say, within $\frac{a}{2}$ of it). Similarly, if $a < 0$ then

$$\lim_{x\to a} \frac{|x| - |a|}{x - a} = \lim_{x\to a} \frac{-x - (-a)}{x - a} = -1.$$

On the other hand, at $a = 0$ we have

$$\frac{|x| - |0|}{x - 0} = \begin{cases} -1, & x < 0 \\ 1, & x > 0 \end{cases}$$

and so $\lim_{x\to 0} \frac{|x| - 0}{x - 0}$ does not exist. So $f(x) = |x|$ is *not* differentiable at 0, and otherwise we have

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

We can prove that some common functions are differentiable, and compute their derivatives.

Proposition 2.3. *For any integer $n \geq 0$, the function $f(x) = x^n$ has derivative $f'(x) = nx^{n-1}$.*

Proof. We use the quotient

$$\frac{x^n - a^n}{x - a} = \sum_{i=0}^{n-1} x^{n-1-i}a^i = x^{n-1} + x^{n-2}a + \cdots + xa^{n-2} + a^{n-1}.$$

As $x \to a$, each of the n terms on the right side approaches a^{n-1}, so we have

$$\lim_{x\to a} \frac{x^n - a^n}{x - a} = na^{n-1}.$$

□

Proposition 2.4. *The function $f(x) = e^x$ has derivative $f'(x) = e^x$.*

Proof. We recall the identity $f(x + y) = f(x)f(y)$, which implies that

$$\frac{e^x - e^a}{x - a} = e^a \left(\frac{e^{x-a} - 1}{x - a} \right).$$

Now we set $h = x - a$, and note that $x \to a$ is the same as $h \to 0$, so that

$$\lim_{x\to a} \frac{e^x - e^a}{x - a} = e^a \lim_{h\to 0} \frac{e^h - 1}{h}.$$

But we have

$$\begin{aligned} \frac{e^h - 1}{h} &= \frac{\left(\sum_{n=0}^{\infty} \frac{h^n}{n!}\right) - 1}{h} = \sum_{n=1}^{\infty} \frac{h^{n-1}}{n!} \\ &= 1 + \frac{h}{2!} + \frac{h^2}{3!} + \frac{h^3}{4!} + \ldots, \end{aligned}$$

and this is a convergent series with limit 1 as $h \to 0$. Putting this all together, we have

$$\lim_{x\to a} \frac{e^x - e^a}{x - a} = e^a \lim_{h\to 0} \frac{e^h - 1}{h} = e^a,$$

exactly as claimed. □

Being differentiable is in fact a stronger condition than being continuous. It's certainly true that continuous functions need not be differentiable – we've already seen $f(x) = |x|$ at $x = 0$ – but differentiable functions are always continuous.

Proposition 2.5. *If $f(x)$ is differentiable at a, then it is continuous at a.*

Proof. If f is differentiable at a then we have

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

which means that

$$\forall \epsilon > 0\ \exists \delta > 0 \text{ such that } 0 < |x - a| < \delta\ \Rightarrow \left| \frac{f(x) - f(a)}{x - a} - f'(a) \right| < \epsilon.$$

The triangle inequality says that

$$|f(x) - f(a)| \le |f(x) - f(a) + (x - a)f'(a)| + |(a - x)f'(a)|,$$

so given ϵ and δ as above and $|x - a| < \delta$, we have

$$|f(x) - f(a)| \le \epsilon|x - a| + |a - x||f'(a)| = (\epsilon + |f'(a)|)|x - a|.$$

In particular, if we let $\delta' = \min(\delta, \frac{\epsilon}{\epsilon+|f'(a)|})$ then it follows that

$$|x - a| < \delta'\ \Rightarrow\ |f(x) - f(a)| < (\epsilon + |f'(a)|)\frac{\epsilon}{\epsilon + |f'(a)|} = \epsilon,$$

and so f must be continuous at a. □

We could have also proved this using the algebra of limits: we have

$$f(x) = f(a) + (x - a)\left(\frac{f(x) - f(a)}{x - a}\right),$$

and as $x \to a$ the right side converges to $f(a) + 0 \cdot f'(a) = f(a)$, so then $\lim\limits_{x\to a} f(x)$ exists and is equal to $f(a)$ as well.

Note that even if f is continuous and differentiable everywhere, its derivative need not be continuous.

Example 2.6. Let $f(x) = \begin{cases} x^2 \sin(1/x), & x \neq 0 \\ 0, & x = 0. \end{cases}$ We'll eventually be able to show that

$$f'(x) = 2x \sin(1/x) - \cos(1/x) \text{ for all } x \neq 0,$$

and this does not converge as $x \to 0$. On the other hand, for nonzero x we have

$$\left|\frac{f(x)-f(0)}{x-0}\right| = \left|\frac{x^2\sin(1/x)}{x}\right| = |x\sin(1/x)| \le |x|,$$

so $\left|\frac{f(x)-f(0)}{x-0}\right| \to 0$ as $x \to 0$, hence f is differentiable at 0 and $f'(0) = 0$. It follows that $f(x)$ is differentiable everywhere, but that $f'(x)$ is not continuous.

2.1 Basic properties

Now we will see the effect of some common operations on derivatives.

Proposition 2.7. *If $f(x)$ and $g(x)$ are both differentiable at $x = a$, then $h(x) = f(x) + g(x)$ is also differentiable at $x = a$, and*

$$h'(a) = f'(a) + g'(a).$$

Proof. We use the algebra of limits to compute that

$$\begin{aligned}\lim_{x\to a}\frac{h(x)-h(a)}{x-a} &= \lim_{x\to a}\frac{(f(x)+g(x))-(f(a)+g(a))}{x-a}\\ &= \lim_{x\to a}\frac{f(x)-f(a)}{x-a} + \lim_{x\to a}\frac{g(x)-g(a)}{x-a}\\ &= f'(a)+g'(a).\end{aligned}$$

□

Theorem 2.8: Product rule

Suppose that $f(x)$ and $g(x)$ are both differentiable at $x = a$. Then $h(x) = f(x)g(x)$ is also differentiable at a, and

$$h'(a) = f'(a)g(a) + f(a)g'(a).$$

Proof. We can break the usual limit up into several pieces and evaluate them separately, after we first add and subtract the same term from it:

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

□

Proposition 2.9. *If $f(x)$ is differentiable at $x = a$ and $f(a) \neq 0$, then $g(x) = \frac{1}{f(x)}$ is differentiable at $x = a$, and*

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

Proof. We evaluate the usual limit:

$$\begin{aligned}
\lim_{x\to a} \frac{g(x)-g(a)}{x-a} &= \lim_{x\to a} \frac{\frac{1}{f(x)} - \frac{1}{f(a)}}{x-a} \\
&= \lim_{x\to a} \frac{f(a)-f(x)}{f(x)f(a)(x-a)} \\
&= \lim_{x\to a} \left(\frac{-1}{f(x)f(a)} \right) \left(\frac{f(x)-f(a)}{x-a} \right) \\
&= \frac{-f'(a)}{(f(a))^2}.
\end{aligned}$$

The last step is mostly algebra of limits, except for one subtle detail: we need to know that f being differentiable at a makes it continuous at a, so that $\lim_{x\to a} f(x)$ exists and is equal to $f(a)$. □

Example 2.10. For any $n \in \mathbb{N}$ we have seen that if $f(x) = x^n$ then $f'(x) = nx^{n-1}$. Letting $g(x) = \frac{1}{f(x)} = x^{-n}$, if $x \neq 0$ then g is differentiable at x and we have

$$g(x) = \frac{-f'(x)}{(f(x))^2} = \frac{-nx^{n-1}}{x^{2n}} = (-n)x^{(-n)-1}.$$

So in fact the claim that x^n has derivative nx^{n-1} holds for *all* integers n.

Theorem 2.11: Quotient rule

Suppose that $f(x)$ and $g(x)$ are both differentiable at $x = a$, and that $g(a) \neq 0$. Then $h(x) = \frac{f(x)}{g(x)}$ is differentiable at a, and

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

Proof. We write $h(x) = f(x)r(x)$, where $r(x) = \frac{1}{g(x)}$, and apply the previous proposition: $r(x)$ is differentiable at $x = a$ with derivative $r'(a) = -\frac{g'(a)}{(g(a))^2}$, so by the product rule, h is differentiable at a and

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

□

Theorem 2.12: Chain rule

Let $f(x)$ and $g(x)$ be functions such that g is differentiable at $x = a$ and f is differentiable at $x = g(a)$. Then $h(x) = f(g(x))$ is differentiable at $x = a$, and

$$h'(a) = f'(g(a))g'(a).$$

Proof. We can define functions r and s such that

$$\begin{aligned} f(y) - f(g(a)) &= (y - g(a))(f'(g(a)) + r(y)) \\ g(x) - g(a) &= (x - a)(g'(a) + s(x)), \end{aligned}$$

satisfying $r(g(a)) = s(a) = 0$, and the definition of the derivative tells us that

$$\lim_{y \to g(a)} r(y) = \lim_{x \to a} s(x) = 0.$$

Then we compute from the definitions of r and s that

$$\begin{aligned} f(g(x)) - f(g(a)) &= (g(x) - g(a))(f'(g(a)) + r(g(x))) \\ &= (x - a)(g'(a) + s(x))(f'(g(a)) + r(g(x))). \end{aligned}$$

Rearranging this slightly, when $x \neq a$ we have

$$\frac{f(g(x)) - f(g(a))}{x - a} = (f'(g(a)) + r(g(x)))(g'(a) + s(x)).$$

As $x \to a$ we have $s(x) \to 0$; and $g(x) \to g(a)$ since g is continuous at a, so $r(g(x)) \to 0$ as well. Thus

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

and this is by definition the value of $h'(a)$. $\square$

Remark 2.13. What we'd really like to have done is take the usual limit definition, and multiply and divide the top and bottom by the same term:

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

Certainly $\frac{g(x)-g(a)}{x-a} \to g'(a)$ as $x \to a$, and with a little more work we would hope to prove that the first factor converges to $f'(g(x))$ as well. This is nearly true, but it requires us to know that $g(x)$ approaches $g(a)$ *without being equal to it.* Since we can't guarantee this – $g(x)$ might be constant near $x = a$, for example – we either have to treat that case separately, or find a different proof like the one we gave above.

2.2 The mean value theorem

Derivatives give us a way to identify extreme values of functions.

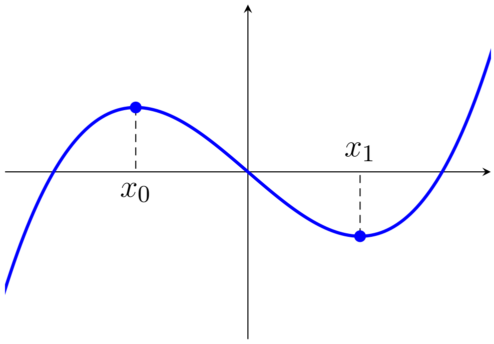
Definition. Let $f : S \to \mathbb{R}$ be a function. We say that f has a *local minimum* at $x \in S$ if and only if

$$\exists \delta > 0 \text{ such that } |y - x| < \delta \;\Rightarrow\; f(y) \geq f(x).$$

Likewise, we say that f has a *local maximum* at $x \in S$ if and only if

$$\exists \delta > 0 \text{ such that } |y - x| < \delta \;\Rightarrow\; f(y) \leq f(x).$$

The following picture shows a function with a local maximum at x_0 and a local minimum at x_1.



The "local" part means that f does not have to take its minimum or maximum value over the whole domain at x, just the minimum or maximum value on some small neighborhood of x.

Mentimeter question 7. Let $f(x) = \begin{cases} 0, & x \notin \mathbb{Q} \\ x, & x \in \mathbb{Q}. \end{cases}$

Where are the local maxima of f?

1. Negative irrational numbers. ✓
2. Negative rational numbers.
3. Positive irrational numbers.
4. Positive rational numbers.
5. 1 and 4.
6. 2 and 3.
7. Nowhere.

Proposition 2.14. *Let $f : [a,b] \to \mathbb{R}$ be a function. If f has a local minimum or a local maximum at some point $x \in (a,b)$, and if f is differentiable at x, then $f'(x) = 0$.*

Proof. Suppose f has a local minimum at x; the local maximum case is nearly identical. Then there is some $\delta > 0$ such that $(x-\delta, x+\delta) \subset (a,b)$ and

$$|y - x| < \delta \ \Rightarrow \ f(y) \geq f(x).$$

(Once the second condition is satisfied, we can take an even smaller δ if needed to satisfy the first one as well.) Using this, we compute that

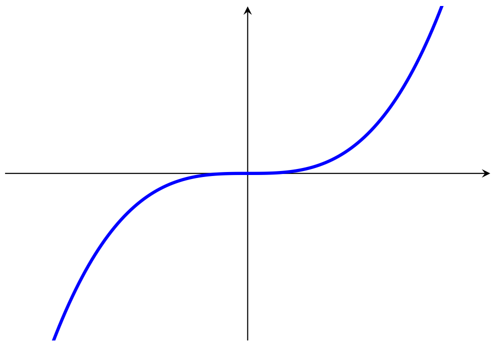
$$\frac{f(y) - f(x)}{y - x} \leq 0 \text{ for } x - \delta < y < x \implies \lim_{y \uparrow x} \frac{f(y) - f(x)}{y - x} \leq 0,$$

because the numerator is positive or zero while the denominator is negative. Similarly, we have

$$\frac{f(y)-f(x)}{y-x} \geq 0 \text{ for } x < y < x+\delta \implies \lim_{y \downarrow x} \frac{f(y)-f(x)}{y-x} \geq 0,$$

because the numerator is nonnegative while the denominator is positive. Now by assumption $f'(x) = \lim_{y \to x} \dfrac{f(y)-f(x)}{y-x}$ exists, so it must be equal to both the limit as $y \uparrow x$ and the limit as $y \downarrow x$, and this is only possible if $f'(x) = 0$. $\square$

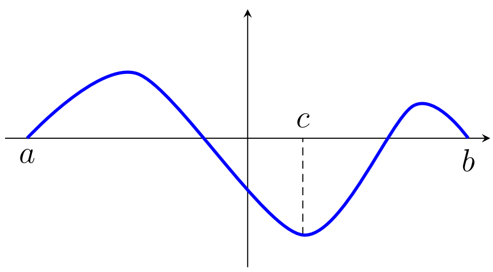
The converse is not true: a function f can be differentiable and satisfy $f'(x) = 0$ at a point x which is neither a local minimum nor a local maximum. For example, when $f(x) = x^3$ we know that $f'(x) = 3x^2$, so $f'(0) = 0$. But $f(x)$ cannot have a local maximum at 0, because $f(y) > 0$ for all positive y, and similarly it does not have a local minimum at 0 because $f(y) < 0$ for all negative y.



Theorem 2.15: Rolle's theorem

Let f be a function which is continuous on $[a, b]$ and differentiable on (a, b). If $f(a) = f(b)$, then there is some $c \in (a, b)$ such that $f'(c) = 0$.

A sketch of some typical f suggests that we should try to prove this by looking for local minima or maxima, and that's exactly what we'll do.



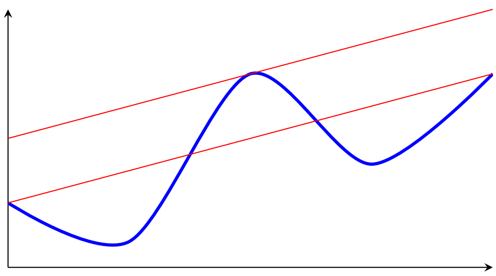
Proof. The extreme value theorem says that f attains both a maximum value and a minimum value on the interval $[a, b]$. If either of these happen at some point $c \in (a, b)$, then the previous proposition says that $f'(c) = 0$ and we are done.

This gives us the desired c in all cases except when f attains its minimum and maximum values at a and b in some order. But if that happens, then since $f(a) = f(b)$, the minimum and maximum values of f are the same, and so f must be constant on $[a, b]$, in which case $f'(c) = 0$ for all $c \in (a, b)$ anyway. □

Theorem 2.16: Mean value theorem

Let f be continuous on $[a, b]$ and differentiable on (a, b). Then there is a point $c \in (a, b)$ such that
$$f'(c) = \frac{f(b) - f(a)}{b - a}.$$

The statement of this theorem says roughly that there's a point c where the slope of the line tangent to the graph of f is the same as the slope of the line through $(a, f(a))$ and $(b, f(b))$:



Our strategy will be to change the function f so that we shear the graph vertically, turning these lines into horizontal ones, and then apply Rolle's theorem.

Proof. We define a function $g : [a, b] \to \mathbb{R}$ by
$$g(x) = f(x) - \frac{f(b) - f(a)}{b - a}(x - a).$$

This is designed to be continuous on $[a, b]$ and differentiable on (a, b), since the same is true of $f(x)$ and $x - a$, and to satisfy $g(a) = g(b) = f(a)$. Rolle's theorem tells us that there is a point $c \in (a, b)$ such that $g'(c) = 0$, and we have
$$g'(c) = f'(c) - \frac{f(b) - f(a)}{b - a} = 0 \implies f'(c) = \frac{f(b) - f(a)}{b - a}.$$

So this is the desired c. □

Proposition 2.17. *Let f be continuous on $[a, b]$ and differentiable on (a, b). If $f'(x) \geq 0$ for all $x \in (a, b)$, then f is monotone increasing on $[a, b]$.*

Moreover, if the stronger inequality $f'(x) > 0$ holds for all $x \in (a, b)$, then f is strictly *monotone increasing on $[a, b]$.*

Proof. For any $x, y \in [a, b]$, with $x < y$, we can apply the mean value theorem to f on the interval $[x, y]$ to find $c \in (x, y)$ such that

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

We multiply both sides by $y - x > 0$ to conclude that $f(y) - f(x) \geq 0$, or $f(y) \geq f(x)$. And if f' were strictly positive, then we would have $f'(c) > 0$ and so the same inequality becomes $f(y) - f(x) = f'(c)(y - x) > 0$, hence $f(y) > f(x)$. $\square$

Of course, the same proof shows that if $f'(x) \leq 0$ or $f'(x) < 0$ for all $x \in (a, b)$, then f is monotone decreasing or strictly monotone decreasing on $[a, b]$. And then we can conclude the following:

Proposition 2.18. *Let f be continuous on $[a, b]$ and differentiable on (a, b). If $f'(x) = 0$ for all $x \in (a, b)$, then f is constant on $[a, b]$.*

Proof. Since $f'(x) \geq 0$ we know that f is monotone increasing, and since $f'(x) \leq 0$ it's also monotone decreasing. So for every $x < y$ in the interval $[a, b]$ we have both $f(x) \leq f(y)$ and $f(x) \geq f(y)$, and therefore $f(x) = f(y)$. $\square$

And this can be applied to tell us that if a function is differentiable, then it's determined up to an additive constant by its derivative.

Proposition 2.19. *Let f and g be continuous on $[a, b]$ and differentiable on (a, b). If $f'(x) = g'(x)$ for all $x \in (a, b)$, then there exists some $c \in \mathbb{R}$ such that $f(x) = g(x) + c$ for all $x \in [a, b]$.*

Proof. The continuous function $h(x) = f(x) - g(x)$ on $[a, b]$ satisfies $h'(x) = 0$ on (a, b), so $h(x) = c$ for some constant c and then $f(x) = g(x) + c$ on all of $[a, b]$. $\square$

2.3 L'Hôpital's rule

Occasionally we come across a limit of the form

$$\lim_{x \to a} \frac{f(x)}{g(x)}, \quad f(a) = g(a) = 0.$$

The algebra of limits doesn't help with this indeterminate form, but if f and g are differentiable near a then the following is often useful.

Theorem 2.20: L'Hôpital's rule, one-sided version

Suppose that f and g are differentiable on an interval (a,b), with $g'(x) \neq 0$ on this interval. If

$$\lim_{x\downarrow a} f(x) = \lim_{x\downarrow a} g(x) = 0 \text{ and } \lim_{x\downarrow a} \frac{f'(x)}{g'(x)} = L,$$

then $\lim\limits_{x\downarrow a} \dfrac{f(x)}{g(x)} = L$.

The proof of l'Hôpital's rule will make use of a stronger version of the mean value theorem, which we present first.

Proposition 2.21. *Let $f, g : [a,b] \to \mathbb{R}$ be continuous functions which are both differentiable on (a,b). Then there is some $c \in (a,b)$ such that*

$$(f(b) - f(a))g'(c) = (g(b) - g(a))f'(c).$$

Proof. Consider $h(x) = (f(b) - f(a))g(x) - (g(b) - g(a))f(x)$. This is continuous on $[a,b]$ and differentiable on (a,b), and we can compute that

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

so Rolle's theorem tells us that $h'(c) = 0$ for some $c \in (a,b)$, and $h'(c) = 0$ is equivalent to the desired condition. $\square$

Note that by taking $g(x) = x$ in this proposition, we recover the original mean value theorem: there is some $c \in (a,b)$ such that $f(b) - f(a) = (b-a)f'(c)$, or $f'(c) = \frac{f(b)-f(a)}{b-a}$.

Proof of l'Hôpital's rule, one-sided version. By the definition of a one-sided limit, given any $\epsilon > 0$ there is a $\delta > 0$ such that

$$a < x < a + \delta \ \Rightarrow \ \left| \frac{f'(x)}{g'(x)} - L \right| < \frac{\epsilon}{2}.$$

If we pick $x, y \in (a, a+\delta)$, say with $a < y < x < a + \delta$, then the generalized mean value theorem says that there is $c \in (x,y)$ such that

$$\frac{f(x) - f(y)}{g(x) - g(y)} = \frac{f'(c)}{g'(c)} \in \left(L - \frac{\epsilon}{2}, L + \frac{\epsilon}{2} \right).$$

Here we use the hypothesis that $g' \neq 0$ on all of (a,b) in order to divide by $g(x) - g(y)$, because if $g(x) = g(y)$ then the mean value theorem would provide $z \in (x,y)$ such that $g'(z) = \frac{g(x)-g(y)}{x-y} = 0$ and this is assumed not to happen.

Since $\left|\frac{f(x)-f(y)}{g(x)-g(y)} - L\right| < \frac{\epsilon}{2}$ whenever $a < y < x < a+\delta$, we take limits as $y \downarrow a$ to get

$$a < x < a+\delta \;\Rightarrow\; \left|\frac{f(x)}{g(x)} - L\right| \leq \frac{\epsilon}{2} < \epsilon.$$

We can find such a $\delta > 0$ for any $\epsilon > 0$, so we have

$$\lim_{x\downarrow a} \frac{f(x)}{g(x)} = L$$

by the definition of a one-sided limit. □

Of course, the same proof works if everything is defined on an interval (b, a) with $b < a$ and we take limits as $x \uparrow a$ instead. So if we combine the two one-sided limits, we get:

Theorem 2.22: L'Hôpital's rule

Suppose that f and g are differentiable on an interval (c, d), except possibly at some point $a \in (c, d)$, and that $g'(x) \neq 0$ on $(c, d) \setminus \{a\}$. If

$$\lim_{x\to a} f(x) = \lim_{x\to a} g(x) = 0 \text{ and } \lim_{x\to a} \frac{f'(x)}{g'(x)} = L,$$

then $\lim_{x\to a} \dfrac{f(x)}{g(x)} = L$.

Mentimeter question 8. In which of the following situations does l'Hôpital's rule also apply, meaning that we can evaluate the limit by replacing f and g with f' and g'? There may be several correct answers.

1. $\lim_{x\to a} \dfrac{f(x)}{g(x)}$, with $f, g \to \infty$ as $x \to a$ ✓
2. $\lim_{x\to\infty} \dfrac{f(x)}{g(x)}$, with $f, g \to 0$ as $x \to \infty$ ✓
3. $\lim_{x\to\infty} (f(x) - g(x))$, with $f, g \to \infty$ as $x \to \infty$
4. $\lim_{x\to\infty} f(x)^{g(x)}$, with $f, g \to 0$ as $x \to \infty$

The first choice is another form of l'Hôpital's rule whose proof is very similar but just a little bit trickier. For the second one, note that

$$\lim_{x\to\infty} \frac{f(x)}{g(x)} = \lim_{y\downarrow 0} \frac{f(1/y)}{g(1/y)} = \lim_{y\downarrow 0} \frac{f'(1/y)(-1/y^2)}{g'(1/y)(-1/y^2)} = \lim_{y\downarrow 0} \frac{f'(1/y)}{g'(1/y)} = \lim_{x\to\infty} \frac{f'(x)}{g'(x)}.$$

The third one cannot be true, because $\lim_{x\to\infty}(2x-x)=\infty$ is not the same as $\lim_{x\to\infty}(2-1)=1$. And for a counterexample to the fourth option, we have

$$\lim_{x\to\infty}\left(-e^{-x}\right)^{2/x}=\lim_{x\to\infty}e^{-2}=\frac{1}{e^2},\qquad \lim_{x\to\infty}\left(e^{-x}\right)^{-2/x^2}=\lim_{x\to\infty}e^{2/x}=1.$$

Example 2.23. We use l'Hôpital's rule to evaluate

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

Both the numerator and denominator approach zero as $x\to 0$, and they are differentiable on $\mathbb{R}$ with derivatives $2\sin(x)\cos(x)$ (via the chain rule) and $\sin(x)$, respectively, according to a problem sheet. We have

$$\lim_{x\to 0}\frac{2\sin(x)\cos(x)}{\sin(x)}=\lim_{x\to 0}2\cos(x)=2,$$

since $\cos(0)=1$, so it follows that

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

as well.

We can apply l'Hôpital's rule multiple times if needed to evaluate a limit.

Example 2.24. We wish to evaluate $\lim_{x\to 0}\frac{e^x-1-x}{x^2}$. By l'Hôpital's rule, we have

$$\lim_{x\to 0}\frac{e^x-1-x}{x^2}=\lim_{x\to 0}\frac{e^x-1}{2x}$$

provided that the limit on the right exists. Again by l'Hôpital's rule, we compute that

$$\lim_{x\to 0}\frac{e^x-1}{2x}=\lim_{x\to 0}\frac{e^x}{2}=\frac{e^0}{2}=\frac{1}{2},$$

so the desired limit does exist and thus $\lim_{x\to 0}\frac{e^x-1-x}{x^2}=\frac{1}{2}$.

In general, if $f'(x)$ exists and is differentiable then we call its derivative $f''(x)$. Note that in this case, $f'(x)$ must be continuous because it is differentiable.

Proposition 2.25. *If $f''(x)$ exists on a neighborhood of $x=a$ and is continuous at $x=a$, then*

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

Proof. The numerator and denominator are differentiable functions of h which both approach 0 as $h \to 0$, and the derivative $2h$ of h^2 is nonzero away from $h = 0$, so l'Hôpital's rule says that

$$\begin{aligned}
\lim_{h\to 0} \frac{f(a+h) - 2f(a) + f(a-h)}{h^2} &= \lim_{h\to 0} \frac{\frac{d}{dh}(f(a+h) - 2f(a) + f(a-h))}{\frac{d}{dh}(h^2)} \\
&= \lim_{h\to 0} \frac{f'(a+h) - f'(a-h)}{2h}
\end{aligned}$$

if the limit on the right exists. But $f'(a+h) - f'(a-h)$ is differentiable when $|h|$ is small, and it approaches 0 as $h \to 0$ since f' is continuous at $a = 0$, so another application of l'Hôpital's rule says that

$$\begin{aligned}
\lim_{h\to 0} \frac{f'(a+h) - f'(a-h)}{2h} &= \lim_{h\to 0} \frac{\frac{d}{dh}(f'(a+h) - f'(a-h))}{\frac{d}{dh}(2h)} \\
&= \lim_{h\to 0} \frac{f''(a+h) + f''(a-h)}{2} \\
&= f''(a)
\end{aligned}$$

by the fact that f'' is continuous at $x = a$, hence $f''(a \pm h) \to f''(a)$ as $h \to 0$. Since this limit exists, the original limit must exist as well, and it is equal to $f''(a)$. □

2.4 Higher derivatives

If a function $f(x)$ is differentiable, then its derivative $f'(x)$ may be differentiable, and we call the derivative of $f'(x)$ the *second derivative* of f, denoted $f''(x)$ or $f^{(2)}(x)$. If $f''(x)$ is differentiable then its derivative is called the *third derivative* and written $f'''(x)$ or $f^{(3)}(x)$. We can repeat this as often as we like; these higher derivatives $f^{(n)}(x)$, also written as $\frac{d^n f}{dx^n}$, carry important information about the original function f.

Note that in order for the nth derivative $f^{(n)}(x)$ to exist at $x = a$, the $(n-1)$st derivative $f^{(n-1)}(x)$ must exist in a neighborhood of a and be differentiable at $x = a$.

Theorem 2.26: Taylor's theorem

Suppose that $f : [c,d] \to \mathbb{R}$ has continuous derivatives $f^{(i)}(x)$ for all $i \leq n$, and that $f^{(n+1)}(x)$ exists for all $x \in (c,d)$. For $a \in [c,d]$, define the *Taylor* *polynomial* of order n at $x = a$ by

$$\begin{aligned} P_n(x) &= f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n \\ &= \sum_{i=0}^{n} \frac{f^{(i)}(a)}{i!}(x-a)^i. \end{aligned}$$

Then for any $b \in [c, d]$ with $b \neq a$, there is a point t between a and b such that

$$f(b) = P_n(b) + \frac{f^{(n+1)}(t)}{(n+1)!}(b-a)^{n+1}.$$

When we take $n = 0$, Taylor's theorem asserts that there is a t between a and b such that

$$f(b) = f(a) + f'(t)(b-a) \iff f'(t) = \frac{f(b) - f(a)}{b-a},$$

which is exactly the mean value theorem. So we can consider Taylor's theorem a massive generalization of the mean value theorem.

In the proof of Taylor's theorem we will repeatedly use the fact that if k is a non-negative integer, then the ith derivative of $(x-a)^k$ at $x = a$ is

$$k(k-1)\ldots(k+1-i)\,(x-a)^{k-i}\Big|_{x=a} = \begin{cases} k!, & i = k \\ 0, & i \neq k. \end{cases}$$

Try to prove this by induction on i.

Proof of Taylor's theorem. We let $k = \frac{f(b)-P_n(b)}{(b-a)^{n+1}}$, and we define an auxiliary function $g : [c, d] \to \mathbb{R}$ by

$$g(x) = f(x) - P_n(x) - k(x-a)^{n+1},$$

noting by our choice of k that

$$g(b) = f(b) - P_n(b) - \left(\frac{f(b) - P_n(b)}{(b-a)^{n+1}}\right)(b-a)^{n+1} = 0.$$

(This may seem unmotivated, but up to a constant term, the $n = 0$ version of this is exactly the same polynomial we used to prove the mean value theorem!)

Since $P_n(x)$ is a polynomial of degree n, we have $P_n^{(n+1)}(x) = 0$, and so

$$g^{(n+1)}(x) = f^{(n+1)}(x) - (n+1)!k.$$

Thus we want to find some t between a and b such that $g^{(n+1)}(t) = 0$.

We observe that for $0 \le i \le n$ we have

$$\begin{aligned} g^{(i)}(a) &= f^{(i)}(a) - P_n^{(i)}(a) - \frac{d^i}{dx^i}\, k(x-a)^{n+1}\Big|_{x=a} \\ &= f^{(i)}(a) - f^{(i)}(a) = 0. \end{aligned}$$

So to summarize, $g(a) = g'(a) = g''(a) = \cdots = g^{(n)}(a) = 0$ and $g(b) = 0$.

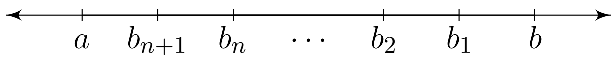
We now apply Rolle's theorem as many times as possible. Since $g(a) = g(b) = 0$, there is some b_1 strictly between a and $b_0 = b$ such that

$$g'(b_1) = 0.$$

Then, since $g'(a) = 0$, there is some b_2 strictly between a and b_1 such that

$$g''(b_2) = 0.$$

We repeat this process $n+1$ times, getting $b_1, b_2, \ldots, b_n, b_{n+1}$ such that b_i is strictly between a and b_{i-1} and such that $\boxed{g^{(i)}(b_i) = 0}$ for each $i \le n+1$.



(Note that this requires $g^{(i-1)}$ to be continuous on $[a, b_{i-1}]$ and $g^{(i)}$ to exist on (a, b_{i-1}) for $1 \le i \le n+1$, and this is guaranteed by the hypotheses of the theorem.) Since $g^{(n+1)}(b_{n+1}) = 0$, we take $t = b_{n+1}$ and we are done. $\square$

Example 2.27. We saw in a problem sheet that $\sin(x)$ has derivative $\cos(x)$, and $\cos(x)$ has derivative $-\sin(x)$, so that

$$\frac{d^n}{dx^n}\sin(x) = \begin{cases} \sin(x), & n = 4k \\ \cos(x), & n = 4k+1 \\ -\sin(x), & n = 4k+2 \\ -\cos(x), & n = 4k+3. \end{cases}$$

The fourth order Taylor polynomial for $\sin(x)$ at $x = a$ is then

$$\begin{aligned} P_4(x) &= \sin(0) + \frac{\cos(0)}{1!}x + \frac{-\sin(0)}{2!}x^2 + \frac{-\cos(0)}{3!}x^3 + \frac{\sin(0)}{4!}x^4 \\ &= x - \frac{x^3}{6}. \end{aligned}$$

Taylor's theorem says that there is some $t \in (0, x)$ such that

$$\sin(x) = P_4(x) + \frac{\frac{d^5}{dx^5}\sin(x)\Big|_{x=t}}{5!}\, x^5 = x - \frac{x^3}{6} + \frac{\cos(t)}{120}x^5.$$

But $|\cos(t)| \leq 1$, and if $0 < x \leq \frac{1}{3}$ then $|x^5| \leq \frac{1}{243}$, so for $0 < x \leq \frac{1}{3}$ the approximation $\boxed{\sin(x) \approx x - \frac{x^3}{6}}$ is accurate to within $\frac{1}{243} \cdot \frac{1}{120} = \frac{1}{29160} < 0.000035$.

Of course, there's no reason why we have to stop a Taylor polynomial at some finite order n.

Definition. Suppose that $f^{(n)}(a)$ exists for all $n \geq 0$. The *Taylor series* for f at $x = a$ is

$$\sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x-a)^n = f(a) + \frac{f'(a)}{1!}(x-a) + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n + \ldots.$$

Mentimeter question 9. Let $P_n(x)$ be the nth order Taylor polynomial for $f(x)$ at $a = 0$, and $P(x)$ the Taylor series at $a = 0$. Which of the following is true?

1. $P(x)$ has infinite radius of convergence.
2. $f(x) = P(x)$ on $(-1, 1)$.
3. The error $|P_{n+1}(x) - f(x)|$ is strictly smaller than $|P_n(x) - f(x)|$.
4. More than one of these.
5. None of these. ✓

None of the first three options is true: (1) If we take $f(x) = \frac{1}{1-x}$, then $f^{(n)}(x) = n!(1-x)^{-n-1}$, so the Taylor series is

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

whose radius of convergence is only 1. (2) We cannot expect that $f(x) = P(x)$ on a fixed interval $(-1, 1)$, because we could start with $f = 0$ and add a little "bump" to the graph on the interval $(\frac{1}{3}, \frac{2}{3})$, and the new function would still have $f^{(n)}(0) = 0$ for all n, so its Taylor series would be $P(x) = 0$. (3) We may have $|P_{n+1}(x) - f(x)| = |P_n(x) - f(x)|$ if $f^{(n+1)}(x) = 0$.

Example 2.28. Since e^x has derivative e^x, we have $\frac{d^n}{dx^n}e^x = e^x$ for all $n \geq 0$. The Taylor series for e^x at $x = 0$ is

$$\sum_{n=0}^{\infty} \frac{\frac{d^n}{dx^n}e^x\Big|_{x=0}}{n!}x^n = \sum_{n=0}^{\infty} \frac{e^0}{n!}x^n = \sum_{n=0}^{\infty} \frac{x^n}{n!},$$

which we recognize as not only a convergent series but the very definition of e^x. So e^x is equal to its own Taylor series as a function $\mathbb{R} \to \mathbb{R}$.

Example 2.29. Let $f : \mathbb{R} \to \mathbb{R}$ be a function such that $f^{(n+1)}(x) = 0$ for all x. If $P_n(x)$ is its nth order Taylor polynomial, say at $a = 0$, then Taylor's theorem says for any $x \neq 0$ that for some t between 0 and x we have

$$f(x) = P_n(x) + \frac{f^{(n+1)}(t)}{(n+1)!}x^{n+1} = P_n(x).$$

So in this case we have

$$f(x) = \sum_{i=0}^{n} \frac{f^{(i)}(0)}{i!}x^i,$$

and in particular f must be a polynomial of degree at most n.

Despite the last few examples, it is definitely not true that every function is equal to its own Taylor series at a point. When this is true, we say that f is *analytic*.

Example 2.30. Let $f(x) = \begin{cases} e^{-1/x^2}, & x \neq 0 \\ 0, & x = 0. \end{cases}$ The chain rule says that

$$f'(x) = \frac{2}{x^3}e^{-1/x^2} \text{ for all } x \neq 0,$$

and we work directly with the definition of the derivative to compute $f'(0)$:

$$\lim_{x\to 0}\left|\frac{e^{-1/x^2} - 0}{x - 0}\right| = \lim_{y\to\infty} \frac{e^{-y^2}}{1/y} = \lim_{y\to\infty} \frac{y}{e^{y^2}}$$

via the substitution $y = \frac{1}{x}$. (We can just take $y \to \infty$ and not worry about $y \to -\infty$, which happens as $x \uparrow 0$, because $\lim\limits_{y\to-\infty}\left|\frac{e^{-y^2}}{1/y}\right|$ is the same.) Since

$e^t \geq 1 + t$ for all $t \geq 0$, we have

$$0 < \frac{y}{e^{y^2}} \leq \frac{y}{y^2 + 1}$$

and the right side goes to zero as $y \to \infty$, so $\frac{y}{e^{y^2}}$ does as well and hence

$$\lim_{x \to 0} \left| \frac{e^{-1/x^2} - 0}{x - 0} \right| = 0 \;\Rightarrow\; f'(0) = 0.$$

Therefore

$$f'(x) = \begin{cases} \frac{2}{x^3} e^{-1/x^2}, & x \neq 0 \\ 0, & x = 0. \end{cases}$$

Similar but more complicated arguments show that $f^{(n)}(x)$ exists for all n, and $f^{(n)}(0) = 0$ for all n, so that $f(x)$ has Taylor series

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

But clearly $f(x)$ is not actually zero anywhere except at $x = 0$.

2.5 Second derivatives and convexity

The first derivative $f'(a)$ can be thought of as the slope of a tangent line to the graph of $f(x)$ at $x = a$. What do the higher derivatives mean? Here we'll try to at least understand the second derivative a little better.

To start, remember that if $f(x)$ has a local maximum or a local minimum at $x = a$, and if f is differentiable at a, then $f'(a) = 0$. The second derivative gives us a converse to this statement in many situations.

Proposition 2.31. *If $f'(a) = 0$ and $f''(a) > 0$, then $f(x)$ has a local minimum at $x = a$. If $f'(a) = 0$ and $f''(a) < 0$, then $f(x)$ has a local maximum at $x = a$.*

Proof. We only prove the case $f''(a) > 0$, since the other one is nearly identical.

By definition, we have

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

Since the limit is strictly positive, we have $\frac{f'(x)}{x-a} > 0$ for all $x \neq a$ in a neighborhood

$(a-\delta, a+\delta)$, where $\delta > 0$. This means that

$$f'(x) < 0 \text{ for all } x \in (a-\delta, a)$$
$$f'(x) > 0 \text{ for all } x \in (a, a+\delta).$$

So f is strictly monotone decreasing on $(a-\delta, a]$ and strictly monotone increasing on $[a, a+\delta)$, and together these imply that

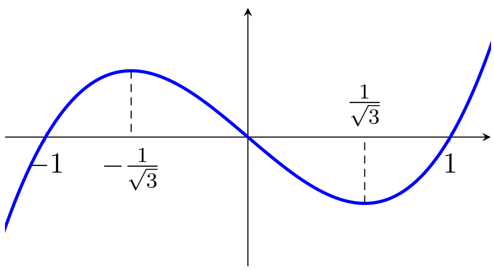
$$f(x) \geq f(a) \text{ for all } a-\delta < x < a+\delta,$$

with equality if and only if $x = a$. In other words, $f(x)$ has a local minimum at $x = a$. □

Example 2.32. Let $f(x) = x^3 - x$. Then $f'(x) = 3x^2 - 1$ is zero at $x = \pm\frac{1}{\sqrt{3}}$, and since $f''(x) = 6x$ we have

$$f''(-1/\sqrt{3}) = -2\sqrt{3} < 0, \qquad f''(1/\sqrt{3}) = 2\sqrt{3} > 0.$$

So $f(x)$ has a local maximum at $x = -\frac{1}{\sqrt{3}}$ and a local minimum at $x = \frac{1}{\sqrt{3}}$.



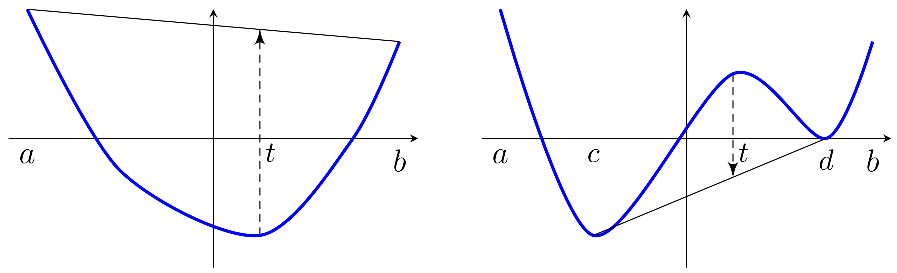
Remark 2.33. If a function $f(x)$ satisfies $f'(a) = f''(a) = 0$, then the second derivative test is *inconclusive*: we cannot say that f has a local minimum or a local maximum. We have already seen the example $f(x) = x^3$ at $x = 0$, where (since $f'(x) = 3x^2$ and $f''(x) = 6x$) we have $f'(0) = f''(0) = 0$, and f has neither a local minimum nor a local maximum there.

Finally, we apply Taylor's theorem to understand when a function is convex.

Definition. We say a function $f : [a,b] \to \mathbb{R}$ is *convex* if for all $c < t < d$ in the domain, we have

$$f(c) + \frac{f(d) - f(c)}{d - c}(t - c) \geq f(t).$$

In other words, the height of the line through $(c, f(c))$ and $(d, f(d))$ at $x = t$ is at least $f(t)$.



Convex Not convex

An equivalent way of stating this is that the region

$$S = \{(x, y) \mid x \in [a, b],\ y \geq f(x)\}$$

lying above the graph of f is a convex subset of $\mathbb{R}^2$: the line segment between any two points of S lies entirely within R.

We can rearrange the definition by writing $t = sc + (1-s)d$ for some $s \in [0, 1]$. Then the left side becomes

$$\begin{aligned} f(c) + \frac{f(d) - f(c)}{d - c}((1-s)(d-c)) &= f(c) + (1-s)(f(d) - f(c)) \\ &= sf(c) + (1-s)f(d) \end{aligned}$$

and so f is convex if and only if for all $c < d$ in the domain and all $s \in (0, 1)$, we have

$$sf(c) + (1-s)f(d) \geq f(sc + (1-s)d).$$

Mentimeter question 10. Which of the following functions $f : \mathbb{R} \to \mathbb{R}$ are convex? Select all that apply.

1. $f(x) = x$ ✓
2. $f(x) = x^2$ ✓
3. $f(x) = x^3$
4. $f(x) = e^x$ ✓

The line through any two points on the graph of $f(x) = x$ coincides with the graph, so it is convex. For $f(x) = x^2$ and $f(x) = e^x$ the convexity should be clear from looking at their graphs, though we can actually prove it using the next proposition. And for $f(x) = x^3$, the line through $(-1, -1)$ and $(1, 1)$ is $y = x$, and the point $(-\frac{1}{2}, -\frac{1}{8})$ on the graph sits above this line, so it isn't convex.

Proposition 2.34. *Let $f : [a,b] \to \mathbb{R}$ have a second derivative $f''(x)$ which is continuous on (a,b). Then f is convex if and only if $f''(x) \geq 0$ for all $x \in (a,b)$.*

Proof. The direction ($\Longrightarrow$) is the easier of the two, so we'll do it first. Suppose f is convex. Then for any $h > 0$ and $x - h < x < x + h$ in the domain, we have

$$\frac{f(x-h) + f(x+h)}{2} \geq f(x),$$

or after some rearranging,

$$\frac{f(x+h) - 2f(x) + f(x+h)}{h^2} \geq 0.$$

Since $f''(x)$ is continuous at $x \in (a,b)$, we have seen that

$$f''(x) = \lim_{h \to 0} \frac{f(x+h) - 2f(x) + f(x+h)}{h^2},$$

and since this ratio is always nonnegative we must have $f''(x) \geq 0$ as well.

In order to prove the direction ($\Longleftarrow$), we now suppose that $f''(x) \geq 0$ for all $x \in (a,b)$. Fix $c < t < d$ in the interval $[a,b]$ and write $t = sc+(1-s)d$, where $0 < s < 1$. By Taylor's theorem, we can write $f(c)$ in terms of the first-order Taylor polynomial of f at $x = t$: there is some $z \in (c,t)$ such that

$$f(c) = f(t) + f'(t)(c-t) + \frac{f''(z)}{2}(c-t)^2,$$

and since $f''(z) \geq 0$ this gives us an inequality

$$\begin{aligned} f(c) &\geq f(t) + f'(t)(c-t) \\ &= f(t) + f'(t)\big((1-s)(c-d)\big). \end{aligned}$$

Likewise for $f(d)$, there is some $w \in (t,d)$ such that

$$f(d) = f(t) + f'(t)(d-t) + \frac{f''(w)}{2}(d-t)^2 \geq f(t) + f'(t)(d-t),$$

so that

$$f(d) \geq f(t) + f'(t)\big(s(d-c)\big).$$

Combining these inequalities, we have

$$\begin{aligned} sf(c) + (1-s)f(d) &\geq s\left(f(t) + f'(t)\big((1-s)(c-d)\big)\right) \\ &\qquad + (1-s)\left(f(t) + f'(t)\big(s(d-c)\big)\right) \\ &= \big(s + (1-s)\big)\, f(t) + f'(t)\,\big(s(1-s)(c-d) + s(1-s)(d-c)\big) \\ &= f(t) \\ &= f(sc + (1-s)d), \end{aligned}$$

and this works for any $t \in (c,d)$, hence for any $s \in (0,1)$, so f is convex. $\square$

We note that in proving that $f''(x) \geq 0$ implies convexity, we did not actually require $f''(x)$ to be continuous; it's enough to know that $f''(x)$ exists on (a,b).

2.6 Limits of differentiable functions

We have already seen that a sequence $\big(f_n\big)$ of continuous functions can converge pointwise to a discontinuous function; introducing the notion of *uniform convergence* gave us a way to ensure that the limit is continuous. We can ask similar questions about the derivative of a pointwise limit.

Example 2.35. Define a sequence $f_n : [0,1] \to \mathbb{R}$ for all $n \in \mathbb{N}$ by

$$f_n(x) = \frac{x^n}{n}.$$

Then $f_n \to 0$ uniformly, because for any $\epsilon > 0$ and all $x \in [0,1]$ we have

$$|f_n(x) - 0| = \left|\frac{x^n}{n}\right| \leq \left|\frac{1}{n}\right| < \epsilon$$

as long as $n > \frac{1}{\epsilon}$. But the f_n are all differentiable, and the familiar sequence

$$f_n'(x) = x^{n-1}$$

does not converge to a continuous function on $[0,1]$. In particular, we have

$$\lim_{n\to\infty} f_n'(1) = 1$$

even though the derivative of the limiting function $f(x) = 0$ satisfies $f'(1) = 0$.

The problem in this example is that the derivatives $f_n'(x)$ do not converge uniformly. When they do, the outcome is much nicer.

Theorem 2.36

Let $f_n : [a,b] \to \mathbb{R}$ be a sequence of differentiable functions, and suppose there is some $c \in [a,b]$ such that $\lim\limits_{n\to\infty} f_n(c)$ exists. If the sequence $\big(f_n'(x)\big)$ converges uniformly on $[a,b]$, then $\big(f_n\big)$ converges uniformly to a function $f : [a,b] \to \mathbb{R}$, and

$$f'(x) = \lim_{n\to\infty} f_n'(x).$$

The proof is a bit long, so we'll break it up into two steps.

Step 1: the functions f_n converge uniformly. Fixing some $\epsilon > 0$, the convergence of $\big(f_n(c)\big)$ and the uniform convergence of the functions $\big(f_n'(x)\big)$ guarantees that there is some $N \geq 0$ such that for all $m, n \geq N$,

$$|f_n(c) - f_m(c)| < \frac{\epsilon}{2} \qquad \text{and} \qquad |f_n'(x) - f_m'(x)| < \frac{\epsilon}{2} \cdot \frac{1}{b-a}$$

for all $x \in [a, b]$. The left side is simply the fact that $\big(f_n(c)\big)$ is a Cauchy sequence; for the right side, if $f_n'(x) \to g(x)$ then we take n large enough such that $|f_n'(x) - g(x)| < \frac{\epsilon}{4(b-a)}$ for all x, and apply the triangle inequality

$$|f_n'(x) - f_m'(x)| \leq |f_n'(x) - g(x)| + |g(x) - f_m'(x)| < \frac{\epsilon}{4(b-a)} + \frac{\epsilon}{4(b-a)}.$$

With $m, n \geq N$ as above, we apply the mean value theorem to the function $f_n(x) - f_m(x)$ to see that for $x \neq c$, there is some t between x and c such that

$$\frac{(f_n(x) - f_m(x)) - (f_n(c) - f_m(c))}{x - c} = f_n'(t) - f_m'(t).$$

We have $|f_n'(t) - f_m'(t)| \cdot |x - c| \leq \frac{\epsilon}{2(b-a)}(b - a) = \frac{\epsilon}{2}$, so by the triangle inequality

$$\begin{aligned} |f_n(x) - f_m(x)| &\leq |(f_n(x) - f_m(x)) - (f_n(c) - f_m(c))| + |f_n(c) - f_m(c)| \\ &< \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon. \end{aligned}$$

This says that for any $x \in [a, b]$, the sequence $\big(f_n(x)\big)$ is Cauchy, hence convergent. We define $f : [a, b] \to \mathbb{R}$ to be the pointwise limit of the f_n, and taking $m \to \infty$ above gives

$$|f_n(x) - f(x)| \leq \epsilon \text{ for all } n \geq N.$$

So in fact the functions f_n converge uniformly to f. □

Step 2: the derivatives f_n' converge to f'. We now fix a point $y \in [a, b]$ and consider the functions

$$g_n(x) = \frac{f_n(x) - f_n(y)}{x - y}$$

defined on $[a, b] \setminus \{y\}$. The uniform convergence of $f_n \to f$ says that

$$g_n(x) \xrightarrow{\text{uniformly}} g(x) = \frac{f(x) - f(y)}{x - y},$$

and we also observe that by definition $\lim\limits_{x \to y} f_n(x) = f_n'(y)$. We want to show that $\lim\limits_{n \to \infty} f_n'(y) = f'(y)$.

Since $\{f_n'(x)\}$ converges uniformly, we know that $L = \lim_{n\to\infty} f_n'(y)$ exists. By the triangle inequality,

$$|g(x) - L| \leq |g(x) - g_n(x)| + |g_n(x) - f_n'(y)| + |f_n'(y) - L|$$

for any n. We observe that

- there is $N_1 \geq 0$ such that if $n \geq N_1$, then $|g(x) - g_n(x)| < \frac{\epsilon}{3}$ for all x;
- for any given n, there is $\delta > 0$ such that $|g_n(x) - f_n'(y)| < \frac{\epsilon}{3}$ as long as $0 < |x - y| < \delta$;
- there is $N_2 \geq 0$ such that if $n \geq N_2$, then $|f_n'(y) - L| < \frac{\epsilon}{3}$.

The first claim uses the uniform convergence of $g_n \to g$, and the second and third are the facts that $\lim_{x\to y} g_n(x) = f_n'(y)$ and $\lim_{n\to\infty} f_n'(y) = L$ respectively. Putting these together, we can fix $n \geq \max(N_1, N_2)$ and then take $\delta > 0$ such that

$$0 < |x - y| < \delta \;\Rightarrow\; |g(x) - L| < \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon.$$

But then this says precisely that

$$\lim_{x\to y} g(x) = L = \lim_{n\to\infty} f_n'(y),$$

or equivalently that

$$\lim_{n\to\infty} f_n'(y) = \lim_{x\to y} \frac{f(x) - f(y)}{x - y} = f'(y).$$

□

The proof of step 2 is nearly identical to the proof that a uniformly convergent sequence of continuous functions has continuous limit. If we repeat the argument, taking $g_n \to g$ to be any uniformly convergent sequence of functions defined on a neighborhood of y (though not necessarily at y itself) and replacing $f_n'(y)$ with $\lim_{x\to y} g_n(y)$, what it really shows is that

$$\lim_{x\to y} \lim_{n\to\infty} g_n(x) = \lim_{n\to\infty} \lim_{x\to y} g_n(x).$$

In other words, the order of the limits does not matter as long as the g_n converge uniformly on $(y - \delta, y + \delta) \setminus \{y\}$.

Theorem 2.37: Differentiation of power series

Let $f(x) = \sum_{n=0}^{\infty} a_n x^n$ be a power series with radius of convergence $R > 0$. Then

f has a continuous derivative on $(-R, R)$, and

$$f'(x) = \sum_{n=0}^{\infty} n a_n x^{n-1}$$

for all $|x| < R$.

Proof. We recall that $f(x)$ converges absolutely for any $|x| < R$. Given any positive $r < R$, and letting $t = \frac{r+R}{2}$, we also proved by the Weierstrass M-test that the partial sums

$$f_n(x) = \sum_{i=0}^{n} a_i x^i$$

converge uniformly to $f(x)$ on the interval $[-t, t] \subset (-R, R)$. Thus $f_n(0) \to f(0)$.

We have $f_n'(x) = \displaystyle\sum_{i=0}^{n} i a_i x^{i-1}$. If $|x| < t$, then for all sufficiently large i we have

$$\left| i a_i x^{i-1} \right| = \frac{|a_i t^i|}{t} \left| i \left(\frac{x}{t}\right)^{i-1} \right| \leq \frac{|a_i t^i|}{t}.$$

The key observation here, left as an exercise, is that if $0 \leq s < 1$ then $i s^{i-1} \leq 1$ for i large enough. So we apply the Weierstrass M-test with

$$|i a_i x^{i-1}| \leq M_i = \frac{a_i t^i}{t} \text{ for all large enough } i,$$

using the fact that $\frac{1}{t} \displaystyle\sum_{i=0}^{\infty} a_i t^i$ converges absolutely, to see that

$$f_n'(x) \to \sum_{i=0}^{\infty} i a_i x^{i-1} \text{ uniformly on } (-t, t),$$

with the limit series being continuous. (Why is it not a problem that we may have $|i a_i x^{i-1}| > M_i$ for finitely many i?) Since the derivatives $f_n'(x)$ converge uniformly on $[-r, r] \subset (-t, t)$, and $f_n(0) \to f(0)$, the above theorem tells us that

$$f'(x) = \lim_{n\to\infty} f_n'(x) = \sum_{i=0}^{\infty} i a_i x^{i-1}$$

on the interval $[-r, r]$, and this works for any $r < R$. $\square$

Example 2.38. Let $f(x) = \frac{1}{1-x}$ for $|x| < 1$. Then f is equal to the power series $\dfrac{1}{1-x} = \displaystyle\sum_{n=0}^{\infty} x^n$, which has radius of convergence 1. We can differentiate both sides to get

$$\frac{1}{(1-x)^2} = \sum_{n=0}^{\infty} nx^{n-1}$$

and then multiply by x to deduce for all $x \in (-1, 1)$ the identity

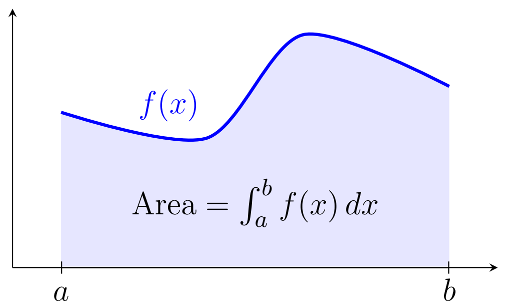
$$\frac{x}{(1-x)^2} = \sum_{n=0}^{\infty} nx^n = x + 2x^2 + 3x^3 + 4x^4 + \dots.$$

3 Integration

No one's fast like Gaston
Good at maths like Gaston
Finds the area under a graph like Gaston

- "Gaston Darboux", definitely not from Beauty and the Beast

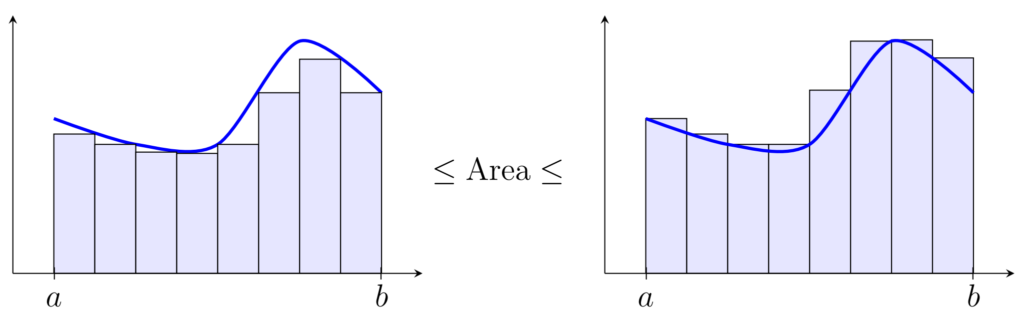
A definite integral $\int_a^b f(x)\,dx$ is supposed to measure the area between the x-axis and the graph of a function $f : [a, b] \to \mathbb{R}$.



In the next few sections we'll see how to make this precise, using the *Darboux integral* – this is equivalent to the more familiar Riemann integral, so we may also call it "Riemann–Darboux integration" – and develop some key properties.

3.1 Darboux sums

The rough idea behind the Darboux integral is that we can estimate the area under the graph of $f(x)$ by approximating it with rectangles, since we know the area of a rectangle. If the rectangles all lie under the graph of $f(x)$, then we'll get a lower bound; if they cover the whole area between the x-axis and the graph, then we'll get an upper bound.



In order to do this, we split the interval $[a, b]$ into finitely pieces as follows.

Definition. A *partition* of the interval $[a,b]$ is a finite sequence of real numbers $P = (x_0, x_1, x_2, \dots, x_k)$ such that

$$a = x_0 < x_1 < x_2 < \cdots < x_k = b.$$

We can view a partition $P = (x_0, \dots . x_k)$ of $[a,b]$ as splitting it up into closed intervals

$$[x_0, x_1],\ [x_1, x_2],\ [x_2, x_3],\ \dots,\ [x_{k-1}, x_k],$$

and we write $\Delta x_i = x_{i+1} - x_i > 0$ for the length of the interval $[x_i, x_{i+1}]$. If $f : [a,b] \to \mathbb{R}$ is a *bounded* function (but not necessarily continuous!), then we can define

$$m_i = \inf_{x_i \le t \le x_{i+1}} f(t), \qquad M_i = \sup_{x_i \le t \le x_{i+1}} f(t)$$

for $0 \le i < k$. We define the *lower Darboux sum* of f with respect to P as

$$L(f,P) = \sum_{i=0}^{k-1} m_i \Delta x_i$$

and the *upper Darboux sum* of f with respect to P is similarly

$$U(f,P) = \sum_{i=0}^{k-1} M_i \Delta x_i.$$

Note that the $m_i \Delta x_i$ terms are the areas of rectangles lying just under the graph of f, as pictured above at left, and the $M_i \Delta x_i$ terms are areas of rectangles lying just above the graph, as shown above at right.

Mentimeter question 11. Which of the following is true for any function $f : [a,b] \to \mathbb{R}$ and any partition $P = (x_0, x_1, \dots, x_n)$ of $[a,b]$?

1. At least one of $L(f,P)$ and $U(f,P)$ exists.
2. If f is continuous then both $L(f,P)$ and $U(f,P)$ exist. ✓
3. We always have $L(f,P) < U(f,P)$ if both are defined.
4. The value of $L(f,P)$ does not depend on P.

If f is continuous then the extreme value theorem says that it is bounded on $[a,b]$, so both $L(f,P)$ and $U(f,P)$ exist. It is possible for f to not be bounded above or below on any interval, and then $L(f,P)$ and $U(f,P)$ are not defined because $\inf f$ and $\sup f$ are not defined on any interval: for example, consider

$$f(x) = \begin{cases} 0, & x \notin \mathbb{Q} \\ (-1)^p q, & x = \frac{p}{q} \in \mathbb{Q}. \end{cases}$$

Also, if f is constant then we have $L(f,P) = U(f,P)$, and nearly any example will show that $L(f,P)$ depends on P (take $f(x) = x$ on $[0,1]$ and $P = (0,1)$ but $P' = (0, \frac{1}{2}, 1)$).

Example 3.1. Suppose that $f(x) = c$ is constant on $[a,b]$. Then for any partition $P = (x_0, \ldots, x_k)$ of $[a,b]$, we have

$$m_i = \inf_{x_i \le t \le x_{i+1}} f(t) = c, \qquad M_i = \sup_{x_i \le t \le x_{i+1}} f(t) = c$$

for all i. The corresponding lower and upper Darboux sums are

$$\begin{aligned} L(f,P) = \sum_{i=0}^{k-1} m_i \Delta x_i &= c(x_1 - x_0) + c(x_2 - x_0) + \cdots + c(x_k - x_{k-1}) \\ &= c(x_k - x_0) = c(b-a) \end{aligned}$$

and similarly

$$U(f,P) = \sum_{i=0}^{k-1} M_i \Delta x_i = c(x_k - x_0) = c(b-a).$$

So $L(f,P) = U(f,P) = c(b-a)$ for all P.

Example 3.2. Let $f(x) = \begin{cases} 0, & x \in \mathbb{Q} \\ 1, & x \notin \mathbb{Q}, \end{cases}$ and fix a partition $P = (x_0, \ldots, x_k)$ of $[a,b]$. Then any interval $[x_i, x_{i+1}]$ contains both rational and irrational numbers, so we have

$$m_i = \inf_{x_i \le t \le x_{i+1}} f(t) = 0, \qquad M_i = \sup_{x_i \le t \le x_{i+1}} f(t) = 1.$$

Thus for any P we have

$$\begin{aligned} L(f,P) &= \sum_{i=0}^{k-1} 0 \cdot \Delta x_i = 0, \\ U(f,P) &= \sum_{i=0}^{k-1} 1 \cdot \Delta x_i = b-a. \end{aligned}$$

Example 3.3. Let $f(x) = x$ on the interval $[0, 1]$, and consider the partition $P_n = (0, \frac{1}{n}, \frac{2}{n}, \dots, \frac{n-1}{n}, 1)$ with $x_i = \frac{i}{n}$ for $0 \le i \le n$. We have $\Delta x_i = \frac{1}{n}$ for all $i < n$, and

$$m_i = \inf_{\frac{i}{n} \le t \le \frac{i+1}{n}} t = \frac{i}{n}, \qquad M_i = \sup_{\frac{i}{n} \le t \le \frac{i+1}{n}} t = \frac{i+1}{n},$$

so we compute that

$$L(f, P_n) = \sum_{i=0}^{n-1} \frac{i}{n} \cdot \frac{1}{n} = \frac{1}{n^2} \sum_{i=0}^{n-1} i = \frac{1}{n^2} \frac{n(n-1)}{2} = \frac{n-1}{2n}$$

and similarly

$$U(f, P_n) = \sum_{i=0}^{n-1} \frac{i+1}{n} \cdot \frac{1}{n} = \frac{1}{n^2} \sum_{i=0}^{n-1} (i+1) = \frac{1}{n^2} \frac{n(n+1)}{2} = \frac{n+1}{2n}.$$

Lemma 3.4. *Given a bounded function $f : [a, b] \to \mathbb{R}$ and a partition P of $[a, b]$, we have*

$$L(f, P) \le U(f, P).$$

Proof. For all $i = 0, 1, \dots, k-1$ we have $\Delta x_i > 0$ and

$$m_i = \inf_{x_i \le t \le x_{i+1}} f(t) \le \sup_{x_i \le t \le x_{i+1}} f(t) = M_i,$$

and so

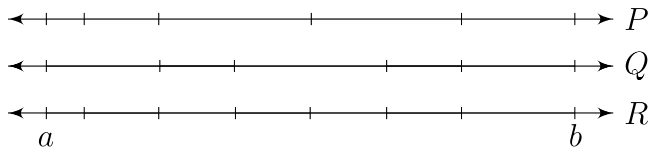
$$L(f, P) = \sum_{i=0}^{k-1} m_i \Delta x_i \le \sum_{i=0}^{k-1} M_i \Delta x_i = U(f, P). \qquad \square$$

Of course, there's no reason to think that $L(f, P)$ and $U(f, P)$ should be equal, but we might hope that the more points we add to our partition, the closer they get to the actual area we want to measure, and this turns out to be true.

Definition. A partition Q is a *refinement* of P, written $P \prec Q$, if and only if every point of P is also a point of Q.

The *common refinement* R of any two partitions P and Q is the partition whose points are precisely those points belonging to either of P and Q. It satisfies

$P \prec R$ and $Q \prec R$.



Mentimeter question 12. Let P, Q, R, S be partitions of $[a, b]$. Which of the following is *not* true?

1. If $P \prec Q$ and $Q \prec R$, then $P \prec R$.
2. Either $P \prec Q$ or $Q \prec P$. $\checkmark$
3. If R is the common refinement of P and Q, then $P \prec S$ and $Q \prec S$ implies $R \prec S$.
4. $\forall \epsilon > 0$, P has a refinement $Q = (x_0, \dots, x_n)$ with $\max_i (x_{i+1} - x_i) < \epsilon$.

The second option is false: for example, if $P = (0, \frac{1}{3}, 1)$ and $Q = (0, \frac{2}{3}, 1)$ then $\frac{1}{3} \in P$ does not belong to Q, and $\frac{2}{3} \in Q$ does not belong to P. The first and third options follow from the definitions, and the fourth is less obvious but we can first take

$$R = \left(a, a + \frac{1}{n}(b-a), a + \frac{2}{n}(b-a), \dots, \frac{n-1}{n}(b-a), b \right)$$

for n large enough – here we have $\max_i (x_{i+1} - x_i) = \dfrac{b-a}{n} < \epsilon$ if $n > \frac{b-a}{\epsilon}$ – and then let Q be the common refinement of P and R.

Proposition 3.5. *If Q is a refinement of P, then we have*

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

Proof. We'll reduce this to the case where $Q \setminus P$ is a single point. In general, if $Q \setminus P$ consists of m points $a_1, \dots, a_m$, then we can define a sequence of partitions

$$P = P_0 \prec P_1 \prec P_2 \prec \dots \prec P_m = Q,$$

where for $1 \le k \le m$ we build P_k by adding the point a_k to P_{k-1}. Since $|P_k \setminus P_{k-1}| = 1$ for all k, we repeatedly apply the $m = 1$ case of the proposition to get

$$L(f, P) \le L(f, P_1) \le L(f, P_2) \le \dots \le L(f, P_m) = L(f, Q)$$

and likewise

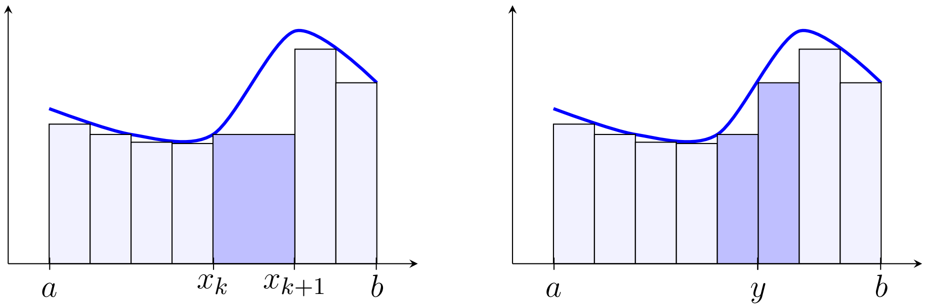
$$U(f, Q) = U(f, P_m) \le U(f, P_{m-1}) \le \dots \le U(f, P_1) \le U(f, P).$$

Putting these together, we conclude the general case from the case where $m = 1$.

To prove the $m = 1$ case, suppose that $P = (x_0, x_1, \ldots, x_n)$ and that

$$Q = (x_0, \ldots, x_k, y, x_{k+1}, \ldots, x_n)$$

for some k and some $y \in (x_k, x_{k+1})$.



Then almost all terms in the lower Darboux sums for P and Q are the same, because they mostly compute the areas of the same rectangles: the difference happens entirely on the interval $[x_k, x_{k+1}]$, and so we compute that

$$\begin{aligned} L(f,Q) - L(f,P) = &\left(\left(\inf_{x_k \le t \le y} f(t)\right)(y - x_k) + \left(\inf_{y \le t \le x_{k+1}} f(t)\right)(x_{k+1} - y)\right) \\ &- \left(\inf_{x_k \le t \le x_{k+1}} f(t)\right)(x_{k+1} - x_k). \end{aligned}$$

Some rearranging, using the fact that $x_{k+1} - x_k = (x_{k+1} - y) + (y - x_k)$, gives us

$$\begin{aligned} L(f,Q) - L(f,P) = &\left(\left(\inf_{x_k \le t \le y} f(t)\right) - \left(\inf_{x_k \le t \le x_{k+1}} f(t)\right)\right)(y - x_k) \\ &+ \left(\left(\inf_{y \le t \le x_{k+1}} f(t)\right) - \left(\inf_{x_k \le t \le x_{k+1}} f(t)\right)\right)(x_{k+1} - y). \end{aligned}$$

Both $y - x_k$ and $x_{k+1} - y$ are positive, and we have

$$\left(\inf_{x_k \le t \le y} f(t)\right), \left(\inf_{y \le t \le x_{k+1}} f(t)\right) \ge \left(\inf_{x_k \le t \le x_{k+1}} f(t)\right)$$

because a lower bound for $f(x)$ on $[x_k, x_{k+1}]$ is certainly also a lower bound for $f(x)$ on each of $[x_k, y]$ and $[y, x_{k+1}]$, so each difference of infima above is nonnegative and we conclude that

$$L(f,Q) - L(f,P) \ge 0.$$

The same argument with sup instead of inf shows that $U(f,Q) - U(f,P) \le 0$. $\square$

Using these inequalities, we can prove that any lower Darboux sum for f is less than or equal to any upper Darboux sum of f, regardless of the partitions we use.

Proposition 3.6. *If $f : [a, b] \to \mathbb{R}$ is bounded, and P and Q are any partitions of $[a, b]$, then*

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

Proof. We let R be the common refinement of P and Q. Then $P \prec R$ and $Q \prec R$, so

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

by applying the last proposition twice. These imply that $L(f, P) \leq U(f, Q)$. □

3.2 The Darboux integral

The inequality $L(f, P) \leq U(f, Q)$ for any P and Q says that the set

$$\{L(f, P) \mid P \text{ is a partition of } [a, b]\}$$

is bounded above by any upper Darboux sum $U(f, Q)$, and likewise the set

$$\{U(f, P) \mid P \text{ is a partition of } [a, b]\}$$

is bounded below by any lower Darboux sum $L(f, Q)$.

Definition. Let $f : [a, b] \to \mathbb{R}$ be a bounded function. We define the *lower* and *upper Darboux integrals* of f on $[a, b]$ by

$$\underline{\int_a^b} f(x)\, dx = \sup_P L(f, P), \qquad \overline{\int_a^b} f(x)\, dx = \inf_P U(f, P).$$

Lemma 3.7. *If $f : [a, b] \to \mathbb{R}$ is bounded, then* $\underline{\int_a^b} f(x)\, dx \leq \overline{\int_a^b} f(x)\, dx.$

Proof. For any partitions P and Q, we have $L(f, P) \leq U(f, Q)$, and so $U(f, Q)$ is an upper bound for the set of all lower Darboux sums $L(f, P)$, or

$$\underline{\int_a^b} f(x)\, dx = \sup_P L(f, P) \leq U(f, Q).$$

Then $\underline{\int_a^b} f(x)\,dx$ is a lower bound for the set of all upper Darboux sums $U(f,Q)$, so

$$\underline{\int_a^b} f(x)\,dx \le \inf_Q U(f,Q) = \overline{\int_a^b} f(x)\,dx$$

as claimed. □

Definition. If the upper and lower Darboux integrals of f on $[a,b]$ are equal, then we say that f is *(Darboux) integrable* on $[a,b]$, and we define

$$\int_a^b f(x)\,dx \stackrel{\text{def}}{=} \underline{\int_a^b} f(x)\,dx = \overline{\int_a^b} f(x)\,dx.$$

Remark 3.8. The "dx" part of the integral is a bit of notation that says we're integrating a function of x. We could change the name of the variable, and the definition would still be the same: it makes perfect sense to say that

$$\int_a^b f(t)\,dt = \int_a^b f(x)\,dx.$$

Example 3.9. When $f(x) = c$ is constant on $[a,b]$, we computed that $L(f,P) = U(f,P) = c(b-a)$ for all P. Thus $f(x) = c$ is integrable, with

$$\int_a^b c\,dx = c(b-a).$$

Example 3.10. We computed for $f(x) = \begin{cases} 0, & x \in \mathbb{Q} \\ 1, & x \notin \mathbb{Q} \end{cases}$ on the interval $[a,b]$ that $L(f,P) = 0$ and $U(f,P) = b-a$ for all P. Thus

$$\underline{\int_a^b} f(x)\,dx = 0 < b-a = \overline{\int_a^b} f(x)\,dx$$

and since the lower and upper Darboux integrals are not equal, $f(x)$ is *not* integrable on $[a,b]$.

Example 3.11. Let $f(x) = x$ on the interval $[0,1]$, and let $P_n = (0, \frac{1}{n}, \frac{2}{n}, \dots, \frac{n-1}{n}, 1)$. We computed that

$$L(f, P_n) = \frac{n-1}{2n} \text{ for all } n \;\Rightarrow\; \underline{\int_0^1} x\,dx \geq \sup_{n\in\mathbb{N}} \frac{n-1}{2n} = \frac{1}{2},$$

and similarly

$$U(f, P_n) = \frac{n+1}{2n} \text{ for all } n \;\Rightarrow\; \overline{\int_0^1} x\,dx \leq \inf_{n\in\mathbb{N}} \frac{n+1}{2n} = \frac{1}{2}.$$

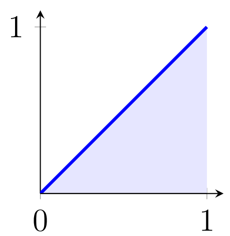
Since the lower Darboux integral is less than or equal to the upper one, we have

$$\frac{1}{2} \leq \underline{\int_0^1} x\,dx \leq \overline{\int_0^1} x\,dx \leq \frac{1}{2},$$

and hence both of them must equal $\frac{1}{2}$. So $f(x) = x$ is integrable on $[0,1]$, with

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

Note that this coincides with the area of the triangle $0 \leq y \leq x$, $0 \leq x \leq 1$.



This last example illustrates an important principle: if we want to show that $f(x)$ is integrable on $[a,b]$ then we don't really need to consider *all* partitions of $[a,b]$, just some well-chosen sequence of partitions for which the lower and upper Darboux sums converge to the same value. To make this precise:

Proposition 3.12. *A bounded function $f : [a,b] \to \mathbb{R}$ is integrable if and only if for every $\epsilon > 0$, there is a partition P of $[a,b]$ such that*

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

Proof. ($\Longrightarrow$) Since f is integrable we have

$$\sup_P L(f, P) = \underline{\int_a^b} f(x)\,dx = \int_a^b f(x)\,dx = \overline{\int_a^b} f(x)\,dx = \inf_P U(f, P),$$

so given $\epsilon > 0$ we can find partitions Q and R of $[a,b]$ such that

$$L(f,Q) > \int_a^b f(x)\,dx - \frac{\epsilon}{2}, \qquad U(f,R) < \int_a^b f(x)\,dx + \frac{\epsilon}{2}.$$

Let P be the common refinement of Q and R; then $Q \prec P$ and $R \prec P$, so we have

$$L(f,Q) \le L(f,P) \le U(f,P) \le U(f,R)$$

and therefore

$$\begin{aligned} U(f,P) - L(f,P) &\le U(f,R) - L(f,Q) \\ &< \left(\int_a^b f(x)\,dx + \frac{\epsilon}{2}\right) - \left(\int_a^b f(x)\,dx - \frac{\epsilon}{2}\right) \\ &= \epsilon. \end{aligned}$$

($\Longleftarrow$) Take $\epsilon > 0$ and a partition P with $U(f,P) - L(f,P) < \epsilon$. Then we have

$$\begin{aligned} 0 \le \overline{\int_a^b} f(x)\,dx - \underline{\int_a^b} f(x)\,dx &= \left(\inf_Q U(f,Q)\right) - \left(\sup_Q L(f,Q)\right) \\ &\le U(f,P) - L(f,P) < \epsilon. \end{aligned}$$

Since the difference between the upper and lower Darboux integrals lies in $[0,\epsilon)$ for all $\epsilon > 0$, it must be 0, and so f is integrable. □

By thinking a little harder about this argument, we can also extract the value of $\int_a^b f(x)\,dx$ from the Darboux sums of any sequence of partitions that were used to prove the integrability of $f(x)$.

Proposition 3.13. *Given a sequence $\big(P_n\big)$ of partitions of the interval $[a,b]$ such that $\lim_{n\to\infty} \big(U(f,P_n) - L(f,P_n)\big) = 0$, we have*

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

Proof. The previous proposition shows that $f(x)$ is integrable on $[a,b]$. We have a sequence of inequalities

$$L(f,P_n) \le \underline{\int_a^b} f(x)\,dx = \int_a^b f(x)\,dx = \overline{\int_a^b} f(x)\,dx \le U(f,P_n)$$

for any n, so from $L(f,P_n) \le \int_a^b f(x)\,dx \le U(f,P_n)$ it follows immediately that

$$0 \le \int_a^b f(x)\,dx - L(f,P_n) \le U(f,P_n) - L(f,P_n).$$

The right side goes to zero as $n \to \infty$, hence so does $\int_a^b f(x)\,dx - L(f,P_n)$. The same argument shows that $U(f,P_n) - \int_a^b f(x)\,dx \to 0$. $\square$

Example 3.14. Let $f(x) = x^2$ on $[0,1]$, and consider the partitions $P_n = (0, \frac{1}{n}, \frac{2}{n}, \dots, \frac{n-1}{n}, 1)$. Then

$$\begin{aligned} L(f,P_n) &= \sum_{i=0}^{n-1} \left(\inf_{\frac{i}{n} \le t \le \frac{i+1}{n}} t^2 \right) \Delta x_i = \sum_{i=0}^{n-1} \left(\frac{i}{n} \right)^2 \frac{1}{n} = \frac{1}{n^3} \sum_{i=0}^{n-1} i^2 \\ U(f,P_n) &= \sum_{i=0}^{n-1} \left(\sup_{\frac{i}{n} \le t \le \frac{i+1}{n}} t^2 \right) \Delta x_i = \sum_{i=0}^{n-1} \left(\frac{i+1}{n} \right)^2 \frac{1}{n} = \frac{1}{n^3} \sum_{j=1}^{n} j^2 \end{aligned}$$

and so $U(f,P_n) - L(f,P_n) = \frac{1}{n^3}(n^2 - 0^2) = \frac{1}{n} \to 0$. Thus

$$\int_0^1 x^2\,dx = \lim_{n\to\infty} U(f,P_n) = \lim_{n\to\infty} \frac{1}{n^3} \frac{n(n+1)(2n+1)}{6} = \frac{1}{3}.$$

Example 3.15. Let $f(x) = \frac{1}{x}$ on $[1,b]$ for some integer $b > 1$, and let

$$P_n = \left(1, 1+\frac{1}{n}, 1+\frac{2}{n}, \dots, b-\frac{1}{n}, b \right).$$

Then we compute that

$$\begin{aligned} L(f,P_n) &= \sum_{i=0}^{(b-1)n-1} \left(\inf_{1+\frac{i}{n} \le t \le 1+\frac{i+1}{n}} \frac{1}{t} \right) \Delta x_i \\ &= \sum_{i=0}^{(b-1)n-1} \frac{1}{1+\frac{i+1}{n}} \cdot \frac{1}{n} = \sum_{i=0}^{(b-1)n-1} \frac{1}{n+i+1} \\ &= \frac{1}{n+1} + \frac{1}{n+2} + \cdots + \frac{1}{bn} \\ &= H_{bn} - H_n, \end{aligned}$$

where $H_k = \frac{1}{1} + \frac{1}{2} + \cdots + \frac{1}{k}$ is the kth harmonic sum. The same computation shows that

$$U(f,P_n) = \frac{1}{n} + \cdots + \frac{1}{bn-1} \;\Rightarrow\; U(f,P_n) - L(f,P_n) = \frac{1}{n} - \frac{1}{bn} \to 0,$$

so $f(x)$ is integrable, and

$$\int_1^b \frac{1}{x}\,dx = \lim_{n\to\infty} L(f,P_n) = \lim_{n\to\infty} \left(H_{bn} - H_n\right).$$

On a problem sheet we showed that $\gamma = \lim_{k\to\infty} \big(H_k - \log(k)\big)$ exists, so we have

$$\begin{aligned}
\lim_{n\to\infty} (H_{bn} - H_n) &= \lim_{n\to\infty} \Big(\big(H_{bn} - \log(bn)\big) - \big(H_n - \log(n)\big) + \log(b)\Big) \\
&= \lim_{n\to\infty} (H_{bn} - \log(bn)) - \lim_{n\to\infty} (H_n - \log(n)) + \log(b) \\
&= \gamma - \gamma + \log(b) = \log(b)
\end{aligned}$$

by the algebra of limits, and thus $\int_1^b \frac{1}{x}\,dx = \log(b)$.

The same criterion for integrability lets us prove the following important theorem, showing that the vast majority of familiar functions are integrable.

Theorem 3.16

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

Proof. We know that f is bounded by the extreme value theorem, and it is uniformly continuous since $[a,b]$ is compact. The latter says that given any $\epsilon > 0$, there is a $\delta > 0$ such that for all $x, y \in [a,b]$,

$$|x-y| < \delta \;\Rightarrow\; |f(x) - f(y)| < \frac{\epsilon}{b-a}.$$

We can choose a partition $P = (x_0, x_1, \ldots, x_n)$ of $[a,b]$ such that $\Delta x_i = x_{i+1} - x_i < \delta$ for all i, say by taking $x_0 = a$ and $x_n = b$ where $n = \lfloor \frac{2(b-a)}{\delta} \rfloor$, and then letting $x_i = a + \frac{i\delta}{2}$ for $1 \le i \le n-1$. Then for all i we have

$$y, z \in [x_i, x_{i+1}] \;\Rightarrow\; |y - z| \le \Delta x_i < \delta \;\Rightarrow\; |f(y) - f(z)| < \frac{\epsilon}{b-a}.$$

By the extreme value theorem, we can find $y, z \in [x_i, x_{i+1}]$ such that

$$M_i - m_i = \left(\sup_{x_i \le t \le x_{i+1}} f(t)\right) - \left(\inf_{x_i \le t \le x_{i+1}} f(t)\right) = f(y) - f(z),$$

and then $|y - z| < \delta$ implies that

$$0 \leq M_i - m_i < \frac{\epsilon}{b-a}.$$

We now estimate the difference between the upper and lower Darboux sums as

$$\begin{aligned} U(f,P) - L(f,P) &= \sum_{i=0}^{n-1} (M_i - m_i)\Delta x_i \\ &< \sum_{i=0}^{n-1} \frac{\epsilon}{b-a}(x_{i+1} - x_i) \\ &= \frac{\epsilon}{b-a} \sum_{i=0}^{n-1} (x_{i+1} - x_i) \\ &= \frac{\epsilon}{b-a}(x_n - x_0) = \frac{\epsilon}{b-a}(b-a) = \epsilon. \end{aligned}$$

Since such a P exists for any $\epsilon > 0$, we conclude that $f(x)$ is integrable on $[a,b]$. $\square$

The converse to this theorem is false, though: an integrable function need not be continuous.

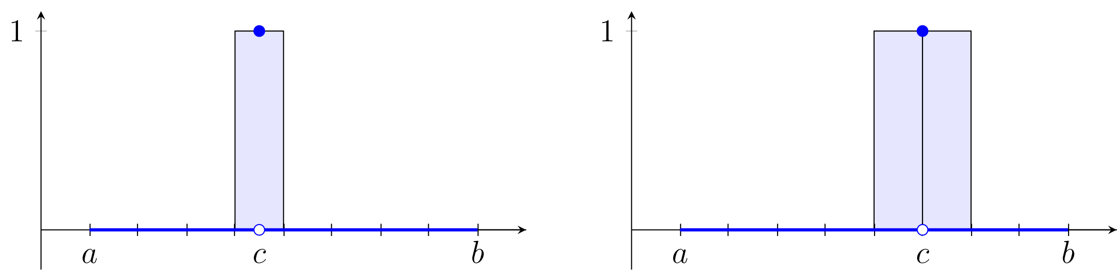
Example 3.17. Pick some $c \in [a,b]$ and define $f : [a,b] \to \mathbb{R}$ by

$$f(x) = \begin{cases} 0, & x \neq c \\ 1, & x = c. \end{cases}$$

Then for any partition P of $[a,b]$, we have $L(f,P) = 0$, because every interval $[x_i, x_{i+1}]$ contains a point $x \neq c$ where $f(t) = 0$. Thus $\underline{\int_a^b} f(x)\,dx = 0$. We also have

$$U(f,P) = \sum_{c \in [x_i, x_{i+1}]} 1 \cdot \Delta x_i,$$

and either one or two such closed intervals contain c – it's one if c is in the interior of an interval, and two if it's a common endpoint of two of them.



> We choose a partition P_n such that $\Delta x_i < \frac{1}{2n}$ for all i, and then we have $U(f, P_n) < 2\left(\frac{1}{2n}\right) = \frac{1}{n}$. It follows that $\overline{\int_a^b} f(x)\,dx \leq 0$, so in fact $f(x)$ is integrable and $\int_a^b f(x)\,dx = 0$.

Mentimeter question 13. Under which of the following circumstances must a bounded function $f : [0,1] \to \mathbb{R}$ be integrable?

1. f is differentiable on $(0,1)$.
2. f is monotone increasing.
3. f is discontinuous at finitely many points.
4. All of these. ✓
5. None of these.

If f is differentiable on $(0,1)$ then it is continuous there, so it is integrable. For monotone increasing functions, we take a partition

$$P_n = \left(0, \frac{1}{n}, \frac{2}{n}, \dots, \frac{n-1}{n}, 1\right)$$

and then we compute that

$$U(f, P_n) - L(f, P_n) = \sum_{i=0}^{n-1} \left(f\left(\frac{i+1}{n}\right) - f\left(\frac{i}{n}\right) \right) \frac{1}{n} = \frac{f(1) - f(0)}{n}$$

which goes to 0 as $n \to \infty$, so f is integrable. If f has finitely many discontinuities, say k of them, then we can try to repeat the proof that continuous functions are integrable, and the additional contribution to $U(f,P) - L(f,P)$ from the subintervals where f is discontinuous will be at most $k \cdot (\sup f - \inf f)\delta$, which vanishes as $\delta \to 0$.

3.3 Basic properties

In this section we'll establish some basic properties of the Darboux integral.

Proposition 3.18. *If $f, g : [a,b] \to \mathbb{R}$ are integrable and $f(x) \leq g(x)$ for all $x \in [a,b]$, then*

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

Proof. The inequality $f(x) \le g(x)$ implies that $L(f,P) \le L(g,P)$ for all partitions P of a, b, so then

$$\underline{\int_a^b} f(x)\,dx = \sup_P L(f,P) \le \sup_P L(g,P) = \underline{\int_a^b} g(x)\,dx.$$

The lower Darboux integrals on either side are equal to $\int_a^b f(x)\,dx$ and $\int_a^b g(x)\,dx$ respectively, so the proof is complete. □

The next theorem asserts that integration is a *linear* operator.

Theorem 3.19

If f and g are integrable on $[a,b]$, then

$$\int_a^b \big(cf(x) + dg(x)\big)\,dx = c\int_a^b f(x)\,dx + d\int_a^b g(x)\,dx$$

for any constants $c, d \in \mathbb{R}$.

The proof follows immediately from combining the next two propositions. Their proofs are a bit tedious, but they each follow the same general outline: take something we already know to be integrable, find a sequence of partitions so that the lower and upper Darboux sums converge to the integral, and then manipulate these Darboux sums to show that something else of interest is also integrable.

Proposition 3.20. *Let $f : [a,b] \to \mathbb{R}$ be integrable. Then $cf(x)$ is integrable on $[a,b]$ for any $c \in \mathbb{R}$, and*

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

Proof. We pick partitions P_n of $[a,b]$ with $U(f,P_n) - L(f,P_n) < \frac{1}{n}$ for all n. If $c \ge 0$ then we have

$$L(cf, P_n) = cL(f,P_n), \qquad U(cf,P_n) = cU(f,P_n)$$

and so $U(cf,P_n) - L(cf,P_n) < \frac{c}{n}$ for all n. Since this difference goes to zero as $n \to \infty$, we see that $cf(x)$ is integrable and

$$\int_a^b cf(x)\,dx = \lim_{n\to\infty} L(cf,P_n) = \lim_{n\to\infty} cL(f,P_n) = c\int_a^b f(x)\,dx.$$

If $c < 0$ then nearly the same argument applies, except we notice that

$$\inf_{x_i \le t \le x_{i+1}} cf(x) = c \left(\sup_{x_i \le t \le x_{i+1}} f(x) \right)$$

for all i, and this implies that $L(cf, P_n) = cU(f, P_n)$; similarly $U(cf, P_n) = cL(f, P_n)$. But we still have

$$U(cf, P_n) - L(cf, P_n) = -c\big(U(f, P_n) - L(f, P_n)\big) \to 0,$$

so $cf(x)$ is still integrable, and

$$\int_a^b cf(x)\,dx = \lim_{n\to\infty} L(cf, P_n) = \lim_{n\to\infty} cU(f, P_n) = c\int_a^b f(x)\,dx. \qquad \square$$

Proposition 3.21. *Let $f, g : [a, b] \to \mathbb{R}$ be integrable. Then $f + g$ is integrable on $[a, b]$, and*

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

Proof. We check the inequalities

$$\begin{aligned} \inf_{x\in S} \big(f(x) + g(x)\big) &\ge \left(\inf_{x\in S} f(x)\right) + \left(\inf_{x\in S} g(x)\right) \\ \sup_{x\in S} \big(f(x) + g(x)\big) &\le \left(\sup_{x\in S} f(x)\right) + \left(\sup_{x\in S} g(x)\right), \end{aligned}$$

which immediately imply for any partition P of $[a, b]$ that

$$L(f, P) + L(g, P) \le L(f + g, P), \qquad U(f + g, P) \le U(f, P) + U(g, P).$$

For any $n > 0$ there are partitions P_n and Q_n of $[a, b]$ such that

$$U(f, P_n) - L(f, P_n) < \frac{1}{2n}, \qquad U(g, Q_n) - L(g, Q_n) < \frac{1}{2n},$$

and if R_n is a common refinement of both P_n and Q_n then it follows that

$$\begin{aligned} U(f, R_n) - L(f, R_n) &\le U(f, P_n) - L(f, P_n) < \frac{1}{2n}, \\ U(g, R_n) - L(g, R_n) &\le U(g, Q_n) - L(g, Q_n) < \frac{1}{2n}. \end{aligned}$$

So then

$$\begin{aligned} U(f + g, R_n) - L(f + g, R_n) &\le \big(U(f, R_n) + U(g, R_n)\big) - \big(L(f, R_n) - L(g, R_n)\big) \\ &= \big(U(f, R_n) - L(f, R_n)\big) + \big(U(g, R_n) - L(g, R_n)\big) \\ &< \frac{1}{2n} + \frac{1}{2n} = \frac{1}{n}. \end{aligned}$$

Since we can do this for any n, it proves that $f+g$ is integrable. We have

$$\begin{aligned}\int_a^b \big(f(x)+g(x)\big)\,dx &= \lim_{n\to\infty} L(f+g,R_n)\\ &\geq \lim_{n\to\infty} L(f,R_n) + \lim_{n\to\infty} L(g,R_n)\\ &= \int_a^b f(x)\,dx + \int_a^b g(x)\,dx,\end{aligned}$$

and the same argument with upper Darboux sums instead of lower sums shows that

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

so the two sides are equal. □

Theorem 3.22

Let $f : [a,b] \to \mathbb{R}$ be integrable, and choose $c \in (a,b)$. Then f is integrable on each of $[a,c]$ and $[c,b]$, and

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

Proof. Since $f(x)$ is integrable on $[a,b]$, given any $n > 0$, we can find a partition P_n of $[a,b]$ such that

$$U(f,P_n) - L(f,P_n) < \frac{1}{n}.$$

We refine P_n to a partition Q_n which also contains c, and then

$$U(f,Q_n) - L(f,Q_n) \leq U(f,P_n) - L(f,P_n) < \frac{1}{n}.$$

Now $Q_n = (a, x_1, x_2, \ldots, x_{k-1}, c, x_{k+1}, \ldots, x_{m-1}, b)$ gives us partitions

$$\begin{aligned}Q_{1,n} &= (a, x_1, \ldots, x_{k-1}, c),\\ Q_{2,n} &= (c, x_{k+1}, \ldots, x_{m-1}, b)\end{aligned}$$

of $[a,c]$ and $[c,b]$ respectively, and by definition we have

$$\begin{aligned}L(f,Q_n) &= L(f,Q_{1,n}) + L(f,Q_{2,n}),\\ U(f,Q_n) &= U(f,Q_{1,n}) + U(f,Q_{2,n}).\end{aligned}$$

Each $U(f, Q_{i,n}) - L(f, Q_{i,n})$ is nonnegative, and their sum over $i = 1, 2$ is equal to $U(f, Q_n) - L(f, Q_n) < \frac{1}{n}$, so we must have

$$0 \leq U(f, Q_{i,n}) - L(f, Q_{i,n}) < \frac{1}{n}, \qquad i = 1, 2.$$

Since we can do this for any $n > 0$, it follows that $f(x)$ is integrable on each interval. We also have

$$\lim_{n\to\infty} L(f, Q_{1,n}) = \int_a^c f(x)\,dx, \qquad \lim_{n\to\infty} L(f, Q_{2,n}) = \int_c^b f(x)\,dx$$

and so the algebra of limits says that

$$\begin{aligned}\int_a^b f(x)\,dx &= \lim_{n\to\infty} L(f, Q_n) \\ &= \lim_{n\to\infty} \big(L(f, Q_{1,n}) + L(f, Q_{2,n})\big) \\ &= \int_a^c f(x)\,dx + \int_c^b f(x)\,dx.\end{aligned}$$

□

We can also compose continuous functions with integrable ones and the result will be integrable, even if we can't say much about the actual value of the integral.

Theorem 3.23

Let $f : [a, b] \to \mathbb{R}$ be integrable, with $m \leq f(x) \leq M$ for all $x \in [a, b]$, and let $g : [m, M] \to \mathbb{R}$ be a continuous function. Then

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

is also integrable on $[a, b]$.

Mentimeter question 14. Let $f : [a, b] \to [m, M]$ be integrable and let $g : [m, M] \to \mathbb{R}$ be continuous. Which must be true of $h(x) = g(f(x))$?

1. $h(x)$ is bounded. ✓
2. $h(x)$ is continuous.
3. $h(x)$ is monotone if $g(x)$ is.
4. More than one of these.
5. None of these.

We know that g is bounded because it is continuous on a compact interval, hence so is h. But if we take $g(x) = x$ then h won't be continuous or monotone unless f is.

Proof. Fix $\epsilon > 0$. Since $g(x)$ is continuous on a compact interval, we know the following:

- The extreme value theorem provides $x_{\min}, x_{\max} \in [m, M]$ which satisfy $g(x_{\min}) \leq g(x) \leq g(x_{\max})$ for all $x \in [m, M]$, and we set

$$C = g(x_{\max}) - g(x_{\min}) + 1 > 0;$$

- $g(x)$ is uniformly continuous, so there is a $\delta > 0$ such that

$$|x - y| < \delta \;\Rightarrow\; |g(x) - g(y)| < \frac{\epsilon}{2(b-a)}.$$

  We are allowed to replace δ with a smaller value, so we insist that $\delta < \frac{\epsilon}{2C}$.

Since $f(x)$ is integrable, we pick a partition $P = (x_0, x_1, \dots, x_n)$ of $[a, b]$ such that

$$U(f, P) - L(f, P) < \delta^2.$$

Letting $m_i = \inf_{x_i \leq t \leq x_{i+1}} f(t)$ and $M_i = \sup_{x_i \leq t \leq x_{i+1}} f(t)$, this means that

$$\sum_{i=0}^{n-1} (M_i - m_i)\Delta x_i < \delta^2.$$

We call an index i *good* if $M_i - m_i < \delta$ and *bad* otherwise. Our goal is to bound $U(h, P) - L(h, P)$ by separating the contributions from good and bad indices into two different sums. Each good interval will contribute a small amount by itself; we can't say this about the bad intervals, but we'll show instead that their total length is very small.

First, if i is good then for all $x, y \in [x_i, x_{i+1}]$ we have

$$|f(y) - f(x)| < \delta \;\Rightarrow\; |g(f(y)) - g(f(x))| < \frac{\epsilon}{2(b-a)},$$

and so

$$\left(\sup_{x_i \leq t \leq x_{i+1}} h(t)\right) - \left(\inf_{x_i \leq t \leq x_{i+1}} h(t)\right) \leq \frac{\epsilon}{2(b-a)}.$$

On the other hand, if i is bad then we only know that $\big(\sup h(t)\big) - \big(\inf h(t)\big) < C$, but summing over all bad i gives

$$\begin{aligned}\sum_{i \text{ bad}} \delta \cdot \Delta x_i &\leq \sum_{i \text{ bad}} (M_i - m_i)\Delta x_i \\ &\leq \sum_{i=0}^{n-1} (M_i - m_i)\Delta x_i = U(f,P) - L(f,P) < \delta^2\end{aligned}$$

and so $\displaystyle\sum_{i \text{ bad}} \Delta x_i < \delta < \frac{\epsilon}{2C}$.

Combining these two bounds, we get

$$\begin{aligned}U(h,P) - L(h,P) = &\sum_{i \text{ good}} \left(\left(\sup_{x_i \leq t \leq x_{i+1}} h(t)\right) - \left(\inf_{x_i \leq t \leq x_{i+1}} h(t)\right)\right)\Delta x_i \\ &+ \sum_{i \text{ bad}} \left(\left(\sup_{x_i \leq t \leq x_{i+1}} h(t)\right) - \left(\inf_{x_i \leq t \leq x_{i+1}} h(t)\right)\right)\Delta x_i \\ \leq &\sum_{i \text{ good}} \left(\frac{\epsilon}{2(b-a)}\right)\Delta x_i + \sum_{i \text{ bad}} C \cdot \Delta x_i \\ < &\left(\frac{\epsilon}{2(b-a)}\right)(b-a) + C\left(\frac{\epsilon}{2C}\right) \\ = &\ \epsilon.\end{aligned}$$

We have thus found for any $\epsilon > 0$ a partition P with $U(h,P) - L(h,P) < \epsilon$, and it follows that $h(x)$ is integrable on $[a,b]$. $\square$

This is a very general result, but the following special cases are interesting.

Proposition 3.24. *If $f : [a,b] \to \mathbb{R}$ is integrable, then $|f|$ is also integrable on $[a,b]$, and we have a* triangle inequality *for integrals:*

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

Proof. The previous theorem tells us that $|f|$ is integrable, since we get it by composing f with the continuous function $|x|$. The inequality

$$-|f(x)| \leq f(x) \leq |f(x)|$$

for all $x \in [a,b]$ then immediately implies that

$$-\int_a^b |f(x)|\,dx = \int_a^b (-|f(x)|)\,dx \leq \int_a^b f(x)\,dx \leq \int_a^b |f(x)|\,dx,$$

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

Proposition 3.25. *If $f, g : [a,b] \to \mathbb{R}$ are integrable, then so is $fg : [a,b] \to \mathbb{R}$.*

Proof. We note that if a function $h(x)$ is integrable on a given domain then so is $h(x)^2$, since it is the composition of $h(x)$ with the continuous function x^2. Now both $\frac{f+g}{2}$ and $\frac{f-g}{2}$ are integrable on $[a,b]$, as linear combinations of the integrable functions $f(x)$ and $g(x)$, and so we use the identity

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

to see that $f(x)g(x)$ is integrable as well. □

3.4 The fundamental theorem of calculus

We've developed a long list of basic properties of integrals, but we haven't managed to compute that many so far. The fundamental theorem of calculus is our best tool for doing so, and it also illustrates the strong relationship between derivatives and integrals.

Theorem 3.26: Fundamental theorem of calculus, first version

Let $f : [a,b] \to \mathbb{R}$ be a continuous function, and define $F : [a,b] \to \mathbb{R}$ by

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

Then F is continuous on $[a,b]$ and differentiable on (a,b), and $F'(x) = f(x)$ for all $x \in (a,b)$.

Proof. For any $x \in (a,b)$ and $h > 0$, we use basic properties of the Darboux integral to see that

$$\begin{aligned}\frac{F(x+h)-F(x)}{h} &= \frac{1}{h}\left(\int_a^{x+h} f(t)\,dt - \int_a^x f(t)\,dt\right) \\ &= \frac{1}{h}\int_x^{x+h} f(t)\,dt.\end{aligned}$$

Similarly, if $h < 0$ then $\frac{F(x+h)-F(x)}{h} = -\frac{1}{h}\int_{x+h}^{x} f(t)\,dt$.

Fixing $\epsilon > 0$, we know since f is continuous at x that there is $\delta > 0$ such that

$$|t - x| < \delta \;\Rightarrow\; |f(t) - f(x)| < \epsilon,$$

or equivalently

$$|t - x| < \delta \;\Rightarrow\; f(x) - \epsilon < f(t) < f(x) + \epsilon.$$

If $0 < h < \delta$ then this holds for all $t \in [x, x+h]$, and so

$$\int_x^{x+h} \big(f(x) - \epsilon\big)\, dt < \int_x^{x+h} f(t)\,dt < \int_x^{x+h} \big(f(x) + \epsilon\big)\, dt,$$

and since $f(x) \pm \epsilon$ is constant *as a function of* t this simplifies to

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

The middle term is $F(x+h) - F(x)$, so upon dividing by h and subtracting $f(x)$ from each side, we have shown that

$$0 < h < \delta \;\Rightarrow\; \left|\frac{F(x+h) - F(x)}{h} - f(x)\right| < \epsilon.$$

The same argument applies when $-\delta < h < 0$, and this works for any $\epsilon > 0$, so

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

Since F is differentiable on (a, b) it is continuous there, though we still need to prove continuity at a and b. Using the continuity of f at a, the same argument as before shows us that for any $\epsilon > 0$ there is a $\delta > 0$ such that

$$\begin{aligned} 0 < h < \delta \;&\Rightarrow\; \left|\frac{F(a+h) - F(a)}{h} - f(a)\right| < \epsilon \\ &\Rightarrow\; |F(a+h) - F(a) - hf(a)| < h\epsilon. \end{aligned}$$

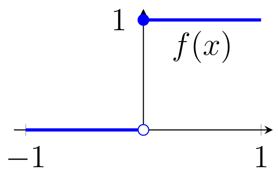
If we take $0 < h < \delta' = \min(\delta, \frac{\epsilon}{|f(a)|+\epsilon})$ and apply the triangle inequality then

$$|F(a+h) - F(a)| < h(|f(a)| + \epsilon) < \epsilon,$$

and we can find such a $\delta' > 0$ for every $\epsilon > 0$, so F is continuous at a; continuity at b follows from the same argument. $\square$

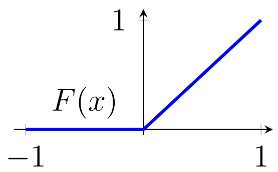
The assumption that f is continuous is crucial here. For example, consider the function $f : [-1, 1] \to \mathbb{R}$ given by

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



If we let $F(x) = \int_{-1}^{x} f(t)\, dt$, then we have

$$\begin{aligned} x < 0 \;&\Rightarrow\; F(x) = \int_{-1}^{x} 0\, dt = 0 \\ x \geq 0 \;&\Rightarrow\; F(x) = \int_{-1}^{0} 0\, dt + \int_{0}^{x} 1\, dt = x. \end{aligned}$$



Here $f(x)$ is not continuous at 0, and as a result F is not differentiable at 0 since

$$\lim_{x \uparrow 0} \frac{F(x) - F(0)}{x - 0} = 0 \quad \text{but} \quad \lim_{x \downarrow 0} \frac{F(x) - F(0)}{x - 0} = \lim_{x \downarrow 0} \frac{x - 0}{x - 0} = 1.$$

Theorem 3.27: Fundamental theorem of calculus, second version

Let $f : [a, b] \to \mathbb{R}$ be a continuous function which has a continuous derivative on (a, b). Then

$$\int_a^b f'(x)\, dx = f(b) - f(a).$$

Proof. Let $F(x) = \int_a^x f'(t)\, dt$. Then the first version of the fundamental theorem of calculus says that

$$F'(x) = f'(x) \text{ for all } x \in (a, b),$$

and both F and f are continuous on $[a, b]$, so there is some constant c such that

$$f(x) = F(x) + c$$

on $[a, b]$. Setting $x = a$ shows us that $c = f(a)$, since $F(a) = 0$, and so

$$\int_a^b f'(x)\, dx = F(b) = f(b) - f(a). \qquad \square$$

Mentimeter question 15. Let $f, g : [a, b] \to \mathbb{R}$ be continuous. Which of the following is a consequence of the fundamental theorem of calculus?

1. There is an $h : [a, b] \to \mathbb{R}$ such that $f = h''$ on (a, b). ✓
2. If $f'(x)$ exists on (a, b), then f' is integrable.
3. If f and g are continuously differentiable on (a, b) and $f'(x) = g'(x)$ for all x, then $f(x) = g(x)$ for all x.
4. More than one of these.

For the first option, we can take $F(x) = \int_a^x f(t)\,dt$ and $g(x) = \int_a^x F(t)\,dt$, and then $g'' = F' = f$ on (a, b). The second is false because we need $f'(x)$ to be continuous: if we take

$$f(x) = \begin{cases} x^2 \sin(\frac{1}{x^2}), & x \neq 0 \\ 0, & x = 0 \end{cases} \Rightarrow f'(x) = \begin{cases} 2x \sin(\frac{1}{x^2}) - \frac{2}{x}\cos(\frac{1}{x^2}), & x \neq 0 \\ 0, & x = 0 \end{cases}$$

then f' exists on $(-1, 1)$ but is discontinuous at $x = 0$, and it is not integrable on $[-1, 1]$ because it is unbounded. The third option is false because we could have $g = f + 1$.

Example 3.28. Suppose we wish to compute $\int_a^b x^n\,dx$ for some integer $n \neq -1$, and that if $n < 0$ then $0 \notin [a, b]$. Since $\frac{x^{n+1}}{n+1}$ has derivative x^n, the fundamental theorem of calculus tells us that

$$\int_a^b x^n\,dx = \frac{b^{n+1}}{n+1} - \frac{a^{n+1}}{n+1}.$$

For the case $n = -1$, assuming that $0 < a < b$, we note that $\log(x)$ has derivative $\frac{1}{x}$, and so

$$\int_a^b \frac{1}{x}\,dx = \log(b) - \log(a).$$

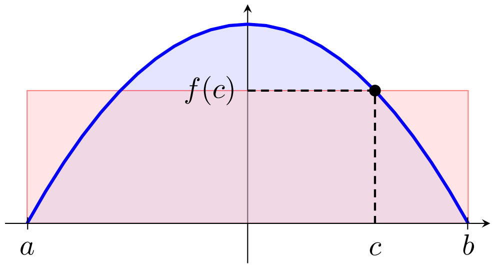
The first version of the fundamental theorem of calculus gives us the following new version of the mean value theorem.

Theorem 3.29: Mean value theorem for integrals

Let $f : [a, b] \to \mathbb{R}$ be a continuous function. Then there is some $c \in (a, b)$ such

that
$$\int_a^b f(x)\,dx = f(c)(b-a).$$

In other words, the two overlapping shaded regions below have the same area:



Proof. Let $F(x) = \int_a^x f(t)\,dt$. Then $F(x)$ is continuous on $[a,b]$ and differentiable on (a,b), so the original mean value theorem gives us some $c \in (a,b)$ such that
$$F'(c) = \frac{F(b)-F(a)}{b-a}.$$

But we know that $F'(c) = f(c)$, so we conclude that
$$\int_a^b f(x)\,dx = F(b)-F(a) = F'(c)(b-a) = f(c)(b-a). \qquad \square$$

3.5 More properties of integrals

The fundamental theorem of calculus allows us to develop some new tools for evaluating integrals.

Theorem 3.30: Integration by parts

If $f,g : [a,b] \to \mathbb{R}$ are continuous with continuous first derivatives, then
$$\int_a^b f(x)g'(x)\,dx = f(b)g(b) - f(a)g(a) - \int_a^b f'(x)g(x)\,dx.$$

Proof. We start with the product rule for derivatives:
$$\frac{d}{dx}(f(x)g(x)) = f'(x)g(x) + f(x)g'(x).$$

Since f' and g' are continuous, so are $f'g$, fg', and $(fg)' = f'g + fg'$, so they're all integrable and the second fundamental theorem of calculus says that

$$\int_a^b (f(x)g(x))'\,dx = f(b)g(b) - f(a)g(a),$$

and hence

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

We subtract $\int_a^b f'(x)g(x)\,dx$ from both sides to complete the proof. □

In practice, given a function $f : [a,b] \to \mathbb{R}$ we may write $f(x)\big|_{x=a}^{x=b}$ as shorthand for $f(b) - f(a)$, as in

$$\int_a^b f'(x)\,dx = f(x)\big|_{x=a}^{x=b}.$$

Example 3.31. We can integrate $\log(x)$ by parts, using $f(x) = \log(x)$ and $g(x) = x$:

$$\begin{aligned}\int_a^b \log(x)\cdot 1\,dx &= x\log(x)\big|_{x=a}^{x=b} - \int_a^b \frac{1}{x}\cdot x\,dx \\ &= (b\log(b) - a\log(a)) - (b-a) \\ &= (b\log(b) - b) - (a\log(a) - a).\end{aligned}$$

This computation shows that $x\log(x) - x$ is an *antiderivative* of $\log(x)$, meaning that its derivative is $\log(x)$.

From here on we may sometimes accidentally write down an integral of the form $\int_a^b f(x)\,dx$ where $a > b$. In this case we make sense of it by defining

$$\int_a^b f(x)\,dx = -\int_b^a f(x)\,dx,$$

and then it is not hard to check that all of the properties we have developed for the Darboux integral remain true even when $a > b$. For example, this can happen in the following theorem statement when ϕ is monotone decreasing.

Theorem 3.32: Integration by substitution

Let $f : [a,b] \to \mathbb{R}$ be a continuous function, and suppose that $\phi : [c,d] \to [a,b]$

has a continuous derivative on (c, d). Then

$$\int_{\phi(c)}^{\phi(d)} f(x)\,dx = \int_c^d f(\phi(t))\phi'(t)\,dt.$$

Proof. Let $F(x) = \int_a^x f(t)\,dt$, which is an antiderivative of f. Then

$$\frac{d}{dx}(F(\phi(x))) = F'(\phi(x))\phi'(x) = f(\phi(x))\phi'(x),$$

which is continuous on (c, d), so by the fundamental theorem of calculus we have

$$\begin{aligned}\int_c^d f(\phi(t))\phi'(t)\,dt &= F(\phi(d)) - F(\phi(c)) \\ &= \int_a^{\phi(d)} f(t)\,dt - \int_a^{\phi(c)} f(t)\,dt \\ &= \int_{\phi(c)}^{\phi(d)} f(t)\,dt.\end{aligned}$$

□

We can use substitution to find antiderivatives of several new functions.

Example 3.33. We evaluate $\int_e^x \frac{1}{t\log(t)}\,dt$ by substituting $t = e^s$. Then

$$\int_e^x \frac{1}{t\log(t)}\,dt = \int_1^{\log(x)} \frac{1}{e^s\log(e^s)} \cdot e^s\,ds = \int_1^{\log(x)} \frac{1}{s}\,ds = \log(s)\Big|_{s=1}^{s=\log(x)},$$

so $\log(\log(x))$ is an antiderivative of $\frac{1}{x\log(x)}$.

Example 3.34. We evaluate $\int_0^x \frac{1}{\sqrt{1-t^2}}\,dt$ for $x \in (-1, 1)$ by substituting $t = \sin(\theta)$. Then we have

$$\begin{aligned}\int_0^x \frac{1}{\sqrt{1-t^2}}\,dt &= \int_0^{\sin^{-1}(x)} \frac{1}{\sqrt{1-\sin^2(\theta)}} \cdot \cos(\theta)\,d\theta \\ &= \int_0^{\sin^{-1}(x)} \frac{\cos(\theta)}{\cos(\theta)}\,d\theta \\ &= \int_0^{\sin^{-1}(x)} 1\,d\theta = \sin^{-1}(x).\end{aligned}$$

So $\sin^{-1} : [-1, 1] \to [-\frac{\pi}{2}, \frac{\pi}{2}]$ is differentiable on $(-1, 1)$, with derivative $\frac{1}{\sqrt{1-x^2}}$.

Example 3.35. We evaluate $\int_0^x \frac{1}{1+t^2}\,dt$ by substituting $t = \tan(\theta)$. Then

$$\frac{dt}{d\theta} = \frac{d}{d\theta}\left(\frac{\sin(\theta)}{\cos(\theta)}\right) = \frac{\cos^2(\theta) + \sin^2(\theta)}{\cos^2(\theta)} = \frac{1}{\cos^2(\theta)}$$

by the quotient rule for derivatives, so we have

$$\begin{aligned}\int_0^x \frac{1}{1+t^2}\,dt &= \int_0^{\tan^{-1}(x)} \frac{1}{1+\tan^2(\theta)} \cdot \frac{1}{\cos^2(\theta)}\,d\theta \\ &= \int_0^{\tan^{-1}(x)} \frac{1}{\cos^2(\theta) + \sin^2(\theta)}\,d\theta \\ &= \int_0^{\tan^{-1}(x)} d\theta = \tan^{-1}(x),\end{aligned}$$

and so $\tan^{-1} : \mathbb{R} \to (-\frac{\pi}{2}, \frac{\pi}{2})$ is differentiable, with derivative $\frac{1}{1+x^2}$.

3.6 Limits of integrable functions

We have seen that very few things commute with pointwise convergence of functions: a sequence of continuous functions need not converge to something continuous, and even if a sequence of differentiable functions f_n converges pointwise to a differentiable f then $f'(x)$ can be different from the pointwise limit $\lim\limits_{n\to\infty} f_n'(x)$. The same is true for integration.

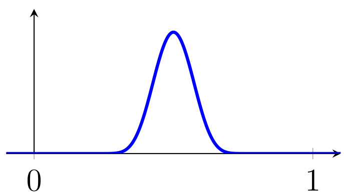
Example 3.36. We will construct a sequence of functions $f_n : [0, 1] \to \mathbb{R}$ such that

$$\lim_{n\to\infty} \int_a^b f_n(x)\,dx \neq \int_a^b \left(\lim_{n\to\infty} f_n(x)\right) dx,$$

even though the functions f_n and their pointwise limit $f(x) = \lim\limits_{n\to\infty} f_n(x)$ are all not just continuous but infinitely differentiable.

We showed in a problem sheet that the function

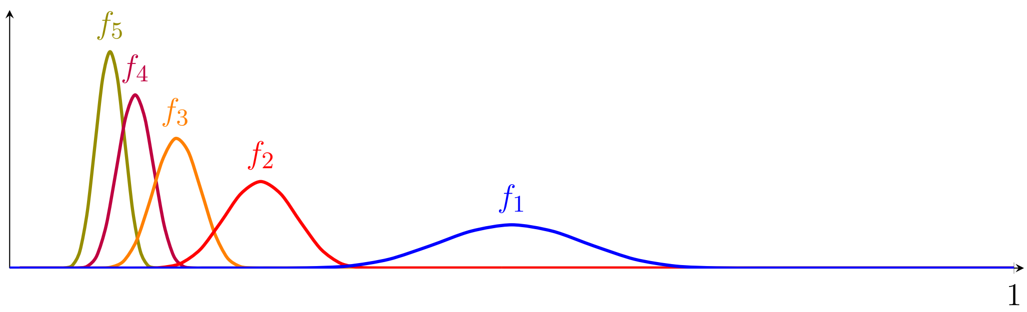
$$\phi(x) = \begin{cases} e^{-\frac{1}{x^2} - \frac{1}{(1-x)^2}}, & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$



is infinitely differentiable, and it is nonzero iff $0 < x < 1$. Let $c = \int_0^1 \phi(x)\,dx$. For each $n \in \mathbb{N}$ we define a function $f_n : [0,1] \to \mathbb{R}$ by

$$f_n(x) = \frac{n}{c} \cdot \phi(nx),$$

so $f_n(x) = 0$ except on the interval $(0, \frac{1}{n})$.



We compute by substitution that for all n,

$$\begin{aligned}\int_0^1 f_n(x)\,dx &= \frac{1}{c}\int_0^1 \phi(nx)\cdot n\,dx \\ &= \frac{1}{c}\int_0^n \phi(y)\,dy \\ &= \frac{1}{c}\int_0^1 \phi(y)\,dy = 1.\end{aligned}$$

But at the same time, the pointwise limit $f(x) = \lim\limits_{n\to\infty} f_n(x)$ is identically zero: this is clear for $x = 0$ since $f_n(0) = 0$ for all n, and if $x > 0$ instead then we have $f_n(x) = 0$ for all $n > \frac{1}{x}$. So

$$\int_0^1 f(x)\,dx = 0 \neq 1 = \lim_{n\to\infty} \int_0^1 f_n(x)\,dx.$$

In fact, we can do even worse: the functions $nf_n(x)$ converge pointwise to $f(x) = 0$ for the same reason, but $\int_0^1 nf_n(x)\,dx = n \to \infty$.

Just as with continuity and differentiation, however, it turns out that uniform convergence saves the day.

Theorem 3.37

Let $f_n : [a,b] \to \mathbb{R}$ be a sequence of integrable functions which converges uniformly to $f : [a,b] \to \mathbb{R}$. Then f is integrable, and

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

Proof. The key idea is that if f_n is very close to f when n is large, then the upper and lower Darboux sums for f must be very close to those for f_n, and so the upper and lower Darboux integrals for f get arbitrarily close to $\int_a^b f_n(x)\,dx$ as $n \to \infty$.

Fix $\epsilon > 0$. Uniform convergence gives us some $N > 0$ such that if $n \geq N$, then

$$|f_n(x) - f(x)| < \frac{\epsilon}{2(b-a)} \quad \forall x \in [a,b].$$

For any $n \geq N$ and any partition $P = (x_0, \ldots, x_k)$ of $[a,b]$, we have

$$\inf_{x_i \leq t \leq x_{i+1}} f(t) \geq \left(\inf_{x_i \leq t \leq x_{i+1}} f_n(t) \right) - \frac{\epsilon}{2(b-a)},$$

for all i, so then the lower Darboux sums of f and f_n satisfy

$$\begin{aligned} L(f,P) &= \sum_{i=0}^{k-1} \left(\inf_{x_i \leq t \leq x_{i+1}} f(t) \right) \Delta x_i \\ &\geq \sum_{i=0}^{k-1} \left(\left(\inf_{x_i \leq t \leq x_{i+1}} f_n(t) \right) - \frac{\epsilon}{2(b-a)} \right) \Delta x_i \\ &= L(f_n,P) - \frac{\epsilon}{2(b-a)} \sum_{i=0}^{k-1} \Delta x_i \\ &= L(f_n,P) - \frac{\epsilon}{2}. \end{aligned}$$

Taking suprema over all P gives us

$$\underline{\int_a^b} f(x)\,dx \geq \underline{\int_a^b} f_n(x)\,dx - \frac{\epsilon}{2}$$

for all $n \geq N$. We can also apply the same argument with the bound

$$\sup_{x_i \leq t \leq x_{i+1}} f(t) \leq \left(\sup_{x_i \leq t \leq x_{i+1}} f_n(t) \right) + \frac{\epsilon}{2(b-a)},$$

taking infima of the corresponding upper Darboux sums over all P, to conclude that

$$\overline{\int_a^b} f(x)\,dx \leq \overline{\int_a^b} f_n(x)\,dx + \frac{\epsilon}{2}$$

for all $n \geq N$. But since the f_n are integrable, their lower and upper Darboux integrals are equal, and we now have

$$\int_a^b f_n(x)\,dx - \frac{\epsilon}{2} \leq \underline{\int_a^b} f(x)\,dx \leq \overline{\int_a^b} f(x)\,dx \leq \int_a^b f_n(x)\,dx + \frac{\epsilon}{2}$$

for all $n \geq N$.

This last chain of inequalities lets us conclude two things. First, we have

$$0 \leq \overline{\int_a^b} f(x)\,dx - \underline{\int_a^b} f(x)\,dx \leq \epsilon,$$

and we can show this for any $\epsilon > 0$, so in fact the upper and lower Darboux integrals must be equal. This proves that f is integrable. Second, given $\epsilon > 0$ we have found N such that

$$n \geq N \;\Rightarrow\; \left| \int_a^b f(x)\,dx - \int_a^b f_n(x)\,dx \right| \leq \frac{\epsilon}{2} < \epsilon.$$

Since we can do this for any $\epsilon > 0$, this is precisely what is needed to show that

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

as $n \to \infty$. $\square$

One useful application of this theorem is to finding antiderivatives of power series.

Proposition 3.38. *Let $f(x) = \sum_{n=0}^{\infty} a_n x^n$ be a power series with radius of convergence $R > 0$. Then f is integrable on any closed subinterval of $(-R, R)$, with*

$$\int_0^x f(t)\,dt = \sum_{n=0}^{\infty} \frac{a_n}{n+1} x^{n+1}$$

for any $x \in (-R, R)$.

Proof. Fix $x \in (-R, R)$. We know that the partial sums

$$f_n(t) = \sum_{i=0}^{n} a_i t^i$$

converge uniformly to f on the interval $[-x, x]$ and thus on the subinterval $[0, x]$, so it follows that f is integrable on $[0, x]$ and that

$$\sum_{i=0}^{\infty} \frac{a_i}{i+1} x^{i+1} = \lim_{n\to\infty} \sum_{i=0}^{n} \frac{a_i}{i+1} x^{i+1} = \lim_{n\to\infty} \int_0^x f_n(t)\,dt = \int_0^x f(t)\,dt,$$

where the last equality uses the uniform convergence of $f_n \to f$ on $[0, x]$. $\square$

Example 3.39. We integrate the power series $\frac{1}{1-x} = \sum_{n=0}^{\infty} x^n$, with radius of convergence 1, to get

$$\log\left(\frac{1}{1-x}\right) = \sum_{n=0}^{\infty} \frac{x^{n+1}}{n+1} = \frac{x}{1} + \frac{x^2}{2} + \frac{x^3}{3} + \frac{x^4}{4} + \dots$$

for all $x \in (-1, 1)$. Taking $x = \frac{1}{2}$ gives us

$$\log(2) = \sum_{n=1}^{\infty} \frac{1}{n \cdot 2^n},$$

which converges very quickly compared to the alternating harmonic series $\log(2) = \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n}$, since the error after the first k terms is

$$\sum_{n=k+1}^{\infty} \frac{1}{n \cdot 2^n} \leq \sum_{n=k+1}^{\infty} \frac{1}{(k+1)2^{k+1}} \cdot \frac{1}{2^{n-(k+1)}} = \frac{1}{(k+1)2^k}.$$

Taking just the first five terms gives

$$\log(2) \approx \frac{1}{2} + \frac{1}{8} + \frac{1}{24} + \frac{1}{64} + \frac{1}{160} = \frac{661}{960} = 0.688541\overline{6}$$

to within at most $\frac{1}{6 \cdot 2^5} = \frac{1}{192} = 0.005208\overline{3}$. (The actual value is $0.6931\dots$.)

3.7 Improper integrals

Occasionally we would like to evaluate integrals $\int_a^b f(x)\,dx$ which don't quite have the form we're used to: either f is unbounded on $[a, b]$, or $[a, b]$ has infinite length because either $a = -\infty$ or $b = \infty$ (or both). In this case, we can still try to make sense of this using limits.

Definition. Suppose $f : (a, b] \to \mathbb{R}$ is integrable on every subinterval $[c, b] \subset (a, b]$. Then we define the *improper integral*

$$\int_a^b f(x)\,dx = \lim_{c \downarrow a} \int_c^b f(x)\,dx$$

if the limit exists, and otherwise we say that it diverges or fails to exist.

Likewise, if $f : [a, b) \to \mathbb{R}$ is integrable on every $[a, c] \subset [a, b)$ then we define

$$\int_a^b f(x)\,dx = \lim_{c\uparrow b} \int_a^c f(x)\,dx.$$

If $a, b \in \mathbb{R}$ and f extends to an integrable function $\tilde{f} : [a, b] \to \mathbb{R}$, then this is no different from the usual integral. To see this, we take an $M > 0$ such that $|\tilde{f}(x)| \leq M$ for all $x \in [a, b]$, and then

$$\left|\int_a^c \tilde{f}(x)\,dx\right| \leq \int_a^c |\tilde{f}(x)|\,dx \leq \int_a^c M\,dx = M(c-a),$$

which implies that $\lim_{c\downarrow a} \int_a^c \tilde{f}(x)\,dx = 0$. Then we have

$$\begin{aligned}\lim_{c\downarrow a} \int_c^b f(x)\,dx &= \int_a^b \tilde{f}(x)\,dx - \lim_{c\downarrow a}\int_a^c \tilde{f}(x)\,dx \\ &= \int_a^b \tilde{f}(x)\,dx.\end{aligned}$$

Improper integrals are more interesting when $\tilde{f}$ doesn't exist, such as when either f or its domain is unbounded.

Example 3.40. If $0 < r < 1$ then $\frac{1}{x^r}$ is unbounded as $x \downarrow 0$ and undefined at $x = 0$. But on any interval $[\epsilon, 1]$ with $\epsilon > 0$, we have

$$\int_\epsilon^1 \frac{1}{x^r}\,dx = \frac{x^{1-r}}{1-r}\bigg|_{x=\epsilon}^{x=1} = \frac{1-\epsilon^{1-r}}{1-r},$$

and this converges as $\epsilon \downarrow 0$, so we have the improper integral

$$\int_0^1 \frac{1}{x^r}\,dx = \lim_{\epsilon\downarrow 0}\int_\epsilon^1 \frac{1}{x^r}\,dx = \frac{1}{1-r}.$$

On the other hand, if $r > 1$ then the limit as $\epsilon \downarrow 0$ does not exist, and likewise if $r = 1$ then

$$\lim_{\epsilon\downarrow 0}\int_\epsilon^1 \frac{1}{x}\,dx = \lim_{\epsilon\downarrow 0}\big(-\log(\epsilon)\big) = \infty.$$

So $\int_0^1 \frac{1}{x^r}\,dx$ exists for $r > 0$ if and only if $0 < r < 1$.

Example 3.41. We define the improper integral $\int_0^\infty \frac{1}{1+x^2}\,dx$ as a limit:

$$\begin{aligned}\int_0^\infty \frac{1}{1+x^2}\,dx &= \lim_{b\to\infty}\int_0^b \frac{1}{1+x^2}\,dx \\ &= \lim_{b\to\infty} \tan^{-1}(x)\Big|_{x=0}^{x=b} \\ &= \lim_{b\to\infty} \tan^{-1}(b).\end{aligned}$$

Since the limit exists and is equal to $\frac{\pi}{2}$, we have $\displaystyle\int_0^\infty \frac{1}{1+x^2}\,dx = \frac{\pi}{2}.$

Mentimeter question 16. For which of the following $f : [1,\infty) \to \mathbb{R}$ does $\int_1^\infty f(x)\,dx$ exist? Choose all that apply.

1. $f(x) = e^{-x}$ ✓
2. $f(x) = \cos(x)$
3. $f(x) = x^{-3}$ ✓
4. $f(x) = \frac{1}{x\log(x)}$

Likewise, we may be interested in an integral $\int_a^b f(x)\,dx$, where f is unbounded at some interior point $c \in (a,b)$ or where both $a = -\infty$ and $b = \infty$. In this case we define the improper integral

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

as a sum of two improper integrals, and we require that both of them exist; we leave it as an exercise to check that the answer does not depend on c.

If f has multiple improprieties (not an official term), meaning points in the domain where f is unbounded or limits of integration at $\pm\infty$, then we similarly split it into a sum of improper integrals $\int_c^d f(x)\,dx$ with at most one impropriety each, and then we take the sum of each of these if they are defined.

Example 3.42. The function $f(x) = \frac{1}{\sqrt{|x|}}$ is unbounded as $x \to 0$, so we define

$$\int_{-1}^1 \frac{1}{\sqrt{|x|}}\,dx = \int_{-1}^0 \frac{1}{\sqrt{|x|}}\,dx + \int_0^1 \frac{1}{\sqrt{|x|}}\,dx.$$

The second of these summands is

$$\lim_{a\downarrow 0}\int_a^1 \frac{1}{\sqrt{x}}\,dx = \lim_{a\downarrow 0} 2\sqrt{x}\Big|_{x=a}^{x=1} = \lim_{a\downarrow 0}\left(2-2\sqrt{a}\right) = 2,$$

and substituting $y=-x$ shows that the first summand is the limit as $a\downarrow 0$ of

$$\int_{-1}^{-a}\frac{1}{\sqrt{-x}}\,dx = -\int_1^a \frac{1}{\sqrt{y}}\,dy = \int_a^1 \frac{1}{\sqrt{y}}\,dy \to 2,$$

so the original improper integral exists and we have

$$\int_{-1}^1 \frac{1}{\sqrt{|x|}}\,dx = 4.$$

Example 3.43. We have

$$\int_{-\infty}^{\infty}\frac{1}{1+x^2}\,dx = \int_{-\infty}^{0}\frac{1}{1+x^2}\,dx + \int_0^{\infty}\frac{1}{1+x^2}\,dx.$$

We have already evaluated the second summand as $\frac{\pi}{2}$, and the first is also $\frac{\pi}{2}$ by essentially the same argument, so

$$\int_{-\infty}^{\infty}\frac{1}{1+x^2}\,dx = \frac{\pi}{2}+\frac{\pi}{2} = \pi.$$

Example 3.44. For $r>0$ we can consider the integral

$$\int_{-1}^1 \frac{1}{x}\,dx = \int_{-1}^0 \frac{1}{x}\,dx + \int_0^1 \frac{1}{x}\,dx.$$

This does not exist, because neither improper integral on the right converges.

Note that if we tried to evaluate this using a single limit for both summands, the answer would depend on how we approached 0 from either side. We have

$$\lim_{\epsilon\downarrow 0}\left(\int_{-1}^{-\epsilon}\frac{1}{x}\,dx + \int_{\epsilon}^1 \frac{1}{x}\,dx\right) = \lim_{\epsilon\downarrow 0}\left(\log(\epsilon) + (-\log(\epsilon))\right) = 0,$$

but

$$\lim_{\epsilon \downarrow 0} \left(\int_{-1}^{-2\epsilon} \frac{1}{x}\, dx + \int_{\epsilon}^{1} \frac{1}{x}\, dx \right) = \lim_{\epsilon \downarrow 0} \big(\log(2\epsilon) + (-\log(\epsilon)) \big)$$
$$= \lim_{\epsilon \downarrow 0} \log(2) = \log(2)$$

and similarly

$$\lim_{\epsilon \downarrow 0} \left(\int_{-1}^{-\epsilon^2} \frac{1}{x}\, dx + \int_{\epsilon}^{1} \frac{1}{x}\, dx \right) = \lim_{\epsilon \downarrow 0} \big(\log(\epsilon^2) + (-\log(\epsilon)) \big)$$
$$= \lim_{\epsilon \downarrow 0} \log(\epsilon) = -\infty.$$

Thus the only sensible thing to do if we want the original improper integral $\int_{-1}^{1} \frac{1}{x}\, dx$ to exist and be well-defined is to insist that each individual limit exists and then take their sum. In this case, neither $\int_{-1}^{0} \frac{1}{x}\, dx$ nor $\int_{0}^{1} \frac{1}{x}\, dx$ exists, so we can't hope to make sense of $\int_{-1}^{1} \frac{1}{x}\, dx$.

Virtually all of our properties of Darboux integrals apply to improper integrals as well; we just need to check that all of the limits involved are well-defined. For example, we have the second fundamental theorem of calculus: if $f : [a, \infty) \to \mathbb{R}$ is continuous and has a continuous derivative on (a, ∞), then for all $b > a$ we have

$$\int_a^b f'(x)\, dx = f(b) - f(a),$$

so $\int_a^\infty f'(x)\, dx$ exists if and only if $\lim\limits_{b \to \infty} f(b)$ exists, and then

$$\int_a^\infty f'(x)\, dx = f(x)\Big|_{x=a}^{x=\infty} = \left(\lim_{x \to \infty} f(x) \right) - f(a).$$

Here we introduce the notation "$\Big|^{x=\infty}$" to mean that we take the limit as $x \to \infty$.

We also have the following useful criterion for integrability.

Proposition 3.45. *Let $f, g : [a, \infty) \to \mathbb{R}$ be functions satisfying $0 \le f(x) \le g(x)$ for all $x \ge a$, and which are both integrable on any interval $[a, b]$. If $\int_a^\infty g(x)\, dx$ exists, then $\int_a^\infty f(x)\, dx$ also exists and*

$$0 \le \int_a^\infty f(x)\, dx \le \int_a^\infty g(x)\, dx.$$

Proof. We know that $\int_a^b g(x)\,dx$ is a monotone increasing function of b, and its limit as $b \to \infty$ exists, so for all $b \geq a$ we have

$$0 \leq \int_a^b f(x)\,dx \leq \int_a^b g(x)\,dx \leq \int_a^\infty g(x)\,dx.$$

But $\int_a^b f(x)\,dx$ is also increasing as a function of b, and it is bounded above by $\int_a^\infty g(x)\,dx$, so it converges as $b \to \infty$ and

$$0 \leq \lim_{b\to\infty} \int_a^b f(x)\,dx \leq \int_a^\infty g(x)\,dx.$$

The middle term is equal to $\int_a^\infty f(x)\,dx$, so this completes the proof. $\square$

Example 3.46. The integral $\int_2^\infty \frac{1}{x^2+\sin(x)}\,dx$ exists, because the integrand satisfies

$$0 \leq \frac{1}{x^2+\sin(x)} \leq \frac{1}{x^2-1} < \frac{2}{x^2} \qquad \text{for all } x \geq 2$$

and the upper bound is integrable on $[2, \infty)$:

$$\int_2^\infty \frac{2}{x^2}\,dx = -\frac{2}{x}\Big|_{x=2}^{x=\infty} = 1.$$

As another example of a property which continues to work for improper integrals, we still have integration by parts: letting $b \to \infty$ in the formula

$$\int_a^b f(x)g'(x)\,dx = f(x)g(x)\big|_{x=a}^{x=b} - \int_a^b f'(x)g(x)\,dx$$

and applying the algebra of limits, we see that if both

$$\int_a^\infty f'(x)g(x)\,dx \qquad \text{and} \qquad \lim_{x\to\infty} f(x)g(x)$$

exist, then $f(x)g'(x)$ is integrable over $[a, \infty)$ as well, and

$$\int_a^\infty f(x)g'(x)\,dx = f(x)g(x)\big|_{x=a}^{x=\infty} - \int_a^\infty f'(x)g(x)\,dx$$

where $x = \infty$ on the right means that we take the limit as $x \to \infty$.

3.8 Lebesgue's criterion for integrability

We have seen that continuous functions $f : [a,b] \to \mathbb{R}$ are integrable, and some but not all discontinuous, bounded functions are as well. In this non-examinable section we'll determine exactly which functions are integrable.

Theorem 3.47: Lebesgue's criterion for integrability

A bounded function $f : [a,b] \to \mathbb{R}$ is integrable if and only if the set

$$D(f) = \{x \in [a,b] \mid f \text{ is not continuous at } x\}$$

has *measure zero*.

We will define "measure zero" momentarily, but for now we claim that

- countable sets have measure zero, and
- a set which contains an interval $[c,d]$ with $c < d$ does not have measure zero.

So a function which is continuous at all irrational $x \in [a,b]$ is integrable, and if f is integrable then the set of points where f is continuous must be dense in $[a,b]$.

The measure zero condition says roughly that a given set is contained in the union of some open intervals whose total length can be made arbitrarily small. We introduce some terminology to make this precise:

Definition. An *open cover* of $S \subset \mathbb{R}$ is a collection of open intervals $\{U_\alpha = (a_\alpha, b_\alpha)\}$ such that

$$S \subset \bigcup_\alpha U_\alpha.$$

Lemma 3.48. *Let $\{U_\alpha = (a_\alpha, b_\alpha)\}$ be an open cover of $S \subset \mathbb{R}$. Then*

1. *$\{U_\alpha\}$ has a countable subcover, meaning a collection of countably many $U_i \in \{U_\alpha\}$ such that $\{U_1, U_2, U_3, \dots\}$ also covers S; and*
2. *If S is compact, then in fact $\{U_\alpha\}$ has a finite subcover.*

Proof. (1) The set of intervals of the form (p,q) with p,q both rational is countable, since there is an injection

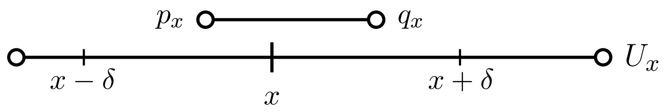
$$\{(p,q) \mid p,q \in \mathbb{Q}\} \hookrightarrow \mathbb{Q} \times \mathbb{Q}$$

by sending the interval (p,q) to the ordered pair (p,q). Every point $x \in S$ belongs to some open set U_x, which by definition contains some open interval $(x-\delta, x+\delta)$

with $\delta > 0$, and if we pick rational p_x, q_x with

$$x - \delta < p_x < x < q_x < x + \delta$$

then x belongs to (p_x, q_x), which is in turn a subset of U_x.



The collection

$$\{(p_x, q_x) \mid x \in S\}$$

is then an open cover of S, and it only has countably many distinct intervals, so we enumerate them as (p_i, q_i) where $i \in \mathbb{N}$. Each (p_i, q_i) was a subset of some U_{α_i} by construction, and we have

$$S \subset \bigcup_i (p_i, q_i) \subset \bigcup_i U_{\alpha_I}$$

so the countable subcover $\{U_{\alpha_i}\}$ of our original $\{U_\alpha\}$ also covers S.

(2) Now suppose that S is compact, and that we have an infinite cover of S; by the above argument we can pass to a countable subcover $\{U_i = (a_i, b_i) \mid i \in \mathbb{N}\}$. We define a function $f : S \to \mathbb{N}$ by

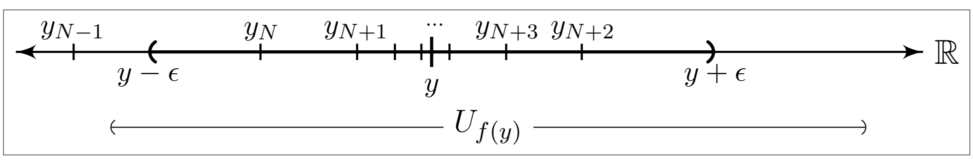
$$f(x) = \min\{n | x \in U_n\};$$

this is well defined because x is in $\bigcup\limits_i U_i$, and so it must belong to some U_i. If no finite union $\bigcup\limits_{i=1}^{m} U_i$ contains all of S, then for every n we can find an $x_n \in S$ such that $f(x_n) \geq n$. Since S is compact, the Bolzano–Weierstrass theorem says that the sequence $\big(x_n\big)$ has a convergent subsequence whose limit lies in S; we write this as

$$y_n \to y \in S, \qquad \lim_{n\to\infty} f(y_n) = \infty.$$

But we have $y \in U_{f(y)}$, so $U_{f(y)}$ contains some open interval $(y - \epsilon, y + \epsilon)$. Since $y_n \to y$, there is an $N > 0$ such that

$$n \geq N \;\Rightarrow\; |y_n - y| < \epsilon \;\Rightarrow\; y_n \in (y - \epsilon, y + \epsilon) \subset U_{f(y)}.$$



This says that $f(y_n) \leq f(y)$ for all $n \geq N$, contradicting $f(y_n) \to \infty$, and so it must be the case that some finite union $\bigcup_{i=1}^{m} U_i$ contains all of S after all. $\square$

Definition. We say that a set $S \subset \mathbb{R}$ has *(outer) measure zero* if for every $\epsilon > 0$, there is a finite or countable open cover $\{U_i = (a_i, b_i)\}$ of S with

$$\sum_i (b_i - a_i) < \epsilon.$$

Example 3.49. If $S = \{x\}$ is a single point, then for any $\epsilon > 0$ it can be covered by the single open interval

$$\left(x - \frac{\epsilon}{3}, x + \frac{\epsilon}{3}\right)$$

of length $\frac{2\epsilon}{3} < \epsilon$. Thus a point has measure zero.

Example 3.50. If $S_1, S_2, \dots$ are countably many sets of measure zero and we are given $\epsilon > 0$, then each S_i admits a countable open cover

$$\{U_{i,j} = (a_{i,j}, b_{i,j}) \mid j \in \mathbb{N}\} \text{ such that } \sum_j (b_{i,j} - a_{i,j}) < \frac{\epsilon}{2^i}.$$

The collection of all open intervals $U_{i,j}$ is countable, since it is a countable union of countably many intervals, and it covers $\bigcup S_i$, with

$$\sum_{i,j} (b_{i,j} - a_{i,j}) = \sum_{i=1}^{\infty} \left(\sum_j (b_{i,j} - a_{i,j}) \right) < \sum_{i=1}^{\infty} \frac{\epsilon}{2^i} = \epsilon.$$

Thus a countable union of sets of measure zero also has measure zero. In particular, $\mathbb{Q}$ has measure zero.

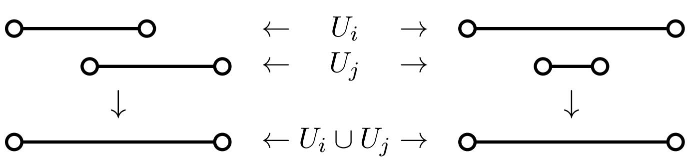
Example 3.51. If S has measure zero, then any subset $T \subset S$ also has measure zero, since any open cover of S with total length ϵ is also an open cover of T.

Proposition 3.52. *If $S = [a, b]$ with $a < b$, then S does* not *have measure zero.*

Proof. Let $\{U_i\}$ be a countable open cover of $[a, b]$, and use the compactness of S to pass to a finite subcover

$$\{U_i \mid 1 \le i \le n\}$$

whose total length does not exceed that of the original cover. If any two intervals U_i and U_j overlap then we can replace the two of them with their union,



which decreases the number of intervals without changing $\bigcup U_i$ or increasing the length of the cover, and we repeat until all of the U_i are disjoint.

We now assume the U_i are labeled so that $U_1 = (c_1, d_1)$ contains a. Then no other $U_j = (c_j, d_j)$ can contain d_1, because as an open set it would have to contain a whole neighborhood $(d_1 - \delta, d_1 + \delta)$ for some $\delta > 0$, and if we take δ small enough then

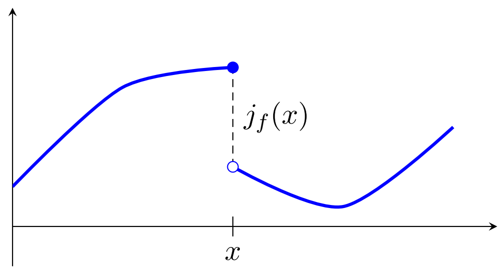
$$(d_1 - \delta, d_1) \subset U_1 \cap U_j,$$

contradicting the fact that U_1 and U_j are disjoint. Since the U_i cover $[a, b]$, it follows that $d_1 \notin [a, b]$, and since $c_1 < a < d_1$ we must have $d_1 > b$. So in fact $[a, b] \subset U_1$, and hence the total length of the cover $\{U_1, \ldots, U_n\}$ is at least $d_1 - c_1 > b - a$. We conclude that $[a, b]$ cannot have measure zero. $\square$

Corollary 3.53. *If S contains an interval $[a, b]$ with $a < b$, then it does not have measure zero.*

We now want to figure out what it takes for a function $f : [a, b] \to \mathbb{R}$ to be integrable, so we'll begin by quantifying how discontinuous it can be at a point. We measure how much f "jumps" near x by defining

$$j_f(x) = \inf_{\delta > 0} \left(\sup_{|y-x|<\delta} f(y) - \inf_{|y-x|<\delta} f(y) \right).$$



Given $x \in [a, b]$ and any fixed $\delta > 0$, we have

$$\sup_{|y-x|<\delta} f(y) \geq f(x) \geq \inf_{|y-x|<\delta} f(y),$$

and f is continuous at x if and only if the left and right sides both approach $f(x)$ as $\delta \to 0$. So $j_f(x) \geq 0$, with equality if and only if f is continuous at x.

Given the function $f : [a, b] \to \mathbb{R}$, we label its set of discontinuous points by

$$D(f) = \{x \in [a, b] \mid f \text{ is not continuous at } x\} = \{x \mid j_f(x) > 0\},$$

and for every $c > 0$ we define

$$D_c(f) = \{x \in [a, b] \mid j_f(x) \geq c\}.$$

Then we can write $D(f)$ as a countable union

$$D(f) = \bigcup_{n\in\mathbb{N}} D_{1/n}(f),$$

so $D(f)$ has measure zero if and only if $D_{1/n}(f)$ has measure zero for all $n \in \mathbb{N}$.

Proposition 3.54. *If f is integrable, then $D_c(f)$ has measure zero for all $c > 0$, and so $D(f) = \bigcup_{n\in\mathbb{N}} D_{1/n}(f)$ has measure zero.*

Proof. Since f is integrable, given $c > 0$ and any $\epsilon > 0$ we can pick a partition $P = (x_0, \dots, x_n)$ of $[a, b]$ such that $U(f, P) - L(f, P) < c\epsilon$, or equivalently

$$\sum_{i=0}^{n-1} \left(\sup_{x_i \le t \le x_{i+1}} f(t) - \inf_{x_i \le t \le x_{i+1}} f(t) \right) \Delta x_i < c\epsilon.$$

If the open interval (x_i, x_{i+1}) contains a point of $D_c(f)$, then the corresponding term contributes at least $c\Delta x_i$ to the sum on the left, and the sum of all such terms is less than $c\epsilon$, so the sum of the lengths of the open intervals (x_i, x_{i+1}) which contain points of $D_c(f)$ is less than ϵ. These intervals provide an open cover of

$$D_c(f) \setminus \{x_0, x_1, \dots, x_n\}$$

with total length less than ϵ, so this set has measure zero. But $D_c(f)$ is the union of this and at most finitely many points, so it has measure zero as well. $\square$

This proves one direction of Lebesgue's criterion. For the converse, we first need to understand a little more about the sets $D_c(f)$.

Lemma 3.55. *Each set $D_c(f)$ is compact.*

Proof. We know that $D_c(f)$ is bounded since it is a subset of $[a, b]$, so we only need to prove that it is closed. Let $\big(x_n\big) \subset D_c(f)$ be a sequence which converges to some $x \in \mathbb{R}$. Given any $\delta > 0$, there is some x_n such that $|x_n - x| < \frac{\delta}{2}$, and then since $j_f(x_n) \ge c$ we have

$$\sup_{|y-x_n|<\frac{\delta}{2}} f(y) - \inf_{|y-x_n|<\frac{\delta}{2}} f(y) \ge c.$$

But if $|y - x_n| < \frac{\delta}{2}$ then by the triangle inequality

$$|y - x| \le |y - x_n| + |x_n - x| < \frac{\delta}{2} + \frac{\delta}{2} = \delta,$$

so we have

$$\sup_{|y-x|<\delta} f(y) - \inf_{|y-x|<\delta} f(y) \geq \sup_{|y-x_n|<\frac{\delta}{2}} f(y) - \inf_{|y-x_n|<\frac{\delta}{2}} f(y) \geq c.$$

Since this works for any $\delta > 0$, we have $j_f(x) \geq c$ and so $x \in D_c(f)$ after all, proving that $D_c(f)$ is closed. $\square$

Proposition 3.56. *If f is bounded and $D(f)$ has measure zero, then f is integrable.*

Proof. Choose M such that $|f(x)| \leq M$ for all $x \in [a,b]$, and pick $\epsilon > 0$. Let $c = \frac{\epsilon}{2(b-a)}$; since $D_c(f) \subset D(f)$ is compact and has measure zero, we can cover it by finitely many open intervals

$$U_i = (c_i, d_i), \qquad 1 \leq i \leq n$$

of total length $\displaystyle\sum_{i=1}^{n}(d_i - c_i) < \frac{\epsilon}{4M}$.

Let $S = [a,b] \setminus \displaystyle\bigcup_{i=1}^{n} U_i$; this is closed, as the intersection of the closed sets $[a,b]$ and $\mathbb{R} \setminus \displaystyle\bigcup_{i=1}^{n} U_i$, and it is bounded as a subset of $[a,b]$, so S is compact. For every $x \in S$ we have $j_f(x) < c$, so there is some $\delta_x > 0$ such that

$$\sup_{|y-x|<\delta_x} f(y) - \inf_{|y-x|<\delta_x} f(y) < c,$$

and in particular

$$\sup_{x-\frac{\delta_x}{2} \leq t \leq x+\frac{\delta_x}{2}} f(t) - \inf_{x-\frac{\delta_x}{2} \leq t \leq x+\frac{\delta_x}{2}} f(t) < c.$$

The intervals $(x-\frac{\delta_x}{2}, x+\frac{\delta_x}{2})$ for all $x \in S$ form an open cover of S, so by compactness there is some finite subcollection

$$V_j = (e_j, f_j), \qquad 1 \leq j \leq m$$

which suffices to cover S.

We now form a partition $P = (x_0, x_1, \ldots, x_k)$ of $[a,b]$ which contains all of the numbers c_i, d_i, e_j, f_j. Every interval $[x_i, x_{i+1}]$ either contains a point of $D_c(f)$, in which case it lies in some $[c_i, d_i]$ and we call it *bad*, or it is disjoint from $D_c(f)$ but lies in some $[x - \frac{\delta_x}{2}, x + \frac{\delta_x}{2}] \subset (x - \delta_x, x + \delta_x)$, in which case we call it *good*. Then

$$
\begin{aligned}
U(f,P) - L(f,P) = &\sum_{[x_i, x_{i+1}] \text{ good}} \left(\sup_{x_i \le t \le x_{i+1}} f(t) - \inf_{x_i \le t \le x_{i+1}} f(t) \right) \Delta x_i \\
&+ \sum_{[x_i, x_{i+1}] \text{ bad}} \left(\sup_{x_i \le t \le x_{i+1}} f(t) - \inf_{x_i \le t \le x_{i+1}} f(t) \right) \Delta x_i.
\end{aligned}
$$

On the good intervals we have $\sup f(t) - \inf f(t) < c$, and on the bad intervals we only have $\sup f(t) - \inf f(t) \le 2M$ but their total length is less than $\frac{\epsilon}{4M}$. Thus

$$
\begin{aligned}
U(f,P) - L(f,P) &\le \sum_{[x_i, x_{i+1}] \text{ good}} c\Delta x_i + \sum_{[x_i, x_{i+1}] \text{ bad}} 2M\Delta x_i \\
&< c(b-a) + 2M \left(\frac{\epsilon}{4M} \right) \\
&= \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.
\end{aligned}
$$

Since we can find such a P for any $\epsilon > 0$, we conclude that f is integrable. □

We have now shown that if f is integrable then $D(f)$ has measure zero, and conversely that if f is bounded and has measure zero then it is integrable, so this completes the proof of Lebesgue's criterion. □