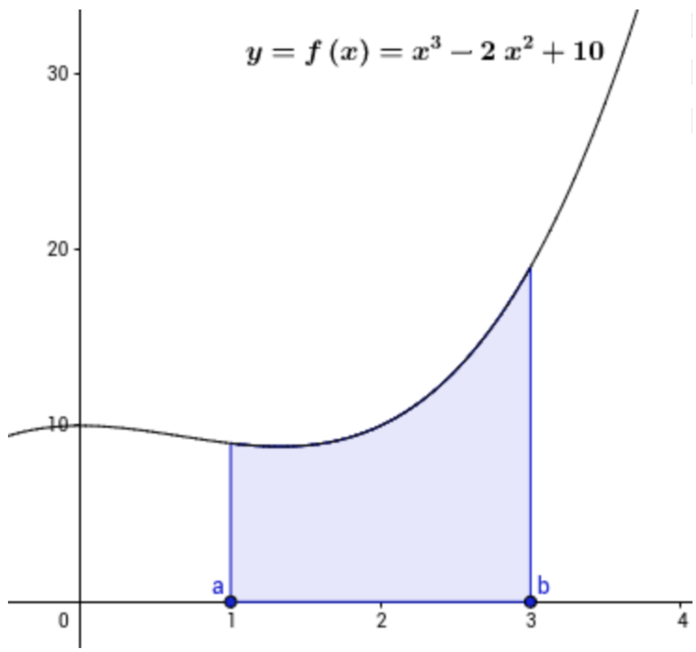
10 The Darboux integral

The main focus of the spring semester is on the integral. In contrast to the derivative, whose motivation was dynamic — measuring rate of change — the main motivation for the integral is static. Although there will be no real similarities between the definitions of the integral and the derivative, the two will quickly turn out to be intimately related.

Note: In this version of the notes, there are very few figures. It will be helpful as you read the notes, to supply graphs to illustrate the various examples, as we will do in class.

10.1 Motivation via area

Suppose $f : [a, b] \to \mathbb{R}$ is a non-negative function. What is the area of the region bounded by the curve $y = f(x)$, the x-axis, the lines $x = a$ and the line $x = b$? (See picture below.)

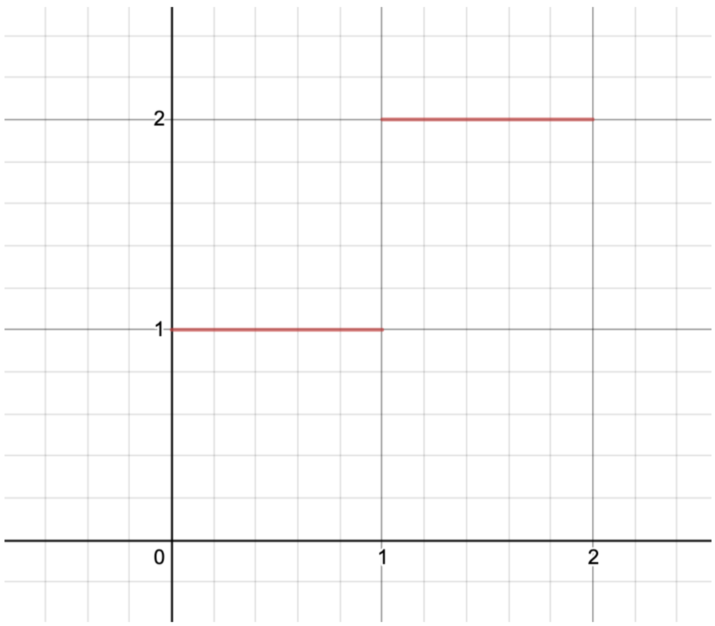


If f is the constant function, $f(x) = c$, the area is clearly $c(b-a)$ (the region is a rectangle with dimensions c and $b - a$). If f is the linear function $f(x) = mx + c$ then the area is equally easy to calculate, since the region is made up of a rectangle and a right triangle. More generally if f is such that the region we are trying to compute the area of is a polygon, then it is relatively straightforward (if somewhat tedious) to compute the area using standard geometric facts.

For general f, however, it is not clear how to calculate "area bounded by curve $y = f(x)$, x-axis, lines $x = a$, $x = b$", or even to interpret what we might mean by that phrase. If, for example, $f(x) = x^2 + 2$ with $a = 0$, $b = 1$, then one of the four lines bounding the relevant area is not a line as such, but a curve, and there are no obvious geometric rules to determine areas bounded by curves. And if f is defined piecewise, say

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

(see picture below) then it is a little unclear what region is under discussion when, say, $a = 0$, $b = 2$. The obvious region (the polygon with vertices $(0,0)$, $(0,1)$, $(1,1)$, $(1,2)$, $(2,2)$ and $(2,0)$ (which has easily calculable area) isn't really the region "bounded" by $y = f(x)$, the x-axis, and the lines $x = a$, $x = b$, since the bounding line from $(1,1)$ to $(1,2)$ is missing.



And when f is something pathological like the stars over Babylon function, questions about area are more than just "a little" unclear.

We get over these issues by the following process, which doesn't calculate exact area, but rather approximates it. More precisely, the process we are about to describe approximates the region we are trying to study by a collection of disjoint rectangles, then calculates the *exact* area of this approximate region, and then attempts to understand what happens to this approximation in some suitable limit.

Mark a, b and a number of points, say $x_1 < \ldots < x_{n-1}$ that are between a and b. Draw vertical lines at $x = x_1$, $x = x_2$, et cetera. This divides the region under discussion into columns, with possibly curved caps. Replace each column by a rectangle, by replacing the curved cap by a straight line that is at approximately the same height as the cap. Then sum up the areas of the rectangles.

If n is small, and/or the gaps between the x_i's are large, then this collection of rectangular columns may not give a good approximation to the area being considered: the curved caps may have quite variable height, and so "replacing the curved cap by a straight line that is at approximately the same height as the cap" may lead to a substantial discrepancy between the area of the approximating rectangle, and the area of the column being approximated.

If n is large, however, and the gaps between the x_i's are small, then it is quite reasonable to expect that the curved caps are reasonably uniform, and that there is little error made in the approximation. In particular, it seems reasonable to imagine that in some kind of limit as n goes to infinity, and the gaps between the x_i's go to 0, the collection of approximate areas converge to something that can be declared to be the actual area.

The goal of this section is to formalize this idea, and to determine large classes of functions for which the idea can be successfully implemented.

Our formalization will involving something called the Darboux integral. It turns out that, unlike the notion of derivative, there are many competing definitions for the integral: there's the Darboux integral, the Riemann integral, the Lebesgue integral, the Riemann-Stieltjes integral, the Henstock integral, and many more. The definitions are all different, and each can be applied to a different class of functions[154]. If you have seen a definition of the integral before, it is almost certainly the Riemann integral. The Darboux integral is defined similarly, but there are significant differences. The advantage of the Darboux integral over the Riemann integral is that the notation is a lot simpler, and this leads to much simpler proofs of all the basic properties of the integral.[155]

10.2 Definition of the Darboux integral

Definition of a partition For real numbers $a < b$, a *partition* P of $[a, b]$ is a sequence of real numbers $(t_0, \ldots, t_n)$ with $a = t_0 < t_1 < \ldots < t_n = b$ (here $n \geq 1$). We refer to the numbers t_i as the *points* of the partition.

Definition of lower and upper Darboux sums Let $f : [a, b] \to \mathbb{R}$ (with $a < b$) be a **bounded** function. For any partition $P = ((t_0, \ldots, t_n))$ of $[a, b]$ the

- *lower Darboux sum* $L(f, P)$

is defined to be

$$L(f, P) = \sum_{i=1}^{n} m_i \Delta_i$$

where

$$m_i = \inf\{f(x) : x \in [t_{i-1}, t_i]\}$$

and

$$\Delta_i = t_i - t_{i-1},$$

and the

- *upper Darboux sum* $U(f, P)$

[154]Of course, whenever two of the definitions can be applied to the same function, they should give the same answer.

[155]It turns out the only difference between the Darboux and the Riemann integral is in the language of their definitions. The set of functions that the two definitions can be applied to end up being exactly the same.

is defined to be

$$U(f,P)=\sum_{i=1}^{n} M_i(t_i-t_{i-1})$$

where

$$M_i=\sup\{f(x) : x\in[t_{i-1},t_i]\}.$$

Some remarks are in order.

- Recall that for a non-empty set A, $\inf A=\alpha$ means that α is a lower bound for A ($\alpha\leq a$ for all $a\in A$), and is the greatest such (if $\alpha'\leq a$ for all $a\in A$, then $\alpha'\leq\alpha$). The completeness axiom asserts that every non-empty set that has a lower bound, has a greatest lower bound. The set $A=\{f(x) : x\in[t_{i-1},t_i]\}$ is non-empty, and by the assumption that f is a bounded function, A has a lower bound (any lower bound for f on $[a,b]$ is also a lower bound for A). So by the completeness axiom the number m_i exists, and the lower Darboux sum exists as a finite number. Similarly, the upper Darboux sum exists. If we did not assume that f is bounded, then one or other of $L(f,P)$, $U(f,P)$ might not exist.

- The extreme value theorem tells us that if $f:[a,b]\to\mathbb{R}$ is continuous then it is bounded. We don't initially want to make the restrictive assumption that f is continuous, though; the Darboux integral can handle many more than just continuous functions.

- Both the upper and lower Darboux sums depend on f and P, and this dependence is captured in the notation we use. They also depend on a and b; but since the partition P already determines what a and b are, we don't mention them in the notation.

- We've motivated the need for the integral by considering non-negative f, but in fact nothing we do in the lead-up to the definition of the Darboux integral requires non-negativity; so from here on you should think of f as an arbitrary (bounded) function.

- If you have seen the Riemann integral (Riemann sum) you may be used to the idea of a partition being an *equipartition*, in which all the Δ_i's (the lengths of the sub-intervals in the partition) are the same, namely $\Delta_i=(b-a)/n$. This is *not* (**not**, ***not***) a requirement of the Darboux approach to integration. While we will often consider equipartitions, it will be very helpful for developing general properties of the integral to allow more general partitions.

Convention going forward To avoid unnecessary extra text, from here until further notice we assume that whenever a function f is mentioned without further qualification, it is a bounded function defined on the closed interval $[a,b]$, where $a<b$ are real numbers, that whenever a partition P is mentioned without further qualification, it is the partition $(t_0,\ldots,t_n)$ of $[a,b]$, and that whenever numbers m_i, M_i, Δ_i are mentioned without further qualification, they are

$$m_i=\inf\{f(x) : x\in[t_{i-1},t_i]\},\quad M_i=\sup\{f(x) : x\in[t_{i-1},t_i]\},\quad\text{and}\quad \Delta_i=t_i-t_{i-1}.$$

A basic fact about Darboux sums is that for any f and any partition P,

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

This follows immediately from the fact the infimum of any set is no bigger than the supremum, so $m_i \le M_i$ for all i.[156] A much less basic fact is that for any f and any two partitions P, Q (which might have nothing to do with each other),

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

In other words,

every lower Darboux sum is no bigger than every upper Darboux sum.

This fact, which we will now prove in stages, drives the definition of the integral.

Lemma 10.1. *Suppose Q has one more point than P, say P is $(t_0, \ldots, t_k, t_{k+1}, \ldots, t_n)$ and Q is $(t_0, \ldots, t_k, u, t_{k+1}, \ldots, t_n)$. Then*

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

Proof: We'll show $L(f, P) \le L(f, Q)$, and leave $U(f, Q) \le U(f, P)$ as an exercise (the proof is *very* similar to that of $L(f, P) \le L(f, Q)$). Since we've already observed that $L(f, Q) \le U(f, Q)$, this gives all the claimed inequalities.

We need a basic fact about infima and suprema, that we will use numerous times as we go on, usually without much comment:

$$\text{If } \emptyset \ne B \subseteq A \text{ and } A \text{ is bounded then } \inf A \le \inf B. \tag{8}$$

Indeed, suppose $\alpha = \inf A$ and $\beta = \inf B$ (α and β both exist; exercise!). We have $\alpha \le a$ for all $a \in A$, so $\alpha \le a$ for all $a \in B$, so α is a lower bound for B, so $\alpha \le \beta$ (since β is the greatest lower bound for B). [157]

In $L(f, P)$ there is the summand

$$\inf\{f(x) : x \in [t_k, t_{k+1}]\}\Delta_{k+1}.$$

[156] Is it perfectly clear why the infimum of any set is no bigger than the supremum? If it is perfectly clear, that's great. If it's not, that's fine too. Treat verifying this fact (and other, similar facts that will get thrown around later) from the definitions as an exercise.

[157] Informally (8) is the (intuitively obvious) fact that "the infimum of a set can't get smaller if we make the set smaller". Similarly, the supremum can't get bigger if we make the set smaller, that is, if $\emptyset \ne B \subseteq A$ and A is bounded then $\sup A \ge \sup B$. This analogous statement is left as an exercise. In fact, as you'll see, almost every fact we encounter going forward will have an "infimum" part and a "supremum" part, and we will only proof the infima parts, leaving the suprema parts as exercises. Going through the entire derivation of the definition of the integral, only verifying the *suprema* parts (the parts that are not verified in these notes) is a great way of checking that you understand the derivation.

This equals

$$\inf\{f(x) : x \in [t_k, t_{k+1}]\}(t_{k+1} - u) + \inf\{f(x) : x \in [t_k, t_{k+1}]\}(u - t_k).$$

By two applications of (8) this is less than or equal to

$$\inf\{f(x) : x \in [u, t_{k+1}]\}(t_{k+1} - u) + \inf\{f(x) : x \in [t_k, u]\}(u - t_k).$$

This is the sum of two summands that appear in $L(f, Q)$. All other summands in $L(f, P)$ appear in $L(f, Q)$, unchanged, and $L(f, Q)$ has no other summands. It follows that $L(f, P) \leq L(f, Q)$. □

Lemma 10.1 says that if *one* point is added to a partition, it brings the upper and lower Darboux sums closer together. It easily follows that the same is true if *many* points are added.

Lemma 10.2. *If Q has all the points of P, and some more, then*

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

Proof: Apply Lemma 10.1 multiple times, adding one new point at a time. (The details are left as an exercise). □

We are now in a position to verify $(\star)$.

Lemma 10.3. *If P and Q are* any *partitions of $[a, b]$, then $L(f, P) \leq U(f, Q)$.*

Proof: Let $P \cup Q$ be the partition of $[a, b]$ that includes every point that is either in P, or in Q (or in both). By Lemma 10.2 (applied twice)

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

□

We now draw a key corollary.

Corollary 10.4. *Let $f : [a, b] \to \mathbb{R}$ be bounded, with $a < b$. The set of all possible lower Darboux sums for f (as P varies over all possible partitions of $[a, b]$) has a supremum,*

$$L(f) := \sup\{L(f, P) : P \text{ a partition of } [a, b]\},$$

the set of all possible upper Darboux sums for f has an infimum,

$$U(f) := \inf\{U(f, P) : P \text{ a partition of } [a, b]\},$$

and, for any partition P of $[a, b]$,

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

Proof: The set $\{L(f,P) : P \text{ a partition of } [a,b]\}$ of lower Darboux sums is non-empty (any partition gives an element), and by Corollary 10.4 it has an upper bound (any upper Darboux sum is an upper bound). So by the completeness axiom $L(f)$ exists. Similarly, $U(f)$ exists.

The inequality $L(f,P) \leq L(f)$ follows immediately from the definition of $L(f)$, and $U(f) \leq U(f,P)$ follows immediately from the definition of $U(f)$, so we just need to verify $L(f) \leq U(f)$. This follows from the basic fact that if non-empty sets A and B are such that everything in A is less than or equal to everything in B, then $\sup A \leq \inf B$ (that everything in $\{L(f,P) : P \text{ a partition of } [a,b]\}$ is less than or equal to everything in $\{U(f,P) : P \text{ a partition of } [a,b]\}$ follows from Corollary 10.4; the basic fact is left as an exercise[158]). $\square$

We are now ready to define the (Darboux) integral.

Definition of (Darboux) integral Let $f : [a,b] \to \mathbb{R}$ be bounded, with $a < b$. Say that f is *(Darboux) integrable* on $[a,b]$ if

$$L(f) = U(f),$$

in which case the common value is the *integral* of f on $[a,b]$, denoted

$$\int_a^b f$$

or

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

If $L(f) < U(f)$ we say that f is *not integrable* on $[a,b]$.

We declare every function defined at a to be integral on the trivial interval $[a,a]$, and set

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

Note that if $f(x) \geq 0$ for all $x \in [a,b]$, then we can think of $\int_a^b f$ the *area* of the region bounded by under $y = f(x)$, the x-axis, $x = a$ and $x = b$. As we will shortly see, if f is not always non-negative then $\int_a^b f$ can still be thought of as area, except now those parts of the region that fall below the x-axis contribute negatively.

Note also that in the expression $\int_a^b f(x)\ dx$, x is a *dummy variable* — the numerical value of the expression is unchanged if all occurrences of x are replaced by r, or t, or ♕:

$$\int_a^b f(x)\ dx = \int_a^b f(r)\ dr = \int_a^b f(t)\ dt = \int_a^b f(♕)\ d♕.$$

The definition of the integral is somewhat involved, and not so easy to use for an arbitrary function (certainly, it is much less easy to work with than the definition of the derivative). We will develop some methods that allow for fairly easy calculations of integrals, but for now we just give two very simple examples direct from the definition.

[158] A familiar one: it was Question 4 in Homework 8 of the fall semester.

Example 1: constant function Let f be the constant function $f(x) = c$ for some constant c. For any $a < b$, and any partition $P = (t_0, \ldots, t_n)$ of $[a, b]$, we have (since f is constant on $[t_{i-1}, t_i]$)

$$m_i = M_i = c,$$

and so

$$L(f, P) = \sum_{i=1}^{n} c\Delta_i = c\sum_{i=1}^{n} \Delta_i = c(b-a)$$

and $U(f, P) = c(b-a)$ also[159]. So $L(f) = U(f) = c(b-a)$, f is integrable on $[a, b]$, and

$$\int_a^b f = c(b-a)$$

(exactly as we would expect).

Example 2: the Dirichlet function Recall that the Dirichlet function $f : \mathbb{R} \to \mathbb{R}$ is defined by

$$f(x) = \begin{cases} 1 & \text{if } x \text{ is rational} \\ 0 & \text{if } x \text{ is irrational.} \end{cases}$$

Given $a < b$, and any partition $P = (t_0, \ldots, t_n)$ of $[a, b]$, we have (since both the rationals and the irrationals are dense in $\mathbb{R}$) that in each interval $[t_{i-1}, t_i]$ there is both a rational and an irrational, and so

$$m_i = 0, \qquad M_i = 1.$$

It follows that $L(f, P) = 0$ and $U(f, P) = 1$, so that $L(f) = 0 < 1 = U(f)$. The conclusion is that the Dirichlet function is *not* integrable on any interval $[a, b]$ with $a < b$.[160]

[159]The equality $\sum_{i=1}^{n} \Delta_i = (b-a)$ can be interpreted geometrically: the partition P divides the interval $[a, b]$ into n non-overlapping subintervals, of lengths $\Delta_1, \Delta_2, \ldots, \Delta_n$, and the sum of the lengths of these subintervals is the length of the whole interval, of $b - a$. This is slightly non-rigorous, as we haven't really talked about "length" in any precise way. But this intuitively obvious result can be verified perfectly rigourously, via:

$$\begin{aligned}
\sum_{i=1}^{n} \Delta_i &= \Delta_1 + \Delta_2 + \cdots + \Delta_{n-1} + \Delta_n \\
&= (t_1 - t_0) + (t_2 - t_1) + \cdots + (t_{n-1} - t_{n-2}) + (t_n - t_{n-1}) \\
&= (-t_0 + t_1) + (-t_1 + t_2) + \cdots + (-t_{n-2} + t_{n-1}) + (-t_{n-1} + t_n) \\
&= -t_0 + (t_1 - t_1) + (t_2 - t_2) + \cdots + (t_{n-1} - t_{n-1}) + t_n \\
&= -t_0 + t_n \\
&= b - a.
\end{aligned}$$

This is an example of a *telescoping sum*.

[160]Is this as we might expect? Maybe, or maybe not. The question of the integrability of the Dirichlet function will return in Honors Analysis 1, as the Lebesgue integral is introduced.

To illustrate the issues surrounding using the definition of the Darboux integral to actually *calculate* an integral, consider an example only slightly more complicated that the two discussed above: $f(x) = x$ on $[0, 1]$. For a partition $P = (t_0, \ldots, t_n)$ of $[0, 1]$, we clearly have

$$m_i = \inf\{f(x) : x \in [t_{i-1}, t_i]\} = t_{i-1}, \text{[161]}$$

so

$$L(f, P) = \sum_{i=1}^{n} t_{i-1}(t_i - t_{i-1}) = \left(\sum_{i=1}^{n} t_{i-1}t_i\right) - \left(\sum_{i=1}^{n} t_{i-1}^2\right).$$

To compute $L(f)$, we have to maximize the above expression, over all choices of $(t_0, \ldots, t_n)$ satisfying $0 = t_0 < t_1 < \cdots < t_n = 1$; not an easy optimization problem!

On the other hand, the definition of the integral does give us an easy way of putting *bounds* on the value of an integral, once we know it exists: if m and M are lower and upper bounds, respectively, for f on $[a, b]$, then we certainly have $m \leq m_i \leq M_i \leq M$ for all i, and so if $\int_a^b f$ exists then

$$m(b - a) \leq \int_a^b f \leq M(b - a).$$

We now develop a useful criterion for integrability, that allows us to verify integrability not by considering all possible partitions, but instead by focusing on a few carefully chosen ones.

Lemma 10.5.

- *Bounded f is integrable on $[a, b]$ if and only if for every $\varepsilon > 0$, there is a partition P_ε with*

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

- *Let f be bounded on $[a, b]$. Suppose there is a number I such that for all $\varepsilon > 0$ there is a partition P_ε such that*

$$L(f, P_\varepsilon) \geq I - \varepsilon$$

and

$$U(f, P_\varepsilon) \leq I + \varepsilon$$

then f is integrable on $[a, b]$, and

$$\int_a^b f = I.$$

Proof: We start with item 1. Suppose that f is integrable, with integral I. There are partitions P_1 and P_2 with $I - \varepsilon/2 < L(f, P_1) \leq I$ and $I \leq U(f, P_2) < I + \varepsilon/2$. Let P be the partition $P_1 \cup P_2$. By Lemma 10.2 we have

$$I - \varepsilon/2 < L(f, P_1) \leq L(f, P) \leq I \leq U(f, P) \leq U(f, P_2) < I + \varepsilon/2$$

[161] There's a simple general fact here, that gets used again and again as we go on, mostly without comment: if f is an increasing function (weakly or strictly) on an interval $[a, b]$, and f is continuous on $[a, b]$, then $\inf\{f(x) : x \in [a, b]\} = f(a)$ and $\sup\{f(x) : x \in [a, b]\} = f(b)$. Similarly if f is decreasing and continuous then $\inf\{f(x) : x \in [a, b]\} = f(b)$ and $\sup\{f(x) : x \in [a, b]\} = f(a)$.

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

Conversely, suppose f is *not* integrable on $[a, b]$. Then $U(f) - L(f) = \delta > 0$, and for any $\varepsilon < \delta$ and any P we have $U(f, P) - L(f, P) \geq \delta > \varepsilon$.

We now move on to item 2. From the hypotheses, for every $\varepsilon > 0$ there is a partition P_ε such that

$$L(f, P_\varepsilon) \geq I - \varepsilon/3 \quad \text{and} \quad U(f, P_\varepsilon) \leq I + \varepsilon/3.$$

Combining these inequalities yields

$$U(f, P_\varepsilon) - L(f, P_\varepsilon) \leq 2\varepsilon/3 < \varepsilon,$$

so by item 1 f is integrable.

Since for every $\varepsilon > 0$ there is a partition P_ε with $U(f, P_\varepsilon) \leq I + \varepsilon$, it follows that $U(f, P) \leq I + \varepsilon$ for every $\varepsilon > 0$, so[162] $U(f, P) \leq I$ and $\int_a^b f \leq I$; but by a similar argument $\int_a^b f \geq I$, so $\int_a^b f = I$.

□

We can do something a little bit better than item 1 of Lemma 10.5.

Corollary 10.6. *Bounded f is integrable on $[a, b]$ if and only if for every natural number n there is a partition P_n with*

$$U(f, P_n) - L(f, P_n) < 1/n.$$

Proof: To see that Corollary 10.6 follows from Lemma 10.5 (item 1), first notice that $1/n > 0$, so if f is integrable we can apply the forward direction of Lemma 10.5 (item 1) with $\varepsilon = 1/n$ to get the forward direction of Corollary 10.6.

To get the backward direction of Corollary 10.6, suppose that for every natural number n there is a partition P_n of $[a, b]$ with $U(f, P_n) - L(f, P_n) < 1/n$. Let $\varepsilon > 0$ be given. There's an n with $1/n \leq \varepsilon$; taking P_ε to be P_n we get $U(f, P_\varepsilon) - L(f, P_\varepsilon) < 1/n \leq \varepsilon$, so we can use the backward direction of Lemma 10.5 (item 1) to get the backward direction of Corollary 10.6. □

Corollary 10.6 is nice because it allows us to check integrability by exhibiting just one (carefully chosen) partition for each natural number — a much easier task, in general, than calculating suprema and infima over uncountably infinitely many partitions!

We illustrate this lemma with a few examples.

Example 1: linear function Consider $f(x) = x$ on $[0, 1]$. For each n let P_n be the partition $(0, 1/n, 2/n, \ldots, 1)$. We have (using, half way down, an identity that is easy to prove by induction: for natural numbers m,

$$\sum_{k=1}^{m} k = \frac{m(m+1)}{2}.)$$

[162] easy exercise.

$$\begin{aligned}
L(f, P_n) &= 0 \cdot \left(\frac{1}{n}\right) + \frac{1}{n} \cdot \left(\frac{1}{n}\right) + \cdots + \frac{n-1}{n}\left(\frac{1}{n}\right) \\
&= \left(\frac{1}{n^2}\right)(0 + 1 + 2 + \cdots + (n-1)) \\
&= \frac{n(n-1)}{2n^2} \\
&= \frac{n-1}{2n} \\
&= \frac{1}{2} - \frac{1}{2n}.
\end{aligned}$$

Note that to get the second line, we factored $1/n^2$ out of each term. In the first term, we rewrote "0" as "$0/n^2$" to put it on the same footing as all the other terms. Similarly

$$U(f, P_n) = \frac{1}{n} \cdot \left(\frac{1}{n}\right) + \frac{2}{n} \cdot \left(\frac{1}{n}\right) + \cdots + \frac{n}{n}\left(\frac{1}{n}\right) = \frac{1}{2} + \frac{1}{2n}.$$

So $U(f, P_n) - L(f, P_n) = 1/n$. Given any $\varepsilon > 0$ there is n with $1/n < \varepsilon$, and so for every $\varepsilon > 0$ there is a partition P with $U(f, P) - L(f, P) < \varepsilon$. From Lemma 10.5 it follows that $\int_0^1 x\, dx$ exists.[163]

What is the value of the integral? Since $L(f, P_n) = 1/2 - 1/(2n)$, and this can be made arbitrarily close to $1/2$ by choosing n large enough, we get that $L(f) \geq 1/2$[164]. Similarly, $U(f) \leq 1/2$. Since $L(f) = U(f)$, the only possible common value is $1/2$, and so $\int_0^1 x\, dx = 1/2$[165]. (We have essentially re-proved item 2 of Lemma 10.5 here.)

More generally, we can easily show that for any $b > 0$,

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

Example 2: quadratic function Consider $f(x) = x^2$ on $[0, b]$, $b > 0$. Take P_n to be the partition that divides $[0, b]$ into n equal intervals: $P_n = (0, b/n, 2b/n, \ldots, (n-1)b/n, b)$. We have (using another identity that is easily proved by induction: for any natural number m,

$$\sum_{k=1}^{m} k^2 = \frac{m(m+1)(2m+1)}{6})$$

[163] Or, we could have directly applied Corollary 10.6 at this point, without introducing the ε — I chose to take the long-winded route to reinforce the proof of Corollary 10.6.

[164] This is another of those statement that might be completely clear to you, or might not. If it is completely clear, that is fine. If it is not, that is fine also. You have the tools to verify it formally — go ahead and do so as an exercise!

[165] As we expected: $\int_0^1 x\, dx$ is supposed to be the area of the triangle with vertices $(0, 0)$, $(0, 1)$ and $(1, 1)$, and this is half the unit square.

$$\begin{aligned}
L(f,P_n) &= 0^2 \cdot \left(\frac{b}{n}\right) + \frac{b^2}{n^2}\cdot\left(\frac{b}{n}\right) + \cdots + \frac{(n-1)^2b^2}{n^2}\left(\frac{b}{n}\right) \\
&= \left(\frac{b^3}{n^3}\right)(0^2+1^2+2^2+\cdots+(n-1)^2) \\
&= b^3\frac{(n-1)n(2n-1)}{6n^3} \\
&= b^3\frac{2n^3-3n^2+n}{6n^3} \\
&= \frac{b^3}{3} - \frac{b^3}{2n} + \frac{b^3}{6n^2}.
\end{aligned}$$

Similary

$$U(f,P_n) = \frac{b^2}{n^2}\cdot\left(\frac{b}{n}\right) + \frac{2b^2}{n^2}\cdot\left(\frac{b}{n}\right) + \cdots + \frac{n^2b^2}{n^2}\left(\frac{b}{n}\right) = \frac{b^3}{3} + \frac{b^3}{2n} + \frac{b^3}{6n^2}.$$

So

$$U(f,P_n) - L(f,P_n) = \frac{b^3}{n}.$$

Exactly as in the case of the linear function, it now follows that $\int_0^b x^2\ dx$ exists and equals $b^3/3$.

Notice that this says something quite non-obvious about an area in the plane that is bounded by some non-linear curves.

More generally, we could push these ideas to show that for any natural number t, and any $b > 0$,

$$\int_0^b x^t\ dx = \frac{b^{t+1}}{t+1}.$$

This would involve a little more computation, and some reasonable expression for

$$\sum_{k=1}^{m} k^t$$

(the sum of the first m perfect tth powers). We'll derive this familiar integral in an easier way shortly, after developing some more theory.

Before doing that, here's one more example. So far all our examples of integrable functions have been continuous; and in fact we will see that *every* continuous function is integrable. But many more functions are integrable. Consider, for example, the function $f : [0,1] \to \mathbb{R}$ defined by

$$f(x) = \begin{cases} 1 & \text{if } x = 1, 1/2, 1/3, \ldots \text{ is rational} \\ 0 & \text{otherwise.} \end{cases}$$

This function is discontinuous at infinitely many points, including at a collection of points that are bunched arbitrarily closely together (around 0). Nevertheless, it is integrable on

$[0,1]$. For any partition P of $[0,1]$we have $L(f,P)=0$ (by density of $\mathbb{R}\setminus\{1,1/2,1/3,\ldots\}$), so $L(f)=0$. So to show that the function is integrable on $[0,1]$, we need to show $U(f)=0$ (and this will for free show that the value of the integral in 0). To show $U(f)=0$, it is enough to find, for every $\varepsilon>0$, a partition P_ε with $U(f,P_\varepsilon)<\varepsilon$) (note that we already know $U(f)\geq 0$, since $U(f)\geq L(f)$).

A partition of $[0,1]$ is determined by a finite collection of distinct points in $[0,1]$, that includes 0 and 1. Consider the following set of points, determining a partition P:

- 0 and 1,
- $1/N+1/2N^2$, for some N that will be determined later,
- $1-1/2N^2$,
- and, for each $k=2,3,\ldots,N-1$, the points $1/k-1/2N^2$ and $1/k+1/2N^2$.

This creates the following intervals in the partition:

- $[0,1/N+1/2N^2]$, of length $1/N+1/2N^2$, on which the supremum of f is 1,
- around 1/2, 1/3, ... $1/(N-1)$, intervals of length $1/N^2$, centered at 1/2, 1/3, et cetera, on each of which the supremum of f is 1. Notice that half of the gap from $1/N$ to $1/(N-1)$ is $1/(2N(N-1))$, which is bigger than $1(2N^2)$, so these intervals don't overlap.
- $[1-1/2N^2,1]$, of length $1/2N^2$, on which the supremum of f is 1,
- and many other intervals ($N-2$ many), on which the supremum of f is 0.

We have

$$U(f,P)=\left(\frac{1}{N}+\frac{1}{2N^2}\right)+\frac{N-2}{N^2}+\frac{1}{2N^2}.$$

This quantity can be made arbitrarily small by choosing N large enough; in particular there is an N such that $U(f,P)<\varepsilon$. This shows that

$$\int_0^1 f=0.$$

Notice that in this example we did not use a partition that divides $[a,b]$ into equal-length subintervals.

10.3 Some basic properties of the integral

In this section we gather together some basic facts about integrals, many of which will allow us to discovered the integrability of some new functions from the integrability of old functions.

Splitting an interval of integration

Lemma 10.7. *Fix $a < c < b$. Suppose that (bounded) f is integrable on $[a, b]$. Then it is integrable on both $[a, c]$ and $[c, b]$, and*

$$\int_a^b f = \int_a^c f + \int_c^b f.$$

Proof: Fix $\varepsilon > 0$. By Lemma 10.5 there is a partition P of $[a, b]$ with $U(f, P) - L(f, P) < \varepsilon$. If c is not one of the partition points of P, then, letting P_c be the partition obtained from P by adding c, we have (by Lemma 10.1)

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

so $U(f, P_c) - L(f, P_c) < \varepsilon$. So we may in fact assume that c is one of the partition points.

We can break P into P', a partition of $[a, c]$ (by taking all the partition points of P between a and c, inclusive), and P'', a partition of $[c, b]$ (by taking all the partition points of P between c and b, inclusive). We have

- $L(f, P) = L(f, P') + L(f, P'')$ and
- $U(f, P) = U(f, P') + U(f, P'')$,

so $[U(f, P') - L(f, P')] + [U(f, P'') - L(f, P'')] < \varepsilon$. Each summand in square brackets is non-negative, so each individually is at most ε. The partitions P' and P'' witness, via Lemma 10.5, that both $\int_a^c f$ and $\int_c^b$ exist.

To get the summation identity ($\int_a^b f = \int_a^c f + \int_c^b f$), let P be any partition. Add c as a partition point (if necessary) to get partitions P_c (of $[a, b]$), P_c' (of $[a, c]$) and P_c'' (of $[c, b]$). By the definition of the integral we have

$$L(f, P_c') \le \int_a^c f \le U(f, P_c') \quad \text{and} \quad L(f, P_c'') \le \int_c^b f \le U(f, P_c'')$$

and so, adding these two inequalities and applying Lemma 10.1 if necessary (i.e., if c was not a partition point of P) it follows that

$$L(f, P) \le L(f, P_c) \le \int_a^c f + \int_c^b f \le U(f, P_c) \le U(f, P)$$

These inequalities are true for any P, so we have

$$\int_a^b f = \sup L(f, P) \le \int_a^c f + \int_c^b f \le \inf U(f, P) = \int_a^b f,$$

and so indeed $\int_a^b f = \int_a^c f + \int_c^b f$. □

This result has a converse.

Lemma 10.8. *Fix $a < c < b$. If (bounded) f integrable on both $[a,c]$ and $[c,b]$ then it is integrable on $[a,b]$ (and so, by Lemma 10.7, again $\int_a^b f = \int_a^c f + \int_c^b f$).*

Proof: Fix $\varepsilon > 0$. There are partitions P' of $[a,c]$ and P'' of $[c,b]$ with

$$U(f,P') - L(f,P') < \varepsilon/2 \quad \text{and} \quad U(f,P'') - L(f,P'') < \varepsilon/2$$

Let $P = P' \cup P''$. This is a partition of $[a,b]$ that satisfies

$$U(f,P) - L(f,P) = [U(f,P') + U(f,P'')] - [(L(f,P') + L(f,P'')] < \varepsilon,$$

so by Lemma 10.5 f is is integrable on $[a,b]$. □

We have already encoded in our definition of the integral that for any function f and any real a, $\int_a^a f = 0$. This allows us to make sense of $\int_a^b$ whenever $a \le b$. To deal with $a > b$, we make a further definition.

For $a > b$, say that $\int_a^b f$ exists if $\int_b^a f$ exists. If $\int_b^a f$ exists, then set $\int_a^b f := -\int_b^a f$.

A consequence of this definition (taken together with Lemmas 10.7 and 10.8) is the following:

Corollary 10.9. *If a, b, c are any three real numbers (distinct or not distinct, and in any relative order), and all the integrals involved exist, then*

$$\int_a^b f = \int_a^c f + \int_c^b f.$$

"Proof": It is a tedious business to verify this in all possible cases[166]. The case $a < c < b$ is exactly Lemmas 10.7 and 10.8. We just verify one other case, and leave the rest as exercises. If $a < b < c$, then $\int_a^c f = \int_a^b f + \int_b^c f$, so $\int_a^b f = \int_a^c f - \int_b^c f$. Now by definition $-\int_b^c f = \int_c^b$, so indeed $\int_a^b f = \int_a^c f + \int_c^b f$. □

Closure under addition

We now move on to some basic closure lemmas for the integral. First, the set of integrable functions is closed under addition.

Lemma 10.10. *If f and g are both integrable on $[a,b]$, then so is $f + g$, and*

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

Proof: Let $P = (t_0, \ldots, t_n)$ be any partition of $[a,b]$. For each i let

- $m_i = \inf\{f(x) + g(x) : x \in [t_{i-1}, t_i]\}$

[166]Question for the curious: how many cases, exactly, have to be considered?

- $m_i^f = \inf\{f(x) : x \in [t_{i-1}, t_i]\}$
- $m_i^g = \inf\{g(x) : x \in [t_{i-1}, t_i]\}$

and define M_i, M_i^f, M_i^g similarly.

It is not unreasonable to expect that $m_i = m_i^f + m_i^g$ and $M_i = M_i^f + M_i^g$, but this is not actually true. What *is* true[167] is that

$$m_i \geq m_i^f + m_i^g \quad \text{and} \quad M_i \leq M_i^f + M_i^g.$$

It follows that

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

Now there is a partition partition P_1 with $U(f, P_1) - L(f, P_1) < \varepsilon/2$, and a partition P_2 with $U(g, P_2) - L(g, P_2) < \varepsilon/2$, so there is a partition P' with $U(f, P') - L(f, P') < \varepsilon/2$ and $U(g, P') - L(g, P') < \varepsilon/2$ (any partition that is a common refinement of P_1 and P_2, that is, that has all the points of P_1 and all the points of P_2, will work). For this partition P' we have

$$U(f + g, P') - L(f + g, P') < \varepsilon$$

so $f + g$ integrable.

What is the value of the integral? For the partition P' created above, we have

$$U(f, P') \leq \int_a^b f + \varepsilon/2$$

($U(f, P') \leq L(f, P') + \varepsilon/2$, and $L(f, P') \leq \int_a^b f$) and

$$U(g, P') \leq \int_a^b g + \varepsilon/2,$$

so

$$U(f, P') + U(g, P') \leq \int_a^b f + \int_a^b g + \varepsilon.$$

Since $U(f + g, P') \leq U(f, P') + U(g, P')$ it follows that

$$U(f + g, P') \leq \int_a^b f + \int_a^b g + \varepsilon.$$

Since $\varepsilon > 0$ was arbitrary, it follows that the infimum of $U(f + g, P)$ over all partitions P is at most $\int_a^b f + \int_a^b g$, that is,

$$\int_a^b (f + g) \leq \int_a^b f + \int_a^b g.$$

[167] Verifying what follows, and showing by example that the inequality can occur strictly, is left as a homework exercise

But a similar argument using lower Darboux sums gives

$$\int_a^b (f+g) \geq \int_a^b f + \int_a^b g,$$

so that indeed

$$\int_a^b (f+g) = \int_a^b f + \int_a^b g$$

as claimed. □

By induction this result extends to the fact that if $f_1, \ldots, f_n$ are all integrable on $[a,b]$, then so is $\sum_{i=1}^n f_i$, and

$$\int_a^b \sum_{i=1}^n f_i = \sum_{i=1}^n \int_a^b f_i.$$

Closure under multiplication by a constant

Next, the set of integrable functions is closed under multiplication by a constant.

Lemma 10.11. *If f is integrable on $[a,b]$ and λ is any real number then λf is integrable on $[a,b]$, and*

$$\int_a^b \lambda f = \lambda \int_a^b f.$$

Proof: If $\lambda = 0$ then the result is easy. For $\lambda > 0$, fix $\varepsilon > 0$ and let P be a partition of $[a,b]$ with $U(f,P) - L(f,P) < \varepsilon/\lambda$. Using

$$\inf\{(\lambda f)(x) : x \in [t_{i-1}, t_i]\} = \lambda \inf\{f(x) : x \in [t_{i-1}, t_i]\} \quad (\star)$$

and

$$\sup\{(\lambda f)(x) : x \in [t_{i-1}, t_i]\} = \lambda \sup\{f(x) : x \in [t_{i-1}, t_i]\} \quad (\star\star)$$

we quickly get that $U(\lambda f, P) - L(\lambda f, P) < \varepsilon$, so λf is integrable on $[a,b]$. The value of the integral can easily be computed, using a trick similar to the one used in the proof of Lemma 10.10: for the partition P introduced above, $U(f,P) \leq \int_a^b f + \varepsilon/\lambda$, so $\lambda U(f,P) \leq \lambda \int_a^b f + \varepsilon$. But $\lambda U(f,P) = U(\lambda f, P)$, so

$$U(\lambda f, P) \leq \lambda \int_a^b f + \varepsilon$$

from which it follows that $\int_a^b \lambda f \leq \lambda \int_a^b f$. A similar argument gives $\int_a^b \lambda f \geq \lambda \int_a^b f$.

The case $\lambda < 0$ is left as an exercise.[168] □

[168] Here one has to be careful with the inequalities, since multiplying an inequality by a negative number flips the direction of the sign. The analogs of $(\star)$ and $(\star\star)$ that need to be used for $\lambda < 0$ are

$$\inf\{(\lambda f)(x) : x \in [t_{i-1}, t_i]\} = \lambda \sup\{f(x) : x \in [t_{i-1}, t_i]\}$$

and

$$\sup\{(\lambda f)(x) : x \in [t_{i-1}, t_i]\} = \lambda \inf\{f(x) : x \in [t_{i-1}, t_i]\}.$$

Closure under changing finitely many values

A nice corollary of Lemmas 10.10 and 10.11 is that if a function is integrable on an interval, and the values of the function are changed at finitely many places, then it remains integrable, and the value of the integral does not change. This is in sharp contrast to the derivative: changing the value of a differentiable function at a single point means that it is no longer differentiable at that point. The corollary presented below is particularly useful for calculating the integrals of piecewise defined functions, as it allows the values of the function at endpoints of the clauses of definition to be changed, arbitrarily. Some examples of this will appear on homework.

Corollary 10.12. *Suppose that f is integrable on $[a,b]$ and g, defined on $[a,b]$, differs from f at only finitely many values. Then g is integrable on $[a,b]$, and $\int_a^b g = \int_a^b f$.*

Proof: Consider the function[169] $\chi_c : \mathbb{R} \to \mathbb{R}$ defined by

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

It is an easy check that χ_c is integrable on any interval, and that the integral is always 0.

Now suppose that g differs from f at the points $c_1, \ldots, c_n$. We have

$$g = f + \sum_{i=1}^{n} (g(c_i) - f(c_i))\chi_{c_i}.$$

Repeated applications of Lemmas 10.10 and 10.11 give that g is integrable on $[a,b]$, with

$$\int_a^b g = \int_a^b f + \sum_{i=1}^{n} (g(c_i) - f(c_i)) \int_a^b \chi_{c_i} = \int_a^b f.$$

□

Integral and area

If f is non-negative on $[a,b]$, then we can interpret $\int_a^b f$ as a measure of the area of the region bounded by the curve $y = f(x)$, the x-axis, and the vertical lines $x = a$ and $x = b$. Indeed, for a large class of curves $y = f(x)$ for which we do not have any known formulae for area, we tend to take $\int_a^b f$ as the *definition* of the area.

What if f is non-positive on $[a,b]$? Then $-f$ is non-negative, and so $\int_a^b (-f)$ is the area of the region bounded by the curve $y = -f(x)$, the x-axis, and the vertical lines $x = a$ and $x = b$. This region is the reflection across the x-axis of the region bounded by the curve $y = -f(x)$, the x-axis, and the vertical lines $x = a$ and $x = b$, and (by Lemma 10.11) the value of $\int_a^b (-f)$ is $-\int_a^b f$. So, in the case where f is non-positive on $[a,b]$,

[169] This is sometimes called an *indicator* function: it indicates whether or not the input is c.

$\int_a^b f$ is the negative area of the region bounded by the curve $y = f(x)$, the x-axis, and the vertical lines $x = a$ and $x = b$.

What if f crosses back and forth over the x-axis many times between a and b? Suppose there is a finite collection of numbers $a = c_0 < c_1 < c_2 < \cdots < c_{n-1} < c_n = b$ such that on each interval $[c_0, c_1], [c_1, c_2], \ldots, [c_{n-1}, c_n]$, f is either always non-negative or always non-positive. By Lemmas 10.7 and 10.8 (and an application of induction) we have

$$\int_a^b f = \sum_{i=1}^n \int_{c_{i-1}}^{c_i} f,$$

and by the discussion above this is the *signed* area bounded by $y = f(x)$, the x-axis, and the lines $x = a$ and $x = b$: area above the axis is counted positively, and area below the axis is counted negatively.

Although our definition of the integral is motivated by calculating area, it doesn't completely settle the issue. For example, in discussing non-positive functions, we implicitly assumed that if a region A below the x-axis is the reflection across the x-axis of a region B above the x-axis, then A and B have the same area. And in saying something quite reasonable like "the area between $y = x^2$ and $y = x^3$, between $x = 0$ and $x = 1$, is $\int_0^1 x^2\, dx - \int_0^1 x^3\, dx$", we're making another implicit assumption about area, namely that sum of areas of two regions that only overlap in a curve is equal to area of union of regions. There are lots of other subtle area/integral issues like this. In truth, the Darboux integral is not the best tool for dealing with areas of regions in the plane. A better tool is the *Lebesgue integral*, that's defined quite differently from the Darboux integral (but agrees with the Darboux integral on every Darboux integrable function) and can handle the integrals of many more functions, for example the Dirichlet function.

Nonetheless the Darboux integral is still incredibly useful, and there is no harm at all in thinking of it of (informally at least) as an area calculator.

Closure under absolute value and multiplication

Our final basic closure lemma for the integral is that the set of integrable functions is closed under multiplication. The proof involves many steps, which will mostly be left as exercises. Some of the steps are themselves useful lemmas that establish the integrability of certain functions that are obtained by modifying known integrable functions.

We first outline some preliminary steps. These will all be homework problems.

- Suppose that A is a bounded, non-empty set of real numbers. Let $|A| = \{|a| : a \in A\}$. Then
$$\sup |A| - \inf |A| \leq \sup A - \inf A. \quad (\star)$$
- From $(\star)$ it follows that
$$\text{if } f \text{ is integrable on } [a, b], \text{ then so is } |f|.$$

- Define $\max\{f, 0\}$ to be the function which at input x takes the value $f(x)$ if $f(x) \geq 0$, and takes value 0 otherwise, and $\min\{f, 0\}$ to be the function which at input x takes the value $f(x)$ if $f(x) \leq 0$, and takes value 0 otherwise. If f is integrable on $[a, b]$, then from the integrability of $|f|$ it follows that both $\max\{f, 0\}$ and $\min\{f, 0\}$ are integrable.

- The *positive part of* f is the function $f^+ = \max\{f, 0\}$. Informally, think of the positive part of f as being obtained from f by pushing all parts of the graph of f that lie below the x-axis, up to the x-axis. The *negative part of* f is the function $f^- = -\min\{f, 0\}$. Note that $f = f^- f^-$ is a representation of f as a linear combination of non-negative functions. It can be shown that f is integrable on $[a, b]$ if and only if f^+ and f^- are both integrable on $[a, b]$.

The preliminary steps lead to main point, the closure of integrable functions under multiplication.

Lemma 10.13. *Suppose that f and g are both integrable on $[a, b]$. Then so is fg.*

Proof (sketch): We begin by assuming that $f, g \geq 0$. This case is somewhat similar to the proof of closure under addition (Lemma 10.10). With the notation as in that proof, we begin by establishing

$$M_i \leq M_i^f M_i^g \quad \text{and} \quad m_i^f m_i^g \leq m_i$$

(these are left as exercises).

Now for any partition P we have

$$\begin{aligned}
U(fg, P) - L(fg, P) &= \sum_{i=1}^{n} (M_i - m_i)(t_i - t_{i-1}) \\
&\leq \sum_{i=1}^{n} (M_i^f M_i^g - m_i^f m_i^g)(t_i - t_{i-1}) \\
&= \sum_{i=1}^{n} (M_i^f M_i^g - m_i^f M_i^g + m_i^f M_i^g - m_i^f m_i^g)(t_i - t_{i-1}) \\
&\leq M \left(\sum_{i=1}^{n} (M_i^f - m_i^f)(t_i - t_{i-1}) + \sum_{i=1}^{n} (M_i^g - m_i^g)(t_i - t_{i-1}) \right).
\end{aligned}$$

There's a partition that makes both summands in the last line above at most $\varepsilon/2M$, so fg is integrable on $[a, b]$.

For general f and g, write $f = f^+ - f^-$ where $f^+ = \max\{f, 0\}$, $f^- = -\min\{f, 0\}$, and $g = g^+ - g^-$, so that

$$fg = f^+ g^+ - f^+ g^- - f^- g^+ + f^- g^-.$$

By the preliminary steps, all of f^+, g^+, f^-, g^- are integrable, and they are all non-negative, so by the early case of this proof, the various products of pairs of them are integrable, and so fg, being a linear combination of integrable functions, is integrable. □

Some integral inequalities

Notice that in the last section we asserted that if f is integrable, then so is $|f|$; and that if f and g are integrable, then so is fg. We did not, however, give a way of expressing $\int_a^b |f|$ in terms of $\int_a^b f$, or $\int_a^b fg$ in terms of $\int_a^b f$ and $\int_a^b g$ (as we did with, for example, closure under addition, where we showed $\int_a^b (f+g) = \int_a^b f + \int_a^b g$). This was not an oversight; there is no way of simply expressing $\int_a^b |f|$ in terms of $\int_a^b f$, and there is no equivalent of the product rule for differentiation, allowing us in general to express $\int_a^b fg$ in terms of $\int_a^b f$ and $\int_a^b g$.

There are, however, some *inequalities* that relate $\int_a^b |f|$ to $\int_a^b f$, and $\int_a^b fg$ to $\int_a^b f$ and $\int_a^b g$. We discuss two of them here. The proof of the first (the triangle inequality) is left as an exercise.

Lemma 10.14. *(Triangle inequality for integrals) If f is integrable on $[a,b]$ then*

$$\left|\int_a^b f\right| \leq \int_a^b |f|.$$

The second inequality, the *Cauchy-Schwarz-Bunyakovsky inequality*, is one of the most most important and frequently occurring in analysis.[170]

Lemma 10.15. *if f, and g are both integrable on $[a,b]$ then*

$$\left(\int_a^b fg\right)^2 \leq \left(\int_a^b f^2\right)\left(\int_a^b g^2\right).$$

Proof: Let $\alpha = \int_a^b g^2$, $\beta = \int_a^b fg$ and $\gamma = \int_a^b f^2$ (all of which we now know to exist, by Lemma 10.13). For any real number t

$$0 \leq \int_a^b (f - tg)^2 = \alpha t^2 - 2\beta t + \gamma.$$

Because $g^2 \geq 0$ we have $\alpha \geq 0$. If $\alpha > 0$, then the fact that the quadratic $\alpha t^2 - 2\beta t + \gamma$ is always non-negative means that it must have either repeated (real) roots or complex roots. The only way this can happen is if its discriminant $(-2\beta)^2 - 4\alpha\gamma$ is non-positive; but this says that $\beta^2 \leq \alpha\gamma$, which is exactly the claimed inequality.

The case that remains to be considered is $\alpha = 0$. In this case $0 \leq -2\beta t + \gamma$ for all t, which can only happen if also $\beta = 0$; but then the claimed inequality $\beta^2 \leq \alpha\gamma$ is trivial. □

Here's a special[171] (discrete) case of the Cauchy-Schwarz-Bunyakovsky inequality: if $(a_1, \ldots, a_n)$ and $(b_1, \ldots, b_n)$ are real sequences then

$$\left(\sum_{i=1}^n a_i b_i\right)^2 \leq \left(\sum_{i=1}^n a_i^2\right)\left(\sum_{i=1}^n b_i^2\right).$$

[170] It has it own 300-page book, for example, the wonderful *The Cauchy-Schwarz Master Class (An Introduction to the Art of Mathematical Inequalities)*, by J. Michael Steele, Cambridge University Press, 2004.

[171] Exercise: figure out what choice of integrable functions f, g reduce the integral form to the inequality to the discrete form given here.

Conclusion

Together with the fact that the constant function and the linear function are both integrable, the various lemmas and corollaries presented in this section give a large class of integrable function: for example, all polynomial functions are integrable, as are all functions obtainable from polynomials by any combination of

- altering the function at finitely many values,
- taking absolute value
- taking positive or negative part
- multiplying.

One very large class of integrable functions remains to be explored — the set of continuous functions. That will require a slight digression, which we make in the next section.

10.4 Uniform continuity

Recall what it means for a function f to be *continuous* on an interval I (open, closed, or open at one end, closed at the other): it means that

- for all $c \in I$
- for all $\varepsilon > 0$
- there is $\delta > 0$ such that
- for all $x \in I$[172]

$$x \in (c-\delta, c+\delta) \quad \text{implies} \quad f(x) \in (f(c)-\varepsilon, f(c)+\varepsilon).$$

Notice that δ in this definition depends on *both* ε *and* c, and it may not be the case that the same δ works for a given ε for *all* c in the interval. Consider, for example, the function $f(x) = 1/x$ defined on $(0,1)$. f is continuous on the interval, but for each ε it is not the case that there is a single δ that works for all $c \in (0,1)$ in the definition of continuity. To see this, consider $\varepsilon = 1/10$. What does it take to force $f(x)$ to be in the interval

$$(f(c)-1/10, f(c)+1/10) = ((10-c)/10c, (10+c)/10c)?$$

[172]By saying "$c \in I$" here, we get around the problem of c being one of the endpoints of I. For example, suppose that $I = [c, c']$. In this case, saying $x \in I$ *and* $x \in (c-\delta, c+\delta)$ is the same as saying just $x \in [c, c+\delta)$, and we are where we want to be, asserting the continuity of f at c *from the right*.

If $x \in (c-\delta, c+\delta)$ then $f(x) \in (1/(c+\delta), 1/(c-\delta))$ (and as x ranges over $(c-\delta, c+\delta)$, $f(x)$ ranges over all of $(1/(c+\delta), 1/(c-\delta))$). For $f(x)$ also to always be in the interval $((10-c)/10c, (10+c)/10c)$, it is necessary that

$$\frac{1}{c+\delta} \geq \frac{10+c}{10c},$$

or $\delta < 10c/(10-c) - c$. As c approaches 0, the quantity $10c/(10-c) - c$ approaches 0 also, so as c gets smaller, the required δ to witness that the continuity statement is true (at $\varepsilon = 1/10$) needs to get smaller too, and approaches 0. So no single $\delta > 0$ will work for all c.

Sometimes, a function is continuous on an interval in a more "uniform" way: for every c and every ε there is a δ, but the δ does *not* depend on c, only on ε; it can be chosen after ε has been selected, *without* reference to c.

Consider, for example, $f(x) = x$ on $\mathbb{R}$. Given $\varepsilon > 0$, choose $\delta = \varepsilon$. Now, for each $c \in (0,1)$, suppose $x \in (c-\delta, c+\delta) = (c-\varepsilon, c+\varepsilon)$. Then $f(x) = x \in (c-\varepsilon, c+\varepsilon) = (f(c)-\varepsilon, f(c)+\varepsilon)$. This shows that f is continuous on $\mathbb{R}$, and notice that for each $\varepsilon > 0$ a *single* choice of $\delta > 0$ was enough for all $c \in \mathbb{R}$.

This leads to a new definition, that captures a slightly stronger notion of continuity.

Definition of uniform continuity on an interval Let f be a function defined on an interval I. Say that f is *uniformly continuous* on I if

- for all $\varepsilon > 0$
- there is $\delta > 0$ such that
- for all $c \in I$
- for all $x \in I$

$$x \in (c-\delta, c+\delta) \quad \text{implies} \quad f(x) \in (f(c)-\varepsilon, f(c)+\varepsilon).$$

Symbolically, f is uniformly continuous on I if

$$(\forall \varepsilon > 0)(\exists \delta > 0)(\forall c \in I)(\forall x \in I)([x \in (c-\delta, c+\delta)] \Rightarrow [f(x) \in (f(c)-\varepsilon, f(c)+\varepsilon)]).$$

Some remarks:

- Uniform continuity is a concept that only makes sense *on an interval*: it makes no sense to say that a function is uniformly continuous at a point.
- A function that is uniformly continuous on an interval, is necessarily continuous on that interval (easy exercise), but as we have seen from an example, the converse is not true.
- Compare the definition of "f is continuous on I:

$$(\forall c \in I)(\forall \varepsilon > 0)(\exists \delta > 0)(\forall x \in I)([x \in (c-\delta, c+\delta)] \Rightarrow [f(x) \in (f(c)-\varepsilon, f(c)+\varepsilon)]).$$

with the definition of "f is uniformly continuous on I:

$$(\forall \varepsilon > 0)(\exists \delta > 0)(\forall c \in I)(\forall x \in I)([x \in (c-\delta, c+\delta)] \Rightarrow [f(x) \in (f(c)-\varepsilon, f(c)+\varepsilon)]).$$

The difference is in the order of the quantifiers.

- There is an alternate form of the definition of uniform continuity, which will be useful when we prove the main theorem of this section (Theorem 10.17). The premise "$x \in (c-\delta, c+\delta)$" of the implication in the definition is exactly the same as the premise "$|x-c| < \delta$" ("x is within δ of c" is the same as "x and c are within δ of each other"). Similarly the conclusion "$f(x) \in (f(c)-\varepsilon, f(c)+\varepsilon$" is the same as "$|f(x)-f(c)| < \varepsilon$". The usefulness of the alternate premise and conclusion is that they are more symmetric than the premise and conclusion they replace. To further highlight this symmetry, we can replace "c" with "y", to get the following, equivalent, formulation of uniform continuity: f is *uniformly continuous* on I if

  – for all $\varepsilon > 0$
  – there is $\delta > 0$ such that
  – for all $x, y \in I$
  $$|x-y| < \delta \quad \text{implies} \quad |f(x)-f(y)| < \varepsilon.$$

  Symbolically,

  $$(\forall \varepsilon > 0)(\exists \delta > 0)(\forall x \in I)(\forall y \in I)([|x-y| < \delta] \Rightarrow [|f(x)-f(y)| < \varepsilon]).$$

Uniformly continuous functions are often easier to work with than continuous functions. Here's an illustration — the proof that if f is uniformly continuous on an interval $[a,b]$, then it is integrable on $[a,b]$.

Theorem 10.16. *If $f : [a,b] \to \mathbb{R}$ is uniformly continuous on $[a,b]$, then f is integrable on $[a,b]$.*

Proof: Let $\varepsilon > 0$ be given. We need to find a partition P of $[a,b]$ with $U(f,P) - L(f,P) < \varepsilon$.

Because f is uniformly continuous on $[a,b]$, there is a $\delta > 0$ such that for all $c \in [a,b]$, if $x \in [a,b]$ and $x \in (c-\delta, c+\delta)$ then $f(x) \in (f(c) - \varepsilon/(3(b-a)), f(c) + \varepsilon/(3(b-a)))$.

Let P be any partition of $[a,b]$ in which the length $t_i - t_{i-1}$ of every interval $[t_{i-1}, t_i]$ in the partition is less than 2δ. Let c_i be the midpoint of $[t_{i-1}, t_i]$. If $x \in [t_{i-1}, t_i]$, then $x \in (c-\delta, c+\delta)$, so $f(x) \in f(c_i) - \varepsilon/(3(b-a)), f(c_i) + \varepsilon/(3(b-a)))$. It follows that

- $m_i \geq f(c_i) - \varepsilon/(3(b-a))$,
- $M_i \leq f(c_i) + \varepsilon/(3(b-a))$, and
- $M_i - m_i \leq 2\varepsilon/(3(b-a))$,

so

$$U(f,P) - L(f,P) = \sum_{i=1}^{n}(M_i - m_i)(t_i - t_{i-1}) \leq (2\varepsilon/(3(b-a)))(b-a) < \varepsilon.$$

$\square$

This raises a natural question: *which* continuous functions on $[a,b]$ are *uniformly* continuous? All of them, it turns out!

Theorem 10.17. *If $f : [a,b] \to \mathbb{R}$ is continuous, it is uniformly continuous.*

Corollary 10.18. *If $f : [a,b] \to \mathbb{R}$ is continuous, it is integrable.*

So, the set of functions on $[a,b]$ that are differentiable everywhere, sits strictly inside the set of functions on $[a,b]$ that are continuous everywhere, and this in turn sits strictly inside the set of functions on $[a,b]$ that are integrable.

Proof (of Theorem 10.17): Let $\varepsilon > 0$ be given. We would like to show that there is a $\delta > 0$ such that if $x, y \in [a,b]$ are within δ of each other, $f(x)$ and $f(y)$ are within ε of each other. Following Spivak, we say that f is ε*-good* on $[a,b]$ if this happens.

Let $A = \{x \in [a,b] : f \text{ is } \varepsilon\text{-good on } [a,x]\}$. We have $a \in A$ (and $\delta > 0$ will work to witness that f is ε-good on $[a,a]$), and b is an upper bound for A, so by the completeness axiom A has a least upper bound, $\alpha = \sup A$. We will argue first that $\alpha = b$, and then that $\alpha \in A$, to conclude that f is ε-good on $[a,b]$; since $\varepsilon > 0$ was arbitrary, this completes the proof.

Suppose $\alpha < b$. Since f is continuous at α, there is a $\delta > 0$ such that $|y - \alpha| < \delta$ implies $|f(y) - f(\alpha)| < \varepsilon/2$ (and we can choose δ small enough that $|y - \alpha| < \delta$ implies $y \in [a,b]$). So if y, z are both in $(\alpha - \delta, \alpha + \delta)$ then

$$|f(y) - f(z)| \leq |f(y) - f(\alpha)| + |f(z) - f(\alpha)| < \varepsilon.$$

From this we can conclude that f is ε-good on $[\alpha - \delta/2, \alpha + \delta/2]$. But also, because $\alpha = \sup A$, there is some $x \in (\alpha - \delta/3, \alpha]$ with f ε-good on $[a,x]$, and so also f is ε-good on $[a, \alpha - \delta/2]$.

We now need a lemma:

> Suppose $p < q < r$ and f is continuous on $[p,r]$. If f is ε-good on $[p,q]$, and on $[q,r]$, then it is ε-good on $[p,r]$.

With this lemma we conclude that f is ε-good on $[a, \alpha + \delta/2]$, contradicting that $\alpha = \sup A$; so we conclude that $\alpha = b$.

We finish the proof of the main theorem, by establishing that $b \in A$, before proving the lemma. Since f is left-continuous at b, there is a $\delta > 0$ with f is ε-good on $[b - \delta/2, b]$; and since $b = \sup A$, f is ε-good on $[a, b - \delta/2]$ (both of these steps are very similar to our arguments about α). From the lemma it follows that f is ε-good on $[a,b]$, as required.

We now turn to the proof of the lemma. There is $\delta_1 > 0$ such that

$$\text{for } x, y \in [p, q],\ |x - y| < \delta_1 \text{ implies } |f(x) - f(y)| < \varepsilon,\ (\star)$$

and there is $\delta_2 > 0$ such that

$$\text{for } x, y \in [q, r],\ |x - y| < \delta_2 \text{ implies } |f(x) - f(y)| < \varepsilon.\ (\star\star)$$

Since f is continuous at q there is also $\delta_3 > 0$ such that $|x-q| < \delta_3$ implies $|f(x)-f(q)| < \varepsilon/2$, so

$$\text{for } x, y \text{ with } |x - q| < \delta_3 \text{ and } |y - q| < \delta_3, \text{ we have } |f(x) - f(y)| < \varepsilon.\ (\star\star\star)$$

Let $\delta = \min\{\delta_1, \delta_2, \delta_3\}$. Suppose we have $x, y \in [p, r]$ with $|x - y| < \delta$.

- If $x, y \in [p, q]$ then $|f(x) - f(y)| < \varepsilon$, by $(\star)$.
- If $x, y \in [q, r]$ then $|f(x) - f(y)| < \varepsilon$, by $(\star\star)$.
- If $x < q < y$ or $y < q < x$ then both $|x-q| < \delta$ and $|y-q| < \delta$ hold, and $|f(x)-f(y)| < \varepsilon$ by $(\star\star\star)$.

□

10.5 The fundamental theorem of calculus, part 1

Suppose that a function f is integrable on some interval I. Fix $a \in I$. We can define a new function $F : I \to \mathbb{R}$ by

$$F(x) = \int_a^x f.$$

We can think of F as a function which "accumulates area": if f is non-negative, then (by definition) $F(a) = 0$, and for $x > a$ $F(x)$ is the area of the region bounded by the graph of f, the x-axis, and vertical lines at a and x. As x increases, so does F, as more and more area is accumulated. On the other hand if $x < a$ then $F(x)$ is a negative area, and as x gets further from a in the negative direction, F gets more negative. In what follows we will (somewhat informally) refer to F as the "integral" of f.

Recall that we saw that often the derivative of a function is less well behaved than the original function. For example, the function $f(x) = |x|$ is continuous everywhere, but not differentiable everywhere; and the function

$$f(x) = \begin{cases} x^2/2 & \text{if } x \geq 0 \\ -x^2/2 & \text{if } x \leq 0 \end{cases}$$

is differentiable everywhere, but it's derivate ($f'(x) = |x|$) is *not* differentiable everywhere.

The integral of a function, on the other hand, tends to be better behaved than the original function, in the sense that it is typically "smoother". Consider, for example,

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

This function is not continuous on any interval that includes 0. It is integrable on any interval, however, and so we can define the function $F(x) = \int_0^x f(t)\ dt$[173]. Without delving into the gritty details of the definition of the integral, we can calculate the values of $F(x)$ by remember the "accumulating area" interpretation. For $x \geq 0$, $F(x)$ is the area of a square whose dimensions are x by 1, so $F(x) = x$. For $x < 0$ we have

$$F(x) = \int_0^x f(t)\ dt = -\int_x^0 f(t)\ dt = \int_x^0 (-f)(t)\ dt.$$

Since $-f \geq 0$ on $[x, 0]$, this last integral is the area of a square whose dimensions are $-x$ by 1, so $F(x) = -x$. Combining we find that $F(x) = |x|$. So in this example

f is not continuous everywhere, but its integral F is.

As a second example, consider $f(x) = |x|$. Again thinking about the integral function as accumulating area, we find that

$$F(x) := \int_0^x f(t)\ dt = \begin{cases} -x^2/2 & \text{if } x < 0 \\ x^2/2 & \text{if } x \geq 0. \end{cases}$$

Notice that f is continuous at 0, but not differentiable, but F is "smoother", being not just continuous at 0 but also differentiable. So in this example

f is not differentiable everywhere, but its integral F is.

These are just two example, and are not much on which to base a general hypothesis. But we could work through many more examples, and always find the same thing happening — if f is integrable, then integration "smooths out" f in the sense that F, the integral of f,

[173] A note on notation: when we say "$f : [a, b] \to \mathbb{R}$ is function", we mean that symbol f is standing for set a of ordered pairs, whose set of first coordinates is exactly $[a, b]$, and with no element of that set appearing twice as a first coordinate. If f is integrable on $[a, b]$, we use $\int_a^b f$ to denote integral.

Sometimes we represent a function by the expression "$f(x)$", which is usually a rule explaining how to compute the second co-ordinate of the pair whose first co-ordinate is x; for example, $f(x) = x^2$. In this case we write $\int_a^b f(x)\ dx$. Here "dx" means *nothing* on its own.

Note that the "x" in here (in the integral "$\int_a^b f(x)\ dx$") is *not* a variable of a function — $\int_a^b f(x)\ dx$ is not a function, it is a *number*. "x" is a *dummy variable*, and can be given any name we like: $\int_a^b f(x)\ dx = \int_a^b f(y)\ dy = \int_a^b f(\text{banana})\ d\text{banana}$.

Sometimes we are *forced* to use a name other than x in the presentation of an integral: For example here we want to define a function F, whose value at input x is the integral of f on the interval $[a, x]$. We could denote that by $F(x) = \int_a^x f$, or, using the "$f(\cdot)d\cdot$" notation, we could denote it by $F(x) = \int_a^x f(t)\ dt$, or $F(x) = \int_a^x f(r)\ dr$, or .... What we *cannot* do is write $F(x) = \int_1^x \frac{dx}{x}$ — we're already using the symbol x for the variable inputed into the function F (and it *is* a variable in this equation), and so we need another, different (and quite arbitrary) name for the dummy variable.

Sometimes a variable in the limits of integration *can* appear sensibly inside integral. For examle, the expression $\int_1^x xt\ dt$ makes perfect sense for all x. At $x = 2$, it is the number $\int_1^2 2t\ dt$, at $x = -1$ it is the number $\int_1^{-1} (-t)\ dt$, and so on.

is continuous even where f is not, and moreover is differentiable whenever f is continuous, even when f is not differentiable.

We now prove that these phenomena occur in general. We start with continuity.

Proposition 10.19. *Suppose $f : I \to \mathbb{R}$ (defined on an interval I) is integrable on any closed interval contained in I. Fix $a \in I$ and define $F : I \to \mathbb{R}$ by*

$$F(x) = \int_a^x f.$$

Then F is continuous at all points in I.

Proof: Fix $a < c < b \in I$. We will argue that F is continuous from the right at c, and from the left. We start with right-continuity.

Let $h > 0$. Define

- $M_h = \sup\{f(x) : x \in [c, c+h]\}$
- $m_h = \inf\{f(x) : x \in [c, c+h]\}$
- M to be any be such that $|f(x)| < M$ for all $x \in [a, b]$.

Observe that $-M \le m_h \le M_h \le M$ (as long as h is small enough that $c + h \le b$), and $M > 0$.

We have

$$F(c+h) - F(c) = \int_c^{c+h} f$$

so

$$-Mh \le m_h h \le F(c+h) - F(c) \le M_h h \le Mh.$$

By the squeeze theorem, as $h \to 0^+$ get $F(c+h) \to F(c)$, so F is right continuous at c.

For left continuity, again let $h > 0$. Re-define

- $M_h = \sup\{f(x) : x \in [c-h, c]\}$
- $m_h = \inf\{f(x) : x \in [c-h, c]\}$

We still have $-M \le m_h \le M_h \le M$ (as long as $c - h \ge a$), and $M > 0$. Now

$$F(c) - F(c-h) = \int_{c-h}^{c} f$$

so

$$-Mh \le hm_h \le F(c) - F(c-h) \le hM_h \le Mh.$$

Again by the squeeze theorem, as $h \to 0^+$ get $F(c-h) \to F(c)$, so F is left continuous at c.

Notice that there might not *be* a number b such that $a < c < b$; this happens if c is the right-endpoint of I. It is left as an exercise to consider what happens in this case (and so only continuity from the left need be considered). The case $c = a$ is also left as an exercise.

What happens if $c < a$? Now $F(c) = \int_a^c f = -\int_c^a f$, and we can run a very similar argument to the one presented above to obtain the desired result. $\square$

Returning to our two examples, recall that if

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

then $F(x) := \int_0^x f = |x|$. The function f is continuous for all x other than 0, and the integral F is *differentiable* at all those points. Moreover, the derivative of f at every non-zero point x, is exactly the value of f at x.

The same holds for our other example, $f(x) = |x|$, for which

$$F(x) := \int_0^x f(t)\, dt = \begin{cases} -x^2/2 & \text{if } x < 0 \\ x^2/2 & \text{if } x \geq 0: \end{cases}$$

f is continuous everywhere (but not differentiable everywhere), and the integral F is differentiable everywhere, with $F'(x) = f(x)$ for all x.

It is not hard to intuitively see that this is a general phenomenon. Suppose that f is a continuous function, and define $F(x) = \int_a^x f$ (for some fixed a). We have, for $h > 0$

$$\frac{F(x+h) - F(x)}{h} = \frac{1}{h}\int_x^{x+h} f \approx f(x)$$

using that, since f is continuous at x its values close to x are close to $f(x)$, so $\int_x^{x+h} f$ is approximately $f(x)$ times the length of the interval $[x, x+h]$, or $hf(x)$. This strongly suggests that F is differentiable[174] with derivative $f(x)$.

That this is in fact true in general is the content of one of two theorems that are significant enough to warrant the adjective "fundamental".

Theorem 10.20. *(fundamental theorem of calculus, Part 1) Suppose $f : I \to \mathbb{R}$ (defined on an interval I) is integrable on any closed interval contained in I. Fix $a \in I$ and define $F : I \to \mathbb{R}$ by*

$$F(x) = \int_a^x f.$$

If f is continuous at some $c \in I$, then F is differentiable at c, and $F'(c) = f(c)$.

Proof: As with the proof of Proposition 10.19, we begin by considering $a < c < b \in I$. What happens when c is an endpoint of I is left as an exercise, as are the cases $c = a$ and $c < a$.

We begin by showing that F is differentiable from the right at c. With the notation from the proof of Proposition 10.19, we have from that proof that

$$m_h \leq \frac{F(c+h) - F(c)}{h} \leq M_h.$$

[174] at least, from the right; but we make make a similar argument to justify differentiability from the left

Now since f is continuous at c, we have

$$\lim_{h\to 0^+} M_h = \lim_{h\to 0^+} m_h = f(c) \quad (\star)$$

(and so, by the squeeze theorem, F is right differentiable at c with derivative $f(c)$). To see $(\star)$ note that:

- Given $\varepsilon > 0$ there's $\delta > 0$ such that $x \in (c-\delta, c+\delta)$ implies $f(x) \in (f(c)-\varepsilon, f(c)+\varepsilon)$.
- Choose $h = \min\{\delta, b-c\}$; for this h, $f(c) \le M_h \le f(c)+\varepsilon$ and $f(c)-\varepsilon \le m_h \le f(c)$, and indeed for all $0 < h' \le h$ we have $f(c) \le M_{h'} \le f(c)+\varepsilon$ and $f(c)-\varepsilon \le m_{h'} \le f(c)$.
- Letting $\varepsilon \to 0$, the squeeze theorem gives $(\star)$.

For left differentiability we have from the proof of Proposition 10.19 that

$$m_h \le \frac{F(c)-F(c-h)}{-h} \le M_h,$$

and the proof proceeds as before. □

We conclude this section with two remarks.

- It is not uncommon to see the Fundamental Theorem of Calculus, part 1 (from here on abbreviated FTOC1) presented as:

  if f is continuous on I, $a \in I$ and F is defined on I by $F(x) = \int_a^x f$, then F is differentiable on I with $F' = f$.

  This *follows* from FTOC1 as we have given it (just apply FTOC1 at each point in I), but it doesn't obviously *imply* our FTOC1, which allows for the possibility that f is *not* continuous at lots of points of I, but still says something about F at those points where f happens to be continuous.

- When dealing with the case $c < a$ in the proofs of Proposition 10.19 and Theorem 10.20, rather than repeating the arguments for continuity and differentiability of F given for $c > a$, it's possible to take the following approach: for input $x < a$, choose a number $b < x$. We have

$$F(x) = \int_a^x f = -\int_x^a f = -\left(\int_b^a f - \int_b^x f\right) = \int_b^x f - \int_b^a f = G(x) - \int_b^a f$$

  where G is defined by $G(x) = \int_b^x f$. Because $x > b$ we know (from the arguments given in the proofs of Proposition 10.19 and Theorem 10.20) that G is continuous at x, so F is also, and that if f is continuous at x then G is differentiable at x with $G'(x) = f(x)$, so F is also differentiable at x with $F'(x) = f(x)$.

10.6 The fundamental theorem of calculus, part 2

FTOC1 says (loosely) that if a new function is defined from an old function by integration, then (under appropriate circumstances) the derivative of that new function is the old function. So "differentiation undoes integration". The connection between differentiation and integration goes the other way, too. We start with a statement that is a fairly immediate corollary of FTOC1, that we will think of as a "weak" version of what we will eventually call the fundamental theorem of calculus, part 2.

Corollary 10.21. *(Weak FTOC2; a corollary of FTOC1) Suppose that f is continuous on $[a,b]$.* If *there is a function $g : [a,b] \to \mathbb{R}$ such that $g' = f$ on $[a,b]$, then*

$$\int_a^b f = g(b) - g(a).$$

Proof: Since f is continuous, it is integrable, so the function $F(x) = \int_a^x f$ exists; and moreover, by FTOC1, F is differentiable on $[a,b]$, with $F' = f = g'$. Since $F' = g'$ it follows that there is a number c such that $F(x) = g(x) + c$. Evaluating at $x = a$ we get $c = -g(a)$, and then evaluating at b we get $F(b) = g(b) - g(a)$, as claimed. □

The power of this corollary is that if a function f is continuous, and if we can find an *antiderivative* or *primitive* of f — a function g such that $g' = f$, then we can *very easily* evaluate integrals involving f.

For example, what is the area under $y = x^n$, above x-axis, between $x = a$ and $x = b$? In other words, what is

$$\int_a^b x^n \, dx?$$

Here $f(x) = x^n$, and since, as long as n is an integer other than -1

$$g(x) = \frac{x^{n+1}}{n+1}$$

has $g'(x) = f(x)$, we get

$$\text{Area} = \int_a^b x^n \, dx = g(b) - g(a) = \frac{b^{n+1} - a^{n+1}}{n+1}. \tag{9}$$

Calculating this directly, using the definition of the integral, would be quite unpleasant.

If $n = -2$, $a = 1$ and $b = 10$, we have

$$\text{Area} = \int_1^{10} \frac{dx}{x^2} = \left(-\frac{1}{10}\right) - \left(-\frac{1}{1}\right) = \frac{9}{10},$$

a reasonable answer. If $n = -2$ and we try to use (9) to calculate the area under $y = x^2$, above the x-axis, and between $x = -a$ and $x = a$ for some positive a, we get

$$\int_{-a}^a \frac{dx}{x^2} = \left(-\frac{1}{a}\right) - \left(-\frac{1}{-a}\right) = -\frac{2}{a},$$

which is clearly wrong, since the function we are integrating is always positive!

What went wrong here? The issue is that the expression $1/x^2$ is not defined on all of $[-a, a]$ (specifically it is not defined at 0), and so the integral expression makes no sense. So we need to insert a caveat into (9): for negative n we have

$$\int_a^b x^n \, dx = g(b) - g(a) = \frac{b^{n+1} - a^{n+1}}{n+1}$$

only if either $0 < a \leq b$ or $a \leq b < 0$.

We specified above that $n \neq -1$; this is because there is no obvious g with $g'(x) = 1/x$. We know, though, that there *must be* such a function g: since $1/x$ is a continuous function away from 0, by FTOC1 if we define (for positive x)

$$g(x) = \int_1^x \frac{dt}{t}$$

then $g'(x) = 1/x$. But, if we know of no better expression for $g(x)$, then Corollary 10.21 (weak FTOC2) tells us nothing about the value of integrals of $1/x$: it tells us that

$$\int_1^a \frac{dx}{x} = \int_1^a \frac{dt}{t} - \int_1^1 \frac{dt}{t} = \int_1^a \frac{dt}{t},$$

a trivial tautology.[175]

We refer to Corollary 10.21 as "weak" because it includes as a hypothesis that f is continuous. As we have seen, this makes it little more than a corollary of FTOC1. We can in fact drop this hypothesis, or rather, replace it with the weaker hypothesis that f is merely integrable. The proof of what we will call the *fundamental theorem of calculus, part 2* cannot appeal to FTOC1 (because no assumption about continuity of f is made), and so has to go back to the definition of the integral. Nonetheless, it is quite a short proof.

Theorem 10.22. *(fundamental theorem of calculus, part 2) Suppose $f : [a, b] \to \mathbb{R}$ is integrable, and that there is some function $g : [a, b] \to \mathbb{R}$ such that $g' = f$. Then*

$$\int_a^b f = g(b) - g(a)$$

.

Proof: g is differentiable on $[a, b]$, so it is continuous. Given a partition P of $[a, b]$, there is an x_i in each $[t_{i-1}, t_i]$ with

$$g(t_i) - g(t_{i-1}) = g'(x_i)(t_i - t_{i-1}) = f(x_i)(t_i - t_{i-1}).$$

(The first equality here is an application of the Mean Value Theorem.) Now $f(x_i) \in [m_i, M_i]$, so from the above we get

$$m_i(t_i - t_{i-1}) \leq g(t_i) - g(t_{i-1}) \leq M_i(t_i - t_{i-1})$$

[175] We will (of course) return to $\int_1^x dt/t$ shortly!

Summing these inequalities from $i = 1$ to n we get

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

This is true for *all* partitions P. Since f is integrable, it follows that $\int_a^b f = g(b) - g(a)$. $\square$

Note that FTOC2 *does not* assert that if f is integrable, then there is a function g with $g' = f$; on the homework, there is example of an $f : [a, b] \to \mathbb{R}$ that is integrable, for which there is no g satisfying $g' = f$. Rather, FTOC2 assert that *if* f is integrable *and also* there is a function g with $g' = f$, then $\int_a^b f$ can be computed using g.

The idea of taking a known function f and creating from it a new function F via $F(x) = \int_a^x f$ is a very valuable one; it will allow us to properly define many of the basic transcendental (beyond ration) functions of mathematics, such as

- the trigonometric functions
- the hyperbolic trigonometric functions $(\sinh, \cosh$, et cetera)
- the logarithmic and exponential functions.

It's important to be able to work with fluently with this construction, and to manipulate functions defined in this way just as we manipulated functions defined in a more standard way. We present some examples here.

- Suppose $f(x) = \int_1^{x^2-x} g(t)\ dt$. What is $f'(x)$?

  Define $F(x) = \int_1^x g(t)\ dt$. Then $f(x) = F(x^2 - x) = (F \circ h)(x)$ where $h(x) = x^2 - x$, so, by the chain rule, $f'(x) = F'(h(x))h'(x)$. Now $F'(x) = g(x)$, so $F'(h(x)) = g(h(x))$, so

  $$f'(x) = g(x^2 - x)(2x - 1).$$

- Suppose $q(x) = \cos\left(\int_{\sqrt{x}}^a g(t)\ dt\right)$. What is $q'(x)$?

  Well, $q(x) = (\cos \circ m \circ F \circ s)(x)$ where $s(x) = \sqrt{x}$, $F(x) = \int_a^x g(t)\ dt$, and $m(x) = -x$, so

  $$\begin{aligned} q'(x) &= \cos'(m \circ F \circ s(x))m'((F \circ s)(x))F'(s(x))s'(x) \\ &= \left[-\sin\left(\int_{\sqrt{x}}^a g(t)\ dt\right)\right] \times [-1] \times \left[g(\sqrt{x})\right] \times \left[\frac{1}{2x^{1/2}}\right]. \end{aligned}$$

These and other similar examples are essentially all exercises in the chain rule.

10.7 Improper integrals

The aim of this section is to make sense of expressions like

- $\int_0^\infty f$,

and

- $\int_a^b f$ when f is not necessarily bounded on $[a, b]$.

We will refer to integrals like this – whose interpretations are derived from the definition of the integral, but do not perfectly fit the conditions of the definition — as *improper integrals*.

We start with integrals like $\int_0^\infty f$, where there is an obvious definition:

Definition of improper integral (I) The expression $\int_a^\infty f$ is defined to mean

$$\int_a^\infty f := \lim_{N\to\infty} \int_a^N f,$$

as long as this limit exists (and so, in particular, as long as f is bounded on $[a, N]$ and $\int_a^N f$ exists, for every large enough N[176]

As an example, we consider $\int_1^\infty \frac{dx}{x^r}$, where $r \geq 2$ is an integer.[177]. Certainly $\int_1^N \frac{dx}{x^r}$ exists for all $N \geq 1$, since $f(x) = 1/x^r$ is continuous on any such interval. Moreover the integral is easy to calculate since f has a simple antiderivative: $g(x) = -1/((r-1)x^{r-1})$. So

$$\int_1^N \frac{dx}{x^r} = g(N) - g(1) = \frac{1}{r-1} - \frac{1}{(r-1)N^{r-1}}.$$

Now for $r \geq 2$ we have $r - 1 \geq 1$ and $1/N^{r-1} \to 0$ as $N \to \infty$, and so

$$\int_1^\infty \frac{dx}{x^r} = \lim_{N\to\infty} \left(\frac{1}{r-1} - \frac{1}{(r-1)N^{r-1}} \right) = \frac{1}{r-1}.$$

This clearly doesn't work when $r = 1$; and in fact the above computation hints that $\int_1^\infty \frac{dx}{x}$ probably does not exist. To see this formally we want to show that

$$\lim_{N\to\infty} \int_1^N \frac{dx}{x}$$

does not exist. We will do this by considering $\int_1^{2^n} \frac{dx}{x}$, for integers $n \geq 0$. We cannot evaluate this integral directly (we do not yet know anything about an antiderivative for $1/x$), but we can put a lower bound on the integral by observing that on the interval $[1, 2]$, the function f

[176]But of course that immediately means, via Lemma 10.7, that f is bounded on $[a, N]$, and $\int_a^N f$ exists, for all $N \geq a$.

[177]We haven't yet properly defined x^r for r *not* an integer — that will come soon.

has $1/2$ as a lower bound; on $[2,4]$ it has $1/4$, and in general on $[2^{k-1}, 2^k]$ there is a lower bound of $1/2^k$. So the partition $P = (1, 2, 4, \ldots, 2^i, \ldots, 2^n)$ has

$$L(f,P) = \sum_{i=1}^{n} m_i(t_i - t_{i-1}) = \sum_{i=1}^{n} \frac{2^{i-1}}{2^i} = \frac{n}{2}.$$

Since $n/2 \to \infty$ as $n \to \infty$, this shows that $L(f)$ does not exist.

Definition of improper integral (II) The expression $\int_{-\infty}^{a} f$ is defined to mean

$$\int_{-\infty}^{a} f := \lim_{N \to -\infty} \int_{N}^{a} f,$$

as long as this limit exists.

What about an expression like $\int_{-\infty}^{\infty} f$? It is tempting to consider $\lim_{N\to\infty} \int_{-N}^{N} f$, and take this limit (if it exists) to be the value of the improper integral. But this approaches leads to some unfortunate oddities. For example, we would have

$$\int_{-\infty}^{\infty} x\ dx \neq \int_{-\infty}^{a} x\ dx + \int_{a}^{\infty} x\ dx$$

since the integral on the left would equal 0, and neither of the two integrals on the right exist.[178]

A much better approach is to notice that the range of integration $(-\infty, \infty)$ has two "problem points" ($+\infty$ and $-\infty$). We can break up the range of integration into two pieces, in such a way that each piece has only one problem point, and then use the previous definitions.

Definition of improper integral (III) The expression $\int_{-\infty}^{\infty} f$ is defined to mean

$$\int_{-\infty}^{\infty} f := \int_{-\infty}^{0} f + \int_{0}^{\infty} f,$$

as long as both of the integrals on the right exist.

The was nothing special about the choice of 0 in this definition. The following easy lemma is left as an exercise.

Lemma 10.23. *If $\int_{-\infty}^{\infty} f$ exists, then for* every *real a, both of $\int_{-\infty}^{a} f, \int_{a}^{\infty} f$ exist, and $\int_{-\infty}^{\infty} f = \int_{-\infty}^{a} f + \int_{a}^{\infty} f$. Conversely, if $\int_{-\infty}^{a} f, \int_{a}^{\infty} f$ both exist for* some *$a \in \mathbb{R}$, then also both $\int_{-\infty}^{0} f, \int_{0}^{\infty} f$ exist, and so $\int_{-\infty}^{\infty} f$ exists.*

[178] One fix might be: since $\int_{-\infty}^{a} x\ dx = -\infty$, and $\int_{a}^{\infty} x = \infty$, we could *declare* "$0 = -\infty + \infty$". But this leads to serious problems; for example, we would have

$$\infty = \lim_{x\to\infty} x = \lim_{x\to\infty} (2x - x) = \infty - \infty = 0.$$

Before giving an example of an integral of this kind, we present a useful "comparison" lemma.

Lemma 10.24. *Suppose*

- $0 \leq g(x) \leq f(x)$ *for all* $x \in [a, \infty)$;
- $\int_a^\infty f$ *exists; and*
- $\int_a^N g$ *exists for all* $N \geq a$.

Then $\int_a^\infty g$ *exists, and* $\int_a^\infty g \leq \int_a^\infty f$.

Proof: Consider the set $A = \{\int_a^N g : N \geq a\}$. This certainly non-empty, and it is bounded above by $\int_a^\infty f$ (since, for all $N \geq a$,

$$\int_a^N g \leq \int_a^N f \leq \int_a^\infty f).$$

It follows that A has a supremum $\alpha \leq \int_a^\infty f$. We claim that $\lim_{N\to\infty} \int_a^N g$ exists and equals α, from which the claim follows.

Fix $\varepsilon > 0$. There is an N_ε such that $\alpha - \varepsilon \leq \int_a^{N_\varepsilon} g \leq \alpha$. Now for all $N > N_\varepsilon$ we have $\int_a^{N_\varepsilon} g \leq \int_a^N g \leq \alpha$ (the latter inequality because $\alpha = \sup A$, the former because $g(x) \geq 0$ so $\int_a^N g$ increases as N increases). So for all $N > N_\varepsilon$ we have $\alpha - \varepsilon \leq \int_a^N g \leq \alpha$. Since this is true for arbitrary $\varepsilon > 0$, it follows that $\lim_{N\to\infty} \int_a^N g = \alpha$, as claimed. □

We illustrate both the definition of $\int_{-\infty}^\infty f$ and the utility of Lemma 10.24 with the example of $\int_{-\infty}^\infty \frac{dt}{1+t^2}$. To see if this integral exists, we need to examine $\int_0^\infty \frac{dt}{1+t^2}$ and $\int_{-\infty}^0 \frac{dt}{1+t^2}$. For the first of these, note that on $(0, \infty)$ we have

$$0 \leq \frac{1}{1+t^2} \leq \frac{1}{t^2}.$$

Unfortunately, this inequality does not extend to $t = 0$. However, we can use it on $[1, \infty)$ to conclude, via Lemma 10.24 (and the fact proved earlier that $\int_1^\infty \frac{dt}{t^2}$ exists), to conclude that $\int_1^\infty \frac{dt}{1+t^2}$ exists; and since $\int_0^1 \frac{dt}{1+t^2}$ clearly exists, it quickly follows that $\int_0^\infty \frac{dt}{1+t^2}$ exists.

For $\int_{-\infty}^0 \frac{dt}{1+t^2}$, we could develop an analog of Lemma 10.24 for the interval $(-\infty, a]$, or we could appeal to symmetry. One half of the following lemma appears in the homework; the other half (the half we'll use here) is left as an exercise.

Lemma 10.25. *Let* $f : [-b, b] \to \mathbb{R}$ *be a function that is integrable on the interval* $[0, b]$.

- *If* f *is an odd function* ($f(-x) = -f(x)$) *then* $\int_{-b}^b f$ *exists, and equals* 0.
- *If* f *is an even function* ($f(-x) = f(x)$) *then* $\int_{-b}^b f$ *exists, and equals* $2\int_0^b f$.

From this it quickly follows that $\int_{-\infty}^{0} \frac{dt}{1+t^2}$ exists and equals $\int_0^\infty \frac{dt}{1+t^2}$, and so $\int_{-\infty}^{\infty} \frac{dt}{1+t^2}$ exists.[179]

We discuss one more kind of improper integral.

Definition of improper integral (IV) Suppose that the function $f : (a, b] \to \mathbb{R}$ is unbounded, but that it is bounded on every interval $[a + \varepsilon, b]$ for $\varepsilon > 0$. Define $\int_a^b f$ by

$$\int_a^b f := \lim_{\varepsilon \to 0^+} \int_{a+\varepsilon}^b f,$$

as long as this limit exists.

As an example, we consider $\int_0^a \frac{dx}{\sqrt{x}}$. We can compute $\int_\varepsilon^a \frac{dx}{\sqrt{x}}$ using the fundamental theorem of calculus (part 2): since $g(x) = 2\sqrt{x}$ has $g'(x) = 1/\sqrt{x}$, we have

$$\int_\varepsilon^a \frac{dx}{\sqrt{x}} = g(a) - g(\varepsilon) = 2\sqrt{a} - 2\sqrt{\varepsilon}.$$

Since $2\sqrt{a} - 2\sqrt{\varepsilon} \to 2\sqrt{a}$ as $\varepsilon \to 0$, we conclude that

$$\int_0^a \frac{dx}{\sqrt{x}} = 2\sqrt{a}.$$

There are many other variants of improper integrals. For example, what would $\int_0^\infty f$ mean if f is bounded on $[\varepsilon, \infty)$ for all $\varepsilon > 0$, but f is unbounded on $(0, \infty)$? The paradigm is to break up range of integration into pieces, in such a way that each piece has only one problem point. So to check if $\int_0^\infty f$ exists in this scenario, we would check each of $\int_0^a f$, $\int_a^\infty f$, using the definitions we have given earlier.

[179] Notice that Lemma 10.24 gives no hint as to what the *value* of the integral is; we will return to this later.