Chapter 1

The Riemann Integral

> I know of some universities in England where the Lebesgue integral is *taught* in the first year of a mathematics degree instead of the Riemann integral, but I know of no universities in England where students *learn* the Lebesgue integral in the first year of a mathematics degree. (Approximate quotation attributed to T. W. Körner)

Let $f : [a, b] \to \mathbb{R}$ be a bounded (not necessarily continuous) function on a compact (closed, bounded) interval. We will define what it means for f to be Riemann integrable on $[a, b]$ and, in that case, define its Riemann integral $\int_a^b f$. The integral of f on $[a, b]$ is a real number whose geometrical interpretation is the signed area under the graph $y = f(x)$ for $a \leq x \leq b$. This number is also called the definite integral of f. By integrating f over an interval $[a, x]$ with varying right end-point, we get a function of x, called the indefinite integral of f.

The most important result about integration is the fundamental theorem of calculus, which states that integration and differentiation are inverse operations in an appropriately understood sense. Among other things, this connection enables us to compute many integrals explicitly.

Integrability is a less restrictive condition on a function than differentiability. Roughly speaking, integration makes functions smoother, while differentiation makes functions rougher. For example, the indefinite integral of every continuous function exists and is differentiable, whereas the derivative of a continuous function need not exist (and generally doesn't).

The Riemann integral is the simplest integral to define, and it allows one to integrate every continuous function as well as some not-too-badly discontinuous functions. There are, however, many other types of integrals, the most important of which is the Lebesgue integral. The Lebesgue integral allows one to integrate unbounded or highly discontinuous functions whose Riemann integrals do not exist, and it has better mathematical properties than the Riemann integral. The definition of the Lebesgue integral requires the use of measure theory, which we will not describe here. In any event, the Riemann integral is adequate for many purposes, and even if one needs the Lebesgue integral, it's better to understand the Riemann integral first.

1.1. Definition of the Riemann integral

We say that two intervals are almost disjoint if they are disjoint or intersect only at a common endpoint. For example, the intervals $[0,1]$ and $[1,3]$ are almost disjoint, whereas the intervals $[0,2]$ and $[1,3]$ are not.

Definition 1.1. Let I be a nonempty, compact interval. A partition of I is a finite collection $\{I_1, I_2, \dots, I_n\}$ of almost disjoint, nonempty, compact subintervals whose union is I.

A partition of $[a,b]$ with subintervals $I_k = [x_{k-1}, x_k]$ is determined by the set of endpoints of the intervals

$$a = x_0 < x_1 < x_2 < \cdots < x_{n-1} < x_n = b.$$

Abusing notation, we will denote a partition P either by its intervals

$$P = \{I_1, I_2, \dots, I_n\}$$

or by the set of endpoints of the intervals

$$P = \{x_0, x_1, x_2, \dots, x_{n-1}, x_n\}.$$

We'll adopt either notation as convenient; the context should make it clear which one is being used. There is always one more endpoint than interval.

Example 1.2. The set of intervals

$$\{[0,1/5], [1/5,1/4], [1/4,1/3], [1/3,1/2], [1/2,1]\}$$

is a partition of $[0,1]$. The corresponding set of endpoints is

$$\{0, 1/5, 1/4, 1/3, 1/2, 1\}.$$

We denote the length of an interval $I = [a,b]$ by

$$|I| = b - a.$$

Note that the sum of the lengths $|I_k| = x_k - x_{k-1}$ of the almost disjoint subintervals in a partition $\{I_1, I_2, \dots, I_n\}$ of an interval I is equal to length of the whole interval. This is obvious geometrically; algebraically, it follows from the telescoping series

$$\begin{aligned}
\sum_{k=1}^{n} |I_k| &= \sum_{k=1}^{n} (x_k - x_{k-1}) \\
&= x_n - x_{n-1} + x_{n-1} - x_{n-2} + \cdots + x_2 - x_1 + x_1 - x_0 \\
&= x_n - x_0 \\
&= |I|.
\end{aligned}$$

Suppose that $f : [a,b] \to \mathbb{R}$ is a bounded function on the compact interval $I = [a,b]$ with

$$M = \sup_I f, \qquad m = \inf_I f.$$

If $P = \{I_1, I_2, \dots, I_n\}$ is a partition of I, let

$$M_k = \sup_{I_k} f, \qquad m_k = \inf_{I_k} f.$$

These suprema and infima are well-defined, finite real numbers since f is bounded. Moreover,

$$m \le m_k \le M_k \le M.$$

If f is continuous on the interval I, then it is bounded and attains its maximum and minimum values on each subinterval, but a bounded discontinuous function need not attain its supremum or infimum.

We define the upper Riemann sum of f with respect to the partition P by

$$U(f;P) = \sum_{k=1}^{n} M_k |I_k| = \sum_{k=1}^{n} M_k (x_k - x_{k-1}),$$

and the lower Riemann sum of f with respect to the partition P by

$$L(f;P) = \sum_{k=1}^{n} m_k |I_k| = \sum_{k=1}^{n} m_k (x_k - x_{k-1}).$$

Geometrically, $U(f;P)$ is the sum of the areas of rectangles based on the intervals I_k that lie above the graph of f, and $L(f;P)$ is the sum of the areas of rectangles that lie below the graph of f. Note that

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

Let $\Pi(a,b)$, or Π for short, denote the collection of all partitions of $[a,b]$. We define the upper Riemann integral of f on $[a,b]$ by

$$U(f) = \inf_{P \in \Pi} U(f;P).$$

The set $\{U(f;P) : P \in \Pi\}$ of all upper Riemann sums of f is bounded from below by $m(b-a)$, so this infimum is well-defined and finite. Similarly, the set $\{L(f;P) : P \in \Pi\}$ of all lower Riemann sums is bounded from above by $M(b-a)$, and we define the lower Riemann integral of f on $[a,b]$ by

$$L(f) = \sup_{P \in \Pi} L(f;P).$$

These upper and lower sums and integrals depend on the interval $[a,b]$ as well as the function f, but to simplify the notation we won't show this explicitly. A commonly used alternative notation for the upper and lower integrals is

$$U(f) = \overline{\int_a^b} f, \qquad L(f) = \underline{\int_a^b} f.$$

Note the use of "lower-upper" and "upper-lower" approximations for the integrals: we take the infimum of the upper sums and the supremum of the lower sums. As we show in Proposition 1.13 below, we always have $L(f) \le U(f)$, but in general the upper and lower integrals need not be equal. We define Riemann integrability by their equality.

Definition 1.3. A bounded function $f : [a, b] \to \mathbb{R}$ is Riemann integrable on $[a, b]$ if its upper integral $U(f)$ and lower integral $L(f)$ are equal. In that case, the Riemann integral of f on $[a, b]$, denoted by

$$\int_a^b f(x)\,dx, \qquad \int_a^b f, \qquad \int_{[a,b]} f$$

or similar notations, is the common value of $U(f)$ and $L(f)$.

An unbounded function is not Riemann integrable. In the following, "integrable" will mean "Riemann integrable, and "integral" will mean "Riemann integral" unless stated explicitly otherwise.

1.2. Examples of the Riemann integral

Let us illustrate the definition of Riemann integrability with a number of examples.

Example 1.4. Define $f : [0, 1] \to \mathbb{R}$ by

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

Then

$$\int_0^1 \frac{1}{x}\,dx$$

isn't defined as a Riemann integral becuase f is unbounded. In fact, if

$$0 < x_1 < x_2 < \cdots < x_{n-1} < 1$$

is a partition of $[0, 1]$, then

$$\sup_{[0,x_1]} f = \infty,$$

so the upper Riemann sums of f are not well-defined.

An integral with an unbounded interval of integration, such as

$$\int_1^\infty \frac{1}{x}\,dx,$$

also isn't defined as a Riemann integral. In this case, a partition of $[1, \infty)$ into finitely many intervals contains at least one unbounded interval, so the corresponding Riemann sum is not well-defined. A partition of $[1, \infty)$ into bounded intervals (for example, $I_k = [k, k + 1]$ with $k \in \mathbb{N}$) gives an infinite series rather than a finite Riemann sum, leading to questions of convergence.

One can interpret the integrals in this example as limits of Riemann integrals, or improper Riemann integrals,

$$\int_0^1 \frac{1}{x}\,dx = \lim_{\epsilon \to 0^+} \int_\epsilon^1 \frac{1}{x}\,dx, \qquad \int_1^\infty \frac{1}{x}\,dx = \lim_{r \to \infty} \int_1^r \frac{1}{x}\,dx,$$

but these are not proper Riemann integrals in the sense of Definition 1.3. Such improper Riemann integrals involve two limits — a limit of Riemann sums to define the Riemann integrals, followed by a limit of Riemann integrals. Both of the improper integrals in this example diverge to infinity. (See Section 1.10.)

Next, we consider some examples of bounded functions on compact intervals.

Example 1.5. The constant function $f(x) = 1$ on $[0,1]$ is Riemann integrable, and

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

To show this, let $P = \{I_1, I_2, \dots, I_n\}$ be any partition of $[0,1]$ with endpoints

$$\{0, x_1, x_2, \dots, x_{n-1}, 1\}.$$

Since f is constant,

$$M_k = \sup_{I_k} f = 1, \quad m_k = \inf_{I_k} f = 1 \qquad \text{for } k = 1, \dots, n,$$

and therefore

$$U(f;P) = L(f;P) = \sum_{k=1}^{n} (x_k - x_{k-1}) = x_n - x_0 = 1.$$

Geometrically, this equation is the obvious fact that the sum of the areas of the rectangles over (or, equivalently, under) the graph of a constant function is exactly equal to the area under the graph. Thus, every upper and lower sum of f on $[0,1]$ is equal to 1, which implies that the upper and lower integrals

$$U(f) = \inf_{P\in\Pi} U(f;P) = \inf\{1\} = 1, \qquad L(f) = \sup_{P\in\Pi} L(f;P) = \sup\{1\} = 1$$

are equal, and the integral of f is 1.

More generally, the same argument shows that every constant function $f(x) = c$ is integrable and

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

The following is an example of a discontinuous function that is Riemann integrable.

Example 1.6. The function

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

is Riemann integrable, and

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

To show this, let $P = \{I_1, I_2, \dots, I_n\}$ be a partition of $[0,1]$. Then, since $f(x) = 0$ for $x > 0$,

$$M_k = \sup_{I_k} f = 0, \quad m_k = \inf_{I_k} f = 0 \qquad \text{for } k = 2, \dots, n.$$

The first interval in the partition is $I_1 = [0, x_1]$, where $0 < x_1 \le 1$, and

$$M_1 = 1, \quad m_1 = 0,$$

since $f(0) = 1$ and $f(x) = 0$ for $0 < x \le x_1$. It follows that

$$U(f;P) = x_1, \qquad L(f;P) = 0.$$

Thus, $L(f) = 0$ and

$$U(f) = \inf\{x_1 : 0 < x_1 \le 1\} = 0,$$

so $U(f) = L(f) = 0$ are equal, and the integral of f is 0. In this example, the infimum of the upper Riemann sums is not attained and $U(f;P) > U(f)$ for every partition P.

A similar argument shows that a function $f : [a,b] \to \mathbb{R}$ that is zero except at finitely many points in $[a,b]$ is Riemann integrable with integral 0.

The next example is a bounded function on a compact interval whose Riemann integral doesn't exist.

Example 1.7. The Dirichlet function $f : [0,1] \to \mathbb{R}$ is defined by

$$f(x) = \begin{cases} 1 & \text{if } x \in [0,1] \cap \mathbb{Q}, \\ 0 & \text{if } x \in [0,1] \setminus \mathbb{Q}. \end{cases}$$

That is, f is one at every rational number and zero at every irrational number.

This function is not Riemann integrable. If $P = \{I_1, I_2, \dots, I_n\}$ is a partition of $[0,1]$, then

$$M_k = \sup_{I_k} f = 1, \qquad m_k = \inf_{I_k} = 0,$$

since every interval of non-zero length contains both rational and irrational numbers. It follows that

$$U(f;P) = 1, \qquad L(f;P) = 0$$

for every partition P of $[0,1]$, so $U(f) = 1$ and $L(f) = 0$ are not equal.

The Dirichlet function is discontinuous at every point of $[0,1]$, and the moral of the last example is that the Riemann integral of a highly discontinuous function need not exist.

1.3. Refinements of partitions

As the previous examples illustrate, a direct verification of integrability from Definition 1.3 is unwieldy even for the simplest functions because we have to consider all possible partitions of the interval of integration. To give an effective analysis of Riemann integrability, we need to study how upper and lower sums behave under the refinement of partitions.

Definition 1.8. A partition $Q = \{J_1, J_2, \dots, J_m\}$ is a refinement of a partition $P = \{I_1, I_2, \dots, I_n\}$ if every interval I_k in P is an almost disjoint union of one or more intervals J_ℓ in Q.

Equivalently, if we represent partitions by their endpoints, then Q is a refinement of P if $Q \supset P$, meaning that every endpoint of P is an endpoint of Q. We don't require that every interval — or even any interval — in a partition has to be split into smaller intervals to obtain a refinement; for example, every partition is a refinement of itself.

Example 1.9. Consider the partitions of $[0,1]$ with endpoints

$$P = \{0, 1/2, 1\}, \qquad Q = \{0, 1/3, 2/3, 1\}, \qquad R = \{0, 1/4, 1/2, 3/4, 1\}.$$

Thus, P, Q, and R partition $[0,1]$ into intervals of equal length $1/2$, $1/3$, and $1/4$, respectively. Then Q is not a refinement of P but R is a refinement of P.

Given two partitions, neither one need be a refinement of the other. However, two partitions P, Q always have a common refinement; the smallest one is $R = P \cup Q$, meaning that the endpoints of R are exactly the endpoints of P or Q (or both).

Example 1.10. Let $P = \{0, 1/2, 1\}$ and $Q = \{0, 1/3, 2/3, 1\}$, as in Example 1.9. Then Q isn't a refinement of P and P isn't a refinement of Q. The partition $S = P \cup Q$, or

$$S = \{0, 1/3, 1/2, 2/3, 1\},$$

is a refinement of both P and Q. The partition S is not a refinement of R, but $T = R \cup S$, or

$$T = \{0, 1/4, 1/3, 1/2, 2/3, 3/4, 1\},$$

is a common refinement of all of the partitions $\{P, Q, R, S\}$.

As we show next, refining partitions decreases upper sums and increases lower sums. (The proof is easier to understand than it is to write out — draw a picture!)

Theorem 1.11. Suppose that $f : [a, b] \to \mathbb{R}$ is bounded, P is a partitions of $[a, b]$, and Q is refinement of P. Then

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

Proof. Let

$$P = \{I_1, I_2, \dots, I_n\}, \qquad Q = \{J_1, J_2, \dots, J_m\}$$

be partitions of $[a, b]$, where Q is a refinement of P, so $m \geq n$. We list the intervals in increasing order of their endpoints. Define

$$M_k = \sup_{I_k} f, \quad m_k = \inf_{I_k} f, \qquad M'_\ell = \sup_{J_\ell} f, \quad m'_\ell = \inf_{J_\ell} f.$$

Since Q is a refinement of P, each interval I_k in P is an almost disjoint union of intervals in Q, which we can write as

$$I_k = \bigcup_{\ell=p_k}^{q_k} J_\ell$$

for some indices $p_k \leq q_k$. If $p_k < q_k$, then I_k is split into two or more smaller intervals in Q, and if $p_k = q_k$, then I_k belongs to both P and Q. Since the intervals are listed in order, we have

$$p_1 = 1, \qquad p_{k+1} = q_k + 1, \qquad q_n = m.$$

If $p_k \leq \ell \leq q_k$, then $J_\ell \subset I_k$, so

$$M'_\ell \leq M_k, \qquad m_k \geq m'_\ell \qquad \text{for } p_k \leq \ell \leq q_k.$$

Using the fact that the sum of the lengths of the J-intervals is the length of the corresponding I-interval, we get that

$$\sum_{\ell=p_k}^{q_k} M'_\ell |J_\ell| \leq \sum_{\ell=p_k}^{q_k} M_k |J_\ell| = M_k \sum_{\ell=p_k}^{q_k} |J_\ell| = M_k |I_k|.$$

It follows that

$$U(f;Q) = \sum_{\ell=1}^{m} M'_\ell |J_\ell| = \sum_{k=1}^{n} \sum_{\ell=p_k}^{q_k} M'_\ell |J_\ell| \le \sum_{k=1}^{n} M_k |I_k| = U(f;P)$$

Similarly,

$$\sum_{\ell=p_k}^{q_k} m'_\ell |J_\ell| \ge \sum_{\ell=p_k}^{q_k} m_k |J_\ell| = m_k |I_k|,$$

and

$$L(f;Q) = \sum_{k=1}^{n} \sum_{\ell=p_k}^{q_k} m'_\ell |J_\ell| \ge \sum_{k=1}^{n} m_k |I_k| = L(f;P),$$

which proves the result. □

It follows from this theorem that all lower sums are less than or equal to all upper sums, not just the lower and upper sums associated with the same partition.

Proposition 1.12. If $f : [a,b] \to \mathbb{R}$ is bounded and P, Q are partitions of $[a,b]$, then

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

Proof. Let R be a common refinement of P and Q. Then, by Theorem 1.11,

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

It follows that

$$L(f;P) \le L(f;R) \le U(f;R) \le U(f;Q).$$

□

An immediate consequence of this result is that the lower integral is always less than or equal to the upper integral.

Proposition 1.13. If $f : [a,b] \to \mathbb{R}$ is bounded, then

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

Proof. Let

$$A = \{L(f;P) : P \in \Pi\}, \qquad B = \{U(f;P) : P \in \Pi\}.$$

From Proposition 1.12, $a \le b$ for every $a \in A$ and $b \in B$, so Proposition 2.9 implies that $\sup A \le \inf B$, or $L(f) \le U(f)$. □

1.4. The Cauchy criterion for integrability

The following theorem gives a criterion for integrability that is analogous to the Cauchy condition for the convergence of a sequence.

Theorem 1.14. A bounded function $f : [a,b] \to \mathbb{R}$ is Riemann integrable if and only if for every $\epsilon > 0$ there exists a partition P of $[a,b]$, which may depend on ϵ, such that

$$U(f;P) - L(f;P) < \epsilon.$$

Proof. First, suppose that the condition holds. Let $\epsilon > 0$ and choose a partition P that satisfies the condition. Then, since $U(f) \le U(f;P)$ and $L(f;P) \le L(f)$, we have

$$0 \le U(f) - L(f) \le U(f;P) - L(f;P) < \epsilon.$$

Since this inequality holds for every $\epsilon > 0$, we must have $U(f) - L(f) = 0$, and f is integrable.

Conversely, suppose that f is integrable. Given any $\epsilon > 0$, there are partitions Q, R such that

$$U(f;Q) < U(f) + \frac{\epsilon}{2}, \qquad L(f;R) > L(f) - \frac{\epsilon}{2}.$$

Let P be a common refinement of Q and R. Then, by Theorem 1.11,

$$U(f;P) - L(f;P) \le U(f;Q) - L(f;R) < U(f) - L(f) + \epsilon.$$

Since $U(f) = L(f)$, the condition follows. □

If $U(f;P) - L(f;P) < \epsilon$, then $U(f;Q) - L(f;Q) < \epsilon$ for every refinement Q of P, so the Cauchy condition means that a function is integrable if and only if its upper and lower sums get arbitrarily close together for all sufficiently refined partitions.

It is worth considering in more detail what the Cauchy condition in Theorem 1.14 implies about the behavior of a Riemann integrable function.

Definition 1.15. The oscillation of a bounded function f on a set A is

$$\operatorname*{osc}_A f = \sup_A f - \inf_A f.$$

If $f : [a,b] \to \mathbb{R}$ is bounded and $P = \{I_1, I_2, \dots, I_n\}$ is a partition of $[a,b]$, then

$$U(f;P) - L(f;P) = \sum_{k=1}^{n} \sup_{I_k} f \cdot |I_k| - \sum_{k=1}^{n} \inf_{I_k} f \cdot |I_k| = \sum_{k=1}^{n} \operatorname*{osc}_{I_k} f \cdot |I_k|.$$

A function f is Riemann integrable if we can make $U(f;P) - L(f;P)$ as small as we wish. This is the case if we can find a sufficiently refined partition P such that the oscillation of f on most intervals is arbitrarily small, and the sum of the lengths of the remaining intervals (where the oscillation of f is large) is arbitrarily small. For example, the discontinuous function in Example 1.6 has zero oscillation on every interval except the first one, where the function has oscillation one, but the length of that interval can be made as small as we wish.

Thus, roughly speaking, a function is Riemann integrable if it oscillates by an arbitrary small amount except on a finite collection of intervals whose total length is arbitrarily small. Theorem 1.87 gives a precise statement.

One direct consequence of the Cauchy criterion is that a function is integrable if we can estimate its oscillation by the oscillation of an integrable function.

Proposition 1.16. Suppose that $f, g : [a,b] \to \mathbb{R}$ are bounded functions and g is integrable on $[a,b]$. If there exists a constant $C \ge 0$ such that

$$\operatorname*{osc}_I f \le C \operatorname*{osc}_I g$$

on every interval $I \subset [a,b]$, then f is integrable.

Proof. If $P = \{I_1, I_2, \dots, I_n\}$ is a partition of $[a, b]$, then

$$\begin{aligned} U(f; P) - L(f; P) &= \sum_{k=1}^{n} \left[\sup_{I_k} f - \inf_{I_k} f \right] \cdot |I_k| \\ &= \sum_{k=1}^{n} \operatorname*{osc}_{I_k} f \cdot |I_k| \\ &\leq C \sum_{k=1}^{n} \operatorname*{osc}_{I_k} g \cdot |I_k| \\ &\leq C \left[U(g; P) - L(g; P) \right]. \end{aligned}$$

Thus, f satisfies the Cauchy criterion in Theorem 1.14 if g does, which proves that f is integrable if g is integrable. □

We can also give a sequential characterization of integrability.

Theorem 1.17. A bounded function $f : [a, b] \to \mathbb{R}$ is Riemann integrable if and only if there is a sequence (P_n) of partitions such that

$$\lim_{n\to\infty} [U(f; P_n) - L(f; P_n)] = 0.$$

In that case,

$$\int_a^b f = \lim_{n\to\infty} U(f; P_n) = \lim_{n\to\infty} L(f; P_n).$$

Proof. First, suppose that the condition holds. Then, given $\epsilon > 0$, there is an $n \in \mathbb{N}$ such that $U(f; P_n) - L(f; P_n) < \epsilon$, so Theorem 1.14 implies that f is integrable and $U(f) = L(f)$.

Furthermore, since $U(f) \leq U(f; P_n)$ and $L(f; P_n) \leq L(f)$, we have

$$0 \leq U(f; P_n) - U(f) = U(f; P_n) - L(f) \leq U(f; P_n) - L(f; P_n).$$

Since the limit of the right-hand side is zero, the 'squeeze' theorem implies that

$$\lim_{n\to\infty} U(f; P_n) = U(f) = \int_a^b f$$

It also follows that

$$\lim_{n\to\infty} L(f; P_n) = \lim_{n\to\infty} U(f; P_n) - \lim_{n\to\infty} [U(f; P_n) - L(f; P_n)] = \int_a^b f.$$

Conversely, if f is integrable then, by Theorem 1.14, for every $n \in \mathbb{N}$ there exists a partition P_n such that

$$0 \leq U(f; P_n) - L(f; P_n) < \frac{1}{n},$$

and $U(f; P_n) - L(f; P_n) \to 0$ as $n \to \infty$. □

Note that if the limits of $U(f; P_n)$ and $L(f; P_n)$ both exist and are equal, then

$$\lim_{n\to\infty} [U(f; P_n) - L(f; P_n)] = \lim_{n\to\infty} U(f; P_n) - \lim_{n\to\infty} L(f; P_n),$$

so the conditions of the theorem are satisfied. Conversely, the proof of the theorem shows that if the limit of $U(f; P_n) - L(f; P_n)$ is zero, then the limits of $U(f; P_n)$

and $L(f; P_n)$ both exist and are equal. This isn't true for general sequences, where one may have $\lim(a_n - b_n) = 0$ even though $\lim a_n$ and $\lim b_n$ don't exist.

Theorem 1.17 provides one way to prove the existence of an integral and, in some cases, evaluate it.

Example 1.18. Let P_n be the partition of $[0, 1]$ into n-intervals of equal length $1/n$ with endpoints $x_k = k/n$ for $k = 0, 1, 2, \dots, n$. If $I_k = [(k-1)/n, k/n]$ is the kth interval, then

$$\sup_{I_k} f = x_k^2, \qquad \inf_{I_k} = x_{k-1}^2$$

since f is increasing. Using the formula for the sum of squares

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

we get

$$U(f; P_n) = \sum_{k=1}^{n} x_k^2 \cdot \frac{1}{n} = \frac{1}{n^3} \sum_{k=1}^{n} k^2 = \frac{1}{6}\left(1 + \frac{1}{n}\right)\left(2 + \frac{1}{n}\right)$$

and

$$L(f; P_n) = \sum_{k=1}^{n} x_{k-1}^2 \cdot \frac{1}{n} = \frac{1}{n^3} \sum_{k=1}^{n-1} k^2 = \frac{1}{6}\left(1 - \frac{1}{n}\right)\left(2 - \frac{1}{n}\right).$$

(See Figure 1.18.) It follows that

$$\lim_{n\to\infty} U(f; P_n) = \lim_{n\to\infty} L(f; P_n) = \frac{1}{3},$$

and Theorem 1.17 implies that x^2 is integrable on $[0, 1]$ with

$$\int_0^1 x^2\, dx = \frac{1}{3}.$$

The fundamental theorem of calculus, Theorem 1.45 below, provides a much easier way to evaluate this integral.

1.5. Integrability of continuous and monotonic functions

The Cauchy criterion leads to the following fundamental result that every continuous function is Riemann integrable. To prove this, we use the fact that a continuous function oscillates by an arbitrarily small amount on every interval of a sufficiently refined partition.

Theorem 1.19. A continuous function $f : [a, b] \to \mathbb{R}$ on a compact interval is Riemann integrable.

Proof. A continuous function on a compact set is bounded, so we just need to verify the Cauchy condition in Theorem 1.14.

Let $\epsilon > 0$. A continuous function on a compact set is uniformly continuous, so there exists $\delta > 0$ such that

$$|f(x) - f(y)| < \frac{\epsilon}{b-a} \qquad \text{for all } x, y \in [a, b] \text{ such that } |x - y| < \delta.$$

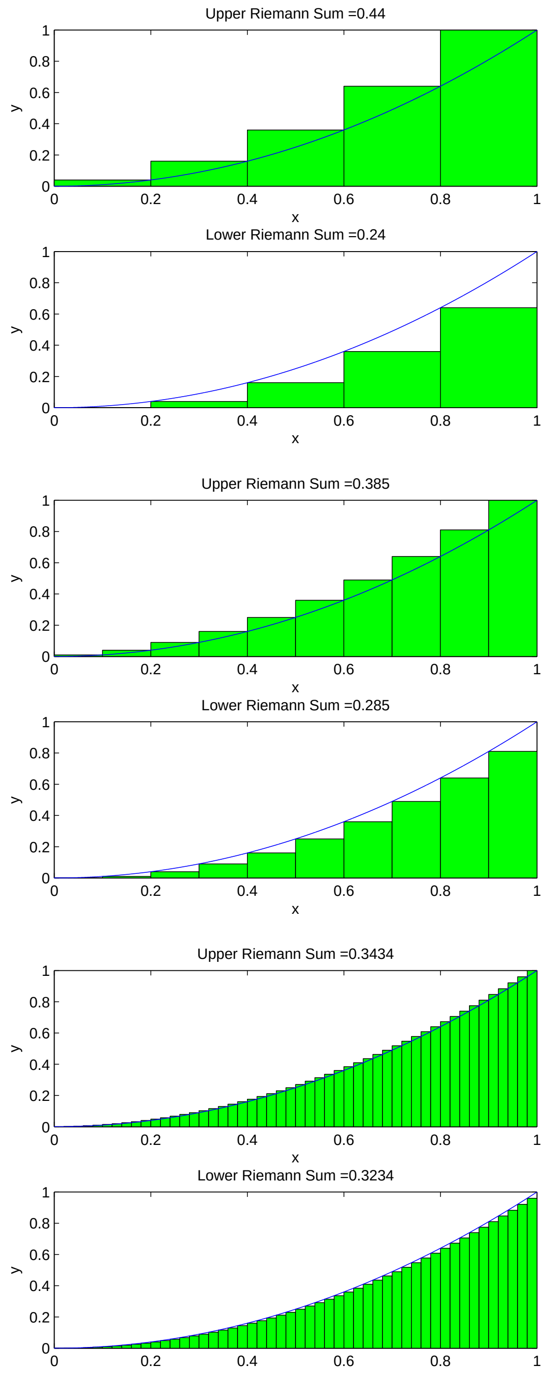

Figure 1. Upper and lower Riemann sums for Example 1.18 with $n = 5, 10, 50$ subintervals of equal length.

Choose a partition $P = \{I_1, I_2, \dots, I_n\}$ of $[a, b]$ such that $|I_k| < \delta$ for every k; for example, we can take n intervals of equal length $(b-a)/n$ with $n > (b-a)/\delta$.

Since f is continuous, it attains its maximum and minimum values M_k and m_k on the compact interval I_k at points x_k and y_k in I_k. These points satisfy $|x_k - y_k| < \delta$, so

$$M_k - m_k = f(x_k) - f(y_k) < \frac{\epsilon}{b-a}.$$

The upper and lower sums of f therefore satisfy

$$\begin{aligned} U(f; P) - L(f; P) &= \sum_{k=1}^{n} M_k |I_k| - \sum_{k=1}^{n} m_k |I_k| \\ &= \sum_{k=1}^{n} (M_k - m_k)|I_k| \\ &< \frac{\epsilon}{b-a} \sum_{k=1}^{n} |I_k| \\ &< \epsilon, \end{aligned}$$

and Theorem 1.14 implies that f is integrable. □

Example 1.20. The function $f(x) = x^2$ on $[0, 1]$ considered in Example 1.18 is integrable since it is continuous.

Another class of integrable functions consists of monotonic (increasing or decreasing) functions.

Theorem 1.21. A monotonic function $f : [a, b] \to \mathbb{R}$ on a compact interval is Riemann integrable.

Proof. Suppose that f is monotonic increasing, meaning that $f(x) \leq f(y)$ for $x \leq y$. Let $P_n = \{I_1, I_2, \dots, I_n\}$ be a partition of $[a, b]$ into n intervals $I_k = [x_{k-1}, x_k]$, of equal length $(b-a)/n$, with endpoints

$$x_k = a + (b-a)\frac{k}{n}, \qquad k = 0, 1, \dots, n-1, n.$$

Since f is increasing,

$$M_k = \sup_{I_k} f = f(x_k), \qquad m_k = \inf_{I_k} f = f(x_{k-1}).$$

Hence, summing a telescoping series, we get

$$\begin{aligned} U(f; P_n) - L(U; P_n) &= \sum_{k=1}^{n} (M_k - m_k)(x_k - x_{k-1}) \\ &= \frac{b-a}{n} \sum_{k=1}^{n} [f(x_k) - f(x_{k-1})] \\ &= \frac{b-a}{n} [f(b) - f(a)]. \end{aligned}$$

It follows that $U(f; P_n) - L(U; P_n) \to 0$ as $n \to \infty$, and Theorem 1.17 implies that f is integrable.

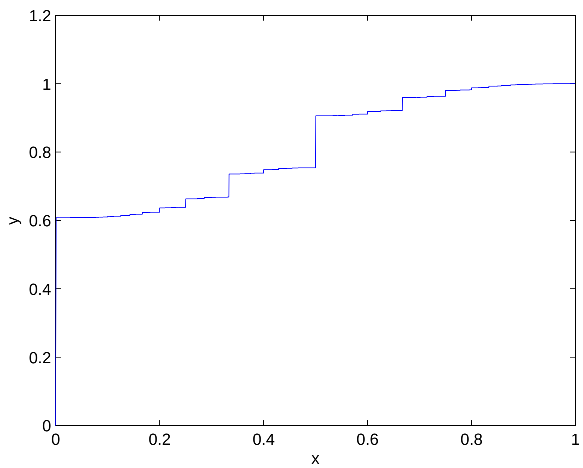

Figure 2. The graph of the monotonic function in Example 1.22 with a countably infinite, dense set of jump discontinuities.

The proof for a monotonic decreasing function f is similar, with

$$\sup_{I_k} f = f(x_{k-1}), \qquad \inf_{I_k} f = f(x_k),$$

or we can apply the result for increasing functions to $-f$ and use Theorem 1.23 below. $\square$

Monotonic functions needn't be continuous, and they may be discontinuous at a countably infinite number of points.

Example 1.22. Let $\{q_k : k \in \mathbb{N}\}$ be an enumeration of the rational numbers in $[0, 1)$ and let (a_k) be a sequence of strictly positive real numbers such that

$$\sum_{k=1}^{\infty} a_k = 1.$$

Define $f : [0, 1] \to \mathbb{R}$ by

$$f(x) = \sum_{k \in Q(x)} a_k, \qquad Q(x) = \{k \in \mathbb{N} : q_k \in [0, x)\}.$$

for $x > 0$, and $f(0) = 0$. That is, $f(x)$ is obtained by summing the terms in the series whose indices k correspond to the rational numbers $0 \leq q_k < x$.

For $x = 1$, this sum includes all the terms in the series, so $f(1) = 1$. For every $0 < x < 1$, there are infinitely many terms in the sum, since the rationals are dense in $[0, x)$, and f is increasing, since the number of terms increases with x. By Theorem 1.21, f is Riemann integrable on $[0, 1]$. Although f is integrable, it has a countably infinite number of jump discontinuities at every rational number in $[0, 1)$, which are dense in $[0, 1]$, The function is continuous elsewhere (the proof is left as an exercise).

Figure 2 shows the graph of f corresponding to the enumeration

$$\{0, 1/2, 1/3, 2/3, 1/4, 3/4, 1/5, 2/5, 3/5, 4/5, 1/6, 5/6, 1/7, \dots\}$$

of the rational numbers in $[0, 1)$ and

$$a_k = \frac{6}{\pi^2 k^2}.$$

1.6. Properties of the Riemann integral

The integral has the following three basic properties.

(1) Linearity:

$$\int_a^b cf = c\int_a^b f, \qquad \int_a^b (f+g) = \int_a^b f + \int_a^b g.$$

(2) Monotonicity: if $f \le g$, then

$$\int_a^b f \le \int_a^b g.$$

(3) Additivity: if $a < c < b$, then

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

In this section, we prove these properties and derive a few of their consequences.

These properties are analogous to the corresponding properties of sums (or convergent series):

$$\sum_{k=1}^n ca_k = c\sum_{k=1}^n a_k, \qquad \sum_{k=1}^n (a_k + b_k) = \sum_{k=1}^n a_k + \sum_{k=1}^n b_k;$$

$$\sum_{k=1}^n a_k \le \sum_{k=1}^n b_k \qquad \text{if } a_k \le b_k;$$

$$\sum_{k=1}^m a_k + \sum_{k=m+1}^n a_k = \sum_{k=1}^n a_k.$$

1.6.1. Linearity. We begin by proving the linearity. First we prove linearity with respect to scalar multiplication and then linearity with respect to sums.

Theorem 1.23. If $f : [a, b] \to \mathbb{R}$ is integrable and $c \in \mathbb{R}$, then cf is integrable and

$$\int_a^b cf = c\int_a^b f.$$

Proof. Suppose that $c \ge 0$. Then for any set $A \subset [a, b]$, we have

$$\sup_A cf = c\sup_A f, \qquad \inf_A cf = c\inf_A f,$$

so $U(cf; P) = cU(f; P)$ for every partition P. Taking the infimum over the set Π of all partitions of $[a, b]$, we get

$$U(cf) = \inf_{P\in\Pi} U(cf; P) = \inf_{P\in\Pi} cU(f; P) = c\inf_{P\in\Pi} U(f; P) = cU(f).$$

Similarly, $L(cf;P) = cL(f;P)$ and $L(cf) = cL(f)$. If f is integrable, then

$$U(cf) = cU(f) = cL(f) = L(cf),$$

which shows that cf is integrable and

$$\int_a^b cf = c\int_a^b f.$$

Now consider $-f$. Since

$$\sup_A(-f) = -\inf_A f, \qquad \inf_A(-f) = -\sup_A f,$$

we have

$$U(-f;P) = -L(f;P), \qquad L(-f;P) = -U(f;P).$$

Therefore

$$\begin{aligned} U(-f) &= \inf_{P\in\Pi} U(-f;P) = \inf_{P\in\Pi}\left[-L(f;P)\right] = -\sup_{P\in\Pi} L(f;P) = -L(f),\\ L(-f) &= \sup_{P\in\Pi} L(-f;P) = \sup_{P\in\Pi}\left[-U(f;P)\right] = -\inf_{P\in\Pi} U(f;P) = -U(f). \end{aligned}$$

Hence, $-f$ is integrable if f is integrable and

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

Finally, if $c < 0$, then $c = -|c|$, and a successive application of the previous results shows that cf is integrable with $\int_a^b cf = c\int_a^b f$. $\square$

Next, we prove the linearity of the integral with respect to sums. If f, g are bounded, then $f + g$ is bounded and

$$\sup_I(f+g) \le \sup_I f + \sup_I g, \qquad \inf_I(f+g) \ge \inf_I f + \inf_I g.$$

It follows that

$$\operatorname*{osc}_I(f+g) \le \operatorname*{osc}_I f + \operatorname*{osc}_I g,$$

so $f+g$ is integrable if f, g are integrable. In general, however, the upper (or lower) sum of $f+g$ needn't be the sum of the corresponding upper (or lower) sums of f and g. As a result, we don't get

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

simply by adding upper and lower sums. Instead, we prove this equality by estimating the upper and lower integrals of $f+g$ from above and below by those of f and g.

Theorem 1.24. If $f, g : [a,b] \to \mathbb{R}$ are integrable functions, then $f+g$ is integrable, and

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

Proof. We first prove that if $f, g : [a, b] \to \mathbb{R}$ are bounded, but not necessarily integrable, then

$$U(f+g) \le U(f) + U(g), \qquad L(f+g) \ge L(f) + L(g).$$

Suppose that $P = \{I_1, I_2, \dots, I_n\}$ is a partition of $[a, b]$. Then

$$\begin{aligned} U(f+g; P) &= \sum_{k=1}^{n} \sup_{I_k}(f+g) \cdot |I_k| \\ &\le \sum_{k=1}^{n} \sup_{I_k} f \cdot |I_k| + \sum_{k=1}^{n} \sup_{I_k} g \cdot |I_k| \\ &\le U(f; P) + U(g; P). \end{aligned}$$

Let $\epsilon > 0$. Since the upper integral is the infimum of the upper sums, there are partitions Q, R such that

$$U(f; Q) < U(f) + \frac{\epsilon}{2}, \qquad U(g; R) < U(g) + \frac{\epsilon}{2},$$

and if P is a common refinement of Q and R, then

$$U(f; P) < U(f) + \frac{\epsilon}{2}, \qquad U(g; P) < U(g) + \frac{\epsilon}{2}.$$

It follows that

$$U(f+g) \le U(f+g; P) \le U(f; P) + U(g; P) < U(f) + U(g) + \epsilon.$$

Since this inequality holds for arbitrary $\epsilon > 0$, we must have $U(f+g) \le U(f) + U(g)$.

Similarly, we have $L(f+g; P) \ge L(f; P) + L(g; P)$ for all partitions P, and for every $\epsilon > 0$, we get $L(f+g) > L(f) + L(g) - \epsilon$, so $L(f+g) \ge L(f) + L(g)$.

For integrable functions f and g, it follows that

$$U(f+g) \le U(f) + U(g) = L(f) + L(g) \le L(f+g).$$

Since $U(f+g) \ge L(f+g)$, we have $U(f+g) = L(f+g)$ and $f+g$ is integrable. Moreover, there is equality throughout the previous inequality, which proves the result. □

Although the integral is linear, the upper and lower integrals of non-integrable functions are not, in general, linear.

Example 1.25. Define $f, g : [0, 1] \to \mathbb{R}$ by

$$f(x) = \begin{cases} 1 & \text{if } x \in [0,1] \cap \mathbb{Q}, \\ 0 & \text{if } x \in [0,1] \setminus \mathbb{Q}, \end{cases} \qquad g(x) = \begin{cases} 0 & \text{if } x \in [0,1] \cap \mathbb{Q}, \\ 1 & \text{if } x \in [0,1] \setminus \mathbb{Q}. \end{cases}$$

That is, f is the Dirichlet function and $g = 1 - f$. Then

$$U(f) = U(g) = 1, \qquad L(f) = L(g) = 0, \qquad U(f+g) = L(f+g) = 1,$$

so

$$U(f+g) < U(f) + U(g), \qquad L(f+g) > L(f) + L(g).$$

The product of integrable functions is also integrable, as is the quotient provided it remains bounded. Unlike the integral of the sum, however, there is no way to express the integral of the product $\int fg$ in terms of $\int f$ and $\int g$.

Theorem 1.26. If $f, g : [a, b] \to \mathbb{R}$ are integrable, then $fg : [a, b] \to \mathbb{R}$ is integrable. If, in addition, $g \neq 0$ and $1/g$ is bounded, then $f/g : [a, b] \to \mathbb{R}$ is integrable.

Proof. First, we show that the square of an integrable function is integrable. If f is integrable, then f is bounded, with $|f| \leq M$ for some $M \geq 0$. For all $x, y \in [a, b]$, we have

$$\left|f^2(x) - f^2(y)\right| = |f(x) + f(y)| \cdot |f(x) - f(y)| \leq 2M|f(x) - f(y)|.$$

Taking the supremum of this inequality over $x, y \in I \subset [a, b]$ and using Proposition 2.19, we get that

$$\sup_I (f^2) - \inf_I (f^2) \leq 2M \left[\sup_I f - \inf_I f\right].$$

meaning that

$$\operatorname*{osc}_I (f^2) \leq 2M \operatorname*{osc}_I f.$$

If follows from Proposition 1.16 that f^2 is integrable if f is integrable.

Since the integral is linear, we then see from the identity

$$fg = \frac{1}{4}\left[(f + g)^2 - (f - g)^2\right]$$

that fg is integrable if f, g are integrable.

In a similar way, if $g \neq 0$ and $|1/g| \leq M$, then

$$\left|\frac{1}{g(x)} - \frac{1}{g(y)}\right| = \frac{|g(x) - g(y)|}{|g(x)g(y)|} \leq M^2 \left|g(x) - g(y)\right|.$$

Taking the supremum of this equation over $x, y \in I \subset [a, b]$, we get

$$\sup_I \left(\frac{1}{g}\right) - \inf_I \left(\frac{1}{g}\right) \leq M^2 \left[\sup_I g - \inf_I g\right],$$

meaning that $\operatorname{osc}_I(1/g) \leq M^2 \operatorname{osc}_I g$, and Proposition 1.16 implies that $1/g$ is integrable if g is integrable. Therefore $f/g = f \cdot (1/g)$ is integrable. □

1.6.2. Monotonicity.

Next, we prove the monotonicity of the integral.

Theorem 1.27. Suppose that $f, g : [a, b] \to \mathbb{R}$ are integrable and $f \leq g$. Then

$$\int_a^b f \leq \int_a^b g.$$

Proof. First suppose that $f \geq 0$ is integrable. Let P be the partition consisting of the single interval $[a, b]$. Then

$$L(f; P) = \inf_{[a,b]} f \cdot (b - a) \geq 0,$$

so

$$\int_a^b f \geq L(f; P) \geq 0.$$

If $f \geq g$, then $h = f - g \geq 0$, and the linearity of the integral implies that

$$\int_a^b f - \int_a^b g = \int_a^b h \geq 0,$$

which proves the theorem. □

One immediate consequence of this theorem is the following simple, but useful, estimate for integrals.

Theorem 1.28. Suppose that $f : [a, b] \to \mathbb{R}$ is integrable and

$$M = \sup_{[a,b]} f, \qquad m = \inf_{[a,b]} f.$$

Then

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

Proof. Since $m \leq f \leq M$ on $[a, b]$, Theorem 1.27 implies that

$$\int_a^b m \leq \int_a^b f \leq \int_a^b M,$$

which gives the result. □

This estimate also follows from the definition of the integral in terms of upper and lower sums, but once we've established the monotonicity of the integral, we don't need to go back to the definition.

A further consequence is the intermediate value theorem for integrals, which states that a continuous function on an interval is equal to its average value at some point.

Theorem 1.29. If $f : [a, b] \to \mathbb{R}$ is continuous, then there exists $x \in [a, b]$ such that

$$f(x) = \frac{1}{b-a} \int_a^b f.$$

Proof. Since f is a continuous function on a compact interval, it attains its maximum value M and its minimum value m. From Theorem 1.28,

$$m \leq \frac{1}{b-a} \int_a^b f \leq M.$$

By the intermediate value theorem, f takes on every value between m and M, and the result follows. □

As shown in the proof of Theorem 1.27, given linearity, monotonicity is equivalent to positivity,

$$\int_a^b f \geq 0 \qquad \text{if } f \geq 0.$$

We remark that even though the upper and lower integrals aren't linear, they are monotone.

Proposition 1.30. If $f, g : [a, b] \to \mathbb{R}$ are bounded functions and $f \leq g$, then

$$U(f) \leq U(g), \qquad L(f) \leq L(g).$$

Proof. From Proposition 2.12, we have for every interval $I \subset [a,b]$ that

$$\sup_I f \le \sup_I g, \qquad \inf_I f \le \inf_I g.$$

It follows that for every partition P of $[a,b]$, we have

$$U(f;P) \le U(g;P), \qquad L(f;P) \le L(g;P).$$

Taking the infimum of the upper inequality and the supremum of the lower inequality over P, we get $U(f) \le U(g)$ and $L(f) \le L(g)$. □

We can estimate the absolute value of an integral by taking the absolute value under the integral sign. This is analogous to the corresponding property of sums:

$$\left| \sum_{k=1}^{n} a_n \right| \le \sum_{k=1}^{n} |a_k|.$$

Theorem 1.31. If f is integrable, then $|f|$ is integrable and

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

Proof. First, suppose that $|f|$ is integrable. Since

$$-|f| \le f \le |f|,$$

we get from Theorem 1.27 that

$$-\int_a^b |f| \le \int_a^b f \le \int_a^b |f|, \qquad \text{or} \qquad \left| \int_a^b f \right| \le \int_a^b |f|.$$

To complete the proof, we need to show that $|f|$ is integrable if f is integrable. For $x, y \in [a,b]$, the reverse triangle inequality gives

$$\big|\, |f(x)| - |f(y)| \,\big| \le |f(x) - f(y)|.$$

Using Proposition 2.19, we get that

$$\sup_I |f| - \inf_I |f| \le \sup_I f - \inf_I f,$$

meaning that $\operatorname{osc}_I |f| \le \operatorname{osc}_I f$. Proposition 1.16 then implies that $|f|$ is integrable if f is integrable. □

In particular, we immediately get the following basic estimate for an integral.

Corollary 1.32. If $f : [a,b] \to \mathbb{R}$ is integrable

$$M = \sup_{[a,b]} |f|,$$

then

$$\left| \int_a^b f \right| \le M(b-a).$$

1.6.3. Additivity. Finally, we prove additivity. This property refers to additivity with respect to the interval of integration, rather than linearity with respect to the function being integrated.

Theorem 1.33. Suppose that $f : [a,b] \to \mathbb{R}$ and $a < c < b$. Then f is Riemann integrable on $[a,b]$ if and only if it is Riemann integrable on $[a,c]$ and $[c,b]$. In that case,

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

Proof. Suppose that f is integrable on $[a,b]$. Then, given $\epsilon > 0$, there is a partition P of $[a,b]$ such that $U(f;P) - L(f;P) < \epsilon$. Let $P' = P \cup \{c\}$ be the refinement of P obtained by adding c to the endpoints of P. (If $c \in P$, then $P' = P$.) Then $P' = Q \cup R$ where $Q = P' \cap [a,c]$ and $R = P' \cap [c,b]$ are partitions of $[a,c]$ and $[c,b]$ respectively. Moreover,

$$U(f;P') = U(f;Q) + U(f;R), \qquad L(f;P') = L(f;Q) + L(f;R).$$

It follows that

$$\begin{aligned} U(f;Q) - L(f;Q) &= U(f;P') - L(f;P') - [U(f;R) - L(f;R)] \\ &\le U(f;P) - L(f;P) < \epsilon, \end{aligned}$$

which proves that f is integrable on $[a,c]$. Exchanging Q and R, we get the proof for $[c,b]$.

Conversely, if f is integrable on $[a,c]$ and $[c,b]$, then there are partitions Q of $[a,c]$ and R of $[c,b]$ such that

$$U(f;Q) - L(f;Q) < \frac{\epsilon}{2}, \qquad U(f;R) - L(f;R) < \frac{\epsilon}{2}.$$

Let $P = Q \cup R$. Then

$$U(f;P) - L(f;P) = U(f;Q) - L(f;Q) + U(f;R) - L(f;R) < \epsilon,$$

which proves that f is integrable on $[a,b]$.

Finally, with the partitions P, Q, R as above, we have

$$\begin{aligned} \int_a^b f &\le U(f;P) = U(f;Q) + U(f;R) \\ &< L(f;Q) + L(f;R) + \epsilon \\ &< \int_a^c f + \int_c^b f + \epsilon. \end{aligned}$$

Similarly,

$$\begin{aligned} \int_a^b f &\ge L(f;P) = L(f;Q) + L(f;R) \\ &> U(f;Q) + U(f;R) - \epsilon \\ &> \int_a^c f + \int_c^b f - \epsilon. \end{aligned}$$

Since $\epsilon > 0$ is arbitrary, we see that $\int_a^b f = \int_a^c f + \int_c^b f$. □

We can extend the additivity property of the integral by defining an oriented Riemann integral.

Definition 1.34. If $f : [a, b] \to \mathbb{R}$ is integrable, where $a < b$, and $a \leq c \leq b$, then

$$\int_b^a f = -\int_a^b f, \qquad \int_c^c f = 0.$$

With this definition, the additivity property in Theorem 1.33 holds for all $a, b, c \in \mathbb{R}$ for which the oriented integrals exist. Moreover, if $|f| \leq M$, then the estimate in Corollary 1.32 becomes

$$\left| \int_a^b f \right| \leq M|b - a|$$

for all $a, b \in \mathbb{R}$ (even if $a \geq b$).

The oriented Riemann integral is a special case of the integral of a differential form. It assigns a value to the integral of a one-form $f\,dx$ on an oriented interval.

1.7. Further existence results for the Riemann integral

In this section, we prove several further useful conditions for the existences of the Riemann integral.

First, we show that changing the values of a function at finitely many points doesn't change its integrability of the value of its integral.

Proposition 1.35. Suppose that $f, g : [a, b] \to \mathbb{R}$ and $f(x) = g(x)$ except at finitely many points $x \in [a, b]$. Then f is integrable if and only if g is integrable, and in that case

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

Proof. It is sufficient to prove the result for functions whose values differ at a single point, say $c \in [a, b]$. The general result then follows by induction.

Since f, g differ at a single point, f is bounded if and only if g is bounded. If f, g are unbounded, then neither one is integrable. If f, g are bounded, we will show that f, g have the same upper and lower integrals because their upper and lower sums differ by an arbitrarily small amount with respect to a partition that is sufficiently refined near the point where the functions differ.

Suppose that f, g are bounded with $|f|, |g| \leq M$ on $[a, b]$ for some $M > 0$. Let $\epsilon > 0$. Choose a partition P of $[a, b]$ such that

$$U(f; P) < U(f) + \frac{\epsilon}{2}.$$

Let $Q = \{I_1, \dots, I_n\}$ be a refinement of P such that $|I_k| < \delta$ for $k = 1, \dots, n$, where

$$\delta = \frac{\epsilon}{8M}.$$

Then g differs from f on at most two intervals in Q. (There could be two intervals if c is an endpoint of the partition.) On such an interval I_k we have

$$\left|\sup_{I_k} g - \sup_{I_k} f\right| \leq \sup_{I_k} |g| + \sup_{I_k} |f| \leq 2M,$$

and on the remaining intervals, $\sup_{I_k} g - \sup_{I_k} f = 0$. It follows that

$$|U(g;Q) - U(f;Q)| < 2M \cdot 2\delta < \frac{\epsilon}{2}.$$

Using the properties of upper integrals and refinements, we obtain

$$U(g) \leq U(g;Q) < U(f;Q) + \frac{\epsilon}{2} \leq U(f;P) + \frac{\epsilon}{2} < U(f) + \epsilon.$$

Since this inequality holds for arbitrary $\epsilon > 0$, we get that $U(g) \leq U(f)$. Exchanging f and g, we see similarly that $U(f) \leq U(g)$, so $U(f) = U(g)$.

An analogous argument for lower sums (or an application of the result for upper sums to $-f$, $-g$) shows that $L(f) = L(g)$. Thus $U(f) = L(f)$ if and only if $U(g) = L(g)$, in which case $\int_a^b f = \int_a^b g$. □

Example 1.36. The function f in Example 1.6 differs from the 0-function at one point. It is integrable and its integral is equal to 0.

The conclusion of Proposition 1.35 can fail if the functions differ at a countably infinite number of points. One reason is that we can turn a bounded function into an unbounded function by changing its values at an infinite number of points.

Example 1.37. Define $f : [0,1] \to \mathbb{R}$ by

$$f(x) = \begin{cases} n & \text{if } x = 1/n \text{ for } n \in \mathbb{N}, \\ 0 & \text{otherwise} \end{cases}$$

Then f is equal to the 0-function except on the countably infinite set $\{1/n : n \in \mathbb{N}\}$, but f is unbounded and therefore it's not Riemann integrable.

The result is still false, however, for bounded functions that differ at a countably infinite number of points.

Example 1.38. The Dirichlet function in Example 1.7 is bounded and differs from the 0-function on the countably infinite set of rationals, but it isn't Riemann integrable.

The Lebesgue integral is better behaved than the Riemann intgeral in this respect: two functions that are equal almost everywhere, meaning that they differ on a set of Lebesgue measure zero, have the same Lebesgue integrals. In particular, two functions that differ on a countable set are equal almost everywhere (see Section 1.12).

The next proposition allows us to deduce the integrability of a bounded function on an interval from its integrability on slightly smaller intervals.

Proposition 1.39. Suppose that $f : [a, b] \to \mathbb{R}$ is bounded and integrable on $[a, r]$ for every $a < r < b$. Then f is integrable on $[a, b]$ and

$$\int_a^b f = \lim_{r \to b^-} \int_a^r f.$$

Proof. Since f is bounded, $|f| \le M$ on $[a, b]$ for some $M > 0$. Given $\epsilon > 0$, let

$$r = b - \frac{\epsilon}{4M}$$

(where we assume ϵ is sufficiently small that $r > a$). Since f is integrable on $[a, r]$, there is a partition Q of $[a, r]$ such that

$$U(f; Q) - L(f; Q) < \frac{\epsilon}{2}.$$

Then $P = Q \cup \{b\}$ is a partition of $[a, b]$ whose last interval is $[r, b]$. The boundedness of f implies that

$$\sup_{[r,b]} f - \inf_{[r,b]} f \le 2M.$$

Therefore

$$\begin{aligned} U(f; P) - L(f; P) &= U(f; Q) - L(f; Q) + \left(\sup_{[r,b]} f - \inf_{[r,b]} f \right) \cdot (b - r) \\ &< \frac{\epsilon}{2} + 2M \cdot (b - r) = \epsilon, \end{aligned}$$

so f is integrable on $[a, b]$ by Theorem 1.14. Moreover, using the additivity of the integral, we get

$$\left| \int_a^b f - \int_a^r f \right| = \left| \int_r^b f \right| \le M \cdot (b - r) \to 0 \qquad \text{as } r \to b^-.$$

□

An obvious analogous result holds for the left endpoint.

Example 1.40. Define $f : [0, 1] \to \mathbb{R}$ by

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

Then f is bounded on $[0, 1]$. Furthemore, f is continuous and therefore integrable on $[r, 1]$ for every $0 < r < 1$. It follows from Proposition 1.39 that f is integrable on $[0, 1]$.

The assumption in Proposition 1.39 that f is bounded on $[a, b]$ is essential.

Example 1.41. The function $f : [0, 1] \to \mathbb{R}$ defined by

$$f(x) = \begin{cases} 1/x & \text{for } 0 < x \le 1, \\ 0 & \text{for } x = 0, \end{cases}$$

is continuous and therefore integrable on $[r, 1]$ for every $0 < r < 1$, but it's unbounded and therefore not integrable on $[0, 1]$.

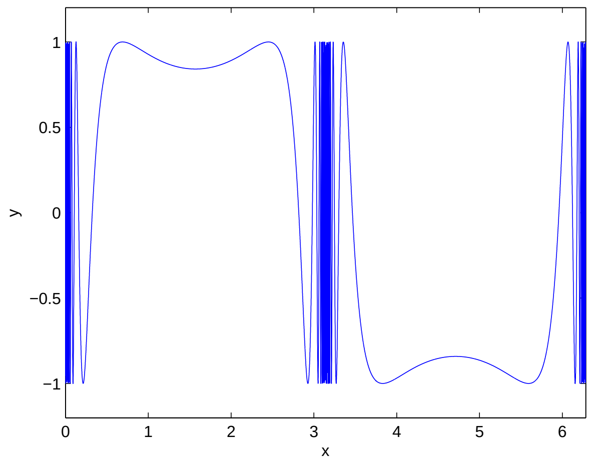

Figure 3. Graph of the Riemann integrable function $y = \sin(1/\sin x)$ in Example 1.43.

As a corollary of this result and the additivity of the integral, we prove a generalization of the integrability of continuous functions to piecewise continuous functions.

Theorem 1.42. If $f : [a, b] \to \mathbb{R}$ is a bounded function with finitely many discontinuities, then f is Riemann integrable.

Proof. By splitting the interval into subintervals with the discontinuities of f at an endpoint and using Theorem 1.33, we see that it is sufficient to prove the result if f is discontinuous only at one endpoint of $[a, b]$, say at b. In that case, f is continuous and therefore integrable on any smaller interval $[a, r]$ with $a < r < b$, and Proposition 1.39 implies that f is integrable on $[a, b]$. □

Example 1.43. Define $f : [0, 2\pi] \to \mathbb{R}$ by

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

Then f is bounded and continuous except at $x = 0, \pi, 2\pi$, so it is integrable on $[0, 2\pi]$ (see Figure 3). This function doesn't have jump discontinuities, but Theorem 1.42 still applies.

Example 1.44. Define $f : [0, 1/\pi] \to \mathbb{R}$ by

$$f(x) = \begin{cases} \operatorname{sgn}[\sin(1/x)] & \text{if } x \neq 1/n\pi \text{ for } n \in \mathbb{N}, \\ 0 & \text{if } x = 0 \text{ or } x \neq 1/n\pi \text{ for } n \in \mathbb{N}, \end{cases}$$

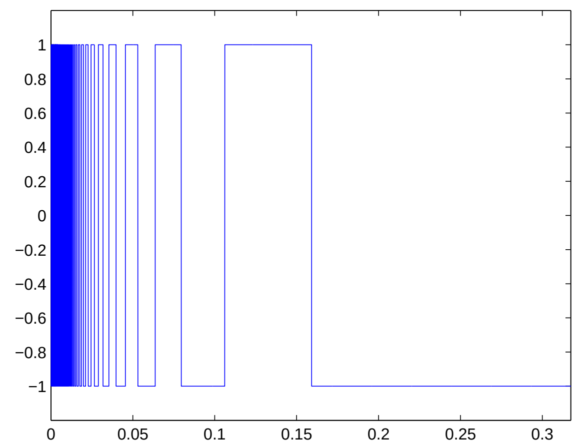

Figure 4. Graph of the Riemann integrable function $y = \operatorname{sgn}(\sin(1/x))$ in Example 1.44.

where sgn is the sign function,

$$\operatorname{sgn} x = \begin{cases} 1 & \text{if } x > 0, \\ 0 & \text{if } x = 0, \\ -1 & \text{if } x < 0. \end{cases}$$

Then f oscillates between 1 and -1 a countably infinite number of times as $x \to 0^+$ (see Figure 4). It has jump discontinuities at $x = 1/(n\pi)$ and an essential discontinuity at $x = 0$. Nevertheless, it is Riemann integrable. To see this, note that f is bounded on $[0, 1]$ and piecewise continuous with finitely many discontinuities on $[r, 1]$ for every $0 < r < 1$. Theorem 1.42 implies that f is Riemann integrable on $[r, 1]$, and then Theorem 1.39 implies that f is integrable on $[0, 1]$.

1.8. The fundamental theorem of calculus

> In the integral calculus I find much less interesting the parts that involve only substitutions, transformations, and the like, in short, the parts that involve the known skillfully applied mechanics of reducing integrals to algebraic, logarithmic, and circular functions, than I find the careful and profound study of transcendental functions that cannot be reduced to these functions. (Gauss, 1808)

The fundamental theorem of calculus states that differentiation and integration are inverse operations in an appropriately understood sense. The theorem has two parts: in one direction, it says roughly that the integral of the derivative is the original function; in the other direction, it says that the derivative of the integral is the original function.

In more detail, the first part states that if $F : [a,b] \to \mathbb{R}$ is differentiable with integrable derivative, then

$$\int_a^b F'(x)\,dx = F(b) - F(a).$$

This result can be thought of as a continuous analog of the corresponding identity for sums of differences,

$$\sum_{k=1}^{n} (A_k - A_{k-1}) = A_n - A_0.$$

The second part states that if $f : [a,b] \to \mathbb{R}$ is continuous, then

$$\frac{d}{dx} \int_a^x f(t)\,dt = f(x).$$

This is a continuous analog of the corresponding identity for differences of sums,

$$\sum_{j=1}^{k} a_j - \sum_{j=1}^{k-1} a_j = a_k.$$

The proof of the fundamental theorem consists essentially of applying the identities for sums or differences to the appropriate Riemann sums or difference quotients and proving, under appropriate hypotheses, that they converge to the corresponding integrals or derivatives.

We'll split the statement and proof of the fundamental theorem into two parts. (The numbering of the parts as I and II is arbitrary.)

1.8.1. Fundamental theorem I.

First we prove the statement about the integral of a derivative.

Theorem 1.45 (Fundamental theorem of calculus I)**.** *If $F : [a,b] \to \mathbb{R}$ is continuous on $[a,b]$ and differentiable in (a,b) with $F' = f$ where $f : [a,b] \to \mathbb{R}$ is Riemann integrable, then*

$$\int_a^b f(x)\,dx = F(b) - F(a).$$

Proof. Let

$$P = \{a = x_0, x_1, x_2, \dots, x_{n-1}, x_n = b\}$$

be a partition of $[a,b]$. Then

$$F(b) - F(a) = \sum_{k=1}^{n} \left[F(x_k) - F(x_{k-1})\right].$$

The function F is continuous on the closed interval $[x_{k-1}, x_k]$ and differentiable in the open interval (x_{k-1}, x_k) with $F' = f$. By the mean value theorem, there exists $x_{k-1} < c_k < x_k$ such that

$$F(x_k) - F(x_{k-1}) = f(c_k)(x_k - x_{k-1}).$$

Since f is Riemann integrable, it is bounded, and it follows that

$$m_k(x_k - x_{k-1}) \le F(x_k) - F(x_{k-1}) \le M_k(x_k - x_{k-1}),$$

where

$$M_k = \sup_{[x_{k-1},x_k]} f, \qquad m_k = \inf_{[x_{k-1},x_k]} f.$$

Hence, $L(f;P) \leq F(b) - F(a) \leq U(f;P)$ for every partition P of $[a,b]$, which implies that $L(f) \leq F(b) - F(a) \leq U(f)$. Since f is integrable, $L(f) = U(f)$ and $F(b) - F(a) = \int_a^b f$. □

In Theorem 1.45, we assume that F is continuous on the closed interval $[a,b]$ and differentiable in the open interval (a,b) where its usual two-sided derivative is defined and is equal to f. It isn't necessary to assume the existence of the right derivative of F at a or the left derivative at b, so the values of f at the endpoints are arbitrary. By Proposition 1.35, however, the integrability of f on $[a,b]$ and the value of its integral do not depend on these values, so the statement of the theorem makes sense. As a result, we'll sometimes abuse terminology, and say that "F' is integrable on $[a,b]$" even if it's only defined on (a,b).

Theorem 1.45 imposes the integrability of F' as a hypothesis. Every function F that is continuously differentiable on the closed interval $[a,b]$ satisfies this condition, but the theorem remains true even if F' is a discontinuous, Riemann integrable function.

Example 1.46. Define $F : [0,1] \to \mathbb{R}$ by

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

Then F is continuous on $[0,1]$ and, by the product and chain rules, differentiable in $(0,1]$. It is also differentiable — but not continuously differentiable — at 0, with $F'(0^+) = 0$. Thus,

$$F'(x) = \begin{cases} -\cos(1/x) + 2x \sin(1/x) & \text{if } 0 < x \leq 1, \\ 0 & \text{if } x = 0. \end{cases}$$

The derivative F' is bounded on $[0,1]$ and discontinuous only at one point ($x = 0$), so Theorem 1.42 implies that F' is integrable on $[0,1]$. This verifies all of the hypotheses in Theorem 1.45, and we conclude that

$$\int_0^1 F'(x)\, dx = \sin 1.$$

There are, however, differentiable functions whose derivatives are unbounded or so discontinuous that they aren't Riemann integrable.

Example 1.47. Define $F : [0,1] \to \mathbb{R}$ by $F(x) = \sqrt{x}$. Then F is continuous on $[0,1]$ and differentiable in $(0,1]$, with

$$F'(x) = \frac{1}{2\sqrt{x}} \qquad \text{for } 0 < x \leq 1.$$

This function is unbounded, so F' is not Riemann integrable on $[0,1]$, however we define its value at 0, and Theorem 1.45 does not apply.

We can, however, interpret the integral of F' on $[0, 1]$ as an improper Riemann integral. The function F is continuously differentiable on $[\epsilon, 1]$ for every $0 < \epsilon < 1$, so

$$\int_{\epsilon}^{1} \frac{1}{2\sqrt{x}}\, dx = 1 - \sqrt{\epsilon}.$$

Thus, we get the improper integral

$$\lim_{\epsilon \to 0^+} \int_{\epsilon}^{1} \frac{1}{2\sqrt{x}}\, dx = 1.$$

The construction of a function with a bounded, non-integrable derivative is more involved. It's not sufficient to give a function with a bounded derivative that is discontinuous at finitely many points, as in Example 1.46, because such a function is Riemann integrable. Rather, one has to construct a differentiable function whose derivative is discontinuous on a set of nonzero Lebesgue measure; we won't give an example here.

Finally, we remark that Theorem 1.45 remains valid for the oriented Riemann integral, since exchanging a and b reverses the sign of both sides.

1.8.2. Fundamental theorem of calculus II. Next, we prove the other direction of the fundamental theorem. We will use the following result, of independent interest, which states that the average of a continuous function on an interval approaches the value of the function as the length of the interval shrinks to zero. The proof uses a common trick of taking a constant inside an average.

Theorem 1.48. *Suppose that $f : [a, b] \to \mathbb{R}$ is integrable on $[a, b]$ and continuous at a. Then*

$$\lim_{h \to 0^+} \frac{1}{h} \int_{a}^{a+h} f(x)\, dx = f(a).$$

Proof. If k is a constant, we have

$$k = \frac{1}{h} \int_{a}^{a+h} k\, dx.$$

(That is, the average of a constant is equal to the constant.) We can therefore write

$$\frac{1}{h} \int_{a}^{a+h} f(x)\, dx - f(a) = \frac{1}{h} \int_{a}^{a+h} \left[f(x) - f(a)\right]\, dx.$$

Let $\epsilon > 0$. Since f is continuous at a, there exists $\delta > 0$ such that

$$|f(x) - f(a)| < \epsilon \quad \text{for} \quad a \le x < a + \delta.$$

It follows that if $0 < h < \delta$, then

$$\left| \frac{1}{h} \int_{a}^{a+h} f(x)\, dx - f(a) \right| \le \frac{1}{h} \cdot \sup_{a \le a \le a+h} |f(x) - f(a)| \cdot h \le \epsilon,$$

which proves the result. □

A similar proof shows that if f is continuous at b, then

$$\lim_{h\to 0^+} \frac{1}{h}\int_{b-h}^{b} f = f(b),$$

and if f is continuous at $a < c < b$, then

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

More generally, if $\{I_h : h > 0\}$ is any collection of intervals with $c \in I_h$ and $|I_h| \to 0$ as $h \to 0^+$, then

$$\lim_{h\to 0^+} \frac{1}{|I_h|}\int_{I_h} f = f(c).$$

The assumption in Theorem 1.48 that f is continuous at the point about which we take the averages is essential.

Example 1.49. Let $f : \mathbb{R} \to \mathbb{R}$ be the sign function

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

Then

$$\lim_{h\to 0^+} \frac{1}{h}\int_{0}^{h} f(x)\,dx = 1, \qquad \lim_{h\to 0^+} \frac{1}{h}\int_{-h}^{0} f(x)\,dx = -1,$$

and neither limit is equal to $f(0)$. In this example, the limit of the symmetric averages

$$\lim_{h\to 0^+} \frac{1}{2h}\int_{-h}^{h} f(x)\,dx = 0$$

is equal to $f(0)$, but this equality doesn't hold if we change $f(0)$ to a nonzero value, since the limit of the symmetric averages is still 0.

The second part of the fundamental theorem follows from this result and the fact that the difference quotients of F are averages of f.

Theorem 1.50 (Fundamental theorem of calculus II)**.** Suppose that $f : [a, b] \to \mathbb{R}$ is integrable and $F : [a, b] \to \mathbb{R}$ is defined by

$$F(x) = \int_{a}^{x} f(t)\,dt.$$

Then F is continuous on $[a, b]$. Moreover, if f is continuous at $a \le c \le b$, then F is differentiable at c and $F'(c) = f(c)$.

Proof. First, note that Theorem 1.33 implies that f is integrable on $[a, x]$ for every $a \le x \le b$, so F is well-defined. Since f is Riemann integrable, it is bounded, and $|f| \le M$ for some $M \ge 0$. It follows that

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

which shows that F is continuous on $[a, b]$ (in fact, Lipschitz continuous).

Moreover, we have

$$\frac{F(c+h)-F(c)}{h} = \frac{1}{h}\int_c^{c+h} f(t)\,dt.$$

It follows from Theorem 1.48 that if f is continuous at c, then F is differentiable at c with

$$F'(c) = \lim_{h\to 0}\left[\frac{F(c+h)-F(c)}{h}\right] = \lim_{h\to 0}\frac{1}{h}\int_c^{c+h} f(t)\,dt = f(c),$$

where we use the appropriate right or left limit at an endpoint. □

The assumption that f is continuous is needed to ensure that F is differentiable.

Example 1.51. If

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

then

$$F(x) = \int_0^x f(t)\,dt = \begin{cases} x & \text{for } x \geq 0, \\ 0 & \text{for } x < 0. \end{cases}$$

The function F is continuous but not differentiable at $x = 0$, where f is discontinuous, since the left and right derivatives of F at 0, given by $F'(0^-) = 0$ and $F'(0^+) = 1$, are different.

1.8.3. Consequences of the fundamental theorem. The first part of the fundamental theorem, Theorem 1.45, is the basic computational tool in integration. It allows us to compute the integral of of a function f if we can find an antiderivative; that is, a function F such that $F' = f$. There is no systematic procedure for finding antiderivatives. Moreover, even if one exists, an antiderivative of an elementary function (constructed from power, trigonometric, and exponential functions and their inverses) may not be — and often isn't — expressible in terms of elementary functions.

Example 1.52. For $p = 0, 1, 2, \dots$, we have

$$\frac{d}{dx}\left[\frac{1}{p+1}x^{p+1}\right] = x^p,$$

and it follows that

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

We remark that once we have the fundamental theorem, we can use the definition of the integral backwards to evaluate a limit such as

$$\lim_{n\to\infty}\left[\frac{1}{n^{p+1}}\sum_{k=1}^{n} k^p\right] = \frac{1}{p+1},$$

since the sum is the upper sum of x^p on a partition of $[0, 1]$ into n intervals of equal length. Example 1.18 illustrates this result explicitly for $p = 2$.

Two important general consequences of the first part of the fundamental theorem are integration by parts and substitution (or change of variable), which come from inverting the product rule and chain rule for derivatives, respectively.

Theorem 1.53 (Integration by parts)**.** Suppose that $f, g : [a, b] \to \mathbb{R}$ are continuous on $[a, b]$ and differentiable in (a, b), and f', g' are integrable on $[a, b]$. Then

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

Proof. The function fg is continuous on $[a, b]$ and, by the product rule, differentiable in (a, b) with derivative

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

Since f, g, f', g' are integrable on $[a, b]$, Theorem 1.26 implies that fg', $f'g$, and $(fg)'$, are integrable. From Theorem 1.45, we get that

$$\int_a^b f g' \, dx + \int_a^b f' g \, dx = \int_a^b f' g \, dx = f(b)g(b) - f(a)g(a),$$

which proves the result. □

Integration by parts says that we can move a derivative from one factor in an integral onto the other factor, with a change of sign and the appearance of a boundary term. The product rule for derivatives expresses the derivative of a product in terms of the derivatives of the factors. By contrast, integration by parts doesn't give an explicit expression for the integral of a product, it simply replaces one integral by another. This can sometimes be used to simplify an integral and evaluate it, but the importance of integration by parts goes far beyond its use as an integration technique.

Example 1.54. For $n = 0, 1, 2, 3, \dots$, let

$$I_n(x) = \int_0^x t^n e^{-t} \, dt.$$

If $n \geq 1$, integration by parts with $f(t) = t^n$ and $g'(t) = e^{-t}$ gives

$$I_n(x) = -x^n e^{-x} + n \int_0^x t^{n-1} e^{-t} \, dt = -x^n e^{-x} + n I_{n-1}(x).$$

Also, by the fundamental theorem,

$$I_0(x) = \int_0^x e^{-t} \, dt = 1 - e^{-x}.$$

It then follows by induction that

$$I_n(x) = n! \left[1 - e^{-x} \sum_{k=0}^{n} \frac{x^k}{k!} \right],$$

where, as usual, $0! = 1$.

Since $x^k e^{-x} \to 0$ as $x \to \infty$ for every $k = 0, 1, 2, \dots$, we get the improper integral

$$\int_0^\infty t^n e^{-t} \, dt = \lim_{r \to \infty} \int_0^r t^n e^{-t} \, dt = n!.$$

This formula suggests an extension of the factorial function to complex numbers $z \in \mathbb{C}$, called the Gamma function, which is defined for $\Re z > 0$ by the improper, complex-valued integral

$$\Gamma(z) = \int_0^\infty t^{z-1} e^{-t}\, dt.$$

In particular, $\Gamma(n) = (n-1)!$ for $n \in \mathbb{N}$. The Gama function is an important special function, which is studied further in complex analysis.

Next we consider the change of variable formula for integrals.

Theorem 1.55 (Change of variable)**.** Suppose that $g : I \to \mathbb{R}$ differentiable on an open interval I and g' is integrable on I. Let $J = g(I)$. If $f : J \to \mathbb{R}$ continuous, then for every $a, b \in I$,

$$\int_a^b f\left(g(x)\right) g'(x)\, dx = \int_{g(a)}^{g(b)} f(u)\, du.$$

Proof. Let

$$F(x) = \int_a^x f(u)\, du.$$

Since f is continuous, Theorem 1.50 implies that F is differentiable in J with $F' = f$. The chain rule implies that the composition $F \circ g : I \to \mathbb{R}$ is differentiable in I, with

$$(F \circ g)'(x) = f\left(g(x)\right) g'(x).$$

This derivative is integrable on $[a, b]$ since $f \circ g$ is continuous and g' is integrable. Theorem 1.45, the definition of F, and the additivity of the integral then imply that

$$\begin{aligned} \int_a^b f\left(g(x)\right) g'(x)\, dx &= \int_a^b (F \circ g)'\, dx \\ &= F\left(g(b)\right) - F\left(g(a)\right) \\ &= \int_{g(a)}^{g(b)} F'(u)\, du, \end{aligned}$$

which proves the result. □

A continuous function maps an interval to an interval, and it is one-to-one if and only if it is strictly monotone. An increasing function preserves the orientation of the interval, while a decreasing function reverses it, in which case the integrals are understood as appropriate oriented integrals. There is no assumption in this theorem that g is invertible, and the result remains valid if g is not monotone.

Example 1.56. For every $a > 0$, the increasing, differentiable function $g : \mathbb{R} \to \mathbb{R}$ defined by $g(x) = x^3$ maps $(-a, a)$ one-to-one and onto $(-a^3, a^3)$ and preserves orientation. Thus, if $f : [-a, a] \to \mathbb{R}$ is continuous,

$$\int_{-a}^{a} f(x^3) \cdot 3x^2\, dx = \int_{-a^3}^{a^3} f(u)\, du.$$

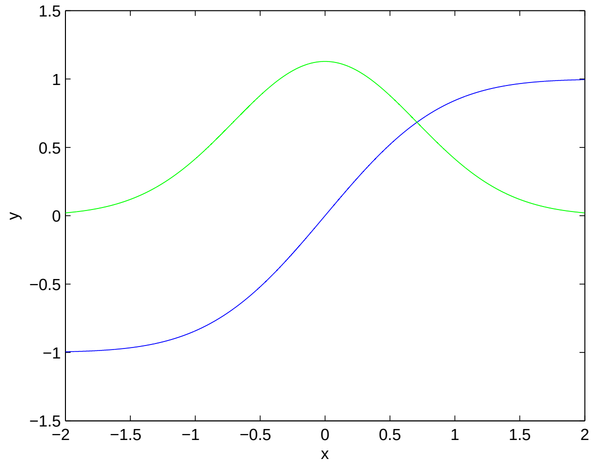

Figure 5. Graphs of the error function $y = F(x)$ (blue) and its derivative, the Gaussian function $y = f(x)$ (green), from Example 1.58.

The decreasing, differentiable function $g : \mathbb{R} \to \mathbb{R}$ defined by $g(x) = -x^3$ maps $(-a, a)$ one-to-one and onto $(-a^3, a^3)$ and reverses orientation. Thus,

$$\int_{-a}^{a} f(-x^3) \cdot (-3x^2)\, dx = \int_{a^3}^{-a^3} f(u)\, du = -\int_{-a^3}^{a^3} f(u)\, du.$$

The non-monotone, differentiable function $g : \mathbb{R} \to \mathbb{R}$ defined by $g(x) = x^2$ maps $(-a, a)$ onto $[0, a^2)$. It is two-to-one, except at $x = 0$. The change of variables formula gives

$$\int_{-a}^{a} f(x^2) \cdot 2x\, dx = \int_{a^2}^{a^2} f(u)\, du = 0.$$

The contributions to the original integral from $[0, a]$ and $[-a, 0]$ cancel since the integrand is an odd function of x.

One consequence of the second part of the fundamental theorem, Theorem 1.50, is that every continuous function has an antiderivative, even if it can't be expressed explicitly in terms of elementary functions. This provides a way to define transcendental functions as integrals of elementary functions.

Example 1.57. One way to define the logarithm $\ln : (0, \infty) \to \mathbb{R}$ in terms of algebraic functions is as the integral

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

The integral is well-defined for every $0 < x < \infty$ since $1/t$ is continuous on the interval $[1, x]$ (or $[x, 1]$ if $0 < x < 1$). The usual properties of the logarithm follow from this representation. We have $(\ln x)' = 1/x$ by definition, and, for example, making the substitution $s = xt$ in the second integral in the following equation,

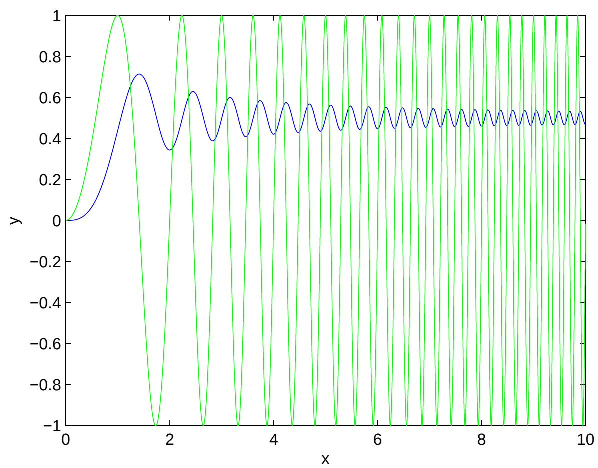

Figure 6. Graphs of the Fresnel integral $y = S(x)$ (blue) and its derivative $y = \sin(\pi x^2/2)$ (green) from Example 1.59.

when $dt/t = ds/s$, we get

$$\ln x + \ln y = \int_1^x \frac{1}{t}\,dt + \int_1^y \frac{1}{t}\,dt = \int_1^x \frac{1}{t}\,dt + \int_x^{xy} \frac{1}{s}\,ds = \int_1^{xy} \frac{1}{t}\,dt = \ln(xy).$$

We can also define many non-elementary functions as integrals.

Example 1.58. The error function

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2}\,dt$$

is an anti-derivative on $\mathbb{R}$ of the Gaussian function

$$f(x) = \frac{2}{\sqrt{\pi}} e^{-x^2}.$$

The error function isn't expressible in terms of elementary functions. Nevertheless, it is defined as a limit of Riemann sums for the integral. Figure 5 shows the graphs of f and F. The name "error function" comes from the fact that the probability of a Gaussian random variable deviating by more than a given amount from its mean can be expressed in terms of F. Error functions also arise in other applications; for example, in modeling diffusion processes such as heat flow.

Example 1.59. The Fresnel sine function S is defined by

$$S(x) = \int_0^x \sin\left(\frac{\pi t^2}{2}\right) dt.$$

The function S is an antiderivative of $\sin(\pi t^2/2)$ on $\mathbb{R}$ (see Figure 6), but it can't be expressed in terms of elementary functions. Fresnel integrals arise, among other places, in analysing the diffraction of waves, such as light waves. From the perspective of complex analysis, they are closely related to the error function through the Euler formula $e^{i\theta} = \cos\theta + i\sin\theta$.

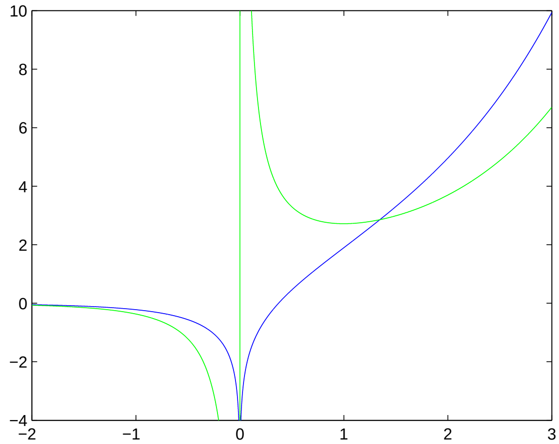

Figure 7. Graphs of the exponential integral $y = \mathrm{Ei}(x)$ (blue) and its derivative $y = e^x/x$ (green) from Example 1.60.

Example 1.60. The exponential integral Ei is a non-elementary function defined by

$$\mathrm{Ei}(x) = \int_{-\infty}^{x} \frac{e^t}{t}\, dt.$$

Its graph is shown in Figure 7. This integral has to be understood, in general, as an improper, principal value integral, and the function has a logarithmic singularity at $x = 0$ (see Example 1.83 below for further explanation). The exponential integral arises in physical applications such as heat flow and radiative transfer. It is also related to the logarithmic integral

$$\mathrm{li}(x) = \int_{0}^{x} \frac{dt}{\ln t}$$

by $\mathrm{li}(x) = \mathrm{Ei}(\ln x)$. The logarithmic integral is important in number theory, and it gives an asymptotic approximation for the number of primes less than x as $x \to \infty$. Roughly speaking, the density of the primes near a large number x is close to $1/\ln x$.

Discontinuous functions may or may not have an antiderivative, and typically they don't. Darboux proved that every function $f : (a, b) \to \mathbb{R}$ that is the derivative of a function $F : (a, b) \to \mathbb{R}$, where $F' = f$ at *every* point of (a, b), has the intermediate value property. That is, if $a < c < d < b$, then for every y between $f(c)$ and $f(d)$ there exists an x between c and d such that $f(x) = y$. A continuous derivative has this property by the intermediate value theorem, but a discontinuous derivative also has it. Thus, functions without the intermediate value property, such as ones with a jump discontinuity or the Dirichlet function, don't have an antiderivative. For example, the function F in Example 1.51 is not an antiderivative of the step function f on $\mathbb{R}$ since it isn't differentiable at 0.

In dealing with functions that are not continuously differentiable, it turns out to be more useful to abandon the idea of a derivative that is defined pointwise everywhere (pointwise values of discontinuous functions are somewhat arbitrary) and introduce the notion of a weak derivative. We won't define or study weak derivatives here.

1.9. Integrals and sequences of functions

A fundamental question that arises throughout analysis is the validity of an exchange in the order of limits. Some sort of condition is always required.

In this section, we consider the question of when the convergence of a sequence of functions $f_n \to f$ implies the convergence of their integrals $\int f_n \to \int f$. There are many inequivalent notions of the convergence of functions. The two we'll discuss here are pointwise and uniform convergence.

Recall that if $f_n, f : A \to \mathbb{R}$, then $f_n \to f$ pointwise on A as $n \to \infty$ if $f_n(x) \to f(x)$ for every $x \in A$. On the other hand, $f_n \to f$ uniformly on A if for every $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that

$$n > N \quad \text{implies that} \quad |f_n(x) - f(x)| < \epsilon \quad \text{for every } x \in A.$$

Equivalently, $f_n \to f$ uniformly on A if $\|f_n - f\| \to 0$ as $n \to \infty$, where

$$\|f\| = \sup\{|f(x)| : x \in A\}$$

denotes the sup-norm of a function $f : A \to \mathbb{R}$. Uniform convergence implies pointwise convergence, but not conversely.

As we show first, the Riemann integral is well-behaved with respect to uniform convergence. The drawback to uniform convergence is that it's a strong form of convergence, and we often want to use a weaker form, such as pointwise convergence, in which case the Riemann integral may not be suitable.

1.9.1. Uniform convergence.

The uniform limit of continuous functions is continuous and therefore integrable. The next result shows, more generally, that the uniform limit of integrable functions is integrable. Furthermore, the limit of the integrals is the integral of the limit.

Theorem 1.61. Suppose that $f_n : [a, b] \to \mathbb{R}$ is Riemann integrable for each $n \in \mathbb{N}$ and $f_n \to f$ uniformly on $[a, b]$ as $n \to \infty$. Then $f : [a, b] \to \mathbb{R}$ is Riemann integrable on $[a, b]$ and

$$\int_a^b f = \lim_{n\to\infty} \int_a^b f_n.$$

Proof. The uniform limit of bounded functions is bounded, so f is bounded. The main statement we need to prove is that f is integrable.

Let $\epsilon > 0$. Since $f_n \to f$ uniformly, there is an $N \in \mathbb{N}$ such that if $n > N$ then

$$f_n(x) - \frac{\epsilon}{b-a} < f(x) < f_n(x) + \frac{\epsilon}{b-a} \qquad \text{for all } a \le x \le b.$$

It follows from Proposition 1.30 that

$$L\left(f_n - \frac{\epsilon}{b-a}\right) \le L(f), \qquad U(f) \le U\left(f_n + \frac{\epsilon}{b-a}\right).$$

Since f_n is integrable and upper integrals are greater than lower integrals, we get that

$$\int_a^b f_n - \epsilon \leq L(f) \leq U(f) \leq \int_a^b f_n + \epsilon,$$

which implies that

$$0 \leq U(f) - L(f) \leq 2\epsilon.$$

Since $\epsilon > 0$ is arbitrary, we conclude that $L(f) = U(f)$, so f is integrable. Moreover, it follows that for all $n > N$ we have

$$\left| \int_a^b f_n - \int_a^b f \right| \leq \epsilon,$$

which shows that $\int_a^b f_n \to \int_a^b f$ as $n \to \infty$. □

Once we know that the uniform limit of integrable functions is integrable, the convergence of the integrals also follows directly from the estimate

$$\left| \int_a^b f_n - \int_a^b f \right| = \left| \int_a^b (f_n - f) \right| \leq \|f_n - f\| \cdot (b-a) \to 0 \qquad \text{as } n \to \infty.$$

Example 1.62. The function $f_n : [0,1] \to \mathbb{R}$ defined by

$$f_n(x) = \frac{n + \cos x}{ne^x + \sin x}$$

converges uniformly on $[0,1]$ to

$$f(x) = e^{-x},$$

since for $0 \leq x \leq 1$

$$\left| \frac{n + \cos x}{ne^x + \sin x} - e^{-x} \right| = \left| \frac{\cos x - e^{-x} \sin x}{ne^x + \sin x} \right| \leq \frac{2}{n}.$$

It follows that

$$\lim_{n\to\infty} \int_0^1 \frac{n + \cos x}{ne^x + \sin x}\, dx = \int_0^1 e^{-x}\, dx = 1 - \frac{1}{e}.$$

Example 1.63. Every power series

$$f(x) = a_0 + a_1 x + a^2 x^2 + \cdots + a_n x^n + \ldots$$

with radius of convergence $R > 0$ converges uniformly on compact intervals inside the interval $|x| < R$, so we can integrate it term-by-term to get

$$\int_0^x f(t)\, dt = a_0 x + \frac{1}{2} a_1 x^2 + \frac{1}{3} a_2 x^3 + \cdots + \frac{1}{n+1} a_n x^{n+1} + \ldots \qquad \text{for } |x| < R.$$

As one example, if we integrate the geometric series

$$\frac{1}{1-x} = 1 + x + x^2 + \cdots + x^n + \ldots \qquad \text{for } |x| < 1,$$

we get a power series for ln,

$$\ln\left(\frac{1}{1-x} \right) = x + \frac{1}{2} x^2 + \frac{1}{3} x^3 \cdots + \frac{1}{n} x^n + \ldots \qquad \text{for } |x| < 1.$$

For instance, taking $x = 1/2$, we get the rapidly convergent series

$$\ln 2 = \sum_{n=1}^{\infty} \frac{1}{n2^n}$$

for the irrational number $\ln 2 \approx 0.6931$. This series was known and used by Euler.

Although we can integrate uniformly convergent sequences, we cannot in general differentiate them. In fact, it's often easier to prove results about the convergence of derivatives by using results about the convergence of integrals, together with the fundamental theorem of calculus. The following theorem provides sufficient conditions for $f_n \to f$ to imply that $f_n' \to f'$.

Theorem 1.64. Let $f_n : (a, b) \to \mathbb{R}$ be a sequence of differentiable functions whose derivatives $f_n' : (a, b) \to \mathbb{R}$ are integrable on (a, b). Suppose that $f_n \to f$ pointwise and $f_n' \to g$ uniformly on (a, b) as $n \to \infty$, where $g : (a, b) \to \mathbb{R}$ is continuous. Then $f : (a, b) \to \mathbb{R}$ is continuously differentiable on (a, b) and $f' = g$.

Proof. Choose some point $a < c < b$. Since f_n' is integrable, the fundamental theorem of calculus, Theorem 1.45, implies that

$$f_n(x) = f_n(c) + \int_c^x f_n' \qquad \text{for } a < x < b.$$

Since $f_n \to f$ pointwise and $f_n' \to g$ uniformly on $[a, x]$, we find that

$$f(x) = f(c) + \int_c^x g.$$

Since g is continuous, the other direction of the fundamental theorem, Theorem 1.50, implies that f is differentiable in (a, b) and $f' = g$. □

In particular, this theorem shows that the limit of a uniformly convergent sequence of continuously differentiable functions whose derivatives converge uniformly is also continuously differentiable.

The key assumption in Theorem 1.64 is that the derivatives f_n' converge uniformly, not just pointwise; the result is false if we only assume pointwise convergence of the f_n'. In the proof of the theorem, we only use the assumption that $f_n(x)$ converges at a single point $x = c$. This assumption together with the assumption that $f_n' \to g$ uniformly implies that $f_n \to f$ pointwise (and, in fact, uniformly) where

$$f(x) = \lim_{n\to\infty} f_n(c) + \int_c^x g.$$

Thus, the theorem remains true if we replace the assumption that $f_n \to f$ pointwise on (a, b) by the weaker assumption that $\lim_{n\to\infty} f_n(c)$ exists for some $c \in (a, b)$. This isn't an important change, however, because the restrictive assumption in the theorem is the uniform convergence of the derivatives f_n', not the pointwise (or uniform) convergence of the functions f_n.

The assumption that $g = \lim f_n'$ is continuous is needed to show the differentiability of f by the fundamental theorem, but the result result true even if g isn't continuous. In that case, however, a different (and more complicated) proof is required.

1.9.2. Pointwise convergence.
On its own, the pointwise convergence of functions is never sufficient to imply convergence of their integrals.

Example 1.65. For $n \in \mathbb{N}$, define $f_n : [0,1] \to \mathbb{R}$ by

$$f_n(x) = \begin{cases} n & \text{if } 0 < x < 1/n, \\ 0 & \text{if } x = 0 \text{ or } 1/n \le x \le 1. \end{cases}$$

Then $f_n \to 0$ pointwise on $[0,1]$ but

$$\int_0^1 f_n = 1$$

for every $n \in \mathbb{N}$. By slightly modifying these functions to

$$f_n(x) = \begin{cases} n^2 & \text{if } 0 < x < 1/n, \\ 0 & \text{if } x = 0 \text{ or } 1/n \le x \le 1, \end{cases}$$

we get a sequence that converges pointwise to 0 but whose integrals diverge to ∞. The fact that the f_n are discontinuous is not important; we could replace the step functions by continuous "tent" functions or smooth "bump" functions.

The behavior of the integral under pointwise convergence in the previous example is unavoidable. A much worse feature of the Riemann integral is that the pointwise limit of integrable functions needn't be integrable at all, even if it is bounded.

Example 1.66. Let $\{q_k : k \in \mathbb{N}\}$ be an enumeration of the rationals in $[0,1]$ and define $f_n : [0,1] \to \mathbb{R}$ by

$$f_n(x) = \begin{cases} 1 & \text{if } x = q_k \text{ for } k \le n, \\ 0 & \text{otherwise.} \end{cases}$$

The each f_n is Riemann integrable since it differs from the zero function at finitely many points. However, $f_n \to f$ pointwise on $[0,1]$ to the Dirichlet function f, which is not Riemann integrable.

This is another place where the Lebesgue integral has better properties than the Riemann integral. The pointwise (or pointwise almost everywhere) limit of Lebesgue integrable functions is Lebesgue integrable. As Example 1.65 shows, we still need conditions to ensure the convergence of the integrals, but there are quite simple and general conditions for the Lebesgue integral (such as the monotone convergence and dominated convergence theorems).

1.10. Improper Riemann integrals

The Riemann integral is only defined for a bounded function on a compact interval (or a finite union of such intervals). Nevertheless, we frequently want to integrate an unbounded function or a function on an infinite interval. One way to interpret such an integral is as a limit of Riemann integrals; this limit is called an improper Riemann integral.

1.10.1. Improper integrals. First, we define the improper integral of a function that fails to be integrable at one endpoint of a bounded interval.

Definition 1.67. Suppose that $f : (a, b] \to \mathbb{R}$ is integrable on $[c, b]$ for every $a < c < b$. Then the improper integral of f on $[a, b]$ is

$$\int_a^b f = \lim_{\epsilon \to 0^+} \int_{a+\epsilon}^b f.$$

The improper integral converges if this limit exists (as a finite real number), otherwise it diverges. Similarly, if $f : [a, b) \to \mathbb{R}$ is integrable on $[a, c]$ for every $a < c < b$, then

$$\int_a^b f = \lim_{\epsilon \to 0^+} \int_a^{b-\epsilon} f.$$

We use the same notation to denote proper and improper integrals; it should be clear from the context which integrals are proper Riemann integrals (i.e., ones given by Definition 1.3) and which are improper. If f is Riemann integrable on $[a, b]$, then Proposition 1.39 shows that its improper and proper integrals agree, but an improper integral may exist even if f isn't integrable.

Example 1.68. If $p > 0$, the integral

$$\int_0^1 \frac{1}{x^p}\, dx$$

isn't defined as a Riemann integral since $1/x^p$ is unbounded on $(0, 1]$. The corresponding improper integral is

$$\int_0^1 \frac{1}{x^p}\, dx = \lim_{\epsilon \to 0^+} \int_\epsilon^1 \frac{1}{x^p}\, dx.$$

For $p \neq 1$, we have

$$\int_\epsilon^1 \frac{1}{x^p}\, dx = \frac{1 - \epsilon^{1-p}}{1 - p},$$

so the improper integral converges if $0 < p < 1$, with

$$\int_0^1 \frac{1}{x^p}\, dx = \frac{1}{p - 1},$$

and diverges to ∞ if $p > 1$. The integral also diverges (more slowly) to ∞ if $p = 1$ since

$$\int_\epsilon^1 \frac{1}{x}\, dx = \ln \frac{1}{\epsilon}.$$

Thus, we get a convergent improper integral if the integrand $1/x^p$ does not grow too rapidly as $x \to 0^+$ (slower than $1/x$).

We define improper integrals on an unbounded interval as limits of integrals on bounded intervals.

Definition 1.69. Suppose that $f : [a, \infty) \to \mathbb{R}$ is integrable on $[a, r]$ for every $r > a$. Then the improper integral of f on $[a, \infty)$ is

$$\int_a^\infty f = \lim_{r \to \infty} \int_a^r f.$$

Similarly, if $f : (-\infty, b] \to \mathbb{R}$ is integrable on $[r, b]$ for every $r < b$, then

$$\int_{-\infty}^{b} f = \lim_{r\to\infty} \int_{-r}^{b} f.$$

Let's consider the convergence of the integral of the power function in Example 1.68 at infinity rather than at zero.

Example 1.70. Suppose $p > 0$. The improper integral

$$\int_{1}^{\infty} \frac{1}{x^p}\,dx = \lim_{r\to\infty} \int_{1}^{r} \frac{1}{x^p}\,dx = \lim_{r\to\infty} \left(\frac{r^{1-p}-1}{1-p}\right)$$

converges to $1/(p-1)$ if $p > 1$ and diverges to ∞ if $0 < p < 1$. It also diverges (more slowly) if $p = 1$ since

$$\int_{1}^{\infty} \frac{1}{x}\,dx = \lim_{r\to\infty} \int_{1}^{r} \frac{1}{x}\,dx = \lim_{r\to\infty} \ln r = \infty.$$

Thus, we get a convergent improper integral if the integrand $1/x^p$ decays sufficiently rapidly as $x \to \infty$ (faster than $1/x$).

A divergent improper integral may diverge to ∞ (or $-\infty$) as in the previous examples, or — if the integrand changes sign — it may oscillate.

Example 1.71. Define $f : [0, \infty) \to \mathbb{R}$ by

$$f(x) = (-1)^n \qquad \text{for } n \le x < n+1 \text{ where } n = 0, 1, 2, \dots.$$

Then $0 \le \int_0^r f \le 1$ and

$$\int_0^n f = \begin{cases} 1 & \text{if } n \text{ is an odd integer,} \\ 0 & \text{if } n \text{ is an even integer.} \end{cases}$$

Thus, the improper integral $\int_0^\infty f$ doesn't converge.

More general improper integrals may be defined as finite sums of improper integrals of the previous forms. For example, if $f : [a, b] \setminus \{c\} \to \mathbb{R}$ is integrable on closed intervals not including $a < c < b$, then

$$\int_a^b f = \lim_{\delta\to 0^+} \int_a^{c-\delta} f + \lim_{\epsilon\to 0^+} \int_{c+\epsilon}^{b} f;$$

and if $f : \mathbb{R} \to \mathbb{R}$ is integrable on every compact interval, then

$$\int_{-\infty}^{\infty} f = \lim_{s\to\infty} \int_{-s}^{c} f + \lim_{r\to\infty} \int_{c}^{r} f,$$

where we split the integral at an arbitrary point $c \in \mathbb{R}$. Note that each limit is required to exist separately.

Example 1.72. If $f : [0, 1] \to \mathbb{R}$ is continuous and $0 < c < 1$, then we define as an improper integral

$$\int_0^1 \frac{f(x)}{|x-c|^{1/2}}\,dx = \lim_{\delta\to 0^+} \int_0^{c-\delta} \frac{f(x)}{|x-c|^{1/2}}\,dx + \lim_{\epsilon\to 0^+} \int_{c+\epsilon}^{1} \frac{f(x)}{|x-c|^{1/2}}\,dx.$$

Integrals like this one appear in the theory of integral equations.

Example 1.73. Consider the following integral, called a Frullani integral,

$$I = \int_0^\infty \frac{f(ax) - f(bx)}{x}\, dx.$$

We assume that $a, b > 0$ and $f : [0, \infty) \to \mathbb{R}$ is a continuous function whose limit as $x \to \infty$ exists; we write this limit as

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

We interpret the integral as an improper integral $I = I_1 + I_2$ where

$$I_1 = \lim_{\epsilon\to 0^+} \int_\epsilon^1 \frac{f(ax) - f(bx)}{x}\, dx, \qquad I_2 = \lim_{r\to\infty} \int_1^r \frac{f(ax) - f(bx)}{x}\, dx.$$

Consider I_1. After making the substitutions $s = ax$ and $t = bx$ and using the additivity property of the integral, we get that

$$I_1 = \lim_{\epsilon\to 0^+} \left(\int_{\epsilon a}^a \frac{f(s)}{s}\, ds - \int_{\epsilon b}^b \frac{f(t)}{t}\, dt \right) = \lim_{\epsilon\to 0^+} \int_{\epsilon a}^{\epsilon b} \frac{f(t)}{t}\, dt - \int_a^b \frac{f(t)}{t}\, dt.$$

To evaluate the limit, we write

$$\begin{aligned} \int_{\epsilon a}^{\epsilon b} \frac{f(t)}{t}\, dt &= \int_{\epsilon a}^{\epsilon b} \frac{f(t) - f(0)}{t}\, dt + f(0) \int_{\epsilon a}^{\epsilon b} \frac{1}{t}\, dt \\ &= \int_{\epsilon a}^{\epsilon b} \frac{f(t) - f(0)}{t}\, dt + f(0) \ln\left(\frac{b}{a}\right). \end{aligned}$$

Assuming for definiteness that $0 < a < b$, we have

$$\left| \int_{\epsilon a}^{\epsilon b} \frac{f(t) - f(0)}{t}\, dt \right| \le \left(\frac{b-a}{a}\right) \cdot \max\{|f(t) - f(0)| : \epsilon a \le t \le \epsilon b\} \to 0$$

as $\epsilon \to 0^+$, since f is continuous at 0. It follows that

$$I_1 = f(0) \ln\left(\frac{b}{a}\right) - \int_a^b \frac{f(t)}{t}\, dt.$$

A similar argument gives

$$I_2 = -f(\infty) \ln\left(\frac{b}{a}\right) + \int_a^b \frac{f(t)}{t}\, dt.$$

Adding these results, we conclude that

$$\int_0^\infty \frac{f(ax) - f(bx)}{x}\, dx = \{f(0) - f(\infty)\} \ln\left(\frac{b}{a}\right).$$

1.10.2. Absolutely convergent improper integrals. The convergence of improper integrals is analogous to the convergence of series. A series $\sum a_n$ converges absolutely if $\sum |a_n|$ converges, and conditionally if $\sum a_n$ converges but $\sum |a_n|$ diverges. We introduce a similar definition for improper integrals and provide a test for the absolute convergence of an improper integral that is analogous to the comparison test for series.

Definition 1.74. An improper integral $\int_a^b f$ is absolutely convergent if the improper integral $\int_a^b |f|$ converges, and conditionally convergent if $\int_a^b f$ converges but $\int_a^b |f|$ diverges.

As part of the next theorem, we prove that an absolutely convergent improper integral converges (similarly, an absolutely convergent series converges).

Theorem 1.75. Suppose that $f, g : I \to \mathbb{R}$ are defined on some finite or infinite interval I. If $|f| \le g$ and the improper integral $\int_I g$ converges, then the improper integral $\int_I f$ converges absolutely. Moreover, an absolutely convergent improper integral converges.

Proof. To be specific, we suppose that $f, g : [a, \infty) \to \mathbb{R}$ are integrable on $[a, r]$ for $r > a$ and consider the improper integral

$$\int_a^\infty f = \lim_{r\to\infty} \int_a^r f.$$

A similar argument applies to other types of improper integrals.

First, suppose that $f \ge 0$. Then

$$\int_a^r f \le \int_a^r g \le \int_a^\infty g,$$

so $\int_a^r f$ is a monotonic increasing function of r that is bounded from above. Therefore it converges as $r \to \infty$.

In general, we decompose f into its positive and negative parts,

$$\begin{aligned} f &= f_+ - f_-, & |f| &= f_+ + f_-, \\ f_+ &= \max\{f, 0\}, & f_- &= \max\{-f, 0\}. \end{aligned}$$

We have $0 \le f_\pm \le g$, so the improper integrals of $f_\pm$ converge by the previous argument, and therefore so does the improper integral of f:

$$\begin{aligned} \int_a^\infty f &= \lim_{r\to\infty} \left(\int_a^r f_+ - \int_a^r f_- \right) \\ &= \lim_{r\to\infty} \int_a^r f_+ - \lim_{r\to\infty} \int_a^r f_- \\ &= \int_a^\infty f_+ - \int_a^\infty f_-. \end{aligned}$$

Moreover, since $0 \le f_\pm \le |f|$, we see that $\int_a^\infty f_+$ and $\int_a^\infty f_-$ converge if $\int_a^\infty |f|$ converges, and therefore so does $\int_a^\infty f$. $\square$

Example 1.76. Consider the limiting behavior of the error function $\operatorname{erf}(x)$ in Example 1.58 as $x \to \infty$, which is given by

$$\frac{2}{\sqrt{\pi}} \int_0^\infty e^{-x^2}\, dx = \frac{2}{\sqrt{\pi}} \lim_{r\to\infty} \int_0^r e^{-x^2}\, dx.$$

The convergence of this improper integral follows by comparison with e^{-x}, for example, since

$$0 \le e^{-x^2} \le e^{-x} \qquad \text{for } x \ge 1,$$

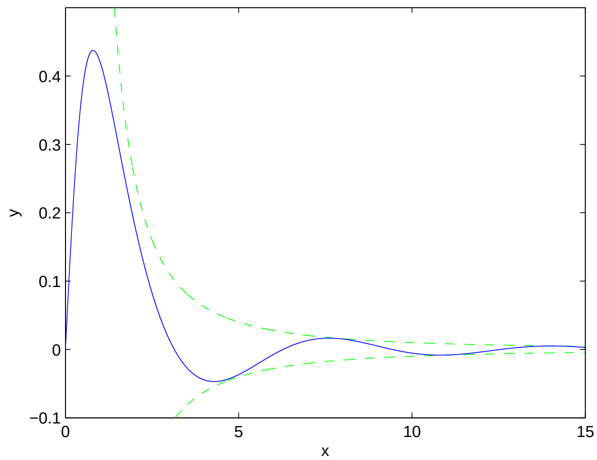

Figure 8. Graph of $y = (\sin x)/(1 + x^2)$ from Example 1.77. The dashed green lines are the graphs of $y = \pm 1/x^2$.

and

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

This argument proves that the error function approaches a finite limit as $x \to \infty$, but it doesn't give the exact value, only an upper bound

$$\frac{2}{\sqrt{\pi}}\int_0^\infty e^{-x^2}\,dx \le M, \qquad M = \frac{2}{\sqrt{\pi}}\int_0^1 e^{-x^2}\,dx + \frac{1}{e}.$$

Numerically, $M \approx 1.2106$. In fact, one can show that

$$\frac{2}{\sqrt{\pi}}\int_0^\infty e^{-x^2}\,dx = 1.$$

The standard trick (apparently introduced by Laplace) uses double integration, polar coordinates, and the substitution $u = r^2$:

$$\begin{aligned}\left(\int_0^\infty e^{-x^2}\,dx\right)^2 &= \int_0^\infty\int_0^\infty e^{-x^2-y^2}\,dxdy \\ &= \int_0^{\pi/2}\left(\int_0^\infty e^{-r^2}\,r\,dr\right)d\theta \\ &= \frac{\pi}{4}\int_0^\infty e^{-u}\,du = \frac{\pi}{4}.\end{aligned}$$

This formal computation can be justified rigorously, but we won't do that here.

Example 1.77. The improper integral

$$\int_0^\infty \frac{\sin x}{1+x^2}\,dx = \lim_{r\to\infty}\int_0^r \frac{\sin x}{1+x^2}\,dx$$

converges absolutely, since

$$\int_0^\infty \frac{\sin x}{1+x^2}\,dx = \int_0^1 \frac{\sin x}{1+x^2}\,dx + \int_1^\infty \frac{\sin x}{1+x^2}\,dx$$

and (see Figure 8)

$$\left|\frac{\sin x}{1+x^2}\right| \le \frac{1}{x^2} \quad \text{for } x \ge 1, \qquad \int_1^\infty \frac{1}{x^2}\,dx < \infty.$$

The value of this integral doesn't have an elementary expression, but by using contour integration from complex analysis one can show that

$$\int_0^\infty \frac{\sin x}{1+x^2}\,dx = \frac{1}{2e}\,\mathrm{Ei}(1) - \frac{e}{2}\,\mathrm{Ei}(-1) \approx 0.6468,$$

where Ei is the exponential integral function defined in Example 1.60.

Improper integrals, and the principal value integrals discussed below, arise frequently in complex analysis, and many such integrals can be evaluated by contour integration.

Example 1.78. The improper integral

$$\int_0^\infty \frac{\sin x}{x}\,dx = \lim_{r\to\infty} \int_0^r \frac{\sin x}{x}\,dx = \frac{\pi}{2}$$

converges conditionally. We leave the proof as an exercise. Comparison with the function $1/x$ doesn't imply absolute convergence at infinity because the improper integral $\int_1^\infty 1/x\,dx$ diverges. There are many ways to show that the exact value of the improper integral is $\pi/2$. The standard method uses contour integration.

Example 1.79. Consider the limiting behavior of the Fresnel sine function $S(x)$ in Example 1.59 as $x \to \infty$. The improper integral

$$\int_0^\infty \sin\left(\frac{\pi x^2}{2}\right)\,dx = \lim_{r\to\infty} \int_0^r \sin\left(\frac{\pi x^2}{2}\right)\,dx = \frac{1}{2}.$$

converges conditionally. This example may seem surprising since the integrand $\sin(\pi x^2/2)$ doesn't converge to 0 as $x \to \infty$. The explanation is that the integrand oscillates more rapidly with increasing x, leading to a more rapid cancelation between positive and negative values in the integral (see Figure 6). The exact value can be found by contour integration, again, which shows that

$$\int_0^\infty \sin\left(\frac{\pi x^2}{2}\right)\,dx = \frac{1}{\sqrt{2}} \int_0^\infty \exp\left(-\frac{\pi x^2}{2}\right)\,dx.$$

Evaluation of the resulting Gaussian integral gives $1/2$.

1.10.3. Principal value integrals. Some integrals have a singularity that is too strong for them to converge as improper integrals but, due to cancelation, they have a finite limit as a principal value integral. We begin with an example.

Example 1.80. Consider $f : [-1,1] \setminus \{0\}$ defined by

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

The definition of the integral of f on $[-1, 1]$ as an improper integral is

$$\begin{aligned}\int_{-1}^{1} \frac{1}{x}\,dx &= \lim_{\delta\to 0^+}\int_{-1}^{-\delta}\frac{1}{x}\,dx + \lim_{\epsilon\to 0^+}\int_{\epsilon}^{1}\frac{1}{x}\,dx \\ &= \lim_{\delta\to 0^+}\ln\delta - \lim_{\epsilon\to 0^+}\ln\epsilon.\end{aligned}$$

Neither limit exists, so the improper integral diverges. (Formally, we get $\infty - \infty$.) If, however, we take $\delta = \epsilon$ and combine the limits, we get a convergent principal value integral, which is defined by

$$\text{p.v.}\int_{-1}^{1}\frac{1}{x}\,dx = \lim_{\epsilon\to 0^+}\left(\int_{-1}^{-\epsilon}\frac{1}{x}\,dx + \int_{\epsilon}^{1}\frac{1}{x}\,dx\right) = \lim_{\epsilon\to 0^+}(\ln\epsilon - \ln\epsilon) = 0.$$

The value of 0 is what one might expect from the oddness of the integrand. A cancelation in the contributions from either side of the singularity is essentially to obtain a finite limit.

The principal value integral of $1/x$ on a non-symmetric interval about 0 still exists but is non-zero. For example, if $b > 0$, then

$$\text{p.v.}\int_{-1}^{b}\frac{1}{x}\,dx = \lim_{\epsilon\to 0^+}\left(\int_{-1}^{-\epsilon}\frac{1}{x}\,dx + \int_{\epsilon}^{b}\frac{1}{x}\,dx\right) = \lim_{\epsilon\to 0^+}(\ln\epsilon + \ln b - \ln\epsilon) = \ln b.$$

The crucial feature if a principal value integral is that we remove a symmetric interval around a singular point, or infinity. The resulting cancelation in the integral of a non-integrable function that changes sign across the singularity may lead to a finite limit.

Definition 1.81. If $f : [a, b]\backslash\{c\} \to \mathbb{R}$ is integrable on closed intervals not including $a < c < b$, then the principal value integral of f is

$$\text{p.v.}\int_a^b f = \lim_{\epsilon\to 0^+}\left(\int_a^{c-\epsilon} f + \int_{c+\epsilon}^{b} f\right).$$

If $f : \mathbb{R} \to \mathbb{R}$ is integrable on compact intervals, then the principal value integral is

$$\text{p.v.}\int_{-\infty}^{\infty} f = \lim_{r\to\infty}\int_{-r}^{r} f.$$

If the improper integral exists, then the principal value integral exists and is equal to the improper integral. As Example 1.80 shows, the principal value integral may exist even if the improper integral does not. Of course, a principal value integral may also diverge.

Example 1.82. Consider the principal value integral

$$\begin{aligned}\text{p.v.}\int_{-1}^{1}\frac{1}{x^2}\,dx &= \lim_{\epsilon\to 0^+}\left(\int_{-1}^{-\epsilon}\frac{1}{x^2}\,dx + \int_{\epsilon}^{1}\frac{1}{x^2}\,dx\right) \\ &= \lim_{\epsilon\to 0^+}\left(\frac{2}{\epsilon} - 2\right) = \infty.\end{aligned}$$

In this case, the function $1/x^2$ is positive and approaches ∞ on both sides of the singularity at $x = 0$, so there is no cancelation and the principal value integral diverges to ∞.

Principal value integrals arise frequently in complex analysis, harmonic analysis, and a variety of applications.

Example 1.83. Consider the exponential integral function Ei given in Example 1.60,

$$\mathrm{Ei}(x) = \int_{-\infty}^{x} \frac{e^t}{t}\,dt.$$

If $x < 0$, the integrand is continuous for $-\infty < t \leq x$, and the integral is interpreted as an improper integral,

$$\int_{\infty}^{x} \frac{e^t}{t}\,dt = \lim_{r\to\infty} \int_{-r}^{x} \frac{e^t}{t}\,dt.$$

This improper integral converges absolutely by comparison with e^t, since

$$\left|\frac{e^t}{t}\right| \leq e^t \qquad \text{for } -\infty < t \leq -1,$$

and

$$\int_{-\infty}^{-1} e^t\,dt = \lim_{r\to\infty} \int_{-r}^{-1} e^t\,dt = \frac{1}{e}.$$

If $x > 0$, then the integrand has a non-integrable singularity at $t = 0$, and we interpret it as a principal value integral. We write

$$\int_{-\infty}^{x} \frac{e^t}{t}\,dt = \int_{-\infty}^{-1} \frac{e^t}{t}\,dt + \int_{-1}^{x} \frac{e^t}{t}\,dt.$$

The first integral is interpreted as an improper integral as before. The second integral is interpreted as a principal value integral

$$\text{p.v.} \int_{-1}^{x} \frac{e^t}{t}\,dt = \lim_{\epsilon\to 0^+} \left(\int_{-1}^{-\epsilon} \frac{e^t}{t}\,dt + \int_{\epsilon}^{x} \frac{e^t}{t}\,dt \right).$$

This principal value integral converges, since

$$\text{p.v.} \int_{-1}^{x} \frac{e^t}{t}\,dt = \int_{-1}^{x} \frac{e^t - 1}{t}\,dt + \text{p.v.} \int_{-1}^{x} \frac{1}{t}\,dt = \int_{-1}^{x} \frac{e^t - 1}{t}\,dt + \ln x.$$

The first integral makes sense as a Riemann integral since the integrand has a removable singularity at $t = 0$, with

$$\lim_{t\to 0} \left(\frac{e^t - 1}{t} \right) = 1,$$

so it extends to a continuous function on $[-1, x]$.

Finally, if $x = 0$, then the integrand is unbounded at the left endpoint $t = 0$. The corresponding improper or principal value integral diverges, and $\mathrm{Ei}(0)$ is undefined.

Example 1.84. Let $f : \mathbb{R} \to \mathbb{R}$ and assume, for simplicity, that f has compact support, meaning that $f = 0$ outside a compact interval $[-r, r]$. If f is integrable, we define the Hilbert transform $Hf : \mathbb{R} \to \mathbb{R}$ of f by the principal value integral

$$Hf(x) = \frac{1}{\pi}\text{p.v.} \int_{-\infty}^{\infty} \frac{f(t)}{x - t}\,dt = \frac{1}{\pi} \lim_{\epsilon\to 0^+} \left(\int_{-\infty}^{x-\epsilon} \frac{f(t)}{x - t}\,dt + \int_{x+\epsilon}^{\infty} \frac{f(t)}{x - t}\,dt \right).$$

Here, x plays the role of a parameter in the integral with respect to t. We use a principal value because the integrand may have a non-integrable singularity at $t = x$. Since f has compact support, the intervals of integration are bounded and there is no issue with the convergence of the integrals at infinity.

For example, suppose that f is the step function

$$f(x) = \begin{cases} 1 & \text{for } 0 \le x \le 1, \\ 0 & \text{for } x < 0 \text{ or } x > 1. \end{cases}$$

If $x < 0$ or $x > 1$, then $t \neq x$ for $0 \le t \le 1$, and we get a proper Riemann integral

$$Hf(x) = \frac{1}{\pi} \int_0^1 \frac{1}{x - t}\, dt = \frac{1}{\pi} \ln \left| \frac{x}{x-1} \right|.$$

If $0 < x < 1$, then we get a principal value integral

$$\begin{aligned} Hf(x) &= \frac{1}{\pi} \lim_{\epsilon \to 0^+} \left(\int_0^{x-\epsilon} \frac{1}{x-t}\, dt + \frac{1}{\pi} \int_{x+\epsilon}^1 \frac{1}{x-t}\, dt \right) \\ &= \frac{1}{\pi} \lim_{\epsilon \to 0^+} \left[\ln\left(\frac{x}{\epsilon}\right) + \ln\left(\frac{\epsilon}{1-x}\right) \right] \\ &= \frac{1}{\pi} \ln\left(\frac{x}{1-x}\right) \end{aligned}$$

Thus, for $x \neq 0, 1$ we have

$$Hf(x) = \frac{1}{\pi} \ln \left| \frac{x}{x-1} \right|.$$

The principal value integral with respect to t diverges if $x = 0, 1$ because $f(t)$ has a jump discontinuity at the point where $t = x$. Consequently the values $Hf(0)$, $Hf(1)$ of the Hilbert transform of the step function are undefined.

1.11. Riemann sums

An alternative way to define the Riemann integral is in terms of the convergence of Riemann sums. This was, in fact, Riemann's original definition, which he gave in 1854 in his *Habilitationsschrift* (a kind of post-doctoral dissertation required of German academics), building on previous work of Cauchy who defined the integral for continuous functions.

It is interesting to note that the topic of Riemann's *Habilitationsschrift* was not integration theory, but Fourier series. Riemann introduced an analytical definition of the integral along the way so that he could state his results more precisely. In fact, almost all of the fundamental developments of rigorous real analysis in the nineteenth century were motivated by problems related to Fourier series and their convergence.

Upper and lower sums were introduced by Darboux, and they simplify the theory. We won't use Riemann sums here, but we will explain the equivalence of the definitions. We'll say, temporarily, that a function is Darboux integrable if it satisfies Definition 1.3.

To give Riemann's definition, we define a tagged partition (P, C) of a compact interval $[a, b]$ to be a partition

$$P = \{I_1, I_2, \dots, I_n\}$$

of the interval together with a set

$$C = \{c_1, c_2, \dots, c_n\}$$

of points such that $c_k \in I_k$ for $k = 1, \dots, n$. (Think of c_k as a "tag" attached to I_k.)

If $f : [a, b] \to \mathbb{R}$, then we define the Riemann sum of f with respect to the tagged partition (P, C) by

$$S(f; P, C) = \sum_{k=1}^{n} f(c_k)|I_k|.$$

That is, instead of using the supremum or infimum of f on the kth interval in the sum, we evaluate f at an arbitrary point in the interval. Roughly speaking, a function is Riemann integrable if its Riemann sums approach the same value as the partition is refined, independently of how we choose the points $c_k \in I_k$.

As a measure of the refinement of a partition $P = \{I_1, I_2, \dots, I_n\}$, we define the mesh (or norm) of P to be the maximum length of its intervals,

$$\mathrm{mesh}(P) = \max_{1 \le k \le n} |I_k| = \max_{1 \le k \le n} |x_k - x_{k-1}|.$$

Definition 1.85. A bounded function $f : [a, b] \to \mathbb{R}$ is Riemann integrable on $[a, b]$ if there exists a number $R \in \mathbb{R}$ with the following property: For every $\epsilon > 0$ there is a $\delta > 0$ such that

$$|S(f; P, C) - R| < \epsilon$$

for every tagged partition (P, C) of $[a, b]$ with $\mathrm{mesh}(P) < \delta$. In that case, $R = \int_a^b f$ is the Riemann integral of f on $[a, b]$.

Note that

$$L(f; P) \le S(f; P, C) \le U(f; P),$$

so the Riemann sums are "squeezed" between the upper and lower sums. The following theorem shows that the Darboux and Riemann definitions lead to the same notion of the integral, so it's a matter of convenience which definition we adopt as our starting point.

Theorem 1.86. A function is Riemann integrable (in the sense of Definition 1.85) if and only if it is Darboux integrable (in the sense of Definition 1.3). In that case, the Riemann and Darboux integrals of the function are equal.

Proof. First, suppose that $f : [a, b] \to \mathbb{R}$ is Riemann integrable with integral R. Then f must be bounded; otherwise f would be unbounded in some interval I_k of every partition P, and we could make its Riemann sums with respect to P arbitrarily large by choosing a suitable point $c_k \in I_k$, contradicting the definition of R.

Let $\epsilon > 0$. There is a partition $P = \{I_1, I_2, \dots, I_n\}$ of $[a, b]$ such that

$$|S(f; P, C) - R| < \frac{\epsilon}{2}$$

for every set of points $C = \{c_k \in I_k : k = 1, \dots, n\}$. If $M_k = \sup_{I_k} f$, then there exists $c_k \in I_k$ such that

$$M_k - \frac{\epsilon}{2(b-a)} < f(c_k).$$

It follows that

$$\sum_{k=1}^{n} M_k |I_k| - \frac{\epsilon}{2} < \sum_{k=1}^{n} f(c_k)|I_k|,$$

meaning that $U(f;P) - \epsilon/2 < S(f;P,C)$. Since $S(f;P,C) < R + \epsilon/2$, we get that

$$U(f) \le U(f;P) < R + \epsilon.$$

Similarly, if $m_k = \inf_{I_k} f$, then there exists $c_k \in I_k$ such that

$$m_k + \frac{\epsilon}{2(b-a)} > f(c_k), \qquad \sum_{k=1}^{n} m_k |I_k| + \frac{\epsilon}{2} > \sum_{k=1}^{n} f(c_k)|I_k|,$$

and $L(f;P) + \epsilon/2 > S(f;P,C)$. Since $S(f;P,C) > R - \epsilon/2$, we get that

$$L(f) \ge L(f;P) > R - \epsilon.$$

These inequalities imply that

$$L(f) + \epsilon > R > U(f) - \epsilon$$

for every $\epsilon > 0$, and therefore $L(f) \ge R \ge U(f)$. Since $L(f) \le U(f)$, we conclude that $L(f) = R = U(f)$, so f is Darboux integrable with integral R.

Conversely, suppose that f is Darboux integrable. The main point is to show that if $\epsilon > 0$, then $U(f;P) - L(f;P) < \epsilon$ not just for some partition but for every partition whose mesh is sufficiently small.

Let $\epsilon > 0$ be given. Since f is Darboux integrable. there exists a partition Q such that

$$U(f;Q) - L(f;Q) < \frac{\epsilon}{4}.$$

Suppose that Q contains m intervals and $|f| \le M$ on $[a,b]$. We claim that if

$$\delta = \frac{\epsilon}{8mM},$$

then $U(f;P) - L(f;P) < \epsilon$ for every partition P with $\operatorname{mesh}(P) < \delta$.

To prove this claim, suppose that $P = \{I_1, I_2, \dots, I_n\}$ is a partition with $\operatorname{mesh}(P) < \delta$. Let P' be the smallest common refinement of P and Q, so that the endpoints of P' consist of the endpoints of P or Q. Since a, b are common endpoints of P and Q, there are at most $m-1$ endpoints of Q that are distinct from endpoints of P. Therefore, at most $m-1$ intervals in P contain additional endpoints of Q and are strictly refined in P', meaning that they are the union of two or more intervals in P'.

Now consider $U(f;P) - U(f;P')$. The terms that correspond to the same, unrefined intervals in P and P' cancel. If I_k is a strictly refined interval in P, then the corresponding terms in each of the sums $U(f;P)$ and $U(f;P')$ can be estimated by $M|I_k|$ and their difference by $2M|I_k|$. There are at most $m-1$ such intervals and $|I_k| < \delta$, so it follows that

$$U(f;P) - U(f;P') < 2(m-1)M\delta < \frac{\epsilon}{4}.$$

Since P' is a refinement of Q, we get

$$U(f;P) < U(f;P') + \frac{\epsilon}{4} \leq U(f;Q) + \frac{\epsilon}{4} < L(f;Q) + \frac{\epsilon}{2}.$$

It follows by a similar argument that

$$L(f;P') - L(f;P) < \frac{\epsilon}{4},$$

and

$$L(f;P) > L(f;P') - \frac{\epsilon}{4} \geq L(f;Q) - \frac{\epsilon}{4} > U(f;Q) - \frac{\epsilon}{2}.$$

Since $L(f;Q) \leq U(f;Q)$, we conclude from these inequalities that

$$U(f;P) - L(f;P) < \epsilon$$

for every partition P with $\operatorname{mesh}(P) < \delta$.

If D denotes the Darboux integral of f, then we have

$$L(f;P) \leq D \leq U(f,P), \qquad L(f;P) \leq S(f;P,C) \leq U(f;P).$$

Since $U(f;P) - L(f;P) < \epsilon$ for every partition P with $\operatorname{mesh}(P) < \delta$, it follows that

$$|S(f;P,C) - D| < \epsilon.$$

Thus, f is Riemann integrable with Riemann integral D. □

Finally, we give a necessary and sufficient condition for Riemann integrability that was proved by Riemann himself (1854). To state the condition, we introduce some notation.

Let $f;[a,b] \to \mathbb{R}$ be a bounded function. If $P = \{I_1, I_2, \dots, I_n\}$ is a partition of $[a,b]$ and $\epsilon > 0$, let $A_\epsilon(P) \subset \{1, \dots, n\}$ be the set of indices k such that

$$\operatorname*{osc}_{I_k} f = \sup_{I_k} f - \inf_{I_k} f \geq \epsilon \qquad \text{for } k \in A_\epsilon(P).$$

Similarly, let $B_\epsilon(P) \subset \{1, \dots, n\}$ be the set of indices such that

$$\operatorname*{osc}_{I_k} f < \epsilon \qquad \text{for } k \in B_\epsilon(P).$$

That is, the oscillation of f on I_k is "large" if $k \in A_\epsilon(P)$ and "small" if $k \in B_\epsilon(P)$. We denote the sum of the lengths of the intervals in P where the oscillation of f is "large" by

$$s_\epsilon(P) = \sum_{k \in A_\epsilon(P)} |I_k|.$$

Fixing $\epsilon > 0$, we say that $s_\epsilon(P) \to 0$ as $\operatorname{mesh}(P) \to 0$ if for every $\eta > 0$ there exists $\delta > 0$ such that $\operatorname{mesh}(P) < \delta$ implies that $s_\epsilon(P) < \eta$.

Theorem 1.87. A bounded function is Riemann integrable if and only if $s_\epsilon(P) \to 0$ as $\operatorname{mesh}(P) \to 0$ for every $\epsilon > 0$.

Proof. Let $f : [a,b] \to \mathbb{R}$ be bounded with $|f| \leq M$ on $[a,b]$ for some $M > 0$.

First, suppose that the condition holds, and let $\epsilon > 0$. If P is a partition of $[a,b]$, then, using the notation above for $A_\epsilon(P)$, $B_\epsilon(P)$ and the inequality

$$0 \leq \operatorname*{osc}_{I_k} f \leq 2M,$$

we get that

$$\begin{aligned} U(f;P) - L(f;P) &= \sum_{k=1}^{n} \operatorname*{osc}_{I_k} f \cdot |I_k| \\ &= \sum_{k \in A_\epsilon(P)} \operatorname*{osc}_{I_k} f \cdot |I_k| + \sum_{k \in B_\epsilon(P)} \operatorname*{osc}_{I_k} f \cdot |I_k| \\ &\le 2M \sum_{k \in A_\epsilon(P)} |I_k| + \epsilon \sum_{k \in B_\epsilon(P)} |I_k| \\ &\le 2M s_\epsilon(P) + \epsilon(b-a). \end{aligned}$$

By assumption, there exists $\delta > 0$ such that $s_\epsilon(P) < \epsilon$ if $\operatorname{mesh}(P) < \delta$, in which case

$$U(f;P) - L(f;P) < \epsilon(2M + b - a).$$

The Cauchy criterion in Theorem 1.14 then implies that f is integrable.

Conversely, suppose that f is integrable, and let $\epsilon > 0$ be given. If P is a partition, we can bound $s_\epsilon(P)$ from above by the difference between the upper and lower sums as follows:

$$U(f;P) - L(f;P) \ge \sum_{k \in A_\epsilon(P)} \operatorname*{osc}_{I_k} f \cdot |I_k| \ge \epsilon \sum_{k \in A_\epsilon(P)} |I_k| = \epsilon s_\epsilon(P).$$

Since f is integrable, for every $\eta > 0$ there exists $\delta > 0$ such that $\operatorname{mesh}(P) < \delta$ implies that

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

Therefore, $\operatorname{mesh}(P) < \delta$ implies that

$$s_\epsilon(P) \le \frac{1}{\epsilon}\left[U(f;P) - L(f;P)\right] < \eta,$$

which proves the result. $\square$

This theorem has the drawback that the necessary and sufficient condition for Riemann integrability is somewhat complicated and, in general, it isn't easy to verify. In the next section, we state a simpler necessary and sufficient condition for Riemann integrability.

1.12. The Lebesgue criterion for Riemann integrability

Although the Dirichlet function in Example 1.7 is not Riemann integrable, it is Lebesgue integrable. Its Lebesgue integral is given by

$$\int_0^1 f = 1 \cdot |A| + 0 \cdot |B|$$

where $A = [0,1] \cap \mathbb{Q}$ is the set of rational numbers in $[0,1]$, $B = [0,1] \setminus \mathbb{Q}$ is the set of irrational numbers, and $|E|$ denotes the Lebesgue measure of a set E. The Lebesgue measure of a set is a generalization of the length of an interval which applies to more general sets. It turns out that $|A| = 0$ (as is true for any countable set of real numbers — see Example 1.89 below) and $|B| = 1$. Thus, the Lebesgue integral of the Dirichlet function is 0.

A necessary and sufficient condition for Riemann integrability can be given in terms of Lebesgue measure. We will state this condition without proof, beginning with a criterion for a set to have Lebesgue measure zero.

Theorem 1.88. A set $E \subset \mathbb{R}$ has Lebesgue measure zero if and only if for every $\epsilon > 0$ there is a countable collection of open intervals $\{(a_k, b_k) : k \in \mathbb{N}\}$ such that

$$E \subset \bigcup_{k=1}^{\infty}(a_k, b_k), \qquad \sum_{k=1}^{\infty}(b_k - a_k) < \epsilon.$$

The open intervals is this theorem are not required to be disjoint, and they may "overlap."

Example 1.89. Every countable set $E = \{x_k \in \mathbb{R} : k \in \mathbb{N}\}$ has Lebesgue measure zero. To prove this, let $\epsilon > 0$ and for each $k \in \mathbb{N}$ define

$$a_k = x_k - \frac{\epsilon}{2^{k+2}}, \qquad b_k = x_k + \frac{\epsilon}{2^{k+2}}.$$

Then $E \subset \bigcup_{k=1}^{\infty}(a_k, b_k)$ since $x_k \in (a_k, b_k)$ and

$$\sum_{k=1}^{\infty}(b_k - a_k) = \sum_{k=1}^{\infty}\frac{\epsilon}{2^{k+1}} = \frac{\epsilon}{2} < \epsilon,$$

so the Lebesgue measure of E is equal to zero.

If $E = [0, 1] \cap \mathbb{Q}$ consists of the rational numbers in $[0, 1]$, then the set $G = \bigcup_{k=1}^{\infty}(a_k, b_k)$ described above encloses the dense set of rationals in a collection of open intervals the sum of whose lengths is arbitrarily small. This isn't so easy to visualize. Roughly speaking, if ϵ is small and we look at a section of $[0, 1]$ at a given magnification, then we see a few of the longer intervals in G with relatively large gaps between them. Magnifying one of these gaps, we see a few more intervals with large gaps between them, magnifying those gaps, we see a few more intervals, and so on. Thus, the set G has a fractal structure, meaning that it looks similar at all scales of magnification.

We then have the following result, due to Lebesgue.

Theorem 1.90. A bounded function on a compact interval is Riemann integrable if and only if the set of points at which it is discontinuous has Lebesgue measure zero.

For example, the set of discontinuities of the Riemann-integrable function in Example 1.6 consists of a single point $\{0\}$, which has Lebesgue measure zero. On the other hand, the set of discontinuities of the non-Riemann-integrable Dirichlet function in Example 1.7 is the entire interval $[0, 1]$, and its set of discontinuities has Lebesgue measure one.

Theorem 1.90 implies that every bounded function with a countable set of discontinuities is Riemann integrable, since such a set has Lebesgue measure zero. A special case of this result is Theorem 1.33 that every bounded function with finitely many discontinuities is Riemann integrable. The monotonic function in Example 1.22 is an explicit example of a Riemann integrable function with a dense, countably infinite set of discontinuities. A set doesn't have to be countable to have Lebesgue measure zero, and there are many uncountable sets whose Lebesgue measure is zero.

Example 1.91. The standard "middle-thirds" Cantor set $K \subset [0, 1]$ is an uncountable set with Lebesgue measure zero. The characteristic function $f : [0, 1] \to \mathbb{R}$ of K, defined by

$$f(x) = \begin{cases} 1 & \text{if } x \in K \\ 0 & \text{if } x \in [0, 1] \setminus K, \end{cases}$$

has K as its set of discontinuities. Therefore, f is Riemann integrable on $[0, 1]$, with integral zero, even though it is discontinuous at uncountably many points.