University of Birmingham
School of Mathematics

Real Analysis – Integration – Spring 2025

Lecture Notes
by Andrew Morris

CONTENTS

WEEK 1: THE RIEMANN–DARBOUX INTEGRAL (PRELIMINARIES)

The first two weeks we develop the Riemann–Darboux theory of integration to define precisely what is meant by an integrable function and its integral. The aim is to develop a theory of integration that is not limited to continuous functions. This will lead us to fundamental questions about the construction of the real numbers.

1.0. **Introduction and motivation.** The definitions of continuity and differentiability for functions are motivated by geometric properties of their graphs. These are, respectively, the connectivity of the graph and the existence of tangent lines on the graph. The definition of integrability is likewise motivated by the geometric notion of the area bound by the graph. The following example serves as a prototype for our construction of the Riemann–Darboux integral.

Example 1.0.1 (Prototype). Suppose that $f : [0,1] \to \mathbb{R}$ is defined by $f(x) := x$. We can use rectangles to approximate the area bound by its graph and the x-axis. To begin, consider dividing the domain $[0,1]$ into three non-overlapping subintervals of equal width

$$[0,1] = [0, x_1] \cup [x_1, x_2] \cup [x_2, 1].$$

It seems convenient to label $x_0 = 0$ and $x_3 = 1$. The quantities $(x_1 - x_0)$, $(x_2 - x_1)$ and $(x_3 - x_2)$ will provide the widths of our rectangles but we have many sensible choices for the heights of our rectangles. For example, on the subinterval $[x_0, x_1]$, we could choose $f(x_0)$, $f(x_1)$ or any $f(y)$ with $y \in [x_0, x_1]$. This was Riemann's original approach but we shall consider a streamlined version developed later by Darboux. The innovation in this approach is to avoid using function values at specific locations by using the extreme values of the function on each subinterval. In our example, these are expressed as follows:

$$m_1 := \min\{f(x) : x \in [x_0, x_1]\} \qquad M_1 := \max\{f(x) : x \in [x_0, x_1]\}$$
$$m_2 := \min\{f(x) : x \in [x_1, x_2]\} \qquad M_2 := \max\{f(x) : x \in [x_1, x_2]\}$$
$$m_3 := \min\{f(x) : x \in [x_2, x_3]\} \qquad M_3 := \max\{f(x) : x \in [x_2, x_3]\}$$

We then consider just two types of approximations for the area bound by the graph. The first uses the minimum values for the heights of the rectangles, which thus underestimates the area as

$$s_3 := m_1(x_1 - x_0) + m_2(x_2 - x_1) + m_3(x_3 - x_2) \leq \text{Area}.$$

The second uses the maximum values for the heights of the rectangles, which thus overestimates the area as

$$S_3 := M_1(x_1 - x_0) + M_2(x_2 - x_1) + M_3(x_3 - x_2) \geq \text{Area}.$$

These sums are labelled as s_3 and S_3 to denote the fact that they are based on dividing $[0,1]$ into three subintervals of equal width. For each $n \in \mathbb{N}$, we could follow the same procedure using n subintervals to obtain an estimate

$$s_n \leq \text{Area} \leq S_n$$

in the hope that s_n and S_n tend to a common value as n tends to infinity. It is this value that we would like to define as the area bound by the graph of the function.

This example raises many issues that we will resolve this week, principally the precise meanings of the limits of s_n and S_n, and the possibility that s_n and S_n might tend to different values. Moreover, by considering a specific function ($f(x) = x$) and dividing its domain into pieces of equal width, we have unwittingly limited the scope of its applicability to more general functions. In particular, if f is not a continuous function, then the formulas for s_n and S_n may not be well-defined, as the following example shows.

Example 1.0.2. Suppose that $f : [0, 1] \to [0, 1]$ is defined by $f(x) := \begin{cases} 1, & x = 0 \\ x, & x \neq 0 \end{cases}$.

Using the notation in Example 1.0.1, we find

$$m_1 := \min\{f(x) : x \in [0, x_1]\} = \min(\{1\} \cup \{x : x \in (0, x_1]\}) = \min(0, x_1]$$

but the minimum of the interval $(0, x_1]$ does not exist because $0 \notin (0, x_1]$. This is because the minimum of a set *must be an element of the set* which is less than all other elements in the set (see Definition 1.1.2 below).

We can develop a theory of integration that is not limited to continuous functions, however, if we begin by replacing the minimum and maximum in Example 1.0.1 with the infimum and supremum.

1.1. **Supremum/Infimum.** We begin our construction of the Riemann–Darboux integral with an analysis of the supremum and infimum for subsets of the real line. To this end, let us introduce the following definition.

Definition 1.1.1. A set $X \subseteq \mathbb{R}$ is called *bounded above* if there exists $M \in \mathbb{R}$ such that $x \leq M$ for all $x \in X$. If $X \subseteq \mathbb{R}$ is bounded above and not empty, then the following terms apply:

(1) An *upper bound* for X is any $M \in \mathbb{R}$ such that $x \leq M$ for all $x \in X$.
(2) The *maximum* of X, denoted $\max X$, is an upper bound for X that is in X.
(3) The *supremum* of X, denoted $\sup X$, is the least upper bound for X, that is:
    (a) $\sup X$ is an upper bound for X;
    (b) $\sup X \leq M$ for all upper bounds M for X.

If $X \subseteq \mathbb{R}$ is not bounded above, then $\sup X := \infty$, whilst $\sup \emptyset := -\infty$ for the empty set $\emptyset$.

It is equivalent in (3) to write $\sup X := \min\{M : M \text{ is an upper bound for } X\}$. There is also the following analogue of these definitions based on lower bounds.

Definition 1.1.2. A set $X \subseteq \mathbb{R}$ is *bounded below* if there exists $m \in \mathbb{R}$ such that $x \geq m$ for all $x \in X$. If $X \subseteq \mathbb{R}$ is bounded below and not empty, then a *lower bound* for X is any such number m; the *minimum* of X, denoted $\min X$, is a lower bound that is in X; and the *infimum* of X, denoted $\inf X$, is the greatest lower bound for X. If $X \subseteq \mathbb{R}$ is not bounded below, then $\inf X := -\infty$, whilst $\inf \emptyset := \infty$.

Let us also highlight the definition of a bounded subset of $\mathbb{R}$.

Definition 1.1.3. A set $X \subseteq \mathbb{R}$ is *bounded* if it is bounded above and bounded below, so there exist real numbers m and M such that $m \leq x \leq M$ for all $x \in X$.

The next lemma provides a useful 'ϵ-type' characterisation for the supremum. It is a good exercise to try to prove an analogous result for the infimum.

Lemma 1.1.4. Suppose that $N \in \mathbb{R}$ is an upper bound for a nonempty set $X \subseteq \mathbb{R}$. The following properties are equivalent:

    (1) $N = \sup X$;
    (2) For each $\epsilon > 0$, there exists x in X such that $N - \epsilon < x$.

Proof. Suppose that $N \in \mathbb{R}$ is an upper bound for a nonempty set $X \subseteq \mathbb{R}$. The equivalence requires us to prove that (1) holds if and only if (2) holds:

First, we prove that (1) implies (2). Assume that (1) holds, so $N = \sup X$. Suppose that $\epsilon > 0$ and observe that $N - \epsilon$ is not an upper bound for X, because $N - \epsilon < N$ and N is the *least* upper bound for X. Now, since $N - \epsilon$ is not an upper bound for X, there exists x in X such that $N - \epsilon < x$, as required.

The proof that (2) implies (1) is posed as a question on the problem sheet. $\quad\square$

Here are some examples that use an 'ϵ-type' approach to find the supremum and infimum of sets of real numbers.

Example 1.1.5. Find, with proof, the supremum and infimum of these sets:

(1) $A = \{x : x \le 0 \text{ and } x^2 + x - 1 < 0\}$

(2) $B = \{\frac{n+1}{n} : n \in \mathbb{N}\}$

Solution. (1) If $x^2 + x - 1 < 0$, then completing the square gives $(x + \frac{1}{2})^2 - \frac{5}{4} < 0$ and $|x + \frac{1}{2}| < \frac{\sqrt{5}}{2}$ (note $\sqrt{x^2} = |x|$). We thus have $A = (-\frac{1}{2} - \frac{\sqrt{5}}{2}, 0]$ and so we claim that $\inf A = -\frac{1}{2} - \frac{\sqrt{5}}{2} =: a$ and $\sup A = 0$. The interval expression $A = (a, 0]$ shows that a is a lower bound for A, and 0 is an upper bound for A. It remains to prove that a is the *greatest* such lower bound, and 0 is the *least* such upper bound. To do this let $\epsilon > 0$, and observe that $a + \epsilon$ is never a lower bound for A (since either $a + \frac{\epsilon}{2} \in A$ or $a + \epsilon > 0 \in A$), and that $-\epsilon$ is never an upper bound for A (since $-\epsilon < 0 \in A$).

(2) The elements of the set $B = \{1 + \frac{1}{n} : n \in \mathbb{N}\} = \{2, 1\frac{1}{2}, 1\frac{1}{3}, \ldots\}$ form a decreasing sequence with $\lim_{n \to \infty}(1 + \frac{1}{n}) = 1$. This shows that 1 is a lower bound for B, and 2 is an upper bound for B. Now let $\epsilon > 0$, and observe that $1 + \epsilon$ is never a lower bound for B, since we can always find $n \in \mathbb{N}$ such that $1 + \frac{1}{n} < 1 + \epsilon$ (choose $n \in \mathbb{N}$ with $n > \frac{1}{\epsilon}$), hence $\inf B = 1$. Finally, observe that $2 - \epsilon$ is never an upper bound for B, since $2 \in B$, hence $\sup B = 2$.

Our discussion so far has purposefully neglected the question of whether the supremum and infimum of a set of real numbers exists. This is, in actual fact, the quintessential property of the real numbers. The real numbers can be constructed rigorously as a mathematical object known as a *complete ordered field*. It can be proved that there exists a unique complete ordered field, so the real numbers can be understood without ambiguity, but we don't have time to present this lengthy and technical construction here.

As a complete ordered field, however, we can understand the real numbers as a set on which certain properties hold. These properties are so fundamental that they are called axioms (it is a good exercise to look up these axioms) and these fall into three categories. The first are algebraic axioms, which say that $\mathbb{R}$ is a set on which appropriate operations of addition ($+$) and multiplication ($\cdot$) are defined. This means that $\mathbb{R}$ is a field. The second are ordering axioms, which say that there is a certain order ($<$) between real numbers. This means that $\mathbb{R}$ is an ordered field.

The rational numbers also form an ordered field, however, and it is only the final Completeness Axiom below that distinguishes $\mathbb{Q}$ from $\mathbb{R}$. This (perhaps) obvious property is all that is required to provide a mathematically rigorous foundation for our intuition that the real numbers 'complete' the rational numbers by 'filling in all the gaps' between them. For example, using this axiom it can be proved that there exists a real number whose square is equal to 2 (so $\sqrt{2} \in \mathbb{R}$), and that given any two rational numbers q_1 and q_2, there exists a real number r such that $q_1 < r < q_2$ (this is known as the Archimedean Property of $\mathbb{R}$).

Axiom 1.1.6 (Completeness Axiom of $\mathbb{R}$). If $X \subseteq \mathbb{R}$ is bounded above and not empty, then the supremum of X exists in $\mathbb{R}$, that is, $\sup X \in \mathbb{R}$.

It is simply a matter of historical choice that this axiom is stated for sets that are bounded above, rather than for sets that are bounded below. The Completeness Axiom implies the following analogue for sets that are bounded below. It is a good exercise to try to prove this result.

Corollary 1.1.7. If $X \subseteq \mathbb{R}$ is not empty and bounded below, then the infimum of X exists in $\mathbb{R}$.

1.2. **Upper/Lower sums.** We will initially be concerned with bounded functions. It is important to distinguish between the definition of a bounded *set* and the definition of a bounded *function* (compare the following definition with Definition 1.1.3).

Definition 1.2.1. A function $f : X \to \mathbb{R}$, with domain $X \subseteq \mathbb{R}$, is *bounded* if there exist real numbers m and M such that $m \leq f(x) \leq M$ for all $x \in X$.

It is also important to recognise that whether or not a function is bounded depends on both the rule defining it *and* its domain. In fact, all continuous functions $f : [a, b] \to \mathbb{R}$ are automatically bounded. This result is known as the Extreme Value Theorem but we do know pursue this here, since we want to develop a theory of integration for functions that may not be continuous.

Example 1.2.2. The following functions are defined by the same rule on different domains. Determine which are bounded functions:

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

(2) $g : [\frac{1}{2}, 1] \to \mathbb{R}$, $f(x) := \frac{1}{x}$

(3) $h : [0, 1] \to \mathbb{R}$, $f(x) := \begin{cases} \frac{1}{x}, & x \in (0, 1] \\ 1, & x = 0 \end{cases}$

Solution. The functions f and h are not bounded, since for each $M \geq 1$, it holds that $f(\frac{1}{2M}) = h(\frac{1}{2M}) = 2M > M$. (In other words, for each $M \geq 1$, there exists x in the domain of f and h such that $f(x) = h(x) > M$, so f and h are not bounded.) The function g is bounded, since $1 \leq \frac{1}{x} \leq 2$ for all $x \in [\frac{1}{2}, 1]$. $\qquad \square$

We now begin our construction of the integral in earnest by restricting attention to $[0, \infty)$-valued bounded functions $f : [a, b] \to [0, \infty)$, where $-\infty < a < b < \infty$. The following convention will reduce the associated burden of notation.

Notation 1.2.3. We henceforth use $[a, b]$ to denote a non-empty, closed and bounded subinterval of $\mathbb{R}$, or more specifically, where $-\infty < a < b < \infty$.

Note that our restriction to non-negative-valued functions is not necessary here but we do so for the convenience of avoiding negative-values in computations of area. It will be a simple matter to extend our theory to $\mathbb{R}$-valued bounded functions in Lecture 8.4. We will also define 'improper integrals' on unbounded intervals and for unbounded functions in Lectures 10.3 and 10.4.

Our prototype for the construction of the Riemann–Darboux integral in Example 1.0.1 was based on dividing the interval $[a, b]$ into n subintervals of equal width. There is no reason, however, to limit our attention to such 'equal-width' divisions. The concept of a partition below allows for much more general divisions.

Definition 1.2.4. A *partition P of an interval* $[a, b]$ is a finite set such that

$$\{a, b\} \subseteq P \subseteq [a, b].$$

A partition can always be expressed in the form $P = \{x_0, x_1, \ldots, x_n\}$, where

$$a = x_0 < x_1 < \ldots < x_n = b,$$

for some $n \in \mathbb{N}$.

Our original 'equal-width' divisions are then a special type of partition.

Example 1.2.5. For each $n \in \mathbb{N}$, we can express the interval $[0, 1]$ as a union of n closed subintervals of equal width $\frac{1}{n}$ by writing

$$[0, 1] = [0, \tfrac{1}{n}] \cup [\tfrac{1}{n}, \tfrac{2}{n}] \cup [\tfrac{2}{n}, \tfrac{3}{n}] \cup \ldots \cup [\tfrac{(n-1)}{n}, 1].$$

The endpoints of these subintervals define the partition

$$P_n := \{0, \tfrac{1}{n}, \tfrac{2}{n}, \ldots, \tfrac{(n-1)}{n}, 1\} = \{\tfrac{i}{n} : i = 0, 1, \ldots, n\}.$$

This can be done for any (closed bounded) interval $[a, b]$ and the resulting partition is called the *partition of $[a, b]$ into n subintervals of equal width.*

We are now in a position to define sums, analogous to the prototypes s_n and S_n in Example 1.0.1, to approximate the area bounded by the graph of a bounded function $f : [a, b] \to [0, \infty)$ and the x-axis.

Definition 1.2.6. Suppose that $f : [a, b] \to [0, \infty)$ is a bounded function. Let $P = \{x_0, x_1, \ldots, x_n\}$ denote a partition of $[a, b]$. The *lower Riemann–Darboux sum* $L(f, P)$ is defined by

$$L(f, P) := \sum_{i=1}^{n} m_i(x_i - x_{i-1}), \text{ where } m_i := \inf\{f(x) : x \in [x_{i-1}, x_i]\}.$$

The *upper Riemann–Darboux sum* $U(f, P)$ is defined by

$$U(f, P) := \sum_{i=1}^{n} M_i(x_i - x_{i-1}), \text{ where } M_i := \sup\{f(x) : x \in [x_{i-1}, x_i]\}.$$

These are commonly referred to, respectively, as the *lower sum* and *upper sum*.

The following example shows how to compute these sums using the 'equal-width' partitions from Example 1.2.5.

Example 1.2.7. Let $f : [0, b] \to [0, \infty)$ be defined by $f(x) := x$.
  (1) For each $n \in \mathbb{N}$, calculate $L(f, P_n)$ and $U(f, P_n)$, where P_n is the partition of $[0, b]$ into n subintervals of equal width.
  (2) Calculate $\sup\{L(f, P_n) : n \in \mathbb{N}\}$ and $\inf\{U(f, P_n) : n \in \mathbb{N}\}$.

Solution. (1) We begin by using the ideas in Example 1.2.5 to write

$$P_n = \{0, \tfrac{b}{n}, \tfrac{2b}{n}, \ldots, \tfrac{(n-1)b}{n}, b\} = \{\tfrac{ib}{n} : i = 0, 1, \ldots, n\}.$$

We substitute $f(x) = x$ and $x_i = \frac{ib}{n}$ into Definition 1.2.6, beginning with

$$m_i = \inf\{x : x \in [\tfrac{(i-1)b}{n}, \tfrac{ib}{n}]\} = \inf[\tfrac{(i-1)b}{n}, \tfrac{ib}{n}] = \tfrac{(i-1)b}{n}$$
$$M_i = \sup\{x : x \in [\tfrac{(i-1)b}{n}, \tfrac{ib}{n}]\} = \sup[\tfrac{(i-1)b}{n}, \tfrac{ib}{n}] = \tfrac{ib}{n}$$

to obtain

$$L(f, P_n) = \sum_{i=1}^{n} m_i(x_i - x_{i-1}) = \sum_{i=1}^{n} \tfrac{(i-1)b}{n}\tfrac{b}{n} = \sum_{i=1}^{n}(i-1)\tfrac{b^2}{n^2} = \tfrac{(n-1)n}{2}\tfrac{b^2}{n^2} = \tfrac{n-1}{n}\tfrac{b^2}{2}$$

and

$$U(f, P_n) = \sum_{i=1}^{n} M_i(x_i - x_{i-1}) = \sum_{i=1}^{n} \tfrac{ib}{n}\tfrac{b}{n} = \sum_{i=1}^{n} i\tfrac{b^2}{n^2} = \tfrac{n(n+1)}{2}\tfrac{b^2}{n^2} = \tfrac{n+1}{n}\tfrac{b^2}{2},$$

where we used the formula for an arithmetic sum in the penultimate equalities.
  (2) Now, recalling Example 1.1.5, we calculate

$$\sup\{L(f, P_n) : n \in \mathbb{N}\} = \sup\{(1 - \tfrac{1}{n})\tfrac{b^2}{2} : n \in \mathbb{N}\} = \tfrac{b^2}{2}$$

and

$$\inf\{U(f, P_n) : n \in \mathbb{N}\} = \inf\{(1 + \tfrac{1}{n})\tfrac{b^2}{2} : n \in \mathbb{N}\} = \tfrac{b^2}{2}$$

as required.

The 'equal-width' partitions from Example 1.2.5 are less useful, however, for functions that have points of discontinuity. The following example introduces an alternative type of partition, often abbreviated P_δ, which can be used to zoom in on points of discontinuity by choosing $\delta > 0$ sufficiently small.

Example 1.2.8. Let $f : [0,2] \to [0,1]$ be defined by $f(x) := \begin{cases} 1, & x \neq 1 \\ 0, & x = 1 \end{cases}$.

(1) Calculate $L(f, P_\delta)$ and $U(f, P_\delta)$ for $P_\delta = \{0, 1-\delta, 1, 1+\delta, 2\}$ and $\delta \in (0,1)$.
(2) Calculate $\sup\{L(f, P_\delta) : \delta \in (0,1)\}$ and $\inf\{U(f, P_\delta) : \delta \in (0,1)\}$.

Solution. We adopt the notation from Definition 1.2.6 by writing

$$\{x_0, x_1, x_2, x_3, x_4\} = \{0, 1-\delta, 1, 1+\delta, 2\}.$$

We have $m_1 = 1$, $m_2 = 0$, $m_3 = 0$, $m_4 = 1$ whilst $M_1 = 1$, $M_2 = 1$, $M_3 = 1$, $M_4 = 1$, so

$$L(f, P_\delta) = \sum_{i=1}^{4} m_i(x_i - x_{i-1}) = 1(1-\delta) + 0(\delta) + 0(\delta) + 1(1-\delta) = 2 - 2\delta$$

and

$$U(f, P_\delta) = \sum_{i=1}^{4} M_i(x_i - x_{i-1}) = 1(1-\delta) + 1(\delta) + 1(\delta) + 1(1-\delta) = 2.$$

2) Now we calculate

$$\sup\{L(f, P_\delta) : \delta \in (0,1)\} = \sup\{2 - 2\delta : \delta \in (0,1)\} = \sup(1, 2) = 2$$

and

$$\inf\{U(f, P_\delta) : \delta \in (0,1)\} = \inf\{2 : \delta \in (0,1)\} = \inf\{2\} = 2$$

as required.

Week 2: The Riemann–Darboux Integral

This week we will complete the construction of the Riemann–Darboux Integral. A technical result, here called the Partition Lemma, shows that there is a certain ordering amongst the lower sums and upper sums for any bounded function. This allows us to define a lower integral and an upper integral for any bounded function. A function will be called integrable precisely when these two formative integrals are equal.

2.1. **The Partition Lemma.** We need to establish some properties about lower sums and upper sums before we can finally define the Riemann–Darboux integral. These are contained in the following useful result.

Lemma 2.1.1 (Partition Lemma). Suppose that $f : [a,b] \to [0,\infty)$ is a bounded function. The following properties hold for partitions P, Q, R and S of $[a,b]$:

(1) $L(f, P) \leq U(f, P)$
(2) $P \subseteq Q \implies \begin{cases} L(f, P) \leq L(f, Q) \\ U(f, P) \geq U(f, Q) \end{cases}$
(3) $L(f, R) \leq U(f, S)$

Proof. Let $P = \{x_0, x_1, \ldots, x_n\}$ with m_i and M_i as in Definitions 1.2.4 and 1.2.6.
(1) Observe that $m_i \leq M_i$, hence

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

(2) We will only prove the stated inequality for lower sums, since an analogous argument implies the reverse inequality for upper sums.

We first consider the special case when $Q = P \cup \{y\}$ for some $y \in (x_0, x_1)$. In that case $Q = \{x_0, y, x_1, \ldots, x_n\}$, so we introduce the notation

$$m_1 := \inf\{f(x) : x \in [x_0, x_1]\},$$
$$m_1' := \inf\{f(x) : x \in [x_0, y]\},$$
$$m_1'' := \inf\{f(x) : x \in [y, x_1]\}.$$

Observe that $m_1 \leq m_1'$ and $m_1 \leq m_1''$, hence

$$\begin{aligned} m_1(x_1 - x_0) &= m_1(x_1 - y + y - x_0) \\ &= m_1(y - x_0) + m_1(x_1 - y) \\ &\leq m_1'(y - x_0) + m_1''(x_1 - y). \end{aligned}$$

This implies the required estimate, since

$$L(f, P) = m_1(x_1 - x_0) + \sum_{i=2}^{n} m_i(x_i - x_{i-1})$$

$$\leq m_1'(y - x_0) + m_1''(x_1 - y) + \sum_{i=2}^{n} m_i(x_i - x_{i-1})$$

$$= L(f, Q).$$

To obtain the result in the general case when $P \subseteq Q$, we make two observations. First, the proof above can be adapted to the case where $y \in [x_i, x_{i-1}]$ for any $i \in \{1, \ldots, n\}$. Second, since partitions are finite sets and $P \subseteq Q$, it must hold that $Q = P \cup \{y_1\} \cup \{y_2\} \cup \ldots \cup \{y_N\}$ for some $\{y_1, \ldots, y_N\} \subseteq [a, b]$ and $N \in \mathbb{N}$. Therefore, iterating the former result a *finite* number of times (at most N) gives the result. (It is a good exercise to write out these details.)

(3) In this case R and S are arbitrary partitions of $[a, b]$, so (2) does not apply. Instead, since $R \subseteq R \cup S$ and $S \subseteq R \cup S$, where $R \cup S$ is a partition of $[a, b]$ (it is a good exercise to prove this), we may apply (2) to obtain

$$L(f, R) \leq L(f, R \cup S) \text{ and } U(f, S) \geq U(f, R \cup S).$$

Moreover, we know from (1) that $L(f, R \cup S) \leq U(f, R \cup S)$, so altogether

$$L(f, R) \leq L(f, R \cup S) \leq U(f, R \cup S) \leq U(f, S),$$

as required. $\square$

2.2. **Integrable functions.** The lower Riemann–Darboux sums and the upper Riemann–Darboux sums provide the following two candidates for the integral of a bounded function $f : [a, b] \to [0, \infty)$.

Definition 2.2.1. Suppose that $f : [a, b] \to [0, \infty)$ is a bounded function. The *lower integral* of f is defined by

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

Thee *upper integral* of f is defined by

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

These are also often abbreviated respectively as $\sup_P L(f, P)$ and $\inf_P U(f, P)$.

The Completeness Axiom of $\mathbb{R}$ guarantees that the lower integral and the upper integral of a bounded function $f : [a, b] \to [0, \infty)$ exist as real numbers. It is a good exercise to prove this. Let us now investigate whether these have to be equal.

Proposition 2.2.2. If $f : [a, b] \to [0, \infty)$ is a bounded function, then $\underline{\int_a^b} f \le \overline{\int_a^b} f$.

Proof. We know from the Partition Lemma (Lemma 2.1.1) that $L(f, R) \le U(f, S)$ for all partitions R and S of $[a, b]$. This means that $U(f, S)$ is an upper bound for the set $\{L(f, R) : R \text{ is a partition of } [a, b]\}$, so by the definition of the supremum

$$\sup_R L(f, R) \le U(f, S),$$

for all partitions S of $[a, b]$. This, in turn, means that the real number $\sup_R L(f, R)$ is a lower bound for the set $\{U(f, S) : S \text{ is a partition of } [a, b]\}$, so by the definition of the infimum

$$\sup_R L(f, R) \le \inf_S U(f, S).$$

This inequality is exactly $\underline{\int_a^b} f \le \overline{\int_a^b} f$ by the definition of the lower integral and the upper integral. $\qquad \square$

The above proposition shows that the lower integral is always less than or equal to the upper integral for a bounded function $f : [a, b] \to [0, \infty)$ but the reverse inequality is not always true (see Example 2.2.4 below). The resolution to this (perhaps) unfortunate reality is to say that a function is integrable precisely when the lower integral and the upper integral are equal, and then the integral of f is defined to be this common value.

Definition 2.2.3. A bounded function $f : [a, b] \to [0, \infty)$ is called *integrable* if $\underline{\int_a^b} f = \overline{\int_a^b} f$. If such f is integrable, then the *integral* of f is $\int_a^b f := \underline{\int_a^b} f = \overline{\int_a^b} f$.

Here is an example of a function that is not integrable according to this definition.

Example 2.2.4 (Non-integrable function)**.** The function $f : [a, b] \to [0, 1]$ with

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

satisfies $\underline{\int_a^b} f < \overline{\int_a^b} f$ (hence f is not integrable).

Proof. To prove this, let $P = \{x_0, x_1, \ldots, x_n\}$ denote a partition of $[a, b]$ with m_i and M_i as in Definitions 1.2.4 and 1.2.6. Observe that

$$m_i := \inf\{f(x) : x \in [x_{i-1}, x_i]\} = 0 \text{ and } M_i := \sup\{f(x) : x \in [x_{i-1}, x_i]\} = 1$$

since the non-empty interval $[x_{i-1}, x_i]$ contains at least one (in fact infinitely many) rational numbers and irrational numbers. Therefore, we have

$$L(f, P) = \sum_{i=1}^n 0(x_i - x_{i-1}) = 0 \text{ and } U(f, P) = \sum_{i=1}^n 1(x_i - x_{i-1}) = b - a,$$

hence

$$\underline{\int_a^b} f := \sup_P L(f, P) = \sup_P\{0\} = 0$$

whilst

$$\overline{\int_a^b} f := \inf_P U(f, P) = \inf\{b - a\} = b - a.$$

The result follows because $0 < b - a$ (recall Notation 1.2.3). $\qquad \square$

We can now prove that the functions in Examples 1.2.7 and 1.2.8 are integrable.

Example 2.2.5. Let $f : [0, b] \to [0, \infty)$ be defined by $f(x) := x$. Prove that f is integrable and compute $\int_0^b f$.

Solution. We continue on from Example 1.2.7, where we proved that

$$\sup\{L(f, P_n) : n \in \mathbb{N}\} = \frac{b^2}{2} \text{ and } \inf\{U(f, P_n) : n \in \mathbb{N}\} = \frac{b^2}{2}.$$

Now, as $\{L(f, P_n) : n \in \mathbb{N}\} \subseteq \{L(f, P) : P \text{ is a partition of } [0, b]\}$, we have

$$\underline{\int_0^b} f := \sup\{L(f, P) : P \text{ is a partition of } [0, b]\} \geq \sup\{L(f, P_n) : n \in \mathbb{N}\} = \frac{b^2}{2}.$$

Moreover, as $\{U(f, P_n) : n \in \mathbb{N}\} \subseteq \{U(f, P) : P \text{ is a partition of } [0, b]\}$, we have

$$\overline{\int_0^b} f := \inf\{U(f, P) : P \text{ is a partition of } [0, b]\} \leq \inf\{U(f, P_n) : n \in \mathbb{N}\} = \frac{b^2}{2}.$$

We also know from Proposition 2.2.2 that $\underline{\int_0^b} f \leq \overline{\int_0^b} f$, so altogether

$$\frac{b^2}{2} \leq \underline{\int_0^b} f \leq \overline{\int_0^b} f \leq \frac{b^2}{2}.$$

Notice that this chain of inequalities can only hold if $\frac{b^2}{2} = \underline{\int_0^b} f = \overline{\int_0^b} f = \frac{b^2}{2}$, hence f is integrable and $\int_0^b f = \frac{b^2}{2}$. $\qquad\square$

The following example is completely analogous to that above but its is included for completeness.

Example 2.2.6. Let $f : [0, 2] \to [0, 1]$ be defined by $f(x) := \begin{cases} 1, & x \neq 1 \\ 0, & x = 1 \end{cases}$. Prove that f is integrable and compute $\int_0^1 f$.

Solution. We continue on from Example 1.2.8, where we proved that

$$\sup\{L(f, P_\delta) : \delta \in (0, 1)\} = 2 \text{ and } \inf\{U(f, P_\delta) : \delta \in (0, 1)\} = 2.$$

Now, as $\{L(f, P_\delta) : \delta \in (0, 1)\} \subseteq \{L(f, P) : P \text{ is a partition of } [0, 2]\}$, we have

$$\underline{\int_0^2} f := \sup\{L(f, P) : P \text{ is a partition of } [0, 2]\} \geq \sup\{L(f, P_\delta) : \delta \in (0, 1)\} = 2.$$

Moreover, as $\{U(f, P_\delta) : \delta \in (0, 1)\} \subseteq \{U(f, P) : P \text{ is a partition of } [0, 2]\}$, we have

$$\overline{\int_0^2} f := \inf\{U(f, P) : P \text{ is a partition of } [0, 2]\} \leq \inf\{U(f, P_\delta) : \delta \in (0, 1)\} = 2.$$

We also know from Proposition 2.2.2 that $\underline{\int_0^2} f \leq \overline{\int_0^2} f$, so altogether

$$2 \leq \underline{\int_0^2} f \leq \overline{\int_0^2} f \leq 2.$$

Notice that this chain of inequalities can only hold if $2 = \underline{\int_0^2} f = \overline{\int_0^2} f = 2$, hence f is integrable and $\int_0^2 f = 2$. $\qquad\square$

Week 3: Riemann's Integrability Criterion

This week we prove Riemann's equivalent characterisation for integrability. This is an 'ϵ-type' characterisation which is often much easier to verify than the definition of integrability we encountered last week. In particular, we will use it to prove that monotonic functions and continuous functions on closed bounded intervals are always integrable.

3.1. **An equivalent definition for integrability.** The 'ϵ-type' characterisation for integrability proved below is known as Riemann's Criterion for Integrability.

Theorem 3.1.1 (Riemann's Criterion). The following properties are equivalent for any bounded function $f : [a, b] \to \mathbb{R}$:

(1) The function f is integrable.
(2) For each $\epsilon > 0$, there exists a partition P of $[a, b]$ such that
$$U(f, P) - L(f, P) < \epsilon.$$

Proof. The equivalence requires us to prove that (1) holds if and only if (2) holds:

First, we prove that (1) implies (2). Assume that (1) holds, so f is integrable with $\int_a^b f := \underline{\int_a^b f} = \overline{\int_a^b f}$. Now suppose that $\epsilon > 0$. The lower integral is the supremum $\underline{\int_a^b f} := \sup_P L(f, P)$, so the 'ϵ-type' characterisation of the supremum in Lemma 1.1.4 implies that there exists a partition P' of $[a, b]$ such that
$$\underline{\int_a^b f} - \tfrac{\epsilon}{2} \le L(f, P').$$

Moreover, the upper integral is the infimum $\overline{\int_a^b f} := \inf_P U(f, P)$, so there exists a partition P'' of $[a, b]$ such that
$$\overline{\int_a^b f} + \tfrac{\epsilon}{2} \ge U(f, P'').$$

We can combine the two estimates above by considering the partition $P' \cup P''$ and using the Partition Lemma (Lemma 2.1.1) to note that $L(f, P'') \le L(f, P' \cup P'')$ and $U(f, P') \ge U(f, P' \cup P'')$, since $P' \subseteq P' \cup P''$ and $P'' \subseteq P' \cup P''$. This gives
$$\underline{\int_a^b f} - \tfrac{\epsilon}{2} \le L(f, P' \cup P'') \quad \text{and} \quad \overline{\int_a^b f} + \tfrac{\epsilon}{2} \ge U(f, P' \cup P'').$$

Now, subtracting the first inequality from the second inequality gives
$$U(f, P' \cup P'') - L(f, P' \cup P'') \le \left(\overline{\int_a^b f} + \tfrac{\epsilon}{2}\right) - \left(\underline{\int_a^b f} - \tfrac{\epsilon}{2}\right) = \epsilon,$$

where we used the integrability assumption $\underline{\int_a^b f} = \overline{\int_a^b f}$ to obtain the final equality. This proves (2).

Second, we prove that (2) implies (1). Assume that (2) holds, and suppose that $\epsilon > 0$, so there exists a partition P_ϵ such that $U(f, P_\epsilon) - L(f, P_\epsilon) < \epsilon$. We use the subscript ϵ here to show the explicit dependence of the partition on ϵ. This is in contrast to the arbitrary partition P considered in the estimate
$$\underline{\int_a^b f} := \sup_P L(f, P) \ge L(f, P_\epsilon),$$

which holds because the supremum is also an upper bound, and the estimate
$$\overline{\int_a^b f} := \inf_P U(f, P) \le U(f, P_\epsilon)$$

which holds because the infimum is also a lower bound. We subtract the two estimates above to obtain
$$0 \le \overline{\int_a^b f} - \underline{\int_a^b f} \le U(f, P_\epsilon) - L(f, P_\epsilon) < \epsilon,$$

where the first inequality uses Proposition 2.2.2, and the last inequality is our assumption above. In summary, we have shown that

$$0 \leq \overline{\int_a^b} f - \underline{\int_a^b} f < \epsilon \quad \text{for all} \quad \epsilon > 0,$$

which is only possible when $\overline{\int_a^b} f - \underline{\int_a^b} f = 0$, hence f is integrable by Definition 2.2.3. This proves (1). $\qquad\square$

Let us verify Riemann's Criterion for the two key examples (Examples 1.2.7 and 1.2.8) from Lecture 7.2.

Example 3.1.2. Use Riemann's Criterion for Integrability to prove that each of the functions below are integrable

(1) $f : [0, b] \to [0, \infty)$, $f(x) := x$.

(2) $f : [0, 2] \to [0, 1]$, $f(x) := \begin{cases} 1, & x = 1 \\ 0, & x \neq 1 \end{cases}$

Solution. (1) Let $\epsilon > 0$. For each $n \in \mathbb{N}$, using the partition P_n of $[0, b]$ into n subintervals of equal width and the values for $L(f, P_n)$ and $U(f, P_n)$ in Example 1.2.7, we have

$$U(f, P_n) - L(f, P_n) = \tfrac{n+1}{n} \tfrac{b^2}{2} - \tfrac{n-1}{n} \tfrac{b^2}{2} = \tfrac{b^2}{n}.$$

To prove Riemann's Criterion, we need to have $\frac{b^2}{n} < \epsilon$ or equivalently $n > \frac{b^2}{\epsilon}$. Therefore, we now choose $N \in \mathbb{N}$ with $N > \frac{b^2}{\epsilon}$, so the above calculations give

$$U(f, P_N) - L(f, P_N) = \tfrac{b^2}{N} < \epsilon.$$

Hence f is integrable by Riemann's Criterion.

(2) Let $\epsilon > 0$. For each $\delta \in (0, 1)$, using the partition P_δ of $[0, 2]$ and the values for $L(f, P_\delta)$ and $U(f, P_\delta)$ in Example 1.2.8, we have

$$U(f, P_\delta) - L(f, P_\delta) = 2 - (2 - 2\delta) = 2\delta.$$

To prove Riemann's Criterion, we need to have $2\delta < \epsilon$ or equivalently $\delta < \frac{\epsilon}{2}$. Therefore, we now choose $\delta_0 \in (0, 1)$ with $\delta_0 < \frac{\epsilon}{2}$, so the above calculations give

$$U(f, P_{\delta_0}) - L(f, P_{\delta_0}) = 2\delta_0 < \epsilon.$$

Hence f is integrable by Riemann's Criterion. $\qquad\square$

In the example above, it is important to observe how $\epsilon > 0$ remains arbitrary throughout the proof, as required by Riemann's Criterion. In contrast, the variables $n \in \mathbb{N}$ and $\delta \in (0, 1)$ are initially allowed to be arbitrary in the calculations but then they are chosen, albeit dependent on ϵ, in the sense that we set $n = N$ and $\delta = \delta_0$ to produce the partitions P_N and P_{δ_0} at the end of each part.

We conclude this lecture with a few remarks and a result concerning Riemann sums and Riemann's original definition of the integral, since you may encounter these in more advanced theory. This material is for information only and will not be assessed.

The Riemann integral, as opposed to the Riemann–Darboux integral, relies on tagged partitions to define Riemann sums. A *tagged partition* (P, T) of $[a, b]$ consists of a partition $P = \{x_0, x_1, \ldots, x_n\}$ of $[a, b]$ and a subset $T = \{t_1, \ldots, t_n\}$ of $[a, b]$ whose elements are called *tags* and must satisfy $t_1 \in [x_0, x_1]$, $\ldots$, $t_n \in [x_{n-1}, x_n]$. The maximum width of the subintervals defined by a such a partition is denoted

$$\Delta P := \max\{(x_1 - x_0), \ldots, (x_n - x_{n-1})\}.$$

If $f : [a, b] \to \mathbb{R}$ is bounded, then the *Riemann sum* $R(f, P, T)$ is defined by

$$R(f, P, T) := \sum_{i=1}^{n} f(t_i)(x_i - x_{i-1}).$$

We are now in a position to state the following result, wherein property (2) is Riemann's original definition of integrability. This result will not be assessed.

Theorem 3.1.3 (Riemann's Integral)**.** The following properties are equivalent for any bounded function $f : [a, b] \to \mathbb{R}$:

(1) The function f is integrable.
(2) There exists $s \in \mathbb{R}$ such that for each $\epsilon > 0$, there exists $\delta > 0$ such that

$$|R(f, P, T) - s| < \epsilon$$

for all tagged partitions (P, T) of $[a, b]$ with $\Delta P < \delta$.

Moreover, if (2) holds, then $s = \int_a^b f$.

In Riemann's formulation, the real number s above is defined to be the integral. This can actually be understood as a limit, written $s = \lim_{\Delta P \to 0} R(f, P, T)$, but this limit must be understood using *nets*, which are a generalisation of sequences beyond the scope of this module. The above theorem shows, however, that the Riemann–Darboux integral and the original Riemann integral are in fact the same.

An advantage of the Riemann integral, however, is that if $f : [a, b] \to \mathbb{R}$ is known to be integrable, then it provides limiting formulas for the integral, such as $\int_a^b f = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} f(t_i)$, where $\{t_1, \ldots, t_n\}$ can be any set of tags associated with a partition of $[a, b]$ into n subintervals of equal width (e.g. left endpoints, right endpoints, midpoints, etc.). These explicit formulas are useful in numerics and advanced analysis but we shall not pursue these here.

3.2. **Integrability of monotonic functions.** The following example will serve as a prototype for a proof that monotonic functions on closed bounded intervals are always integrable.

Example 3.2.1 (Prototype for monotonic functions)**.** Use Riemann's Criterion to prove that $f : [0, \frac{\pi}{2}] \to \mathbb{R}$ given by $f(x) := \sin(x)$ for all $x \in [0, \frac{\pi}{2}]$ is integrable.

Solution. Let $\epsilon > 0$. For each $n \in \mathbb{N}$, consider the partition P_n of $[0, \frac{\pi}{2}]$ into n subintervals of equal width:

$$P_n := \{0, \tfrac{\pi}{2n}, \tfrac{2\pi}{2n}, \ldots, \tfrac{n\pi}{2n}\} = \{\tfrac{i\pi}{2n} : i = 0, 1, \ldots, n\}.$$

The sine function is increasing on $[0, \frac{\pi}{2}]$, so for each $i \in \{1, \ldots, n\}$, we have

$$m_i := \inf\left\{\sin(x) : x \in [\tfrac{(i-1)\pi}{2n}, \tfrac{i\pi}{2n}]\right\} = \sin((i-1)\tfrac{\pi}{2n}),$$
$$M_i := \sup\left\{\sin(x) : x \in [\tfrac{(i-1)\pi}{2n}, \tfrac{i\pi}{2n}]\right\} = \sin(i\tfrac{\pi}{2n}).$$

Together, we have

$$U(f, P_n) - L(f, P_n) = \sum_{i=1}^{n} M_i(x_i - x_{i-1}) - \sum_{i=1}^{n} m_i(x_i - x_{i-1})$$

$$= \sum_{i=1}^{n} \sin(i\tfrac{\pi}{2n})\tfrac{\pi}{2n} - \sum_{i=1}^{n} \sin((i-1)\tfrac{\pi}{2n})\tfrac{\pi}{2n}$$

$$= \left[\sum_{i=1}^{n} \sin(i\tfrac{\pi}{2n}) - \sum_{j=0}^{n-1} \sin(j\tfrac{\pi}{2n})\right]\tfrac{\pi}{2n}$$

$$= \left[\sin(n\tfrac{\pi}{2n}) - \sin(0\tfrac{\pi}{2n})\right]\tfrac{\pi}{2n}$$

$$= \left[\sin(\tfrac{\pi}{2}) - \sin(0)\right]\tfrac{\pi}{2n}$$

$$= \tfrac{\pi}{2n},$$

where in the third line we made the substitution $j := i - 1$.

To prove Riemann's Criterion, we need to have $\frac{\pi}{2n} < \epsilon$ or equivalently $n > \frac{\pi}{2\epsilon}$. Therefore, we now choose $N \in \mathbb{N}$ with $N > \frac{\pi}{2\epsilon}$, so the above calculations give

$$U(f, P_N) - L(f, P_N) = \tfrac{\pi}{2N} < \epsilon.$$

Hence f is integrable by Riemann's Criterion. $\qquad\square$

An inspection of the above proof reveals that the method extends almost immediately to any bounded monotonic function $f : [a, b] \to \mathbb{R}$.

Definition 3.2.2. A function $f : [a, b] \to \mathbb{R}$ is *monotonic* if either of the following conditions hold:

    (1) $f(x) \leq f(y)$ for all $x, y \in [a, b]$ such that $x \leq y$
    (2) $f(x) \geq f(y)$ for all $x, y \in [a, b]$ such that $x \leq y$

More specifically, the function is called *monotonic increasing* when (1) holds, and *monotonic decreasing* when (2) holds.

Theorem 3.2.3. If $f : [a, b] \to \mathbb{R}$ is bounded and monotonic, then f is integrable.

Proof. We will only prove the result in the case of monotonic increasing functions $f : [a, b] \to \mathbb{R}$, since the result for monotonic decreasing functions is analogous. We adapt the approach presented in Example 3.2.1.

Let $\epsilon > 0$. For each $n \in \mathbb{N}$, consider the partition $P_n := \{x_0, x_1, \ldots, x_n\}$ of $[a, b]$ into n subintervals of equal width. For each $i \in \{1, \ldots, n\}$, since f is monotonic increasing, we have

$$m_i := \inf\{f(x) : x \in [x_{i-1}, x_i]\} = f(x_{i-1}),$$
$$M_i := \sup\{f(x) : x \in [x_{i-1}, x_i]\} = f(x_i).$$

Together, we have

$$U(f, P_n) - L(f, P_n) = \sum_{i=1}^{n} (M_i - m_i)(x_i - x_{i-1})$$

$$= \sum_{i=1}^{n} (f(x_i) - f(x_{i-1}))\tfrac{b-a}{n}$$

$$= [(f(x_1) - f(x_0)) + (f(x_2) - f(x_1)) + \ldots + (f(x_n) - f(x_{n-1}))]\tfrac{b-a}{n}$$

$$= [-f(x_0) + (f(x_1) - f(x_1)) + \ldots + (f(x_{n-1}) - f(x_{n-1})) + f(x_n)]\tfrac{b-a}{n}$$

$$= (f(x_n) - f(x_0))\tfrac{b-a}{n}$$

$$= (f(b) - f(a))(b - a)\tfrac{1}{n},$$

where in the fourth line we have simply reordered the terms in the sum to show how they can be grouped into pairs that cancel (this was expressed slightly differently using summation notation in Example 3.2.1).

To prove Riemann's Criterion, we need to have $(f(b) - f(a))(b - a)\frac{1}{n} < \epsilon$ or equivalently $n > (f(b) - f(a))(b - a)\frac{1}{\epsilon}$. Therefore, we now choose $N \in \mathbb{N}$ with $N > (f(b) - f(a))(b - a)\frac{1}{\epsilon}$, so the above calculations give

$$U(f, P_N) - L(f, P_N) = (f(b) - f(a))(b - a)\frac{1}{N} < \epsilon.$$

Hence f is integrable by Riemann's Criterion. $\square$

Week 4: Properties of Integrable Functions

This week we will first use Riemann's Criterion for Integrability to prove that continuous functions on closed bounded intervals are always integrable. It will then be applied to establish a range of useful properties for integrable functions.

4.1. **Integrability of continuous functions.** In this lecture, we use Riemann's Criterion to prove that all continuous functions $f : [a, b] \to \mathbb{R}$ are integrable. In particular, we are only concerned here with continuous functions on closed bounded intervals, so let's recall the definition in that case.

Definition 4.1.1. A function $f : [a, b] \to \mathbb{R}$ is *continuous* if for each $x \in [a, b]$ and each $\epsilon > 0$, there exists $\delta > 0$ such that the following property holds:

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

This is the same as saying that f is continuous at each x in $[a, b]$.

It is a subtle but extremely important observation that the value of δ is allowed to depend on *both x and ϵ* in the above definition. It might help to think of δ as a function $\delta(x, \epsilon)$. The possible dependence on x, however, becomes problematic if we want to use Riemann's Criterion to prove integrability. The following stronger property, known as uniform continuity, is more suited to our purposes.

Definition 4.1.2. A function $f : [a, b] \to \mathbb{R}$ is *uniformly continuous* if for each $\epsilon > 0$, there exists $\delta > 0$ such that the following property holds:

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

The value of δ is only allowed to depend on ϵ in the above definition of uniform continuity, so we can think of it as a function $\delta(\epsilon)$. If we replace $[a, b]$ with an arbitrary set $\Omega \subseteq \mathbb{R}$ in Definitions 4.1.1 and 4.1.2 , then we obtain the definitions of continuity and uniform continuity for functions $f : \Omega \to \mathbb{R}$. In such generality, uniform continuity is a stronger property than continuity (see Example 4.1.4 below). In the case $\Omega = [a, b]$, however, there is no difference! This is the content of the following theorem (the proof will not be assessed).

Theorem 4.1.3. If $f : [a, b] \to \mathbb{R}$ is continuous, then f is uniformly continuous.

Proof. This can be proved directly by using the Completeness Axiom of $\mathbb{R}$ (see the Appendix to Chapter 8 in Spivak). Alternatively, the more common proof relying on the Bolzano–Weierstrass theorem can be found in many real analysis texts. $\square$

The following example shows how to work with uniform continuity. The second part is quite challenging and such questions will not be assessed.

Example 4.1.4. Prove the following facts:
  (1) $f : \mathbb{R} \to \mathbb{R}$, $f(x) := cx$, where $c \in \mathbb{R} \setminus \{0\}$, is uniformly continuous.
  (2) $g : (0, 1] \to \mathbb{R}$, $g(x) := \frac{1}{x}$, is not uniformly continuous.

Proof. (1) Let $\epsilon > 0$ and choose $\delta := \frac{\epsilon}{|c|}$. If $x \in \mathbb{R}$, $y \in \mathbb{R}$ and $|x - y| < \delta$, then

$$|f(x) - f(y)| = |cx - cy| = |c||x - y| < |c|\delta = |c|\frac{\epsilon}{|c|} = \epsilon,$$

so f is uniformly continuous. (It is important to observe here that the choice of δ did not depend on x or y.)

(2) Let $\epsilon = 1$ and assume for a contradiction that g is uniformly continuous, so there exists $\delta > 0$ such that the following property holds:

$$x \in (0, 1], \ y \in (0, 1] \text{ and } |x - y| < \delta \Longrightarrow |\tfrac{1}{x} - \tfrac{1}{y}| < 1 \qquad (4.1.1)$$

We expect that g is not uniformly continuous because of its growth near the origin. In particular, if $y = \frac{x}{2}$, then $|x - y| = \frac{x}{2}$ whilst $|\frac{1}{x} - \frac{1}{y}| = \frac{1}{x}$ (notice that $\frac{1}{x} \to \infty$ as $\frac{x}{2} \to 0$). This will contradict (4.1.1) when $\frac{x}{2} < \delta$ and $\frac{1}{x} \geq 1$, or equivalently, when $x < 2\delta$ and $x \leq 1$. Therefore, we choose $x = \frac{1}{2}\min\{2\delta, 1\}$, to obtain

$$x \in (0, 1], \ y = \tfrac{x}{2} \in (0, 1], \ |x - y| = \tfrac{x}{2} = \tfrac{\min\{2\delta, 1\}}{4} < \delta,$$

whilst

$$|\tfrac{1}{x} - \tfrac{1}{y}| = \tfrac{1}{x} = \tfrac{2}{\min\{2\delta, 1\}} \geq 2,$$

which contradicts (4.1.1). This proves that g is not uniformly continuous. $\qquad\square$

We can now rely on uniform continuity to prove the following result.

Theorem 4.1.5. If $f : [a, b] \to \mathbb{R}$ is continuous, then f is integrable.

Proof. We let $\epsilon > 0$ with the aim of verifying Riemann's Criterion to prove that f is integrable. We know from Theorem 4.1.3 that f is uniformly continuous, so we can choose $\delta > 0$, depending only on ϵ, such that the following property holds:

$$x \in [a, b], \ y \in [a, b] \text{ and } |x - y| < \delta \Longrightarrow |f(x) - f(y)| < \frac{\epsilon}{b - a} \qquad (4.1.2)$$

Now let $P_\delta = \{x_0, x_1, \ldots, x_n\}$ denote a partition of $[a, b]$ into n subintervals of equal width less than δ. This means that $|x_1 - x_0| = |x_2 - x_1| = \ldots = |x_n - x_{n-1}| < \delta$. For each $i \in \{1, \ldots, n\}$, we then have

$$s \in [x_{i-1}, x_i], \ t \in [x_{i-1}, x_i] \Longrightarrow |s - t| \leq |x_i - x_{i-1}| < \delta$$
$$\Longrightarrow |f(s) - f(t)| < \frac{\epsilon}{b - a},$$

where we used (4.1.2) to obtain the final implication. It follows (using the 'ϵ-type' interpretation of the infimum/supremum in Lemma 1.1.4) that

$$\sup\{f(s) : s \in [x_{i-1}, x_i]\} - \inf\{f(t) : t \in [x_{i-1}, x_i]\} < \frac{\epsilon}{b - a}.$$

Altogether, we obtain

$$U(f, P_\delta) - L(f, P_\delta) = \sum_{i=1}^{n} (M_i - m_i)(x_i - x_{i-1}) < \frac{\epsilon}{b - a} \sum_{i=1}^{n} (x_i - x_{i-1}) = \epsilon.$$

This proves that f is integrable by Riemann's Criterion. $\qquad\square$

4.2. **Restriction/extension, linearity, algebra and estimates.** The following result shows that integrability is preserved when the domain of a function is suitably restricted or extended. It is important to know this result but the proof is somewhat tedious and technical, so only a sketch is provided and you are not expected to be familiar with the details.

Theorem 4.2.1 (Restriction and extension of the integral)**.** Suppose that $c \in [a, b]$. The following properties hold for any bounded function $f : [a, b] \to \mathbb{R}$:

(1) If f is integrable on both $[a, c]$ and $[c, b]$, then f is integrable on $[a, b]$.

(2) If f is integrable on $[a, b]$, then f is integrable on both $[a, c]$ and $[c, b]$.

In both cases, the identity $\int_a^b f = \int_a^c f + \int_c^b f$ holds.

Sketch of the proof. This is a technical proof that seems more complicated than it actually is when written down. We only sketch the main details here. It is a good exercise to transform this into a complete proof yourself. We consider three parts:

To prove (1), we use Riemann's Criterion. If f is integrable on $[a, c]$ and $[c, d]$, then there exists a partition P_1 of $[a, c]$ and a partition P_2 of $[c, b]$ such that

$$U(f|_{[a,c]}, P_1) - L(f|_{[a,c]}, P_1) < \epsilon$$
$$U(f|_{[c,b]}, P_2) - L(f|_{[c,b]}, P_2) < \epsilon. \tag{4.2.1}$$

The union $P := P_1 \cup P_2$ is a partition of $[a, b]$ and you can prove that

$$U(f, P) = U(f|_{[a,c]}, P_1) + U(f|_{[c,b]}, P_2)$$
$$L(f, P) = L(f|_{[a,c]}, P_1) + L(f|_{[c,b]}, P_2)$$

to deduce that $U(f, P) - L(f, P) < 2\epsilon$. The required integrability would then follow from Riemann's Criterion if the 2ϵ here was replaced by ϵ, but this can be achieved by reverse engineering the original choice of ϵ in (4.2.1).

To prove (2), we again use Riemann's Criterion. If f is integrable on $[a, b]$, then there exists a partition P of $[a, b]$ such that $U(f, P) - L(f, P) < \epsilon$. We need to use P to extract partitions of $[a, c]$ and $[c, b]$. We can indeed split P into those points contained in $[a, c]$ and those contained in $[c, d]$, but these won't necessarily be partitions, because they may not contain the endpoint c. This is remedied by starting with the partition $P \cup \{c\}$ and noting (by the Partition Lemma) that

$$U(f, P \cup \{c\}) - L(f, P \cup \{c\}) \leq U(f, P) - L(f, P) < \epsilon.$$

It is then possible to write $P \cup \{c\} = P_1 \cup P_2$, where P_1 is a partition of $[a, c]$ and P_2 is a partition of $[c, b]$. The required integrability results will follow from Riemann's Criterion if you can prove that $U(f, P_1) - L(f, P_1) < \epsilon$ and $U(f, P_2) - L(f, P_2) < \epsilon$.

To prove the identity, we recall the definition of the Riemann–Darboux integral as the supremum of lower sums and the infimum of upper sums. In particular, for any partition P of $[a, b]$, writing $P \cup \{c\} = P_1 \cup P_2$ as above, you can prove that

$$L(f|_{[a,c]}, P_1) \leq \int_a^c f \leq U(f|_{[a,c]}, P_1) \tag{4.2.2}$$

$$L(f|_{[c,b]}, P_2) \leq \int_c^b f \leq U(f|_{[c,b]}, P_2) \tag{4.2.3}$$

$$L(f, P) \leq L(f, P \cup \{c\}) \leq \int_a^c f + \int_c^b f \leq U(f, P \cup \{c\}) \leq U(f, P). \tag{4.2.4}$$

The definition of the integral then implies that

$$\int_a^b f = \sup_P L(f, P) \leq \int_a^c f + \int_c^b f \leq \inf_P U(f, P) = \int_a^b f$$

and the required identity follows. $\qquad\square$

The following result shows that the integral is as a linear operator.

Theorem 4.2.2 (Linearity of the integral)**.** If $\alpha \in \mathbb{R}$ and $f, g : [a, b] \to \mathbb{R}$ are both bounded and integrable, then $f + g$ and αf are both bounded and integrable with

(1) $\int_a^b (f + g) = \int_a^b f + \int_a^b g$
(2) $\int_a^b (\alpha f) = \alpha \int_a^b f$

Proof. We only prove the results for $f + g$ here. The proof of the results for αf, which follows a similar approach, is posed as a question on the problem sheet.

Suppose that $f, g : [a, b] \to \mathbb{R}$ are both bounded and integrable. It is left as an exercise to prove that $f + g$ is bounded. Next, let $\epsilon > 0$, so Riemann's Criterion guarantees that there exist partitions P' and P'' of $[a, b]$ such that

$$U(f, P') - L(f, P') < \tfrac{\epsilon}{2} \quad \text{and} \quad U(g, P'') - L(g, P'') < \tfrac{\epsilon}{2}.$$

We can combine the two estimates above by defining the partition $P := P' \cup P''$ and using the Partition Lemma (Lemma 2.1.1) to note that $L(f, P') \le L(f, P)$ and $U(f, P') \ge U(f, P)$, hence $U(f, P) - L(f, P) \le U(f, P') - L(f, P')$, since $P' \subseteq P$ and $P'' \subseteq P$, and similarly $U(g, P) - L(g, P) \le U(f, P'') - L(f, P'')$. This gives

$$U(f, P) - L(f, P) < \tfrac{\epsilon}{2} \quad \text{and} \quad U(g, P) - L(g, P) < \tfrac{\epsilon}{2}.$$

Furthermore, adding these two inequalities, we arrive at

$$[U(f, P) + U(g, P)] - [L(f, P) + L(g, P)] < \tfrac{\epsilon}{2} + \tfrac{\epsilon}{2} = \epsilon. \tag{4.2.5}$$

It is a good exercise to prove that

$$\begin{aligned} U(f + g, P) &\le U(f, P) + U(g, P) \\ L(f + g, P) &\ge L(f, P) + L(g, P). \end{aligned} \tag{4.2.6}$$

It follows that

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

hence $f + g$ is integrable by Riemann's Criterion.

In order to prove the equality of the associated integrals, however, we must recall the definition of the Riemann–Darboux integral as the supremum of lower sums and the infimum of upper sums. In particular, we have

$$L(f, P) \le \int_a^b f \le U(f, P) \tag{4.2.7}$$

$$L(g, P) \le \int_a^b g \le U(g, P) \tag{4.2.8}$$

$$L(f + g, P) \le \int_a^b (f + g) \le U(f + g, P). \tag{4.2.9}$$

These inequalities hold for arbitrary partitions, but here P continues to denote the partition $P' \cup P''$ from above. Using (4.2.6), inequality (4.2.9) becomes

$$L(f, P) + L(g, P) \le \int_a^b (f + g) \le U(f, P) + U(g, P) \tag{4.2.10}$$

Now, subtracting (4.2.7) and (4.2.8) from (4.2.10), we obtain

$$[L(f, P) + L(g, P)] - [U(f, P) + U(g, P)]$$
$$\le \int_a^b (f + g) - \left(\int_a^b f + \int_a^b g \right)$$
$$\le [U(f, P) + U(g, P)] - [L(f, P) + L(g, P)].$$

Using (4.2.5), it follows that

$$-\epsilon < \int_a^b (f + g) - \left(\int_a^b f + \int_a^b g \right) < \epsilon.$$

This holds for all $\epsilon > 0$, which implies that

$$\int_a^b (f + g) - \left(\int_a^b f + \int_a^b g \right) = 0,$$

hence $\int_a^b (f + g) = \int_a^b f + \int_a^b g$, as required. $\qquad \square$

Here is an example of how the extension and linearity properties of the integral from the previous two theorems can be applied.

Example 4.2.3. Prove that $f : [0, 2] \to \mathbb{R}$ given by $f(x) := \begin{cases} \sin(1) + 3, & x = 1 \\ \sin(x), & x \ne 1 \end{cases}$

for all $x \in [0, \pi]$ is integrable.

Proof. In order to use the linearity of the integral, we write $f = g + 3h$, where

$$g(x) := \sin(x) \quad \text{and} \quad h(x) := \begin{cases} 1, & x = 1 \\ 0, & x \neq 1 \end{cases}$$

are both bounded functions on $[0, 2]$. The function g is monotonic increasing $[0, \frac{\pi}{2}]$ and monotonic decreasing on $[\frac{\pi}{2}, 2]$, so g is integrable on both $[0, \frac{\pi}{2}]$ and $[\frac{\pi}{2}, 2]$ by the integrability of monotonic functions in Theorem 3.2.3. The extension of integrability in Theorem 4.2.1 then implies that g is integrable on $[0, 2]$. Meanwhile, the function h is integrable on $[0, 2]$ by Example 2.2.6. The linearity of the integral in Theorem 4.2.2 thus implies that $f = g + 3h$ is integrable on $[0, 2]$. $\square$

In order to extend our definition of integrability to the class of bounded $\mathbb{R}$-valued functions $f : [a, b] \to \mathbb{R}$, we write $f = f^+ - f^-$, where f^+ and f^- are referred to respectively as the positive part of f and the negative part of f, defined as

$$f^+(x) := \max\{f(x), 0\} = \begin{cases} f(x), & \text{if } f(x) \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

and

$$f^-(x) := -\min\{f(x), 0\} = \begin{cases} -f(x), & \text{if } f(x) \leq 0 \\ 0, & \text{otherwise} \end{cases}.$$

The linearity of the integral in Theorem 4.2.2 suggests the following extension of the definition of integrability from Definition 2.2.3.

Definition 4.2.4. A bounded function $f : [a, b] \to \mathbb{R}$ is called *integrable* if both f^+ and f^- are integrable (according to Definition 2.2.3). If such f is integrable, then the *integral* of f is $\int_a^b f := \int_a^b f^+ - \int_a^b f^-$.

The following example shows how to work with this definition. The result is often useful when dealing with $\mathbb{R}$-valued integrable functions.

Example 4.2.5. Prove that if $f : [a, b] \to \mathbb{R}$ is bounded and integrable, then $|f| : [a, b] \to \mathbb{R}$, given by $|f|(x) := |f(x)|$ for all $x \in [a, b]$, is bounded and integrable.

Proof. The fact that $|f|$ is bounded follows from the fact that f is bounded. It is a good exercise to write out the details. The integrability of f, meanwhile, means that both f^+ and f^- are integrable. After we observe that $|f| = f^+ + f^-$, the linearity of the integral in Theorem 4.2.2 implies that $|f|$ is integrable. $\square$

The attentive reader will note that all of the results presented so far this week have actually been stated for $\mathbb{R}$-valued functions, whereas until now, our definition of integrability only applied to $[0, \infty)$-valued functions. To correct this, one could state Definition 4.2.4 at the start of the week, and then check that all of our results can be adapted to the case of $\mathbb{R}$-valued functions. We chose to avoid this approach, however, in order to use the linearity of the integral in Theorem 4.2.2 to motivate Definition 4.2.4.

Our next aim is to prove that a product of two integrable functions is itself integrable. The following lemma contains the heart of the matter. The proof is left as a challenging exercise.

Lemma 4.2.6. If $f : [a, b] \to \mathbb{R}$ is bounded and integrable, then f^2 is bounded and integrable.

We can now prove that integrability is preserved by products of functions.

Theorem 4.2.7 (Algebra of integrability). If $f, g : [a, b] \to \mathbb{R}$ are both bounded and integrable, then fg is bounded and integrable.

Proof. In order to apply Lemma 4.2.6, we observe the identity

$$fg = \tfrac{1}{4}(f + g)^2 - \tfrac{1}{4}(f - g)^2,$$

which can be proved by expanding the right-hand side. In particular, if f and g are both integrable, then the linearity of the integral in Theorem 4.2.2 implies that $(f + g)$ and $(f - g)$ are both bounded and integrable. The algebra of integrability in Lemma 4.2.6 then implies that $(f + g)^2$ and $(f - g)^2$ are both bounded and integrable. Another application of the linearity of the integral thus finally implies that $\tfrac{1}{4}(f + g)^2 - \tfrac{1}{4}(f - g)^2 = fg$ is bounded and integrable, as required. $\square$

We conclude by proving some useful bounds for integrals. These will be essential for proving the Fundamental Theorem of Calculus next week.

Theorem 4.2.8 (Estimates for the integral). If $f : [a, b] \to \mathbb{R}$ is integrable and $m \leq f(x) \leq M$ for all $x \in [a, b]$ and some $m, M \in \mathbb{R}$, then

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

Proof. Suppose that f satisfies the hypotheses in the theorem and consider the partition $P_0 := \{a, b\}$ of $[a, b]$. Observe that

$$m_1 := \inf\{f(x) : x \in [a, b]\} \geq m$$

hence

$$L(f, P_0) := \sum_{i=1}^{1}(x_i - x_{i-1}) = m_1(b - a) \geq m(b - a).$$

We combine this with the definition of the integral, since f is integrable, to obtain

$$\int_a^b f := \underline{\int_a^b f} := \sup_P L(f, P) \geq L(f, P_0) \geq m(b - a),$$

as required. A similar argument implies the required upper bound for the integral. It is a good exercise to write out the details. $\square$

WEEK 5: THE FUNDAMENTAL THEOREM OF CALCULUS

The theory of integration culminates this week with the proof of two theorems known collectively as the Fundamental Theorem of Calculus. These results reveal how integration and differentiation are inverse operations. The first allows us to differentiate certain integral expressions whilst the second allows us to compute integrals using so-called antiderivatives. The latter result also remedies the main shortcoming of the theory developed so far by providing a method for computing integrals as routine calculations.

5.1. **The First Fundamental Theorem of Calculus.** This first part of the theorem is concerned with the differentiation of certain integral functions.

Theorem 5.1.1 (First Fundamental Theorem of Calculus)**.** Let $f : [a, b] \to \mathbb{R}$ denote a bounded integrable function and suppose that $F : [a, b] \to \mathbb{R}$ is given by $F(x) := \int_a^x f$ for all $x \in [a, b]$. If f is continuous at c, where $c \in [a, b]$, then F is differentiable at c with derivative $F'(c) = f(c)$.

Proof. Let f and F be defined as in the theorem. We only consider the case when $c \in (a, b)$, in which case we need to prove that

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

It is a good exercise to adjust this proof to the remaining cases when $c \in \{a, b\}$.

In what follows, we only consider when $h > 0$, since the arguments when $h < 0$ are analogous. Using the restriction and extension properties of the integral in Theorem 4.2.1, we obtain

$$F(c + h) - F(c) = \int_a^{c+h} f - \int_a^c f = \left(\int_a^c f + \int_c^{c+h} f \right) - \int_a^c f = \int_c^{c+h} f.$$

Next, we use the estimates for integrals in Theorem 4.2.8 to obtain

$$m_h h = m_h((c + h) - c) \leq \int_c^{c+h} f \leq M_h((c + h) - c) = M_h h,$$

where

$$m_h := \inf\{f(x) : x \in [c, c + h]\} \quad \text{and} \quad M_h := \sup\{f(x) : x \in [c, c + h]\}.$$

Altogether, we have

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

We now use that f is continuous at c to obtain $\lim_{h \to 0^+} m_h = f(c) = \lim_{h \to 0^+} M_h$. A good diagram can help to convince you of these limits but it is a better exercise to try and write out a proof using ϵ-type arguments. The required result then follows from the Sandwich Theorem. $\qquad \square$

The First Fundamental Theorem of Calculus shows how differentiation is an inverse of integration. We can see this more explicitly by using Leibniz notation to express the conclusion of the theorem as

$$\frac{\mathrm{d}}{\mathrm{d}x} \left(\int_a^x f \right) \bigg|_{x=c} = f(c).$$

Moreover, if f is actually continuous (meaning continuous on $[a, b]$ rather than just continuous at a single point c), then this becomes

$$\frac{\mathrm{d}}{\mathrm{d}x} \left(\int_a^x f \right) = f(x) \quad \text{for all} \quad x \in [a, b].$$

It is helpful to record this particular case of the theorem as a separate corollary.

Corollary 5.1.2. If $f : [a, b] \to \mathbb{R}$ is continuous and $F : [a, b] \to \mathbb{R}$ is given by $F(x) := \int_a^x f$ for all $x \in [a, b]$, then F is differentiable with derivative $F'(x) = f(x)$ for all $x \in [a, b]$.

Proof. If $f : [a, b] \to \mathbb{R}$ is continuous, then it is bounded (by the Extreme Value Theorem) and integrable (by Theorem 4.1.5). The hypotheses of Theorem 5.1.1 thus hold *for all $c \in [a, b]$* from which the result follows. $\qquad\square$

The First Fundamental Theorem of Calculus is particularly sensitive to the form of the function $F(x) := \int_a^x f$ which appears in the statement of the theorem. For example, if the variable x appears as the lower endpoint of the integral, or if it is replaced by some other function of x, then the theorem does not give a formula for the derivative. The following example, however, shows how the theorem can be combined effectively with other results in such contexts.

Example 5.1.3. Prove that the following functions are differentiable and find a formula for their derivatives:

(1) $G : [0, 1] \to \mathbb{R}$, $G(x) := \int_0^{x^3} \sqrt{1 - t^4} \, dt$

(2) $H : [0, 1] \to \mathbb{R}$, $H(x) := \int_x^1 \sqrt{1 - t^4} \, dt$

Solution. Let $F : [0, 1] \to \mathbb{R}$ be given by $F(x) := \int_0^x \sqrt{1 - t^4} \, dt$ for all $x \in [0, 1]$. The function $f : [0, 1] \to \mathbb{R}$ given by $f(t) := \sqrt{1 - t^4}$ for all $t \in [0, 1]$ is continuous, as a composition of a nonnegative-valued polynomial and the square root function. The First Fundamental Theorem of Calculus for continuous functions in Corollary 5.1.2 thus implies that F is differentiable with $F'(x) = f(x) = \sqrt{1 - x^4}$ for all $x \in [0, 1]$

(1) Observe that $G(x) = F(x^3)$, so $G = F \circ g$, where $g(x) := x^3$ is differentiable. We also know that F is differentiable by the preceding discussion, so the Chain Rule implies that G is differentiable with derivative

$$G'(x) = F'(g(x))g'(x) = \sqrt{1 - (x^3)^4} \, 3x^2 = \sqrt{1 - x^{12}} \, 3x^2$$

for all $x \in [0, 1]$.

(2) Here we can use the restriction and extension properties of the integral in Theorem 4.2.1 to write

$$H(x) = \int_x^1 \sqrt{1 - t^4} \, dt = \int_0^1 \sqrt{1 - t^4} \, dt - \int_0^x \sqrt{1 - t^4} \, dt = C - F(x)$$

for all $x \in [0, 1]$, where $C := \int_0^1 \sqrt{1 - t^4} \, dt \in \mathbb{R}$. A constant function has derivative identically equal to 0, and we know that F is differentiable by the preceding discussion, so H is differentiable as a linear combination of differentiable functions with derivative $H'(x) = 0 - F'(x) = -\sqrt{1 - x^4}$ for all $x \in [0, 1]$.

5.2. **The Second Fundamental Theorem of Calculus.** This second part of the theorem is concerned with the integration of certain derivative functions and leads ultimately to the consideration of antiderivatives.

Theorem 5.2.1 (Second Fundamental Theorem of Calculus)**.** If $f : [a, b] \to \mathbb{R}$ is a bounded integrable function, and there exists a differentiable function $g : [a, b] \to \mathbb{R}$ such that $f = g'$, then $\int_a^b f = g(b) - g(a)$.

Proof. We only present the proof when $f = g'$ is assumed to be continuous. In that case, it follows from the First Fundamental Theorem of Calculus for continuous functions in Corollary 5.1.2 that if $F : [a, b] \to \mathbb{R}$ is given by $F(x) := \int_a^x f$, then $F'(x) = f(x)$, for all $x \in [a, b]$. We then have $F' = f = g'$, so by Example 5.2.3, there exists $C \in \mathbb{R}$ such that $F(x) = g(x) + C$ for all $x \in [a, b]$. In particular, since $F(a) = \int_a^a f = 0$, we find $0 = g(a) + C$ so $C = -g(a)$, whence $F(x) = g(x) - g(a)$ for all $x \in [a, b]$. The result follows since $\int_a^b f = F(b) = g(b) - g(a)$.

In the case when $f = g'$ is only assumed to be bounded and integrable, this result is not a corollary of the First Fundamental Theorem of Calculus. It relies instead on using the Mean Value Theorem to prove that $L(f, P) \leq g(b) - g(a) \leq U(f, P)$ for all partitions P of $[a, b]$. The result then follows by observing that

$$\int_a^b f = \sup_P L(f, P) \leq g(b) - g(a) \leq \inf_P U(f, P) = \int_a^b f.$$

We do not detail this here, however, since such applications of the Mean Value Theorem will be considered in a second-year analysis module. □

The Second Fundamental Theorem of Calculus shows how integration is the inverse of differentiation, since the conclusion of the theorem implies that

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

The theorem actually asserts more than this fact, however, since it shows that *any differentiable function g* satisfying $f = g'$ can be used to compute the integral of f using the formula $\int_a^b f = g(b) - g(a)$. We call such functions *antiderivatives of f* in accordance with the following definition.

Definition 5.2.2. If $\Omega \subseteq \mathbb{R}$ and $f : \Omega \to \mathbb{R}$, then an *antiderivative of f* is a differentiable function $g : \Omega \to \mathbb{R}$ with the property that $f(x) = g'(x)$ for all $x \in \Omega$. We use the notation $\int f$ to denote an antiderivative of f, which simply means that $\int f$ is a differentiable function such that $(\int f)' = f$.

It is important to note here that a single function f may have many different antiderivatives, so the notation $\int f$ does not represent a unique function associated with f. This is precisely why the definition above refers to *an* antiderivative (and not *the* antiderivative) of f. In the case $\Omega = [a, b]$, however, the example below characterises all antiderivatives for a bounded integrable function $f : [a, b] \to \mathbb{R}$. In particular, if g is an antiderivative of such a function f, then any antiderivative of f must satisfy $\int f = g + C$ for some $C \in \mathbb{R}$.

Example 5.2.3. Suppose that $f : [a, b] \to \mathbb{R}$ is bounded and integrable:

(1) Prove that if g and h are antiderivatives of f, then there exists $C \in \mathbb{R}$ such that $g(x) = h(x) + C$ for all $x \in [a, b]$.
(2) Prove that an antiderivative of f exists when f is continuous.

Proof. (1) If g and h are antiderivatives of f, then $f(x) = g'(x) = h'(x)$ whence $(g - h)'(x) = 0$ for all $x \in [a, b]$. This means that $g - h$ is monotonic increasing (because $(g - h)'(x) \geq 0$) and monotonic decreasing (because $(g - h)'(x) \leq 0$). This is only possible, however, if $(g - h)(x) = (g - h)(a)$ (or $(g - h)(b)$) for all $x \in [a, b]$. The result follows with $C := g(a) - h(a)$.

(2) The First Fundamental Theorem of Calculus for continuous functions in Corollary 5.1.2 shows that the function $F : [a, b] \to \mathbb{R}$ given by $F(x) := \int_a^x f$ satisfies $F'(x) = f(x)$ for all $x \in [a, b]$, hence F is an antiderivative of f. □

The example above shows that any continuous function $f : [a, b] \to \mathbb{R}$ has an antiderivative given by $\int_a^x f$ for all $x \in [a, b]$. This does *not* mean, however, that the antiderivative of any continuous function can be expressed in terms of elementary functions. On the contrary, differential Galois Theory shows that the antiderivatives for most continuous functions (even those given by simple expressions such as e^{-x^2} and $\sqrt{1 - x^4}$) cannot be expressed in terms of elementary functions.

In any assessment for this module, however, when you are asked to find an antiderivative of a given function f, it is always expected that your answer should be in terms of elementary functions (so writing $\int_a^x f$ will not do!).

Example 5.2.4. Find an antiderivative of the function f in each case below:

    (1) $f : \mathbb{R} \to \mathbb{R}$, $f(x) := x^{\alpha}$, $\alpha \in \mathbb{R} \setminus \{-1\}$
    (2) $f : \mathbb{R} \setminus \{0\} \to \mathbb{R}$, $f(x) := x^{-1}$

Solution. (1) The function $g : \mathbb{R} \to \mathbb{R}$ given by $g(x) := x^{\alpha+1}/(\alpha+1)$ for all $x \in [a, b]$ is differentiable with $g'(x) = x^{\alpha} = f(x)$ for all $x \in [a, b]$. This proves that g is an antiderivative of f. (This proof breaks down when $\alpha = -1$, since then $g(x) = 0$ with $g'(x) = 0$ so $g' \neq f$.)

    (2) The function $h : \mathbb{R} \setminus \{0\} \to \mathbb{R}$ given by $h(x) := \log|x|$ for all $x \in \mathbb{R} \setminus \{0\}$ is well-defined (the logarithm function has domain $(0, \infty)$) and differentiable with

$$
h'(x) = \begin{cases} \frac{\mathrm{d}}{\mathrm{d}x}(\log(x)), & \text{if } x > 0 \\ \frac{\mathrm{d}}{\mathrm{d}x}(\log(-x)), & \text{if } x < 0 \end{cases}
$$
$$
= \begin{cases} \frac{1}{x}, & \text{if } x > 0 \\ -(\frac{1}{-x}), & \text{if } x < 0 \end{cases}
$$
$$
= f(x)
$$

for all $x \in \mathbb{R} \setminus \{0\}$, where we used the Chain Rule to obtain the second equality. This proves that h is an antiderivative of f.

Week 7: Integral Calculus

The Fundamental Theorem of Calculus provides a bridge connecting our newly acquired understanding of the Riemann–Darboux integral from analysis with our previous experience of integral calculus. We will use it this week to establish the familiar calculus techniques for integrating products and compositions of functions as theorems known as Integration by Parts and the Substitution Formula. We will also see how the antiderivative of any rational function can, in principle, be found using the method of Partial Fractions.

7.1. **Product and composite functions.** Here we combine the Fundamental Theorem of Calculus with the Product Rule and Chain Rule for derivatives to prove the familiar formulas for integrals of products and compositions of functions.

Theorem 7.1.1 (Integration by Parts). If u and $v : [a, b] \to \mathbb{R}$ are differentiable with bounded integrable derivatives u' and v', then

$$
\int_a^b uv' = \left[uv\right]_a^b - \int_a^b u'v \quad \text{and} \quad \int uv' = uv - \int u'v
$$

where $\left[uv\right]_a^b := u(b)v(b) - u(a)v(a)$.

Proof. The functions u and v are assumed to be differentiable so the Product Rule implies that uv is differentiable with derivative $(uv)' = u'v + uv'$. Moreover, the functions u and v are continuous and hence (bounded) integrable by Theorem 4.1.5. The derivatives u' and v' are also assumed to be bounded integrable, so by the algebra of integrability in Theorem 4.2.7, the products $u'v$ and uv' are bounded integrable, and in turn, by the linearity of the integral in Theorem 4.2.2, the sum $u'v + uv'$ is bounded integrable with $\int_a^b(u'v + uv') = \int_a^b u'v + \int_a^b uv'$. Meanwhile, the Second Fundamental Theorem of Calculus from Theorem 5.2.1 applied with $f = u'v + uv'$ and $g = uv$ implies that

$$
\int_a^b(u'v + uv') = (uv)(b) - (uv)(a).
$$

Altogether, we have $\int_a^b uv' = u(b)v(b) - u(a)v(a) - \int_a^b u'v$, as required.

The equality for antiderivatives $\int uv' = uv - \int u'v$ follows because $(\int uv')' = uv'$ whilst the Product Rule implies that

$$(uv - \int u'v)' = (uv)' - (\int u'v)' = uv' + u'v - u'v = uv'.$$

To understand this more, recall the meaning of the symbol $\int$ in Definition 5.2.2. $\quad\square$

The next example shows some useful strategies involving Integration by Parts.

Example 7.1.2. Find the following integrals and antiderivatives:

(1) $\displaystyle\int_0^{\frac{\pi}{2}} x\cos(x)\ \mathrm{d}x$

(2) An antiderivative of $f : (0,\infty) \to \mathbb{R}$, $f(x) := \log(x)$, $x \in (0,\infty)$.

(3) $\displaystyle\int x^2 e^x\ \mathrm{d}x$

(4) $\displaystyle\int e^x \cos(x)\ \mathrm{d}x$

Solution. (1) We apply Integration by Parts with

$$\begin{cases} u(x) = x \\ v(x) = \sin(x) \end{cases} \quad \text{and} \quad \begin{cases} u'(x) = 1 \\ v'(x) = \cos(x) \end{cases}$$

to obtain

$$\int_0^{\frac{\pi}{2}} x\cos(x)\ \mathrm{d}x = [x\sin(x)]_0^{\frac{\pi}{2}} - \int_0^{\frac{\pi}{2}} \sin(x)\ \mathrm{d}x = \tfrac{\pi}{2} - [-\cos(x)]_0^{\frac{\pi}{2}} = \tfrac{\pi}{2} - 1.$$

(2) We apply Integration by Parts with

$$\begin{cases} u(x) = \log(x) \\ v(x) = x \end{cases} \quad \text{and} \quad \begin{cases} u'(x) = \frac{1}{x} \\ v'(x) = 1 \end{cases}$$

to obtain

$$\int \log(x)\ \mathrm{d}x = x\log(x) - \int \left(\frac{1}{x}\right) x\ \mathrm{d}x = x\log(x) - x.$$

(3) We apply Integration by Parts with

$$\begin{cases} u(x) = x^2 \\ v(x) = e^x \end{cases} \quad \text{and} \quad \begin{cases} u'(x) = 2x \\ v'(x) = e^x \end{cases}$$

to obtain $\int x^2 e^x\ \mathrm{d}x = x^2 e^x - \int 2x e^x\ \mathrm{d}x$. Integration by Parts (again) with

$$\begin{cases} u(x) = x \\ v(x) = e^x \end{cases} \quad \text{and} \quad \begin{cases} u'(x) = 1 \\ v'(x) = e^x \end{cases}$$

gives $\int x e^x\ \mathrm{d}x = x e^x - \int e^x\ \mathrm{d}x = (x-1)e^x$. Altogether, we have

$$\int x^2 e^x\ \mathrm{d}x = x^2 e^x - 2(x-1)e^x = (x^2 - 2x + 2)e^x.$$

(4) We apply Integration by Parts with

$$\begin{cases} u(x) = e^x \\ v(x) = \sin(x) \end{cases} \quad \text{and} \quad \begin{cases} u'(x) = e^x \\ v'(x) = \cos(x) \end{cases}$$

to obtain $\int e^x \cos(x)\ \mathrm{d}x = e^x \sin(x) - \int e^x \sin(x)\ \mathrm{d}x$. Integration by Parts with

$$\begin{cases} u(x) = e^x \\ v(x) = -\cos(x) \end{cases} \quad \text{and} \quad \begin{cases} u'(x) = e^x \\ v'(x) = \sin(x) \end{cases}$$

gives $\int e^x \sin(x) \, \mathrm{d}x = -e^x \cos(x) + \int e^x \cos(x) \, \mathrm{d}x$. Altogether, we have

$$\int e^x \cos(x) \, \mathrm{d}x = e^x \sin(x) - \left(-e^x \cos(x) + \int e^x \cos(x) \, \mathrm{d}x \right)$$

and after rearranging this becomes

$$\int e^x \cos(x) \, \mathrm{d}x = \frac{1}{2}(e^x \sin(x) + e^x \cos(x)).$$

The following proof is somewhat technical and the details will not be assessed since they rely on the Extreme Value Theorem and Intermediate Value Theorem, which will be considered in a second-year analysis module. The main point, however, is to recognise that the theorem provides a rigorous justification for the formal algebraic manipulation "$\frac{\mathrm{d}u}{\mathrm{d}x} \, \mathrm{d}x = \mathrm{d}u$" made when changing variables in integral calculus. In particular, the term $u'(x) \, \mathrm{d}x$ appearing in the formula below becomes $\frac{\mathrm{d}u}{\mathrm{d}x} \, \mathrm{d}x$ in Leibniz notation.

Theorem 7.1.3 (Substitution Formula). If $f : \mathbb{R} \to \mathbb{R}$ is a continuous function[1] and $u : [a, b] \to \mathbb{R}$ is differentiable with bounded integrable derivative u', then

$$\int_a^b f(u(x))u'(x) \, \mathrm{d}x = \int_{u(a)}^{u(b)} f(x) \, \mathrm{d}x.$$

Proof. Let F denote an antiderivative of f on the range of u.[2] The function f is assumed to be continuous so the existence of such an antiderivative is guaranteed by Example 5.2.3. This means that F is differentiable with derivative $F' = f$ on the range of u. Meanwhile, the function u is assumed to be differentiable, so the Chain Rule implies that the composition $F \circ u$ is differentiable with derivative

$$(F \circ u)'(x) = F'(u(x))u'(x) = f(u(x))u'(x).$$

We write this as $(f \circ u)u' = (F \circ u)'$ and apply the Second Fundamental Theorem of Calculus from Theorem 5.2.1 to obtain

$$\begin{aligned}
\int_a^b (f \circ u)u' &= \int_a^b (F \circ u)' \\
&= (F \circ u)(b) - (F \circ u)(a) \\
&= F(u(b)) - F(u(a)) \\
&= \int_{u(a)}^{u(b)} F' \\
&= \int_{u(a)}^{u(b)} f,
\end{aligned}$$

where the Fundamental Theorem of Calculus was used in the second and fourth equalities, and the final equality uses the antiderivative property $F' = f$. □

The following example shows how to apply this result.

Example 7.1.4. Find the value of $\int_1^2 e^{-3x^4 + 1} x^3 \, \mathrm{d}x$.

---

[1]The proof shows that f actually only needs to be defined and continuous on the range of u.

[2]The Extreme Value Theorem and the Intermediate Value Theorem imply that the range of u is the bounded interval $[m, M]$ given by $m := \min\{u(x) : x \in [a, b]\}$ and $M := \max\{u(x) : x \in [a, b]\}$.

Solution. Let $u(x) := -3x^4 + 1$ for all $x \in [1, 2]$, so $u'(x) = -12x^3$, $u(1) = -2$ and $u(2) = -47$. Applying the Substitution Formula from Theorem 7.1.3 we obtain

$$\int_1^2 e^{-3x^4+1} x^3 \, \mathrm{d}x = -\frac{1}{12} \int_1^2 e^{u(x)} u'(x) \, \mathrm{d}x$$

$$= -\frac{1}{12} \int_{-2}^{-47} e^x \, \mathrm{d}x$$

$$= -\frac{1}{12} (e^{-47} - e^{-2}).$$

Notice how the order of the endpoints in the transformed integral conforms with Theorem 7.1.3, which in this case meant integrating "backwards" from -2 to -47.

7.2. **Rational functions.** The method of partial fractions allows us to find an antiderivative $\int \frac{p(x)}{q(x)} \, \mathrm{d}x$ of a rational function $\frac{p}{q}$, where p and q are polynomials, as follows:

  (1) If the degree of p is greater than or equal to that of q, then apply polynomial long division to reduce to an expression in which this is not the case.
  (2) Factorise $q(x)$ into linear factors $(ax + b)$ and irreducible quadratic factors $(ax^2 + bx + c)$ which have no real roots.
  (3) Use Table 7.2 to choose the correct partial fraction expansion for each of the factors in $q(x)$ and then determine the appropriate coefficients.
  (4) Use the partial fraction expansion to find an antiderivative of $\frac{p}{q}$.

Factor	Expression	Partial Fraction Expansion
linear	$ax + b$	$\dfrac{A}{ax + b}$
repeated linear	$(ax + b)^N$	$\dfrac{A_1}{ax + b} + \dfrac{A_2}{(ax + b)^2} + \ldots + \dfrac{A_N}{(ax + b)^N}$
quadratic	$ax^2 + bx + c$	$\dfrac{Ax + B}{ax^2 + bx + c}$
repeated quadratic	$(ax^2 + bx + c)^N$	$\dfrac{A_1 x + B_1}{ax^2 + bx + c} + \ldots + \dfrac{A_N x + B_N}{(ax^2 + bx + c)^N}$

TABLE 7.1. Partial Fraction Expansions

The following example demonstrates the method.

Example 7.2.1. Evaluate $\displaystyle\int \frac{x^4 - 2x^2 + 4x + 1}{x^3 - x^2 - x + 1} \, \mathrm{d}x$.

Solution. We use polynomial long division to obtain

$$\frac{x^4 - 2x^2 + 4x + 1}{x^3 - x^2 - x + 1} = (x + 1) + \frac{4x}{x^3 - x^2 - x + 1},$$

so we have

$$\int \frac{x^4 - 2x^2 + 4x + 1}{x^3 - x^2 - x + 1} \, \mathrm{d}x = \frac{x^2}{2} + x + \int \frac{4x}{x^3 - x^2 - x + 1} \, \mathrm{d}x.$$

The factorisation $x^3 - x^2 - x + 1 = (x - 1)^2(x + 1)$ and Table 7.2 show that the correct partial fraction expansion for the last antiderivative above is

$$\frac{4x}{x^3 - x^2 - x + 1} = \frac{A}{x + 1} + \frac{B}{x - 1} + \frac{C}{(x - 1)^2}.$$

To determine the coefficients A, B and C, we add these fractions to obtain

$$4x = A(x - 1)^2 + B(x + 1)(x - 1) + C(x + 1)$$

and equating the coefficients of x^2, x and x^0 gives the simultaneous equations

$$\begin{cases} A + B & = 0 \\ -2A + C & = 4 \\ A - B + C & = 0 \end{cases}$$

which have the solution $A = -1$, $B = 1$ and $C = 2$. Altogether, we have

$$\int \frac{x^4 - 2x^2 + 4x + 1}{x^3 - x^2 - x + 1} \, \mathrm{d}x = \frac{x^2}{2} + x + \int \left(\frac{-1}{x + 1} + \frac{1}{x - 1} + \frac{2}{(x - 1)^2} \right) \mathrm{d}x$$

$$= \frac{x^2}{2} + x - \log|x + 1| + \log|x - 1| - \frac{2}{x - 2}.$$

The following two antiderivatives often arise in the method of partial fractions. If $a > 0$, then using $\frac{\mathrm{d}}{\mathrm{d}x}(\tan^{-1} x) = (x^2 + 1)^{-1}$ and the Chain Rule, we obtain

$$\int \frac{1}{x^2 + a^2} \, \mathrm{d}x = \frac{1}{a} \tan^{-1} \left(\frac{x}{a} \right) \tag{7.2.1}$$

for all $x \in \mathbb{R}$. If $\Omega \subseteq \mathbb{R}$ and $f : \Omega \to \mathbb{R}$ is a differentiable function with continuous derivative f', then using Example 5.2.4 and the Chain Rule, we obtain

$$\int \frac{f'(x)}{f(x)} \, \mathrm{d}x = \log|f(x)| \tag{7.2.2}$$

for all $x \in \Omega$ such that $f(x) \neq 0$. The following example shows how these arise.

Example 7.2.2. Evaluate $\displaystyle\int \frac{x}{x^3 - 1} \, \mathrm{d}x$.

Solution. The factorisation $x^3 - 1 = (x - 1)(x^2 + x + 1)$ can be obtained by observing that 1 is a root of the cubic and then dividing by the factor $(x - 1)$. The quadratic factor $x^2 + x + 1$ is irreducible, because it does not have any real roots, so Table 7.2 shows that the correct partial fraction expansion for this antiderivative is

$$\frac{x}{x^3 - 1} = \frac{A}{x - 1} + \frac{Bx + C}{x^2 + x + 1}.$$

To determine the coefficients A, B and C, we add these fractions to obtain

$$x = A(x^2 + x + 1) + (Bx + C)(x - 1)$$

and equating the coefficients of x^2, x and x^0 shows that $A = \frac{1}{3}$, $B = -\frac{1}{3}$ and $C = \frac{1}{3}$. We now have

$$\int \frac{x}{x^3 - 1} \, \mathrm{d}x = \int \left(\frac{\frac{1}{3}}{x - 1} + \frac{\frac{-x}{3} + \frac{1}{3}}{x^2 + x + 1} \right) \mathrm{d}x = \frac{1}{3} \log|x - 1| - \frac{1}{3} \int \frac{x - 1}{x^2 + x + 1} \, \mathrm{d}x.$$

To find an expression for the remaining antiderivative, we aim to express it as a linear combination of the antiderivatives in (7.2.1) and (7.2.2) with $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) := x^2 + x + 1 = (x + \frac{1}{2})^2 + \frac{3}{4} \neq 0$ for all $x \in \mathbb{R}$. In particular, since $f'(x) = 2x + 1$, our aim is to find two real numbers a and b such that

$$\frac{x - 1}{x^2 + x + 1} = a \frac{f'(x)}{f(x)} + b \frac{1}{f(x)} = \frac{a(2x + 1) + b}{x^2 + x + 1}.$$

This holds provided that $x - 1 = a(2x + 1) + b$, and equating the coefficients of x and x^0 reveals that $2a = 1$ whilst $a + b = -1$, thus $a = \frac{1}{2}$ and $b = -\frac{3}{2}$. We combine the resulting expression with (7.2.1) and (7.2.2) to obtain

$$\int \frac{x - 1}{x^2 + x + 1} \, \mathrm{d}x = \frac{1}{2} \int \frac{2x + 1}{x^2 + x + 1} \, \mathrm{d}x - \frac{3}{2} \int \frac{1}{(x + \frac{1}{2})^2 + \frac{3}{4}} \, \mathrm{d}x$$

$$= \frac{1}{2} \log |x^2 + x + 1| - \frac{3}{2} \frac{2}{\sqrt{3}} \tan^{-1}\left(\frac{2}{\sqrt{3}}\left(x + \frac{1}{2}\right)\right)$$

so finally

$$\int \frac{x}{x^3 - 1} \, \mathrm{d}x = \frac{1}{3} \log |x - 1| - \frac{1}{6} \log |x^2 + x + 1| + \frac{1}{\sqrt{3}} \tan^{-1}\left(\frac{1}{\sqrt{3}}(2x + 1)\right).$$

WEEK 8: INTEGRAL CALCULUS (CONTINUED)

This week we complete our survey of integral calculus techniques by developing strategies for integrating trigonometric functions, considering recursive formulas for antiderivatives also known as reduction formulas, and introducing trigonometric substitutions for integrating irrational functions.

8.1. **Trigonometric functions.** We developing a general strategy for finding the antiderivative $\int \sin^m(x) \cos^n(x) \, \mathrm{d}x$, where m, $n \in \mathbb{N}$, by considering the following three cases:

(1) If n is an odd number, then $n = 2k + 1$ for some $k \in \mathbb{N} \cup \{0\}$, so we substitute $u(x) = \sin(x)$, noting that $u'(x) = \cos(x)$, to obtain

$$
\begin{aligned}
\int \sin^m(x) \cos^{2k+1}(x) \, \mathrm{d}x &= \int \sin^m(x) \cos^{2k}(x) \cos(x) \, \mathrm{d}x \\
&= \int \sin^m(x)(1 - \sin^2(x))^k \cos(x) \, \mathrm{d}x \\
&= \int u^m (1 - u^2)^k \, \mathrm{d}u,
\end{aligned}
$$

where we used $\cos^2(x) = 1 - \sin^2(x)$ in the second line. This reduces matters to finding the antiderivative of a polynomial.

(2) If m is an odd number, then $m = 2k + 1$ for some $k \in \mathbb{N} \cup \{0\}$, so we substitute $u(x) = \cos(x)$, noting that $u'(x) = -\sin(x)$, to obtain

$$
\begin{aligned}
\int \sin^{2k+1}(x) \cos^n(x) \, \mathrm{d}x &= \int \sin^{2k}(x) \cos^n(x) \sin(x) \, \mathrm{d}x \\
&= -\int (1 - \cos^2(x))^k \cos^n(x)(-\sin(x)) \, \mathrm{d}x \\
&= -\int (1 - u^2)^k u^n \, \mathrm{d}u,
\end{aligned}
$$

where we used $\cos^2(x) = 1 - \sin^2(x)$ in the second line. This reduces matters to finding the antiderivative of a polynomial.

(3) If n and m are both even numbers, then the double-angle identities below can be used either directly or to reduce to the above cases:

$$
\begin{aligned}
\sin^2(x) &= \tfrac{1}{2}(1 - \cos(2x)) \\
\cos^2(x) &= \tfrac{1}{2}(1 + \cos(2x)) \\
\sin(x)\cos(x) &= \tfrac{1}{2}\sin(2x)
\end{aligned}
\tag{8.1.1}
$$

The key idea in this approach is to isolate odd powers of the trigonometric functions and then use appropriate substitutions to reduce to polynomial antiderivatives. The method can also be applied sometimes when n and m are integers, rather than just natural numbers, as the following example shows.

Example 8.1.1. Find the following antiderivatives:

$$
(1) \int \cos^3(x) \, \mathrm{d}x \qquad (2) \int \frac{\sin^7(x)}{\cos^4(x)} \, \mathrm{d}x \qquad (3) \int \cos^4(x) \, \mathrm{d}x.
$$

Solution. (1) We substitute $u(x) = \sin(x)$, noting that $u'(x) = \cos(x)$, to obtain

$$\int \cos^3(x)\ dx = \int (1 - \sin^2(x))\cos(x)\ dx$$
$$= \int (1 - u^2)\ du$$
$$= \sin(x) - \tfrac{1}{3}\sin^3(x).$$

(2) We substitute $u(x) = \cos(x)$, noting that $u'(x) = -\sin(x)$, to obtain

$$\int \frac{\sin^7(x)}{\cos^4(x)}\ dx = \int \sin^6(x)\cos^{-4}(x)\sin(x)\ dx$$
$$= -\int (1 - \cos^2(x))^3 \cos^{-4}(x)(-\sin(x))\ dx$$
$$= -\int (1 - u^2)^3 u^{-4}\ du$$
$$= -\int (1 - 3u^2 + 3u^4 - u^6)u^{-4}\ du$$
$$= \int (-u^{-4} + 3u^{-2} - 3 + u^2)\ du$$
$$= \frac{1}{3\cos^3(x)} - \frac{3}{\cos(x)} - 3\cos(x) + \tfrac{1}{3}\cos^3(x).$$

(3) We use a double-angle identity in (8.1.1) to obtain

$$\int \cos^4(x)\ dx = \int (\tfrac{1}{2}[1 + \cos(2x)])^2\ dx$$
$$= \tfrac{1}{4}\int (1 + 2\cos(2x) + \cos^2(2x))\ dx$$
$$= \tfrac{1}{4}\int (1 + 2\cos(2x) + \tfrac{1}{2}[1 + \cos(4x)])\ dx$$
$$= \tfrac{1}{4}\int (\tfrac{3}{2} + 2\cos(2x) + \tfrac{1}{2}\cos(4x))\ dx$$
$$= \tfrac{1}{4}(\tfrac{3}{2}x + 2\tfrac{1}{2}\sin(2x) + \tfrac{1}{2}\tfrac{1}{4}\sin(4x))$$
$$= \tfrac{3}{8}x + \tfrac{1}{4}\sin(2x) + \tfrac{1}{32}\sin(4x).$$

The following identities somewhat mirror the relation between sine and cosine functions that made the above method work and thus suggest a similar strategy for finding antiderivatives of products of secant and tangent functions:

$$\frac{d}{dx}(\tan(x)) = \sec^2(x)$$
$$\frac{d}{dx}(\sec(x)) = \sec(x)\tan(x)$$
$$\tan^2(x) + 1 = \sec^2(x)$$

We developing a general strategy for finding the antiderivative $\int \tan^m(x)\sec^n(x)\ dx$, where $m, n \in \mathbb{N}$, by considering the following two cases:

(1) If n is a even number, then $n = 2k$ for some $k \in \mathbb{N}$, so we substitute $u(x) = \tan(x)$, noting that $u'(x) = \sec^2(x)$, to obtain

$$\int \tan^m(x) \sec^{2k}(x) \, \mathrm{d}x = \int \tan^m(x) \sec^{2k-2}(x) \sec^2(x) \, \mathrm{d}x$$

$$= \int \tan^m(x)(\tan^2(x) + 1)^{k-1} \sec^2(x) \, \mathrm{d}x$$

$$= \int u^m(u^2 + 1)^{k-1} \, \mathrm{d}u,$$

where we used $\sec^2(x) = \tan^2(x) + 1$ in the second line. This reduces matters to finding the antiderivative of a polynomial.

(2) If m is an odd number, then $m = 2k + 1$ for some $k \in \mathbb{N} \cup \{0\}$, so we substitute $u(x) = \sec(x)$, noting that $u'(x) = \sec(x)\tan(x)$, to obtain

$$\int \tan^{2k+1}(x) \sec^n(x) \, \mathrm{d}x = \int \tan^{2k}(x) \sec^{n-1}(x) \sec(x) \tan(x) \, \mathrm{d}x$$

$$= \int (\sec^2(x) - 1)^k \sec^{n-1}(x) \sec(x) \tan(x) \, \mathrm{d}x$$

$$= \int (u^2 - 1)^k u^{n-1} \, \mathrm{d}u,$$

where we used $\tan^2(x) = \sec^2(x) - 1$ in the second line. This reduces matters to finding the antiderivative of a polynomial.

In the remaining case where n is an odd number and m is an even number, there does not seem to be a general strategy that always works. We often need to think creatively, as the following example shows.

Example 8.1.2. Find the following antiderivatives:

$$(1) \int \tan(x) \, \mathrm{d}x \qquad\qquad (2) \int \sec(x) \, \mathrm{d}x \qquad\qquad (3) \int \sec^3(x) \, \mathrm{d}x.$$

Solution. (1) We substitute $u(x) = \cos(x)$, noting that $u'(x) = -\sin(x)$, to obtain

$$\int \tan(x) \, \mathrm{d}x = \int \frac{\sin(x)}{\cos(x)} \, \mathrm{d}x$$

$$= \int \frac{-1}{u} \, \mathrm{d}u$$

$$= -\log|\cos(x)|$$

$$= \log|\sec(x)|.$$

(2) We use (7.2.2) with $f(x) := \sec(x) + \tan(x)$, and $f'(x) = \sec(x)\tan(x) + \sec^2(x)$, to obtain

$$\int \sec(x) \, \mathrm{d}x = \int \sec(x) \frac{\sec(x) + \tan(x)}{\sec(x) + \tan(x)} \, \mathrm{d}x$$

$$= \int \frac{\sec^2(x) + \sec(x)\tan(x)}{\sec(x) + \tan(x)} \, \mathrm{d}x$$

$$= \int \frac{f'(x)}{f(x)} \, \mathrm{d}x$$

$$= \log|\sec(x) + \tan(x)|.$$

(3) We apply Integration by Parts with

$$\begin{cases} u(x) = \sec(x) \\ v(x) = \tan(x) \end{cases} \quad \text{and} \quad \begin{cases} u'(x) = \sec(x)\tan(x) \\ v'(x) = \sec^2(x) \end{cases}$$

to obtain

$$\int \sec^3(x) \, dx = \sec(x)\tan(x) - \int \sec(x)\tan^2(x) \, dx$$

$$= \sec(x)\tan(x) - \int \sec(x)(\sec^2(x) - 1) \, dx$$

$$= \sec(x)\tan(x) - \int \sec^3(x) \, dx + \int \sec(x) \, dx.$$

We now use the antiderivate from part (2) and combine terms to obtain

$$\int \sec^3(x) \, dx = \tfrac{1}{2}\sec(x)\tan(x) + \tfrac{1}{2}\log|\sec(x) + \tan(x)|.$$

This is an example of a reduction formula which we will consider in the next lecture.

We conclude this lecture with an application to household electronics.

Example 8.1.3. Mains electricity in the UK is supplied as alternating current rated as 230 V at 50 Hz. The means that the voltage alternates through 50 cycles per second with a root-mean-square voltage of 230 V. We can model the voltage using the function

$$V(t) := V_{\text{peak}} \sin(\tfrac{2\pi}{T} t),$$

where $V(t)$ is the voltage (measured in volts) at time t (measured in seconds), whilst T is the length of one period in seconds and V_{peak} is the peak voltage of the supply. The peak voltage is not the same as the root-mean-square voltage ($V_{\text{rms}} = 230$ V) but we can calculate it since the latter is defined by the following formula:

$$V_{\text{rms}} := \left(\frac{1}{T} \int_0^T V(t)^2 \, dt \right)^{1/2}.$$

In particular, we a double-angle identity in (8.1.1), to obtain

$$(V_{\text{rms}})^2 = 50 \int_0^{\frac{1}{50}} V_{\text{peak}}^2 \sin^2(100\pi t) \, dt$$

$$= 50 V_{\text{peak}}^2 \int_0^{\frac{1}{50}} \tfrac{1}{2}[1 - \cos(200\pi t)] \, dt \quad = 25 V_{\text{peak}}^2 \left[t - \frac{1}{200\pi}\sin(200\pi t) \right]_1^{\frac{1}{50}}$$

$$= \tfrac{1}{2} V_{\text{peak}}^2,$$

hence $V_{\text{peak}} = \sqrt{2}\, V_{\text{rms}} = \sqrt{2}\,(230)$ V ≈ 325 V.

8.2. **Reduction formulas and bijective substitution.** An inspection of the method used to find the antiderivative in Example 8.1.2(3) shows that it can be extended to obtain a recursive formula for finding an antiderivative of any higher integer power of the secant function. Integration by Parts is often used to derive such reduction formulas, as the following example shows.

Example 8.2.1. For each $n \in \mathbb{N} \cup \{0\}$, define $S_n := \int \sin^n(x) \, dx$. Prove that

$$nS_n(x) = -\sin^{n-1}(x)\cos(x) + (n-1)S_{n-2}(x)$$

for all $x \in \mathbb{R}$ and all integers $n \geq 2$, and use this formula to find $\int \sin^3(x) \, dx$.

Solution. If $n \geq 2$ is an integer, then we apply Integration by Parts with

$$\begin{cases} u(x) = \sin^{n-1}(x) \\ v(x) = -\cos(x) \end{cases} \quad \text{and} \quad \begin{cases} u'(x) = (n-1)\sin^{n-2}(x)\cos(x) \\ v'(x) = \sin(x) \end{cases}$$

to obtain

$$S_n(x) = -\sin^{n-1}(x)\cos(x) - (n-1)\int \sin^{n-2}(x)\cos(x)(-\cos(x))\ \mathrm{d}x$$

$$= -\sin^{n-1}(x)\cos(x) + (n-1)\int \sin^{n-2}(x)(1-\sin^2(x))\ \mathrm{d}x$$

$$= -\sin^{n-1}(x)\cos(x) + (n-1)(S_{n-2}(x) - S_n(x))$$

We now combine terms to obtain

$$nS_n(x) = (1+n-1)S_n(x) = -\sin^{n-1}(x)\cos(x) + (n-1)S_{n-2}(x).$$

This allows us to calculate

$$\int \sin^3(x)\ \mathrm{d}x = S_3(x) = -\tfrac{1}{3}\sin^2(x)\cos(x) + \tfrac{2}{3}S_1(x) = -\tfrac{1}{3}\sin^2(x)\cos(x) - \tfrac{2}{3}\cos(x),$$

as required.

In the next part of this lecture, we introduce some trigonometric substitutions which can be used to transform certain irrational functions into rational functions for the purposes of integration. The following theorem, which is a reprise of the Substitution Formula from Theorem 7.1.3 in the case of a bijective substitution, will allow us to justify such substitutions. It is not important to be familiar with this proof for the assessment in this module.

Theorem 8.2.2 (Bijective Substitution Formula)**.** If $f : [a, b] \to \mathbb{R}$ is a continuous function and $g : [c, d] \to [a, b]$ is bijective and differentiable with bounded integrable derivative g', then

$$\int_a^b f(x)\ \mathrm{d}x = \int_{g^{-1}(a)}^{g^{-1}(b)} f(g(\theta))g'(\theta)\ \mathrm{d}\theta.$$

Proof. The function g is a bijection, so we may apply the Substitution Formula from Theorem 7.1.3 with (a, b, u) replaced by $(g^{-1}(a), g^{-1}(b), g)$ to obtain

$$\int_a^b f(x)\ \mathrm{d}x = \int_{g(g^{-1}(a))}^{g(g^{-1}(b))} f(x)\ \mathrm{d}x$$

$$= \int_{g^{-1}(a)}^{g^{-1}(b)} f(g(\theta))g'(\theta)\ \mathrm{d}\theta,$$

as required. $\qquad\square$

Table 8.2 lists three trigonometric substitutions which can be use to transform an integral with respect to x into an integral with respect to θ by writing $x = g(\theta)$. These are useful because they transform the irrational expressions in the left column into rational expressions by noting the identities in the right column. It is important to recognise that the Bijective Substitution Formula in Theorem 8.2.2 justifies the formal algebraic manipulation "$\mathrm{d}x = \frac{\mathrm{d}x}{\mathrm{d}\theta}\ \mathrm{d}\theta$" associated with these changes of variable *provided* that g, or specifically the mapping $\theta \mapsto x = g(\theta)$, is a *bijection*.

Example 8.2.3. Find the area of the ellipse enclosed by the equation $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$.

Solution. The equation corresponds to an ellipse centre at the origin in the x-y plane with vertices at $(\pm a, 0)$ and $(0, \pm b)$. Using symmetry, the total area enclosed by the ellipse is equal to

$$4\int_0^a \tfrac{b}{a}\sqrt{a^2 - x^2}\ \mathrm{d}x$$

Irrational Term	Trigonometric Substitution $x = g(\theta)$	Rationalising Identity
$\sqrt{a^2 - x^2}$	$x = a\sin(\theta)$	$1 - \sin^2(\theta) = \cos^2(\theta)$
$\sqrt{x^2 + a^2}$	$x = a\tan(\theta)$	$\tan^2(\theta) + 1 = \sec^2(\theta)$
$\sqrt{x^2 - a^2}$	$x = a\sec(\theta)$	$\sec^2(\theta) - 1 = \tan^2(\theta)$

TABLE 8.2. Trigonometric Substitutions for Rationalising Integrands

We follow Table 8.2, making the substitution $x = a\sin(\theta)$, and apply the Bijective Substitution Formula in Theorem 8.2.2 with the bijection $g : [0, \frac{\pi}{2}] \to [0, a]$ given by $g(\theta) := a\sin(\theta)$, noting that $g'(\theta) = a\cos(\theta)$, $g(0) = 0$ and $g(\frac{\pi}{2}) = a$, to obtain

$$4\int_0^a \frac{b}{a}\sqrt{a^2 - x^2}\, \mathrm{d}x = \frac{4b}{a}\int_0^{\frac{\pi}{2}} \sqrt{a^2(1 - \sin^2(\theta))}\, a\cos(\theta)\, \mathrm{d}\theta$$

$$= 4ab\int_0^{\frac{\pi}{2}} \cos^2(\theta)\, \mathrm{d}\theta$$

$$= \pi ab,$$

where we used a double-angle identity from (8.1.1) to obtain the final equality.

The Bijective Substitution Formula can also be used to find antiderivatives. It is good practice, and required in the assessment for this module, to remove any inverse trigonometric expressions from such antiderivatives. This is because there is no widespread agreement on the domains for inverse trigonometric functions. The following example shows how to achieve this by using some trigonometry.

Example 8.2.4. Find the following antiderivatives

$$(1) \int \frac{1}{x^2\sqrt{x^2 + 4}}\, \mathrm{d}x \qquad\qquad (2) \int \frac{x}{\sqrt{x^2 + x + 1}}\, \mathrm{d}x.$$

Solution. (1) We follow Table 8.2, making the substitution $x = 2\tan(\theta)$ and noting that $g'(\theta) = 2\sec^2(\theta)$, to obtain

$$\int \frac{1}{x^2\sqrt{x^2 + 4}}\, \mathrm{d}x = \int \frac{2\sec^2(\theta)}{4\tan^2(\theta)\sqrt{4(\tan^2(\theta) + 1)}}\, \mathrm{d}\theta$$

$$= \int \frac{\sec(\theta)}{4\tan^2(\theta)}\, \mathrm{d}\theta$$

$$= \int \frac{\cos(\theta)}{4\sin^2(\theta)}\, \mathrm{d}\theta$$

$$= \int \frac{1}{4u^2}\, \mathrm{d}u$$

$$= \frac{-1}{4\sin(\theta)},$$

where we use $\tan(\theta) = \sin(\theta)/\cos(\theta)$ in the third line, and made the substitution $u(\theta) = \sin(\theta)$ with $u'(\theta) = \cos(\theta)$ in the fourth line. To conclude, we avoid writing $\sin(\theta) = \sin(\tan^{-1}(\frac{x}{2}))$ by using trigonometry to reveal the relationship between $\sin(\theta)$ and $\tan(\theta)$. In particular, we have $\tan(\theta) = \frac{x}{2} = \frac{\text{opp}}{\text{adj}}$, where we regard θ as the acute angle in a right-angled triangle with opposite side-length x and adjacent side-length 2, hence $\sin(\theta) = \frac{\text{opp}}{\text{hyp}} = \frac{x}{\sqrt{x^2+4}}$, where we used Pythagoras' Theorem to

calculate the hypotenuse side-length $\sqrt{x^2+4}$. We thus have

$$\int \frac{1}{x^2\sqrt{x^2+4}}\,\mathrm{d}x = \frac{-\sqrt{x^2+4}}{4x}.$$

(2) We complete the square and make the substitution $y = x + \frac{1}{2}$, to obtain

$$\int \frac{x}{\sqrt{x^2+x+1}}\,\mathrm{d}x = \int \frac{x}{\sqrt{(x+\frac{1}{2})^2+\frac{3}{4}}}\,\mathrm{d}x$$

$$= \int \frac{y-\frac{1}{2}}{\sqrt{y^2+\frac{3}{4}}}\,\mathrm{d}y$$

$$= \int \frac{y}{\sqrt{y^2+\frac{3}{4}}}\,\mathrm{d}y + \int \frac{-\frac{1}{2}}{\sqrt{y^2+\frac{3}{4}}}\,\mathrm{d}y$$

$$=: I + J$$

The term I can be found using the substitution $u = y^2 + \frac{3}{4}$ as

$$I = \int \tfrac{1}{2}u^{-\frac{1}{2}}\,\mathrm{d}u = u^{\frac{1}{2}} = (y^2+\tfrac{3}{4})^{\frac{1}{2}} = \sqrt{x^2+x+1}$$

The term II can be found using the substitution $y = \frac{\sqrt{3}}{2}\tan(\theta)$ as

$$J = -\tfrac{1}{2}\int \frac{\frac{\sqrt{3}}{2}\sec^2(\theta)}{\frac{\sqrt{3}}{2}\sec(\theta)}\,\mathrm{d}\theta$$

$$= -\tfrac{1}{2}\int \sec(\theta)\,\mathrm{d}\theta$$

$$= -\tfrac{1}{2}\log|\sec(\theta)+\tan(\theta)|,$$

where we used (8.1.2) to obtain the final equality. Using trigonometry, we have $\tan(\theta) = \frac{y}{\sqrt{3}/2} = \frac{\text{opp}}{\text{adj}}$, hence $\sec(\theta) = \frac{\text{hyp}}{\text{adj}} = \frac{\sqrt{y^2+3/4}}{\sqrt{3}/2}$. We thus have

$$J = -\tfrac{1}{2}\log|\tfrac{2}{\sqrt{3}}(\sqrt{x^2+x+1}+x+\tfrac{1}{2})| = -\tfrac{1}{2}\log|\sqrt{x^2+x+1}+x+\tfrac{1}{2}| - \tfrac{1}{2}\log(\tfrac{2}{\sqrt{3}}).$$

We can choose to ignore the final additive constant above, since we only need to find an antiderivative, in which case

$$\int \frac{x}{\sqrt{x^2+x+1}}\,\mathrm{d}x = I + J = \sqrt{x^2+x+1} - \tfrac{1}{2}\log|\sqrt{x^2+x+1}+x+\tfrac{1}{2}|.$$

Week 9: Improper Integrals

This week we return to the theory of integration to define improper integrals for bounded functions on *unbounded* intervals and for *unbounded* functions on bounded intervals.

9.1. **Unbounded intervals.** The Riemann–Darboux theory of integration is concerned with bounded functions on bounded intervals. It is possible, however, for the area beneath a bounded function on an unbounded domain to be finite. For example, if $A(t)$ denotes the area beneath the bounded function $f(x) := \frac{1}{x^2}$ on the interval $[1,t]$, where $t > 1$, then $A(t) = \int_1^t \frac{1}{x^2}\,\mathrm{d}x = 1 - \frac{1}{t}$, and we notice that $\lim_{t\to\infty} A(t) = 1$. In this way, we interpret the area beneath f on the *unbounded* interval $[1,\infty)$ as a limit of Riemann–Darboux integrals on bounded intervals $[1,t]$. The limit is finite in this particular example because $\frac{1}{x^2}$ decays sufficiently fast as x approaches infinity. This motivates the following definition of improper integrals

for bounded functions on semi-infinite domains of the form $[a, \infty)$ and $(-\infty, b]$, as well as on $\mathbb{R}$.

Definition 9.1.1. The following *improper integrals* are called *convergent* when they exist in $\mathbb{R}$ (otherwise they are called *divergent*):

(1) If $f : [a, \infty) \to \mathbb{R}$ is bounded and integrable on all bounded intervals $[a, t]$, where $t > a$, then the improper integral $\int_a^\infty f$ is defined by

$$\int_a^\infty f := \lim_{t \to \infty} \left(\int_a^t f \right)$$

whenever the limit exists in $\overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty, +\infty\}$.
(2) If $f : (-\infty, b] \to \mathbb{R}$ is bounded and integrable on all bounded intervals $[t, b]$, where $t < b$, then the improper integral $\int_{-\infty}^b f$ is defined by

$$\int_{-\infty}^b f := \lim_{t \to -\infty} \left(\int_t^b f \right)$$

whenever the limit exists in $\overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty, +\infty\}$.
(3) If $f : \mathbb{R} \to \mathbb{R}$ is bounded and there exists $c \in \mathbb{R}$ such that both improper integrals $\int_{-\infty}^c f$ and $\int_c^\infty f$ are convergent, then the improper integral $\int_{-\infty}^\infty f$ is the real number defined by $\int_{-\infty}^\infty f := \int_{-\infty}^c f + \int_c^\infty f$.

The following example shows how to calculate these types of improper integral. The main point is to work in two stages, first computing the integral on a bounded domain, and then taking an appropriate limit.

Example 9.1.2. Calculate the following improper integrals or prove that they are divergent:

$$(1) \int_1^\infty \frac{1}{x} \, dx \qquad\qquad (2) \int_{-\infty}^0 x e^x \, dx \qquad\qquad (3) \int_{-\infty}^\infty e^{-|x|} \, dx$$

Solution. (1) We calculate

$$\begin{aligned}
\int_1^\infty \frac{1}{x} \, dx &= \lim_{t \to \infty} \left(\int_1^t \frac{1}{x} \, dx \right) \\
&= \lim_{t \to \infty} \left[\log |x| \right]_1^t \\
&= \lim_{t \to \infty} \log(t) \\
&= \infty,
\end{aligned}$$

hence this improper integral is divergent.
(2) We apply Integration by Parts with

$$\begin{cases} u(x) = x \\ v(x) = e^{(}x) \end{cases} \quad \text{and} \quad \begin{cases} u'(x) = 1 \\ v'(x) = e^{(}x) \end{cases}$$

to calculate

$$\lim_{t \to -\infty} \left(\int_t^0 x e^x \right) = \lim_{t \to -\infty} \left(\left[x e^x \right]_t^0 - \int_t^0 e^x \right)$$

$$= \lim_{t \to -\infty} \left(0 - t e^t - \left[e^x \right]_t^0 \right)$$

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

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

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

$$= -1,$$

where we used the Algebra of Limits and L'Hôpital's Rule on the indeterminate form $-t/e^{-t}$ to obtain the fourth equality. It is important to note here how the Integration by Parts was performed on the interval $[t, 0]$ *before* the limit was taken.

(3) We calculate

$$\int_0^\infty e^{-|x|} \, dx = \lim_{t \to \infty} \left(\int_0^t e^{-x} \, dx \right) = \lim_{t \to \infty} \left(-e^{-t} + 1 \right) = 1,$$

which is convergent, and

$$\int_\infty^0 e^{-|x|} \, dx = \lim_{t \to -\infty} \left(\int_t^0 e^x \, dx \right) = \lim_{t \to -\infty} \left(1 - e^t \right) = 1,$$

which is also convergent. This proves that $\int_{-\infty}^\infty e^{-|x|} \, dx$ is convergent with

$$\int_{-\infty}^\infty e^{-|x|} \, dx = \int_0^\infty e^{-|x|} \, dx + \int_\infty^0 e^{-|x|} \, dx = 2.$$

It is important to note here how the improper integral on $\mathbb{R}$ needed to be split into two improper integrals on semi-infinite domains. In particular, an improper integral on $\mathbb{R}$ should *not* be calculated as single combined limit, such as $\lim_{t \to \infty} \int_{-t}^t f$, as this is *not* in accordance with (3) in Definition 9.1.1

9.2. Unbounded functions. If $f : [a, b] \to \mathbb{R}$ is an unbounded function, then the sums $L(f, P) = \sum_{i=1}^n m_i (x_i - x_{i-1})$ and $U(f, P) = \sum_{i=1}^n M_i (x_i - x_{i-1})$ may not be well-defined, because $m_i = -\infty$ on any subinterval $[x_{i-1}, x_i] \subseteq [a, b]$ where the function is not bounded below, and $M_i = \infty$ on any subinterval where the function is not bounded above. If the function is bounded except for some asymptotic behaviour at certain points, however, then the Riemann–Darboux integral can be defined on subintervals that exclude those singular points. For example, the function $f(x) := \frac{1}{\sqrt{x}}$ is not bounded on $(0, 1]$, but it is bounded on $(\delta, 1]$ with $\int_\delta^1 \frac{1}{\sqrt{x}} \, dx = 2 - 2\sqrt{\delta}$ for all $\delta \in (0, 1]$, and $\lim_{\delta \to 0} \int_\delta^1 \frac{1}{\sqrt{x}} \, dx = 2$. The improper integrals below are defined by taking such limits of Riemann–Darboux integrals.

Definition 9.2.1. The following *improper integrals* are called *convergent* when they exist in $\mathbb{R}$ (otherwise they are called *divergent*):

(1) If $f : (a, b] \to \mathbb{R}$ is bounded and integrable on all intervals $[a + \delta, b] \subset (a, b]$, where $\delta > 0$, then the improper integral $\int_a^b f$ is defined by

$$\int_a^b f := \lim_{\delta \to 0^+} \left(\int_{a+\delta}^b f \right)$$

whenever the limits exists in $\overline{\overline{\mathbb{R}}} := \mathbb{R} \cup \{-\infty, +\infty\}$.

(2) If $f : [a, b) \to \mathbb{R}$ is bounded and integrable on all intervals $[a, b - \delta] \subset [a, b)$, where $\delta > 0$, then the improper integral $\int_a^b f$ is defined by

$$\int_a^b f := \lim_{\delta \to 0^+} \left(\int_a^{b-\delta} f \right)$$

whenever the limit exists in $\overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty, +\infty\}$.

(3) If $f : [a, c) \cup (c, b] \to \mathbb{R}$ is bounded and integrable on all closed subintervals of $[a, c) \cup (c, b]$, and both improper integrals $\int_a^c f$ and $\int_c^b f$ are convergent, then the improper integral $\int_a^b f$ is the real number defined by $\int_a^b f := \int_a^c f + \int_c^b f$.

The following example shows how to calculate these types of improper integral.

Example 9.2.2. Calculate the following improper integrals or prove that they are divergent:

$$(1) \int_2^5 \frac{1}{\sqrt{x-2}} \; dx \qquad (2) \int_0^{\frac{\pi}{2}} \sec(x) \; dx \qquad (3) \int_0^3 \frac{1}{x-1} \; dx$$

Solution. (1) The integrand has a singularity at $x = 2$, so we calculate

$$\int_2^5 \frac{1}{\sqrt{x-2}} \; dx = \lim_{\delta \to 0^+} \left(\int_{2+\delta}^5 (x-2)^{-1/2} \; dx \right)$$
$$= \lim_{\delta \to 0^+} \left[2(x-2)^{1/2} \right]_{2+\delta}^5$$
$$= \lim_{\delta \to 0^+} (2\sqrt{3} - 2\sqrt{\delta})$$
$$= 2\sqrt{3}.$$

(2) The integrand has a singularity at $x = \frac{\pi}{2}$, so we calculate

$$\int_0^{\frac{\pi}{2}} \sec(x) \; dx = \lim_{\delta \to 0^+} \left(\int_0^{\frac{\pi}{2}-\delta} \sec(x) \; dx \right)$$
$$= \lim_{\delta \to 0^+} \left[\log|\sec(x) + \tan(x)| \right]_0^{\frac{\pi}{2}-\delta}$$
$$= \lim_{\delta \to 0^+} \log|\sec(\tfrac{\pi}{2} - \delta) + \tan(\tfrac{\pi}{2} - \delta)|$$
$$= \infty,$$

hence this improper integral is divergent.

(3) The integrand has a singularity at $x = 1$, so we calculate

$$\int_0^1 \frac{1}{x-1} \; dx = \lim_{\delta \to 0^+} \left(\int_0^{1-\delta} \frac{1}{x-1} \; dx \right)$$
$$= \lim_{\delta \to 0^+} \left[\log|x-1| \right]_0^{1-\delta}$$
$$= \lim_{\delta \to 0^+} \log(\delta)$$
$$= -\infty,$$

hence this improper integral is divergent. It is important to note here how treating this as an (ordinary) integral, that is by writing $\int_0^3 \frac{1}{x-1} \; dx = [\log|x-1|]_0^3 = \log(2)$, gives the *wrong* answer.

The following result provides a convergence test for improper integrals.

Theorem 9.2.3 (Integral Comparison Test). If $f : [a, \infty) \to \mathbb{R}$ and $g : [a, \infty) \to \mathbb{R}$ are bounded functions such that $0 \leq f(x) \leq g(x)$ for all $x \in [a, \infty)$, then the following implications hold:

(1) If $\int_a^\infty g$ is convergent, then $\int_a^\infty f$ is convergent.

(2) If $\int_a^\infty f$ is divergent, then $\int_a^\infty g$ is divergent.

Proof. The improper integrals $\int_a^\infty f$ and $\int_a^\infty g$ will both exist in $[0, +\infty]$. This is because $f(x) \geq 0$ and $g(x) \geq 0$ for all $x \in [a, \infty)$, which implies that the values of both $\int_a^t f$ and $\int_a^t g$ increase as t increases, hence $\lim_{t \to \infty}(\int_a^t f)$ and $\lim_{t \to -\infty}(\int_t^b f)$ will be either finite and non-negative or equal to $+\infty$.

The assumption that $f(x) \leq g(x)$ for all $x \in [a, \infty)$ implies that $\int_a^t f \leq \int_a^t g$ for all $t \in (a, \infty)$. It is a good exercise to write out a proof of this fact by considering $\int_a^t (g - f)$ and combining the estimates for the integral in Theorem 4.2.8 with the linearity properties of the integral in Theorem 4.2.2.

The above facts justify the following implications: (1) If $\int_a^\infty g$ is convergent, then

$$0 \leq \int_a^\infty f = \lim_{t \to \infty}\left(\int_a^t f\right) \leq \lim_{t \to \infty}\left(\int_a^t g\right) = \int_a^\infty g < \infty,$$

hence $\int_a^\infty f$ is convergent.

(2) If $\int_a^\infty f$ is divergent, then

$$\infty = \int_a^\infty f = \lim_{t \to \infty}\left(\int_a^t f\right) \leq \lim_{t \to \infty}\left(\int_a^t g\right) = \int_a^\infty g,$$

hence $\int_a^\infty g$ is divergent. $\qquad\square$

There are similar versions of the Integral Comparison Test for improper integrals on intervals of the form $(-\infty, b]$, and also for improper integrals of unbounded functions on domains $(a, b]$ and $[a, b)$ as in Deifinition 9.2.1. We leave the details to the interested reader. We conclude with an example of how to apply this result.

Example 9.2.4. Determine whether $\displaystyle\int_1^\infty \frac{1 + e^{-x}}{x}\,dx$ is convergent or divergent.

Solution. This improper integral is divergent by the Integral Comparison Test, since

$$\frac{1 + e^{-x}}{x} \geq \frac{1}{x} \geq 0$$

for all $x \in [1, \infty)$, and $\int_1^\infty \frac{1}{x}\,dx$ is divergent by Example 9.1.2.

WEEK 10: FIRST-ORDER ODEs AND APPLICATIONS

This week we develop strategies to solve first-order ordinary differential equations. Direct integration solves the equation $y' = f(x)$, the method of separation of variables applies when $y' = f(x)g(y)$, and an integration factor is used to solve linear equations $y' + g(x)y = f(x)$. We will also see how such equations can be used to model mixing problems in engineering applications.

10.1. **Separable equations.** We will typically be solving differential equations to find a function y in terms of an independent variable x (so $x \mapsto y(x)$). We use $y' = \frac{dy}{dx}$ and $y'' = \frac{d^2y}{dx^2}$ to denote the first and second derivatives of y. We begin by making precise the meaning of a solution to a first-order differential equation and associated initial value problems.

Definition 10.1.1. Suppose that $f : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ and that $I \subseteq \mathbb{R}$ is an open interval. A *solution* of the (first-order) differential equation

$$y' = f(x, y) \text{ on } I$$

is a differentiable function $y : I \to \mathbb{R}$ such that $y'(x) = f(x, y(x))$ for all $x \in I$.

We can extend this definition to the case when $f : \Omega_1 \times \Omega_2 \to \mathbb{R}$, where $\Omega_1 \subseteq \mathbb{R}$ and $\Omega_2 \subseteq \mathbb{R}$, by requiring that $I \subseteq \Omega_1$ and that $y : I \to \Omega_2$. We typically need some solution data to specify a unique solution of a first-order equation, which leads us to consider initial value problems.

Definition 10.1.2. Suppose that $f : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$. Let $x_0, y_0 \in \mathbb{R}$ and $x_1 \in (x_0, \infty]$. A *solution* of the initial value problem

$$y' = f(x, y), \quad y(x_0) = y_0$$

is a continuous function $y : [x_0, x_1) \to \mathbb{R}$ that is differentiable on (x_0, x_1) such that $y'(x) = f(x, y(x))$ for all $x \in (x_0, x_1)$, and such that $y(x_0) = y_0$.

In other words, a solution of the initial value problem is a solution of the ordinary differential equation on (x_0, x_1) that is also continuous on $[x_0, x_1)$ with $y(x_0) = y_0$. We can also extend this definition to the case when $f : \Omega_1 \times \Omega_2 \to \mathbb{R}$, where $\Omega_1 \subseteq \mathbb{R}$ and $\Omega_2 \subseteq \mathbb{R}$, by requiring that $(x_0, x_1) \subseteq \Omega_1$ and that $y : (x_0, x_1) \to \Omega_2$.

The simplest first-order equation $y' = f(x)$, where f is function of x only, can be solved by direct integration when f is continuous. Moreover, we have the explicit expression below for the solution of initial value problems on $[x_0, \infty)$.

Theorem 10.1.3. If $x_0, y_0 \in \mathbb{R}$ and $f : [x_0, \infty) \to \mathbb{R}$ is continuous, then the initial value problem

$$y' = f(x), \quad y(x_0) = y_0$$

has the solution $y : [x_0, \infty) \to \mathbb{R}$ given by $y(x) := \int_{x_0}^{x} f + y_0$ for all $x \in [x_0, \infty)$.

Proof. The function f is continuous, so the first Fundamental Theorem of Calculus implies that $y(x) := \int_{x_0}^{x} f + y_0$ defines a differentiable function with derivative $y'(x) = f(x)$ for all $x \in [x_0, \infty)$, hence $y : [x_0, \infty) \to \mathbb{R}$ is a continuous function, and initial value $y(x_0) = \int_{x_0}^{x_0} f + y_0 = y_0$, as required. $\square$

Here is an example of how to apply this theorem. In general, however, there is no need to quote this theorem, and you may integrate directly to obtain a solution.

Example 10.1.4. Find a solution $y : [0, \infty) \to \mathbb{R}$ of the initial value problem

$$y' = \frac{x}{\sqrt{x^2 + x + 1}}, \quad y(0) = 3.$$

Solution. We combine Theorem 10.1.3 with the antiderivative from Example 8.2.4(2) and the second Fundamental Theorem of Calculus to obtain the solution

$$
\begin{aligned}
y(x) :&= \int_0^x \frac{t}{\sqrt{t^2+t+1}}\,\mathrm{d}t + 3 \\
&= \left[\sqrt{t^2+t+1} - \tfrac{1}{2}\log\left|\sqrt{t^2+t+1}+t+\tfrac{1}{2}\right|\right]_0^x + 3 \\
&= \sqrt{x^2+x+1} - \tfrac{1}{2}\log\left(\sqrt{x^2+x+1}+x+\tfrac{1}{2}\right) + 2 + \tfrac{1}{2}\log\tfrac{3}{2}.
\end{aligned}
$$

for all $x \in [0,\infty)$.

It is not usually the case, however, that solutions to a general class of differential equations can be expressed as explicitly as in Theorem 10.1.3. We are instead led to develop formal strategies for generating solutions. These methods are formal[3] in the sense that we *temporarily* abandon rigour by assuming that all relevant mathematical objects (numbers, functions etc.) have properties that are sufficient to justify our manipulations. In this way, the methods only generate *candidates* for solutions, and we must then analyse these *rigorously* to verify whether they actually satisfy all of the properties required by the relevant definition of a solution.

We will encounter four such formal methods in the remainder of this module. The first, which we shall develop now, is the method of *separation of variables*. This applies to *separable* first-order ordinary differential equations

$$
y' = f(x)g(y)
$$

where $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ (or more generally $f : \Omega_1 \to \mathbb{R}$ and $g : \Omega_2 \to \mathbb{R}$). These are called separable because f and g separate the x and y dependency.

The method is as follows: First, for each $y_0 \in \mathbb{R}$ such that $g(y_0) = 0$, we obtain the constant solution $y(x) := y_0$ for all $x_0 \in \mathbb{R}$. Next, assuming that y is a solution such that $g(y(x)) \neq 0$, a formal application of the Substitution Formula gives

$$
\frac{1}{g(y)}\frac{\mathrm{d}y}{\mathrm{d}x} = f(x) \implies \int \frac{1}{g(y)}\,\mathrm{d}y = \int f(x)\,\mathrm{d}x.
$$

A candidate solution can then be obtained by finding antiderivatives for both sides of this equation and trying to solve to obtain an explicit expression for $y(x)$. These antiderivatives can involve additive constants and it is important to account for these here (as well as in all of the formal methods we develop this week).

The following convention is commonly used to reduce the burden of notation associated with constants in calculations and estimates.

Notation 10.1.5. We use the convention whereby C denotes an arbitrary real number, the value of which can change from one line to the next in calculations and estimates.

The following example shows how to apply the method of separation of variables.

Example 10.1.6. Find two solutions of the differential equation $y' = 15 - 3y$ on $\mathbb{R}$.

Solution. Observe that $15 - 3y = 0$ when $y = 5$, hence $y_1 : \mathbb{R} \to \mathbb{R}$ given by $y_1(x) := 5$ for all $x \in \mathbb{R}$ is a solution on $\mathbb{R}$, since it is differentiable on $\mathbb{R}$ and $y_1'(x) = 0 = 15 - 3y_1(x)$ for all $x \in \mathbb{R}$.

---

[3]In mathematical analysis, a formal computation refers to a manipulation of mathematical symbols based on experience and intuition but without rigorous justification. For example, we use the formal algebraic manipulation "$\mathrm{d}u = \frac{\mathrm{d}u}{\mathrm{d}x}\,\mathrm{d}x$" when applying the Substitution Formula.

Next, assume that y is a solution on $\mathbb{R}$ such that $15 - 3y(x) \neq 0$ for all $x \in \mathbb{R}$, so a formal application of the Substitution Formula gives

$$\frac{1}{15 - 3y} \frac{\mathrm{d}y}{\mathrm{d}x} = 1 \Longrightarrow \int \frac{1}{15 - 3y} \,\mathrm{d}y = \int 1 \,\mathrm{d}x$$

$$\Longrightarrow -\tfrac{1}{3} \log|15 - 3y| = x + C$$

$$\Longrightarrow |15 - 3y| = Ce^{-3x}$$

for some $C \in \mathbb{R}$, where we have used the convention regarding the value of C from Notation 10.1.5. The assumption that y is a solution on $\mathbb{R}$ implies that $15 - 3y$ is differentiable and hence continuous. The continuity of $15 - 3y$ combined with the assumption that $15 - 3y(x) \neq 0$ for all $x \in \mathbb{R}$ implies that either $15 - 3y(x) > 0$ for all $x \in \mathbb{R}$ or $15 - 3y(x) < 0$ for all $x \in \mathbb{R}$. In either case, since the value of C can be different on each line, we have

$$\frac{1}{15 - 3y} \frac{\mathrm{d}y}{\mathrm{d}x} = 1 \Longrightarrow |15 - 3y| = Ce^{-3x}$$

$$\Longrightarrow 15 - 3y = Ce^{-3x}$$

$$\Longrightarrow y = 5 - Ce^{-3x}$$

for some $C \in \mathbb{R}$.

The above manipulations suggest that for any $C \in \mathbb{R}$, the function $y_2 : \mathbb{R} \to \mathbb{R}$ given by $y_2(x) := 5 - Ce^{-3x}$ is a solution on $\mathbb{R}$. This is indeed the case, since y_2 is differentiable on $\mathbb{R}$ (as a linear combination of a constant function and an exponential function) and $y_2'(x) = 15 - 3y_2(x)$ for all $x \in \mathbb{R}$ by inspection. This gives an infinite number of solutions and we recover $y_2 = y_1$ when $C = 0$.

The following example shows how separable first-order differential equations can be used to model mixing problems in engineering applications

Example 10.1.7 (Mixing Problem I). A 5,000 L industrial mixing tank is filled with a salt solution containing 20 kg of dissolved salt. The tank is drained at a rate of 25 L min^{-1} whilst a mixture containing 0.03 kg L^{-1} of salt is pumped into the tank at the same rate. Assume that the mixture is instantly and uniformly mixed:

(1) Let $y(t)$ denote the kilograms of salt in the tank after t minutes of mixing. Formulate an initial value problem to model the flow using the *flow equation*

$$y'(t) = \text{rate in} - \text{rate out}.$$

(2) Find a solution of the initial value problem to determine the mass of salt in the tank after 30 minutes of mixing.
(3) Determine the long-term behaviour of the salt concentration in the tank.

Solution. (1) The flow equation $y'(t) = \text{rate in} - \text{rate out}$ (in kg min^{-1}) becomes

$$y'(t) = \left(0.03 \frac{\text{kg}}{\text{L}}\right)\left(25 \frac{\text{L}}{\text{min}}\right) - \left(\frac{y(t)}{5000} \frac{\text{kg}}{\text{L}}\right)\left(25 \frac{\text{L}}{\text{min}}\right).$$

The flow is thus modelled by solutions $y : [0, \infty) \to \mathbb{R}$ to the initial value problem

$$y'(t) = \frac{150 - y}{200}, \quad y(0) = 20.$$

(2) This is a separable first-order differential equation. Observe that $\frac{150-y}{200} = 0$ when $y = 150$, but we ignore the constant solution $y(t) = 150$ because it does not satisfy the initial condition. Next, assume that y is a solution on $(0, \infty)$ such that

$\frac{150-y(t)}{200} \neq 0$ for all $t \in (0, \infty)$, so proceeding formally we obtain

$$\frac{1}{150 - y(t)} \frac{dy}{dt} = \frac{1}{200} \implies \int \frac{1}{150 - y} \, dy = \int \frac{1}{200} \, dt$$

$$\implies -\log|150 - y| = \tfrac{t}{200} + C$$

$$\implies |150 - y| = Ce^{-\frac{t}{200}}$$

$$\implies 150 - y = Ce^{-\frac{t}{200}}$$

$$\implies y = 150 + Ce^{-\frac{t}{200}}$$

for some $C \in \mathbb{R}$, since either $150 - y(t) > 0$ for all $t \in (0, \infty)$, or $150 - y(t) < 0$ for all $t \in (0, \infty)$, because y is continuous and $150 - y(t) \neq 0$ for all $t \in (0, \infty)$. The initial condition $y(0) = 20$ requires that $150 + C = 20$, so $C = -130$

The function $y : [0, \infty) \to \mathbb{R}$ given by $y(t) := 150 - 130e^{-\frac{t}{200}}$ is differentiable on $(0, \infty)$ and continuous on $[0, \infty)$, since it is the composition of an exponential function and linear functions. It also satisfies the differential equation and initial condition by construction, hence it is a solution of the initial value problem. Therefore, the mass of salt in the tank after 30 minutes of mixing is modelled by $y(30) = 150 - 130e^{-\frac{3}{20}} \approx 38.1$ kg.

(3) The long behaviour of the salt concentration in the tank is modelled by

$$\lim_{t \to \infty} \frac{y(t)}{5000} = \frac{1}{5000} \lim_{t \to \infty} \left(150 - 130e^{-\frac{t}{200}} \right) = \frac{150}{5000} = 0.03 \text{ kg L}^{-1}.$$

In other words, the salt concentration in the tank approaches the concentration of the input flow in the long-term.

10.2. **Linear equations.** We now introduce an integration factor to generate solutions of *linear* first-order differential equations $y' + g(x)y = f(x)$. These equations are called linear because the dependent variable y and its derivative y' appear as linear factors when x is fixed. The following example motivates the definition of the integration factor.

Example 10.2.1. Suppose that $g : [a, b] \to \mathbb{R}$ is continuous and define $I : [a, b] \to \mathbb{R}$ by $I(x) := e^{\int_a^x g}$ for all $x \in [a, b]$. Prove that I is differentiable and that $I' = Ig$.

Proof. The function g is continuous so the First Fundamental Theorem of Calculus implies that $F : [a, b] \to \mathbb{R}$ given by $F(x) := \int_a^x g$ is differentiable with derivative $F'(x) = g(x)$ for all $x \in [a, b]$. Now observe that $I(x) = e^{F(x)} = (\exp \circ F)(x)$, where $\exp(x) := e^x$ is a differentiable function on $\mathbb{R}$, so the Chain Rule implies that I is differentiable with derivative

$$I'(x) = (\exp \circ F)'(x) = \exp'(F(x))F'(x) = e^{F(x)}g(x) = I(x)g(x)$$

for all $x \in [a, b]$, hence $I' = Ig$. $\square$

More generally, if $\Omega \subseteq \mathbb{R}$ and $g : \Omega \to \mathbb{R}$ has an antiderivative $h : \Omega \to \mathbb{R}$, then the above proof also shows that $I(x) := e^{h(x)}$ is differentiable with $I' = Ig$.

The first step to generate solutions of the equation $y' + g(x)y = f(x)$ is to introduce an *integration factor*

$$I := e^{\int g},$$

where $\int g$ denotes (as usual) an antiderivative of g. It is not necessary to introduce an additive constant C here, as we noted above that any antiderivative of g will ensure that I is differentiable with $I' = Ig$. Next, proceeding formally, we multiply

the differential equation by I and combine the identity $I' = Ig$ with the Product Rule $(Iy)' = Iy' + I'y$, to obtain

$$Iy' + (Ig)y = If \implies Iy' + I'y = If$$
$$\implies (Iy)' = If$$
$$\implies Iy = \int If$$
$$\implies y = \frac{1}{I} \int If,$$

where the final division by I is justified because the exponential function is never equal to 0. In this way, we obtain the candidate solutions

$$y(x) := \frac{1}{I(x)} \left(\int I(x)f(x) \, \mathrm{d}x + C \right),$$

where $C \in \mathbb{R}$ is included here to emphasise that it is particularly important to account for such additive constants when substituting an antiderivative of If.

The following examples show how to use an integration factor to obtain a solution to an initial value problem.

Example 10.2.2. Find a solution $y : [0, \infty) \to \mathbb{R}$ of the initial value problem

$$y' - 2xy = 2x, \quad y(0) = 1.$$

Solution. We introduce the integration factor

$$I(x) := e^{\int (-2x) \, \mathrm{d}x} = e^{-x^2}.$$

Next, we multiply the differential equation by I and proceed formally to obtain

$$I(x)[y'(x) - 2xy(x)] = I(x)2x \implies (Iy)'(x) = e^{-x^2}2x$$
$$\implies (Iy)(x) = \int e^{-x^2}2x \, \mathrm{d}x$$
$$\implies e^{-x^2}y(x) = -e^{-x^2} + C$$
$$\implies y(x) = -1 + Ce^{x^2},$$

where $C \in \mathbb{R}$. The initial condition $y(0) = 1$ requires that $-1 + C = 1$, so $C = 2$. The function $y : [0, \infty) \to \mathbb{R}$ given by $y(x) := -1 + 2e^{x^2}$ for all $x \in [0, \infty)$ is differentiable on $(0, \infty)$ and continuous on $[0, \infty)$, since it is the composition of an exponential function and a polynomial. It also satisfies the differential equation and initial condition by construction, hence it is a solution of the initial value problem.

Example 10.2.3. Find a solution $y : [0, \infty) \to \mathbb{R}$ of the initial value problem

$$y' + \frac{1}{x+1}y = x, \quad y(0) = y_0,$$

where $y_0 \in \mathbb{R}$.

Solution. We introduce the integration factor

$$I(x) := e^{\int \frac{1}{x+1} \, \mathrm{d}x} = e^{\log|x+1|} = |x+1| = (x+1),$$

where the final equality is justified because we are only concerned with $x \in [0, \infty)$. Next, we multiply the differential equation by I and proceed formally to obtain

$$(x+1)y'(x) + y(x) = (x+1)x \implies (Iy)'(x) = x^2 + x$$
$$\implies (Iy)(x) = \int (x^2 + x) \, \mathrm{d}x$$
$$\implies (x+1)y(x) = \tfrac{1}{3}x^3 + \tfrac{1}{2}x^2 + C$$
$$\implies y(x) = \tfrac{1}{x+1} \left(\tfrac{1}{3}x^3 + \tfrac{1}{2}x^2 + C \right),$$

where $C \in \mathbb{R}$. The initial condition $y(0) = y_0$ requires that $C = y_0$. The function $y : [0, \infty) \to \mathbb{R}$ given by $y(x) := \frac{1}{x+1} \left(\frac{1}{3}x^3 + \frac{1}{2}x^2 + y_0 \right)$ for all $x \in [0, \infty)$ is differentiable on $(0, \infty)$ and continuous on $[0, \infty)$, since it is a rational function with nonzero denominator. It also satisfies the differential equation and initial condition by construction, hence it is a solution of the initial value problem.

The following example shows how linear first-order differential equations can be used to model mixing problems in engineering applications.

Example 10.2.4 (Mixing Problem II)**.** Consider again the mixing problem from Example 10.1.7. Suppose that the input pump is malfunctioning and only allows the mixture to flow into the tank at a reduced rate of 20 L min^{-1}. Solve an initial value problem to determine the mass of salt in the tank after t minutes of mixing.

Solution. The flow equation $y'(t) = $ rate in $-$ rate out (in kg min^{-1}) now becomes

$$y'(t) = \left(0.03 \frac{\text{kg}}{\text{L}} \right) \left(20 \frac{\text{L}}{\text{min}} \right) - \left(\frac{y(t)}{5000 - 5t} \frac{\text{kg}}{\text{L}} \right) \left(25 \frac{\text{L}}{\text{min}} \right).$$

In particular, since the output flow is now 5 L min^{-1} greater than the input flow, the volume in the tank decreases from the tank's $5{,}000$ L capacity to $5{,}000 - 5t$ at time $t \in (0, 1000)$, until the tank is completely drained at $t = 1000$. The flow is thus modelled by solutions $y : [0, 1000) \to \mathbb{R}$ to the initial value problem

$$y'(t) = \frac{3}{5} - \frac{5}{1000 - t} y, \quad y(0) = 20.$$

This is a linear first-order differential equation, which in standard form becomes

$$y'(t) + \frac{5}{1000 - t} y = \frac{3}{5}, \quad y(0) = 20.$$

We introduce the integration factor

$$I(t) := e^{\int \frac{5}{1000 - t} \, dx} = e^{-5 \log |1000 - t|} = |1000 - t|^{-5} = (1000 - t)^{-5},$$

where the final equality is justified because we are only concerned with $t \in [0, 1000)$. Next, we multiply the differential equation by I and proceed formally to obtain

$$
\begin{aligned}
I(t) \left[y'(t) + \tfrac{5}{1000-t} y(t) \right] &= \tfrac{3}{5}(1000 - t)^{-5} \\
&\implies (Iy)'(t) = \tfrac{3}{5}(1000 - t)^{-5} \\
&\implies (Iy)(t) = \tfrac{3}{5} \int (1000 - t)^{-5} \, dt \\
&\implies (1000 - t)^{-5} y(t) = \tfrac{-3}{-20}(1000 - t)^{-4} + C \\
&\implies y(t) = \tfrac{3}{20}(1000 - t) + C(1000 - t)^5,
\end{aligned}
$$

where $C \in \mathbb{R}$. The initial condition $y(0) = 20$ requires that $C = -1.3 \times 10^{-13}$. The function $y : [0, \infty) \to \mathbb{R}$ given by $y(t) := \frac{3}{20}(1000 - t) - 1.3 \times 10^{-13}(1000 - t)^5$ for all $t \in [0, 1000)$ is differentiable on $(0, 1000)$ and continuous on $[0, 1000)$, since it is a rational function with nonzero denominator. It also satisfies the differential equation and initial condition by construction, hence it is a solution of the initial value problem. More generally, since the tank will be empty when $t \geq 1000$, the function $\tilde{y} : [0, \infty) \to \mathbb{R}$ given by

$$\tilde{y}(t) := \begin{cases} \frac{3}{20}(1000 - t) - 1.3 \times 10^{-13}(1000 - t)^5, & t \in [0, 1000) \\ 0, & t \geq 1000 \end{cases}$$

models the mass of salt in the tank after t minutes of mixing for all $t \geq 0$.

WEEK 11: SECOND-ORDER ODEs AND APPLICATIONS

This week we develop strategies to solve second-order ordinary differential equations. For equations with constant coefficients $ay'' + by' + cy = f(x)$, we will obtain general solutions in the homogeneous case (when $f(x) = 0$) using the method of characteristic equations, and in the inhomogeneous case (when $f(x) \neq 0$) using the method of undetermined coefficients. We will also see how such equations model the motion of vibrating springs in physics.

11.1. **Homogeneous equations.** We now consider second-order equations with constant coefficients $ay'' + by' + cy = f(x)$. These equations are also linear because the dependent variable y and its derivatives y' and y'' appear as linear factors when x is fixed. They are also called *homogeneous* when $f(x) = 0$ because each term is then a homogeneous function of the dependent variable y. This means that the equation has the multiplicative scaling property

$$ay'' + by' + cy = 0 \implies (\alpha y)'' + b(\alpha y)' + c(\alpha y) = \alpha(ay'' + by' + cy) = 0$$

for all $\alpha \in \mathbb{R}$. We begin by making precise the meaning of a solution to these equations and associated initial value problems and boundary value problems.

Definition 11.1.1. Suppose that $f : \mathbb{R} \to \mathbb{R}$, that a, b, $c \in \mathbb{R}$ with $a \neq 0$, and that $I \subseteq \mathbb{R}$ is an open interval. A *solution* of the (second-order) differential equation

$$ay'' + by' + cy = f(x) \text{ on } I$$

is a twice differentiable function $y : I \to \mathbb{R}$ such that $ay''(x) + by'(x) + cy(x) = f(x)$ for all $x \in I$.

We can extend this definition to the case when $f : \Omega \to \mathbb{R}$, where $\Omega \subseteq \mathbb{R}$, by requiring that $I \subseteq \Omega$. We typically need two pieces of solution data to specify a unique solution of a second-order differential equation. This leads use to consider both initial value problems and boundary value problems.

Definition 11.1.2. Suppose that $f : \mathbb{R} \to \mathbb{R}$ and that a, b, $c \in \mathbb{R}$ with $a \neq 0$.
Let x_0, y_0, $y_1 \in \mathbb{R}$ and $x_1 \in (x_0, \infty]$. A *solution* of the initial value problem

$$ay''(x) + by'(x) + cy(x) = f(x), \quad y(x_0) = y_0, \quad y'(x_0) = y_1$$

is a differentiable function $y : [x_0, x_1) \to \mathbb{R}$ that is twice differentiable on (x_0, x_1) such that $ay''(x) + by'(x) + cy(x) = f(x)$ for all $x \in (x_0, x_1)$, and such that $y(x_0) = y_0$ and $y'(x_0) = y_1$.
Let x_0, y_0, $y_1 \in \mathbb{R}$ and $x_1 \in (x_0, \infty)$. A *solution* of the boundary value problem

$$ay''(x) + by'(x) + cy(x) = f(x), \quad y(x_0) = y_0, \quad y(x_1) = y_1$$

is a continuous function $y : [x_0, x_1] \to \mathbb{R}$ that is twice differentiable on (x_0, x_1) such that $ay''(x) + by'(x) + cy(x) = f(x)$ for all $x \in (x_0, x_1)$, and such that $y(x_0) = y_0$ and $y(x_1) = y_1$.

We will now see how to obtain *general* solutions of homogeneous equations by reducing matters to the *algebra* of certain characteristic equations. We begin by proving the linearity property in (1) below for solutions of homogeneous equations. This is the first step toward obtaining the full characterisation of solutions in (2), although this requires techniques beyond the scope of this module and is more suited to a second-year differential equations module. In particular, part (2) below requires the notion that two functions f and g are *linearly independent* when they are not constant multiples of each other (so $f \neq Cg$) but you are not expected to be familiar with this property for the assessment in this module.

Proposition 11.1.3. Suppose that a, b, $c \in \mathbb{R}$ with $a \neq 0$. The following properties concern solutions of the homogeneous differential equation $ay'' + by' + cy = 0$ on $\mathbb{R}$:

(1) If y_1, y_2 are solutions and $C_1, C_2 \in \mathbb{R}$, then $y := C_1 y_1 + C_2 y_2$ is a solution.

(2) If y is a solution and y_1, y_2 are linearly independent solutions, then there exist $C_1, C_2 \in \mathbb{R}$ such that $y = C_1 y_1 + C_2 y_2$.

Proof. (1) We have $ay_1''(x) + by_1'(x) + cy_1(x) = 0$ and $ay_2''(x) + by_2'(x) + cy_2(x) = 0$ for all $x \in \mathbb{R}$. We multiply these equations by C_1 and C_2 respectively to obtain

$$C_1(ay_1''(x) + by_1'(x) + cy_1(x)) + C_2(ay_2''(x) + by_2'(x) + cy_2(x)) = 0$$

for all $x \in \mathbb{R}$. The linearity of the derivatives implies that

$$a(C_1 y_1 + C_2 y_2)''(x) + b(C_1 y_1 + C_2 y_2)'(x) + c(C_1 y_1 + C_2 y_2)(x) = 0$$

for all $x \in \mathbb{R}$, so $y := C_1 y_1 + C_2 y_2$ is a solution on $\mathbb{R}$.

(2) There is no scope to provide this proof here, except to mention that it involves the fact that if the Wronksian $W(y_1, y_2) := y_1 y_2' - y_2 y_1'$ does not vanish identically on $\mathbb{R}$, then the solutions y_1 and y_2 must be linearly independent. $\square$

The main point to recognise from (2) in Proposition 11.1.3 is that *every* solution of the equation $ay'' + by' + cy = 0$ can be characterised as a linear combination of two linearly independent solutions y_1 and y_2. Table 11.1 below shows how to obtain two such solutions from the roots $\lambda \in \mathbb{C}$ of the characteristic equation $a\lambda^2 + b\lambda + c = 0$. The linear combination $y := C_1 y_1 + C_2 y_2$, where $C_1, C_2 \in \mathbb{R}$, is then called the *general solution* of the differential equation.

Roots of the Characteristic Equation $a\lambda^2 + b\lambda + c = 0$, where $\lambda \in \{\lambda_1, \lambda_2\}$	General Solution $y := C_1 y_1 + C_2 y_2$ on $\mathbb{R}$, where $C_1, C_2 \in \mathbb{R}$
$\lambda_1 \neq \lambda_2$ ($\lambda_1, \lambda_2 \in \mathbb{R}$)	$y(x) := C_1 e^{\lambda_1 x} + C_2 e^{\lambda_2 x}$
$\lambda_1 = \lambda_2$ ($\lambda_1, \lambda_2 \in \mathbb{R}$)	$y(x) := C_1 e^{\lambda_1 x} + C_2 x e^{\lambda_1 x}$
$\lambda_1 = \alpha + i\beta, \lambda_2 = \alpha - i\beta$ ($\alpha, \beta \in \mathbb{R}$)	$y(x) := e^{\alpha x}(C_1 \cos(\beta x) + C_2 \sin(\beta x))$

TABLE 11.3. Roots of the characteristic equation $a\lambda^2 + b\lambda + c = 0$ corresponding to general solutions of $ay'' + by' + cy = 0$.

We now use Proposition 11.1.3 to prove these results.

Theorem 11.1.4. Suppose that a, b, $c \in \mathbb{R}$ with $a \neq 0$. The general solution of the equation $ay'' + by' + cy = 0$ on $\mathbb{R}$ is determined by the roots of the characteristic equation $a\lambda^2 + b\lambda + c = 0$ as in Table 11.1.

Proof. First, observe that if $\lambda \in \mathbb{C}$ satisfies $(a\lambda^2 + b\lambda + c)$ and $y(x) := e^{\lambda x}$ for all $x \in \mathbb{R}$, then y is twice differentiable with

$$ay''(x) + by'(x) + cy(x) = a\lambda^2 e^{\lambda x} + b\lambda e^{\lambda x} + ce^{\lambda x} = (a\lambda^2 + b\lambda + c)e^{\lambda x} = 0$$

for all $x \in \mathbb{R}$, so y is a solution of the differential equation on $\mathbb{R}$. We use this result in the following three cases:

Case 1: If the characteristic equation has two distinct real roots $\lambda_1 \neq \lambda_2$, then $y_1(x) := e^{\lambda_1 x}$ and $y_2(x) := e^{\lambda_2 x}$ are two linearly independent solutions, so the result follows from Proposition 11.1.3.

Case 2: If the characteristic equation has only one real root $\lambda = -\frac{b}{2a}$, then $y_1(x) := e^{\lambda x}$ is a solution. In this case, observe that $y_2(x) := xe^{\lambda x}$ is a solution,

since it is twice differentiable with

$$ay_2''(x) + by_2'(x) + cy_2(x) = a(2\lambda e^{\lambda x} + \lambda^2 x e^{\lambda x}) + b(e^{\lambda x} + \lambda x e^{\lambda x}) + cx e^{\lambda x}$$
$$= (2a\lambda + b)e^{\lambda x} + (a\lambda^2 + b\lambda + c)x e^{\lambda x}$$
$$= (2a\tfrac{-b}{2a} + b)e^{\lambda x}$$
$$= 0$$

for all $x \in \mathbb{R}$. The solutions y_1 and y_2 are also linearly independent, so the result follows from Proposition 11.1.3.

Case 3: If the characteristic equation has a pair of complex conjugate roots $\lambda_1 = \alpha + i\beta$ and $\lambda_2 = \alpha - i\beta$, then $y_1(x) := \frac{1}{2}(e^{(\alpha+i\beta)x} + e^{(\alpha-i\beta)x}) = e^{\alpha x}\cos(\beta x)$ and $y_2(x) := \frac{1}{2i}(e^{(\alpha+i\beta)x} - e^{(\alpha-i\beta)x}) = e^{\alpha x}\sin(\beta x)$ are two linearly independent solutions, so the result follows from Proposition 11.1.3. $\qquad\square$

The next example shows how to use characteristic equations to obtain general solutions, as well as solve initial value problems and boundary value problems, for second-order homogeneous equations with constant coefficients.

Example 11.1.5.

(1) Find the general solution of the equation $y'' + 3y' + 2y = 0$ on $\mathbb{R}$.
(2) Find a solution $y : [0, \infty) \to \mathbb{R}$ of the initial value problem

$$y'' + 36y = 0, \quad y(0) = 0, \quad y'(0) = 3.$$

(3) Find a solution $y : [0, 1] \to \mathbb{R}$ of the boundary value problem

$$y'' + 12y' + 36y = 0, \quad y(0) = 0, \quad y(1) = 3.$$

Solution. (1) The characteristic equation $\lambda^2 + 3\lambda + 2 = (\lambda + 1)(\lambda + 2) = 0$ has the two real roots $\lambda = -1$ and $\lambda = -2$. Using Theorem 11.1.4 (and Table 11.1), the general solution is $y(x) := C_1 e^{-x} + C_2 e^{-2x}$ for all $x \in \mathbb{R}$, where $C_1, C_2 \in \mathbb{R}$.

(2) The characteristic equation $\lambda^2 + 36 = 0$ has the complex roots $\lambda = \pm 6i$, so by Theorem 11.1.4 the general solution is $y(x) := C_1 \cos(6x) + C_2 \sin(6x)$ for all $x \in \mathbb{R}$, where $C_1, C_2 \in \mathbb{R}$. The requirement that $y(0) = 0$ implies that $C_1 = 0$, whilst $y'(0) = 3$ requires that $6C_2 = 3$ so $C_2 = \frac{1}{2}$. The differentiable function $y : [0, \infty) \to \mathbb{R}$ given by $y(x) := \frac{1}{2}\sin(6x)$ for all $x \in [0, \infty)$ is thus a solution of the initial value problem.

(3) The characteristic equation $\lambda^2 + 12\lambda + 36 = (\lambda + 6)^2 = 0$ has only the single real root $\lambda = -6$, so by Theorem 11.1.4 the general solution is $y(x) := C_1 e^{-6x} + C_2 x e^{-6x}$ for all $x \in \mathbb{R}$, where $C_1, C_2 \in \mathbb{R}$. The requirement that $y(0) = 0$ implies that $C_1 = 0$, whilst $y(1) = 3$ requires that $C_2 e^{-6} = 3$ so $C_2 = 3e^6$. The continuous function $y : [0, 1] \to \mathbb{R}$ given by $y(x) := 3e^6 x e^{-6x}$ for all $x \in [0, 1]$ is thus a solution of the boundary value problem.

11.2. **Inhomogeneous equations.** The method of undetermined coefficients combines the following theorem with a trial-and-error procedure to find solutions of second-order *inhomogeneous* equations $ay'' + by' + cy = f(x)$. These equations are called inhomogeneous because they are not homogeneous when the function f is not identically zero.

Theorem 11.2.1. Suppose that $f : \mathbb{R} \to \mathbb{R}$ and that $a, b, c \in \mathbb{R}$ with $a \neq 0$. The general solution of the equation $ay'' + by' + cy = f(x)$ on $\mathbb{R}$ is given by

$$y(x) := y_c(x) + y_p(x) \text{ for all } x \in \mathbb{R},$$

where y_c is the general solution of the homogenous equation $ay'' + by' + cy = 0$, and y_p is any particular solution of $ay_p'' + by_p' + cy_p = f(x)$, on $\mathbb{R}$.

Proof. If y denotes the general solution to the equation $ay'' + by' + cy = f(x)$ on $\mathbb{R}$ and y_p is any particular solution to the same equation, then

$$a(y - y_p)''(x) + b(y - y_p)'(x) + c(y - y_p)(x)$$
$$= [ay''(x) + by'(x) + cy(x)] - [ay_p''(x) + by_p'(x) + cy_p(x)]$$
$$= f(x) - f(x)$$
$$= 0$$

for all $x \in \mathbb{R}$, hence $y - y_p$ is a solution of the homogenous equation. This proves that $y - y_p = y_c$, so we have $y(x) = y_p(x) + y_c(x)$ for all $x \in \mathbb{R}$, as required. $\qquad \square$

The main point to recognise from Theorem 11.2.1 is that the general solution to such an inhomogeneous equation can be found by solving the homogenous version of the equation (for which we can use the characteristic equation methods from the preceding lecture) and by finding *one* solution to the inhomogeneous equation.

The method of undetermined coefficients proceeds by guessing that any such particular solution might have a similar form to the inhomogeneous term $f(x)$. For example, Table 11.2 suggests some trial solutions y_p for common inhomogeneous terms f. The coefficients in the right-hand column are determined by substituting the trial solution into the equation and equating coefficients. If these coefficients can be determined, then a particular solution has been obtained. If the coefficients cannot be determined, however, then another trial solution needs to be considered.

Note that if a trial solution is contained in the general homogeneous solution, then it cannot be a solution of the inhomogeneous equation. In that case, it is often useful to try multiplying the trial function by x to obtain a particular solution.

Inhomogeneous Term $f(x)$	Trial Solution $y_p(x)$
polynomial $a_0 + a_1 x + \ldots + a_n x^n$	$A_0 + A_1 x + \ldots + A_n x^n$
exponential $e^{\alpha x}$	$A e^{\alpha x}$
sine $\sin(\alpha x)$ or cosine $\cos(\alpha x)$	$A \cos(\alpha x) + B \sin(\alpha x)$

TABLE 11.4. Trial particular solutions y_p suggested for common inhomogeneous terms f in the method of undetermined coefficients.

The following example shows how to use the method of undetermined coefficients to obtain general solutions of second-order inhomogeneous equations.

Example 11.2.2. Find the general solution of the equation

$$y'' + 3y' + 2y = x^2 - 5x + 2 \text{ on } \mathbb{R}.$$

Solution. We know from Theorem 11.2.1 that $y = y_c + y_p$. We also know from Example 11.1.5(1) that $y_c(x) := C_1 e^{-x} + C_2 e^{-2x}$ for all $x \in \mathbb{R}$, where $C_1, C_2 \in \mathbb{R}$. To find a particular solution we use Table 11.2 and consider $y_p(x) := Ax^2 + Bx + C$. We substitute $y_p'(x) := 2Ax + B$ and $y_p''(x) = 2A$ into the differential equation to obtain

$$2A + 3(2Ax + B) + 2(Ax^2 + Bx + C) = x^2 - 5x + 2.$$

To determine the coefficients A, B and C, we equate the coefficients of x^2, x and x^0 to obtain the simultaneous equations

$$2A = 1$$
$$6A + 2B = -5$$
$$2A + 3B + 2C = 2$$

which have the solution $A = \frac{1}{2}$, $B = -4$ and $C = \frac{13}{2}$. Altogether, we have

$$y(x) = C_1 e^{-x} + C_2 e^{-2x} + \tfrac{1}{2}x^2 - 4x + \tfrac{13}{2}$$

for all $x \in \mathbb{R}$, where $C_1, C_2 \in \mathbb{R}$.

The following example shows how to use the method of undetermined coefficients to solve initial value problems for second-order inhomogeneous equations.

Example 11.2.3. Find a solution $y : [0, \infty) \to \mathbb{R}$ of the initial value problem

$$y'' + 3y' - 18y = 9e^{3x}, \quad y(0) = 0, \quad y'(0) = 0.$$

Solution. We know from Theorem 11.2.1 that $y = y_c + y_p$. To find the homogeneous solution y_c, we solve the characteristic equation $\lambda^2 + 3\lambda - 18 = (\lambda + 6)(\lambda - 3) = 0$ to obtain the two real roots $\lambda = -6$ and $\lambda = 3$. Theorem 11.1.4 then shows that the general solution is $y_c(x) := C_1 e^{-6x} + C_2 e^{3x}$ for all $x \in \mathbb{R}$, where $C_1, C_2 \in \mathbb{R}$.

To find a particular solution y_p, we use Table 11.2, and after noting that e^{3x} appears in y_c, we consider $y_p(x) := Axe^{3x}$. We substitute $y_p'(x) = 3Axe^{3x} + Ae^{3x}$ and $y_p''(x) = 9Axe^{3x} + 6Ae^{3x}$ into the differential equation to obtain

$$(9Axe^{3x} + 6Ae^{3x}) + 3(3Axe^{3x} + Ae^{3x}) - 18Axe^{3x} = 9e^{3x}.$$

We then equate the coefficients of e^{3x} and xe^{3x} to obtain the simultaneous equations $18A - 18A = 0$ and $9A = 9$, which have the solution $A = 1$, so $y_p(x) := xe^{3x}$ is a particular solution to the inhomogeneous equation.

We now have $y(x) = C_1 e^{-6x} + C_2 e^{3x} + xe^{3x}$ for all $x \in \mathbb{R}$, where $C_1, C_2 \in \mathbb{R}$. The requirement that $y(0) = 0$ implies that $C_1 + C_2 = 0$, whilst $y'(0) = 0$ requires that $-6C_1 + 3C_2 + 1 = 0$, so $C_1 = \frac{1}{9}$ and $C_2 = -\frac{1}{9}$. Altogether, the differentiable function $y : [0, \infty) \to \mathbb{R}$ given by

$$y(x) := \tfrac{1}{9}e^{-6x} - \tfrac{1}{9}e^{3x} + xe^{3x}$$

for all $x \in [0, \infty)$ is thus a solution of the boundary value problem.

We conclude with an example to show how second-order differential equations with constant coefficients model the motion of vibrating springs in physics.

Example 11.2.4 (Vibrating Springs)**.** Consider a small object of mass $m > 0$ attached to the end of a horizontal spring. Let $x(t)$ denote the position of the object with respect to its equilibrium position at time $t \geq 0$. According to Hooke's Law, the spring exerts a restorative force on the object equal to $-kx$, where $k > 0$ is known as the spring constant, whilst Newton's Law asserts that this force must be equal to mx''. We thus obtain the second-order differential equation $mx'' + kx = 0$. This is a homogeneous equation and the characteristic equation $m\lambda^2 + k = 0$ has the complex roots $\lambda = \pm\sqrt{\frac{k}{m}}i$, so by Theorem 11.1.4 we obtain the general solution

$$x(t) := C_1 \cos(\sqrt{\tfrac{k}{m}}t) + C_2 \sin(\sqrt{\tfrac{k}{m}}t)$$

for all $t \in (0, \infty)$, where $C_1, C_2 \in \mathbb{R}$. The oscillatory behaviour of the object's motion described by this equation is known as *simple harmonic motion.*

A more realistic model of the object's motion is obtained by incorporating the frictional force $-cx'$ exerted on the object, where $c > 0$ is known as the coefficient of friction, to obtain the equation

$$mx'' + cx' + kx = 0.$$

In this case, the roots of the characteristic equation $m\lambda^2 + c\lambda + k = 0$ are given by $\lambda = \frac{1}{2m}(-c \pm \sqrt{c^2 - 4mk})$, so by Theorem 11.1.4, and in particular Table 11.1, the nature of the motion will be governed by the sign of the term $c^2 - 4mk$. More specifically, we identify the following three types of motion:

(1) If $c^2 - 4mk < 0$, then the *under damped* motion is modelled by
$$x(t) := e^{-\alpha t}(C_1 \cos(\omega t) + C_2 \sin(\omega t))$$
for all $t \in (0, \infty)$, where $\alpha := \frac{c}{2m}$ and $\omega := \frac{1}{2m}\sqrt{|c^2 - 4mk|}$.

(2) If $c^2 - 4mk = 0$, then the *critically damped* motion is modelled by
$$x(t) := C_1 e^{-\alpha t} + C_2 t e^{-\alpha t}$$
for all $t \in (0, \infty)$, where $\alpha := \frac{c}{2m}$.

(3) If $c^2 - 4mk > 0$, then the *over damped* motion is modelled by
$$x(t) := C_1 e^{-(\alpha-\omega)t} + C_2 e^{-(\alpha+\omega)t}$$
for all $t \in (0, \infty)$, where $\alpha := \frac{c}{2m}$ and $\omega := \frac{1}{2m}\sqrt{|c^2 - 4mk|}$.