MA2002 Calculus

Thang Pang Ern

These notes are based off **Dr. Wang Fei's** MA2002 Calculus materials. Additional references are cited in the bibliography. Note that the old course code is MA1102R but the scope was pretty much the same.

Please send me an email at `thangpangern@u.nus.edu` if you would like to contribute a nice discussion to the notes or point out a typo.

Contents

Chapter 1

Functions

1.1 Sets and Operations

A set is a collection of objects. It is usually denoted by a capital letter, say A, B, C, or S. For any set A, the objects a, b, c in A are called the elements of the set A. We can write the elements of a set by listing all its elements regardless of its order. For example, we have the finite set $\{1,2\}$ consisting of 2 elements and the infinite set $\{2,3,5,7,11,\ldots\}$ consisting of prime numbers.

For any set A, if a is an element of A, we write $a \in A$; otherwise, we write $a \notin A$. This is known as the membership relation.

Definition 1.1 (subset). For any two sets A and B, if every element of A is also an element of B, then we say that A is a subset of B, which is denoted by $A \subseteq B$. If A is not a subset of B, we write $A \not\subseteq B$.

Definition 1.2 (equality of sets). We say that two sets A and B are equal if and only if they have the same collection of elements regardless of order. That is to say,

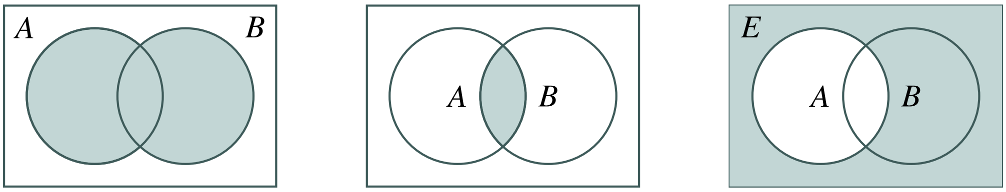
$$A = B \quad \text{if and only if} \quad A \subseteq B \text{ and } B \subseteq A. \tag{1.1}$$

For students who have had exposure to MA1100 Basic Discrete Mathematics, we say that (1.1) is the antisymmetric property of the subset relation.

We now discuss some operations on sets. Let A, B, E be three sets.

- The union $A \cup B$ refers to be the set of all x such that $x \in A$ or $x \in B$ (see **(a)** in Figure 1.1)
- The intersection $A \cap B$ refers to be the set of all x such that $x \in A$ and $x \in B$ (see **(b)** in Figure 1.1)

- The set-theoretic difference $E \setminus A$ denotes the set of all x such that $x \in E$ but $x \notin A$ (see **(c)** in Figure 1.1)
- The Cartesian product $A \times B$ refers to the set of all pairs (x,y) such that $x \in A$ and $y \in B$



(a) $A \cup B$ (b) $A \cap B$ (c) $E \setminus A$

Figure 1.1: Some set operations

Throughout the course, we will adopt the standard notations $\mathbb{Z}, \mathbb{N}, \mathbb{Q}, \mathbb{R}$ to denote the integers, the positive integers/natural numbers, the rational numbers, and the real numbers respectively[1]. The empty set $\emptyset$ uniquely represents the set containing no elements. Similarly, we can define $\mathbb{Z}^+$ to be the positive integers, and $\mathbb{Z}^-, \mathbb{R}^+, \mathbb{R}^-$ etc. are defined similarly. We can also restrict a set by adding a condition as a subscript. For example, $\mathbb{Z}_{\geq 0}$ denotes the set of non-negative integers.

For any $a,b \in \mathbb{R}$ with $a \leq b$, every *interval* in $\mathbb{R}$ is either an open interval (a,b), a closed interval $[a,b]$, a half-open interval (either $(a,b]$ or $[a,b)$), an unbounded open interval (i.e. (a,∞)), an unbounded closed interval (i.e. $(-\infty,b]$), or the entire real line. Some might refer to this as the interval classification theorem in $\mathbb{R}$. For interested readers, refer to p. 84-86 of J. Munkres '*Topology*' [2].

1.2 Functions

Let A and B be two sets. A function $f : A \to B$ is a rule which assigns each element in A to a unique element in B. For any $a \in A$, the unique element in B that is assigned to by f is called the image of a, which is denoted by $f(a)$. Here, we refer to A and B as the domain and codomain of f respectively. The range of $f : A \to B$ is the set of images $f(x)$ where $x \in A$. Note that because $x \in A$ implies $f(x) \in B$, it implies that the range of a function is always a subset of the codomain.

We define the graph of f, $\Gamma(f)$, to be the set

$$G(f) = \{(x, f(x)) : x \in A\}.$$

Suppose $A \subseteq \mathbb{R}$ and $B \subseteq \mathbb{R}$. Then, $\Gamma(f) \subseteq A \times B \subseteq \mathbb{R} \times \mathbb{R} = \mathbb{R}^2$ so we infer that the graph $\Gamma(f)$ is a subset of the Cartesian plane $\mathbb{R}^2$.

[1]Some readers might regard 0 to be a natural number but we will not because we follow the traditional definition where natural numbers are the positive integers used for counting.

Chapter 2

The Formal Definition of a Limit

2.1 Using Intuition

In elementary geometry, a straight line in the plane is uniquely determined by two distinct points. Let $A(x_0, y_0)$ and $B(x_1, y_1)$ be two distinct points in $\mathbb{R}$. Also, let (x, y) be a point on a line ℓ which passes through points A and B. Then,

$$\frac{y - y_0}{x - x_0} = \frac{y_1 - y_0}{x_1 - x_0}$$

denotes the slope (or gradient) of ℓ.

Say we have the function $f(x) = x^2 + 3$. As x approaches 2, the value of $f(x)$ gets closer to 7. As such, we write

$$\lim_{x \to 2} f(x) = 7.$$

Intuitively, if by taking x to be sufficiently close (but not equal) to a, the value of $f(x)$ becomes arbitrarily close to the number L, then we say that the limit of $f(x)$, as x approaches a, equals L. We write

$$\lim_{x \to a} f(x) = L \quad \text{or} \quad \text{as } x \to a \text{ we have } f(x) \to L. \tag{2.1}$$

We will see in Chapter 2.2 that this intuitive definition of a limit is not rigorous at all (that is why it is merely an intuition). Also, note that the value of the mentioned limit in (2.1) only depends on the values of $f(x)$ for x near a and does not depend on the value of $f(x)$ for x at a. For example, consider the limit

$$\lim_{x \to 0} \frac{x^2}{x}. \tag{2.2}$$

One can use an online tool like Desmos to obtain the sketch of the function in (2.2). We say that $x = 0$ is a *removable singularity* of the function even though the limit exists and is finite even though $f(0)$ is not originally defined. The numerator and denominator have a common factor of x, and one can then deduce that the limit is 0.

Consider the parabola $y = x^2$. Fix the point $P(1,1)$ on the parabola and let $Q\left(x,x^2\right)$ be another point on the same curve with $x \neq 1$. The slope of the secant line through P and Q is

$$\frac{x^2-1}{x-1}.$$

This slope depends on the choice of Q, i.e. on the parameter x. Intuitively, if we let Q move along the parabola and approach the point P, the secant line should approach a limiting position, which we call the tangent line to the parabola at P. To make sense of this limiting process, we must understand what it means for a quantity depending on x to approach some value as x approaches a certain point.

As another motivation, consider the motion of a falling object in a uniform gravitational field. Ignoring air resistance, its displacement from the starting point at time t is given by $s(t) = \frac{1}{2}gt^2$, where $g > 0$ is the gravitational constant. The *average velocity* of the object over the time interval $[5, 5+h]$ (with $h \neq 0$) is

$$v = \frac{s(5+h) - s(5)}{(5+h)-5} = \frac{1}{2}g(10+h).$$

As h becomes very small, the average velocity over $[5,5+h]$ should approach the instantaneous velocity at $t = 5$. Again this is a limiting process: we want to understand what happens as $h \to 0$. These two examples (tangent lines and instantaneous velocity) motivate the formal study of the limits of functions (a more rigorous discussion in Chapter 2.2).

We now discuss some limit laws (Propositions 2.1 and 2.2 and Theorem 2.3).

Proposition 2.1 (basic limit laws). We have the following:

(i) Constant function: For any $c \in \mathbb{R}$, the constant function $f(x) = c$ is not affected by x. So for any $a \in \mathbb{R}$,

$$\lim_{x \to a} c = c.$$

(ii) For any $a \in \mathbb{R}$,

$$\lim_{x \to a} x = a$$

(iii) Scalar multiple: Let $c \in \mathbb{R}$ be arbitrary. Then,

$$\lim_{x \to a} (cf(x)) = c \lim_{x \to a} f(x).$$

Proposition 2.2 (basic limit laws). Suppose

$$\lim_{x \to a} f(x) = L \quad \text{and} \quad \lim_{x \to a} g(x) = M.$$

Then, the following hold:

(i) Sum and difference: We have

$$\lim_{x\to a}[f(x)\pm g(x)] = \lim_{x\to a} f(x) \pm \lim_{x\to a} g(x) = L\pm M$$

(ii) Product: We have

$$\lim_{x\to a}[f(x)g(x)] = \lim_{x\to a} f(x) \cdot \lim_{x\to a} g(x) = LM$$

(iii) Quotient: We have

$$\lim_{x\to a}\frac{f(x)}{g(x)} = \frac{\lim_{x\to a} f(x)}{\lim_{x\to a} g(x)} = \frac{L}{M} \quad \text{provided that } \lim_{x\to a} g(x) \neq 0$$

The rules in Proposition 2.2 are natural if we think of replacing $f(x)$ by its limit L and $g(x)$ by its limit M. From **(ii)** of Proposition 2.2, one might naturally extend it to the limit of a product of multiple functions. That is, given functions $f_1(x),\ldots,f_n(x)$, we have

$$\lim_{x\to a}[f_1(x)\ldots f_n(x)] = \lim_{x\to a} f_1(x)\cdot\ldots\cdot\lim_{x\to a} f_n(x).$$

In particular, if all the functions are equal, i.e. $f_1 = \ldots = f_n = f$, then the limit of the product of n copies of f can be written as follows:

$$\lim_{x\to a}[f(x)\ldots f(x)] = \lim_{x\to a} f(x)\cdot\ldots\cdot\lim_{x\to a} f(x)$$

which is equivalent to saying that for any $n\in\mathbb{N}$, we have

$$\lim_{x\to a}(f(x))^n = \left(\lim_{x\to a} f(x)\right)^n.$$

After some discussion on exponentiation, what about taking the n^{th} root, where $n\in\mathbb{N}$? Suppose

$$\lim_{x\to a} f(x) = L \quad \text{and} \quad \lim_{x\to a} g(x) = A.$$

Then,

$$A^n = \left(\lim_{x\to a}\sqrt[n]{f(x)}\right)^n = \lim_{x\to a}\left(\sqrt[n]{f(x)}\right)^n = \lim_{x\to a} f(x) = L,$$

where we could *pull* the limit symbol out by our earlier discussion on exponentiation. So, provided that $\sqrt[n]{f(x)}$ is well-defined, we have $A = \sqrt[n]{L}$.

Theorem 2.1 (direct substitution). Let f be a polynomial or a rational function, and let a be a point in the domain of f. Then

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

The property in Theorem 2.1 is usually referred to as continuity at the point a. Thus, polynomials and rational functions are continuous at every point of their domain. See Chapter 3 for further discussion on the continuity of functions.

Proof. We first prove the result for polynomials. Let

$$f(x) = c_n x^n + c_{n-1} x^{n-1} + \cdots + c_1 x + c_0$$

be a polynomial of degree n. By definition, $c_0, \ldots, c_n \in \mathbb{R}$ and $c_n \neq 0$. By the limit laws in Proposition 2.2, we have

$$\begin{aligned} \lim_{x \to a} f(x) &= \lim_{x \to a} \left(c_n x^n + \cdots + c_1 x + c_0 \right) \\ &= c_n \lim_{x \to a} x^n + \cdots + c_1 \lim_{x \to a} x + \lim_{x \to a} c_0. \end{aligned}$$

By Proposition 2.1, it is easy to deduce that

$$\lim_{x \to a} f(x) = c_n a^n + \cdots + c_1 a + c_0 = f(a).$$

We then prove the result for rational functions. Suppose $f(x) = \frac{P(x)}{Q(x)}$, where P and Q are polynomials and Q is not identically zero. So, the domain of f is $\{x \in \mathbb{R} : Q(x) \neq 0\}$. If a is in the domain of f, then $Q(a) \neq 0$. By the polynomial case,

$$\lim_{x \to a} P(x) = P(a) \quad \text{and} \quad \lim_{x \to a} Q(x) = Q(a) \neq 0.$$

Using the quotient law (Proposition 2.2),

$$\lim_{x \to a} f(x) = \lim_{x \to a} \frac{P(x)}{Q(x)} = \frac{\lim_{x \to a} P(x)}{\lim_{x \to a} Q(x)} = \frac{P(a)}{Q(a)} = f(a)$$

and the result follows. □

We then discuss one-sided limits. Sometimes, the behaviour of a function as x approaches a from the left and from the right can be different. So, we introduce one-sided limits.

Definition 2.1 (right-hand limit, intuitive). We say that the right-hand limit of $f(x)$ as x approaches a equals L, and write

$$\lim_{x \to a^+} f(x) = L,$$

if by taking x sufficiently close to a with $x > a$, the values $f(x)$ become arbitrarily close to L.

Definition 2.2 (left-hand limit, intuitive). We say that the left-hand limit of $f(x)$ as x approaches a equals L, and write

$$\lim_{x \to a^-} f(x) = L,$$

if by taking x sufficiently close to a with $x < a$, the values $f(x)$ become arbitrarily close to L.

Theorem 2.2. The two-sided limit $\lim_{x\to a} f(x) = L$ exists and equals L if and only if

$$\lim_{x\to a^+} f(x) = L \quad \text{and} \quad \lim_{x\to a^-} f(x) = L.$$

Example 2.1 (absolute value). Consider the function $f(x) = |x|$. For $x > 0$, we have $|x| = x$, so

$$\lim_{x\to 0^+} |x| = \lim_{x\to 0^+} x = 0.$$

For $x < 0$, we have $|x| = -x$, so

$$\lim_{x\to 0^-} |x| = \lim_{x\to 0^-} (-x) = 0.$$

Since both one-sided limits exist and are equal to 0, we obtain

$$\lim_{x\to 0} |x| = 0.$$

One affirms the validity of Theorem 2.2.

Example 2.2. Consider the function $\frac{|x|}{x}$ defined for $x \neq 0$. For $x > 0$, $\frac{|x|}{x} = 1$, so

$$\lim_{x\to 0^+} \frac{|x|}{x} = 1.$$

For $x < 0$, $\frac{|x|}{x} = -1$, so

$$\lim_{x\to 0^-} \frac{|x|}{x} = -1.$$

The two one-sided limits are different, hence the two-sided limit does not exist.

Definition 2.3 (infinite limits, intuitive). Let f be defined near a (except possibly at a).

(i) We write

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

if by taking x sufficiently close to a, the values $f(x)$ become arbitrarily large

(ii) We write

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

if by taking x sufficiently close to a, the values $f(x)$ become arbitrarily large negative numbers

Although it is convenient to write expressions such as $\lim_{x\to a} f(x) = \infty$, we should keep in mind that ∞ is not a real number. Thus, strictly speaking, a limit that is equal to ∞ or $-\infty$ is considered to *diverge* (it does not exist as a finite real number). However, infinite limits are very useful in describing vertical asymptotic behaviour of functions.

Example 2.3. Let $f(x) = \frac{1}{x^2}$. As x approaches 0, the values of $f(x)$ become arbitrarily large. Informally, we write

$$\lim_{x\to 0} \frac{1}{x^2} = \infty.$$

Theorem 2.3 (squeeze theorem). Let f, g, h be functions such that $f(x) \leq g(x) \leq h(x)$ for all x near a except possibly at a. If

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

This can be proven using the formal definition of the limit (Definition 2.4) but we will not prove it in this set of notes. For interested readers, we refer you to the proof of Theorem 14.3 on p. 47 of R. Johnsonbaugh's '*Foundations of Mathematical Analysis*' [3].

Example 2.4. Consider the limit

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

For $x \neq 0$, we know that (which is the trick)

$$-1 \leq \sin\left(\frac{1}{x}\right) \leq 1 \quad \text{so} \quad -x^2 \leq x^2 \sin\left(\frac{1}{x}\right) \leq x^2.$$

One can refer to Figure 2.1.

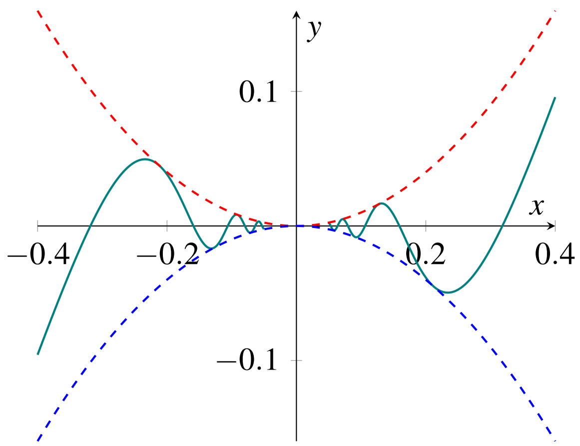


Figure 2.1: Graph of $y = x^2 \sin\left(\frac{1}{x}\right)$ with envelopes $y = \pm x^2$

Since

$$\lim_{x \to 0} \left(-x^2\right) = 0 \quad \text{and} \quad \lim_{x \to 0} x^2 = 0,$$

the squeeze theorem (Theorem 2.3) implies that

$$\lim_{x \to 0} x^2 \sin\left(\frac{1}{x}\right) = 0.$$

2.2 The Formal Definition of a Limit

The intuitive definition of a limit in Chapter 2.1 referred to phrases like 'arbitrarily close' and 'sufficiently close'. To make these ideas mathematically precise, we use the ε-δ formulation.

Definition 2.4 (formal definition of a limit). Let f be a function defined on an open interval containing a, except possibly at a. Then,

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

if for every $\varepsilon > 0$, there exists $\delta > 0$ such that whenever $0 < |x-a| < \delta$, we have $|f(x) - L| < \varepsilon$.

In Definition 2.4, the number δ may depend on ε and on the function f, but it must not depend on the choice of x. Once an $\varepsilon > 0$ is fixed, we choose a single $\delta > 0$ that works for *all* x satisfying $0 < |x-a| < \delta$. We give a pictorial definition of Definition 2.4 (see Figure 2.2). It says the following: given any $\varepsilon > 0$, we can find an interval $(a-\delta, a+\delta)$ with the point a removed, such that whenever x lies in this interval, the value $f(x)$ lies inside the interval $(L-\varepsilon, L+\varepsilon)$. In symbols,

$$0 < |x-a| < \delta \quad \Rightarrow \quad |f(x) - L| < \varepsilon.$$

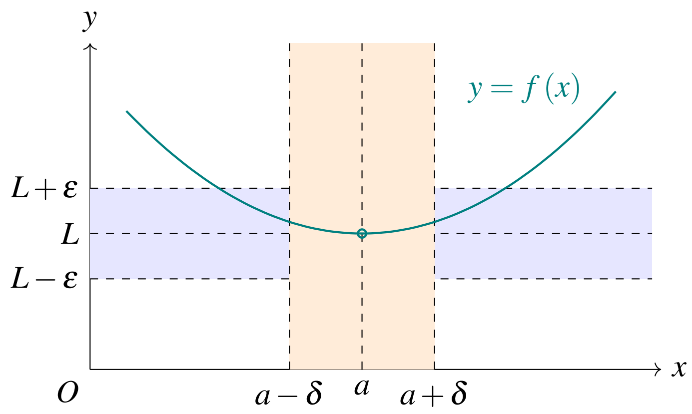


Figure 2.2: Pictorial illustration of the formal definition of a limit

Example 2.5. We now give a proof of **(i)** of Proposition 2.1 using the formal definition of a limit (Definition 2.4). That is, to show that for any constant $c \in \mathbb{R}$ and any $a \in \mathbb{R}$, we have

$$\lim_{x \to a} c = c.$$

Solution. Let c and a be given and let $\varepsilon > 0$ be arbitrary. We need to construct $\delta > 0$ such that

$$0 < |x-a| < \delta \quad \Rightarrow \quad |c - c| < \varepsilon.$$

However, $|c-c| = 0$, which is strictly less than any $\varepsilon > 0$. Thus, any positive δ works; for instance, we may choose $\delta = 1$. This verifies Definition 2.4. $\square$

Example 2.6. Use the formal definition of a limit (Definition 2.4) to prove that

$$\lim_{x \to 3} (4x - 5) = 7.$$

Solution. Let $\varepsilon > 0$ be arbitrary. We compute

$$|(4x-5)-7| = |4x-12| = 4\,|x-3|\,.$$

We want to ensure that this is less than ε. Thus, it suffices to ensure that that $4\,|x-3| < \varepsilon$, or equivalently, $|x-3| < \frac{\varepsilon}{4}$. As such, we choose $\delta = \frac{\varepsilon}{4}$. This implies that

$$0 < |x-3| < \delta \quad \Rightarrow \quad |(4x-5)-7| = 4\,|x-3| < 4\delta = \varepsilon$$

which concludes the proof. □

The ε-δ proofs often require estimates using the following classical result in Theorem 2.4, known as the triangle inequality.

Theorem 2.4 (triangle inequality). For all $a, b \in \mathbb{R}$,

$$|a+b| \le |a| + |b| \quad \text{and} \quad |a| - |b| \le |a+b|\,.$$

Proof. The first inequality $|a+b| \le |a| + |b|$ is standard; one way to show it is to square both sides and use $|ab| \le |a|\,|b|$. We omit the full proof here.

For the second inequality, note that

$$|a| = |(a+b) + (-b)| \le |a+b| + |-b| = |a+b| + |b|\,.$$

Rearranging gives $|a| - |b| \le |a+b|$ which concludes the proof. □

Example 2.7. Use the formal definition of a limit (Definition 2.4) to prove that

$$\lim_{x \to 3} x^2 = 9.$$

Solution. Let $\varepsilon > 0$ be arbitrary. We need to find $\delta > 0$ such that

$$0 < |x-3| < \delta \quad \Rightarrow \quad \left|x^2 - 9\right| < \varepsilon.$$

We first estimate

$$\left|x^2 - 9\right| = |x-3|\,|x+3|\,.$$

To control $|x+3|$, we can impose a preliminary restriction on $|x-3|$. For instance, if we require that $|x-3| < 1$, then $x \in (2,4)$, and therefore $|x+3| \le 7$. Thus, under the conditions $|x-3| < 1$ and $|x-3| < \delta$, we have

$$\left|x^2 - 9\right| = |x-3|\,|x+3| \le |x-3| \cdot 7 \le 7\delta.$$

To guarantee $\left|x^2 - 9\right| < \varepsilon$, it is enough to impose $7\delta \le \varepsilon$, or equivalently, $\delta \le \frac{\varepsilon}{7}$. We must enforce both $|x-3| < 1$ and $|x-3| < \frac{\varepsilon}{7}$. Thus we choose $\delta = \min\left\{1, \frac{\varepsilon}{7}\right\}$. Consequently,

$$0 < |x-3| < \delta \quad \Rightarrow \quad |x-3| < 1 \text{ and } |x-3| < \frac{\varepsilon}{7},$$

and so

$$\left|x^2 - 9\right| \le 7\,|x-3| < 7 \cdot \frac{\varepsilon}{7} = \varepsilon.$$

This completes the proof. □

Example 2.8 (MA2002 AY21/22 Sem 1). Using only the ε-δ definition of a limit (Definition 2.4), prove that

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

Solution. Let $\varepsilon>0$ be arbitrary. We wish to prove that for any $\varepsilon>0$, there exists $\delta>0$ such that

$$|x+1|<\delta \quad\Rightarrow\quad \left|\frac{1}{x^2}-1\right|<\varepsilon.$$

Setting $|x+1|<1/2$, we have

$$-\frac{1}{2}-1<x<\frac{1}{2}-1 \quad\text{so}\quad -\frac{3}{2}<x<-\frac{1}{2}.$$

Hence,

$$\frac{3}{2}<1-x<\frac{5}{2} \quad\text{and}\quad \frac{4}{9}\le\frac{1}{x^2}\le 4.$$

As such, we can choose $\delta=\min\left\{\frac{1}{2},\frac{\varepsilon}{10}\right\}$ so that

$$\left|\frac{1}{x^2}-1\right|=\frac{|1-x^2|}{x^2}=\frac{|1+x||1-x|}{x^2}<\delta(5/2)\cdot 4=10\delta=\varepsilon,$$

thus proving the result. □

We now show how the basic limit laws follow from the precise definition.

Theorem 2.5. Recall **(i)** of Proposition 2.1. That is, suppose $\lim_{x\to a} f(x)=L$ and $c\in\mathbb{R}$. Then,

$$\lim_{x\to a} cf(x)=cL.$$

Proof. If $c=0$, then $cf(x)=0$ and the statement reduces to $\lim_{x\to a}0=0$, which we have already proved. Now, assume that $c\neq 0$. Let $\varepsilon>0$ be arbitrary. We want $|cf(x)-cL|<\varepsilon$. Note that

$$|cf(x)-cL|=|c|\,|f(x)-L|.$$

If we ensure that

$$|f(x)-L|<\frac{\varepsilon}{|c|},$$

then automatically

$$|cf(x)-cL|<\varepsilon.$$

Since $\lim_{x\to a} f(x)=L$, there exists $\delta>0$ such that

$$0<|x-a|<\delta \quad\Rightarrow\quad |f(x)-L|<\frac{\varepsilon}{|c|}.$$

For this choice of δ,

$$0<|x-a|<\delta \quad\Rightarrow\quad |cf(x)-cL|=|c|\,|f(x)-L|<\varepsilon.$$

Thus, $\lim_{x\to a} cf(x)=cL$, which completes the proof. □

Theorem 2.6 (sum law). Recall **(i)** of Proposition 2.2. Suppose $\lim_{x\to a} f(x) = L$ and $\lim_{x\to a} g(x) = M$. Then,

$$\lim_{x\to a}(f(x)+g(x)) = L+M.$$

Proof. Let $\varepsilon > 0$ be given. We want to find $\delta > 0$ such that

$$0<|x-a|<\delta \quad \Rightarrow \quad |(f(x)+g(x))-(L+M)|<\varepsilon.$$

By the triangle inequality (Theorem 2.4),

$$\begin{aligned}|(f(x)+g(x))-(L+M)| &= |(f(x)-L)+(g(x)-M)| \\ &\leq |f(x)-L|+|g(x)-M|.\end{aligned}$$

It is enough to choose δ so that each of the two terms on the right is bounded by $\varepsilon/2$. Since $\lim_{x\to a} f(x) = L$, there exists $\delta_1 > 0$ such that

$$0<|x-a|<\delta_1 \quad \Rightarrow \quad |f(x)-L|<\frac{\varepsilon}{2}.$$

Similarly, since $\lim_{x\to a} g(x) = M$, there exists $\delta_2 > 0$ such that

$$0<|x-a|<\delta_2 \quad \Rightarrow \quad |g(x)-M|<\frac{\varepsilon}{2}.$$

Set $\delta = \min\{\delta_1, \delta_2\}$ (we need δ to be sufficiently small so we take the minimum). Then,

$$0<|x-a|<\delta \quad \Rightarrow \quad |f(x)-L|<\frac{\varepsilon}{2} \text{ and } |g(x)-M|<\frac{\varepsilon}{2}.$$

Hence,

$$|(f(x)+g(x))-(L+M)| \leq \frac{\varepsilon}{2}+\frac{\varepsilon}{2} = \varepsilon.$$

This proves the sum law. □

Using the sum and constant multiple laws, one can prove the difference law

$$\lim_{x\to a}(f(x)-g(x)) = L-M$$

by writing $f-g = f+(-1)g$. The product and quotient laws require more work. We omit the full proofs here, but the key idea is always the same: start from the desired inequality $|\ldots| < \varepsilon$ and work backwards to find suitable constraints on $|x-a|$, then close the argument using the precise definition.

Definition 2.5 (right-hand limit, precise). We say that the right-hand limit of $f(x)$ as x approaches a equals L, and write

$$\lim_{x\to a^+} f(x) = L,$$

if for every $\varepsilon > 0$ there exists $\delta > 0$ such that

$$0<x-a<\delta \quad \Rightarrow \quad |f(x)-L|<\varepsilon.$$

Definition 2.6 (left-hand limit, precise). We say that the left-hand limit of $f(x)$ as x approaches a equals L, and write

$$\lim_{x\to a^-} f(x) = L,$$

if for every $\varepsilon > 0$ there exists $\delta > 0$ such that

$$0 < a - x < \delta \quad \Rightarrow \quad |f(x) - L| < \varepsilon.$$

The equivalence between the two-sided and one-sided definitions can be proved directly from these precise formulations. Figure 2.3 depicts Definition 2.5 by showing how, for a given $\varepsilon > 0$, we can find a $\delta > 0$ such that all points x with $0 < x - a < \delta$ are mapped by f into the horizontal ε-band around L.

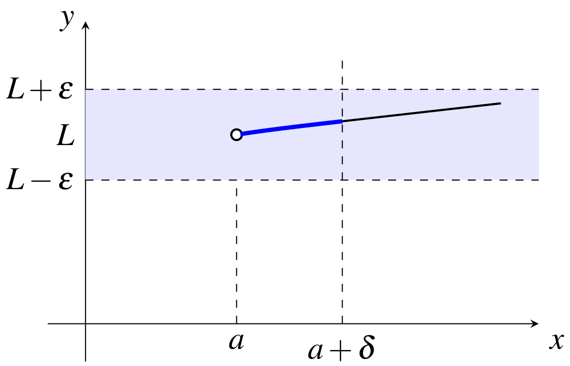


Figure 2.3: Formal definition of right-hand limit

Definition 2.7 (infinite limit, precise). Let f be defined on an open interval containing a, except possibly at a.

(i) We write

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

if for every $M > 0$ there exists $\delta > 0$ such that

$$0 < |x - a| < \delta \quad \Rightarrow \quad f(x) > M.$$

See Figure

(ii) We write

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

if for every $M < 0$ there exists $\delta > 0$ such that

$$0 < |x - a| < \delta \quad \Rightarrow \quad f(x) < M.$$

Figure 2.4 gives a depiction of Definition 2.7. In the picture, the horizontal dashed line at height $y = M$ represents an arbitrary large threshold, while the two vertical dashed lines at $x = a - \delta$ and $x = a + \delta$ mark the endpoints of the interval $(a - \delta, a + \delta)$ around

a. The shaded rectangle above the line $y = M$ between these two vertical lines illustrates the set of points (x, y) with $0 < |x - a| < \delta$ and $y > M$. No matter how large we make M, we can always find a sufficiently small neighbourhood of a (excluding the point a itself) on which all the values of f lie above this horizontal level.

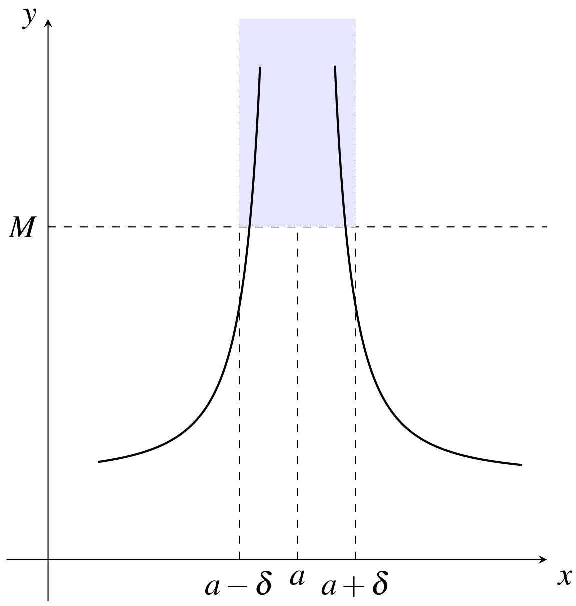


Figure 2.4: Formal definition of an infinite limit: $\lim_{x \to a} f(x) = \infty$

Example 2.9. Show that

$$\lim_{x \to 0} \frac{1}{x^2} = \infty.$$

Solution. Let $M > 0$ be arbitrary. We wish to find $\delta > 0$ such that

$$0 < |x - 0| < \delta \quad \Rightarrow \quad \frac{1}{x^2} > M.$$

The inequality $\frac{1}{x^2} > M$ is equivalent to $0 < |x| < \frac{1}{\sqrt{M}}$. Thus, it suffices to choose $\delta = \frac{1}{\sqrt{M}}$. Then,

$$0 < |x| < \delta \quad \Rightarrow \quad \frac{1}{x^2} > M.$$

Since this works for every $M > 0$, the result follows. □

Theorem 2.7 (equality of limits for coinciding functions). Suppose that $f(x) = g(x)$ for all x in some open interval containing a, except possibly at a itself. If

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

Proof. There exists $r > 0$ such that

$$0 < |x - a| < r \quad \Rightarrow \quad f(x) = g(x).$$

Let $\varepsilon > 0$. Since $\lim_{x \to a} f(x) = L$, there exists $\delta_1 > 0$ such that

$$0 < |x - a| < \delta_1 \quad \Rightarrow \quad |f(x) - L| < \varepsilon.$$

Set $\delta = \min\{r, \delta_1\}$. Then

$$0 < |x-a| < \delta \quad \Rightarrow \quad f(x) = g(x) \text{ and } |f(x) - L| < \varepsilon.$$

Hence,

$$|g(x) - L| = |f(x) - L| < \varepsilon.$$

By the precise definition, $\lim_{x\to a} g(x) = L$. □

Lemma 2.1. Let f be a function such that $f(x) \geq 0$ for all x in some open interval containing a, except possibly at a. If $\lim_{x\to a} f(x) = L$, then $L \geq 0$.

Proof. Suppose on the contrary that $L < 0$. Let $\varepsilon = -L > 0$. By the formal definition of a limit, there exists $\delta > 0$ such that

$$0 < |x-a| < \delta \quad \Rightarrow \quad |f(x) - L| < \varepsilon.$$

The inequality $|f(x) - L| < \varepsilon$ implies

$$L - \varepsilon < f(x) < L + \varepsilon.$$

But here $L - \varepsilon = L - (-L) = 2L < 0$ and $L + \varepsilon = 0$. Thus, for $0 < |x-a| < \delta$,

$$2L < f(x) < 0.$$

In particular, $f(x) < 0$ for x near a. This contradicts the assumption $f(x) \geq 0$ near a. Hence, our assumption $L < 0$ is impossible, and we must have $L \geq 0$. □

Theorem 2.8. Suppose $f(x) \geq g(x)$ for all x in some open interval containing a, except possibly at a. If $\lim_{x\to a} f(x) = L$ and $\lim_{x\to a} g(x) = M$, then $L \geq M$.

Proof. Define $h(x) = f(x) - g(x)$. Then $h(x) \geq 0$ near a, and

$$\lim_{x\to a} h(x) = \lim_{x\to a} f(x) - \lim_{x\to a} g(x) = L - M.$$

By Lemma 2.1, $L - M \geq 0$, hence $L \geq M$. □

Chapter 3

Continuity

3.1 Introduction

In the study of limits, we saw that for many functions the limit of $f(x)$ as $x \to a$ coincides with the value $f(a)$ whenever the latter is defined. Such points are precisely the points where the function is *continuous*.

For polynomials and rational functions, we have the following fundamental property:

Theorem 3.1 (direct substitution property). Let f be a polynomial or a rational function, and let a be a point in the domain of f. Then,

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

Thus, for polynomials and rational functions, finite limits at points in their domain can be computed simply by substituting $x = a$. In particular, any polynomial is continuous on $\mathbb{R}$, and any rational function is continuous on its domain.

We now formulate the general definition.

Definition 3.1 (continuity at a point). Let f be a function and let $a \in \mathbb{R}$. We say that f is continuous at a if

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

If f is not continuous at a, we say that f is discontinuous at a.

We shall unpack Definition 3.1. As such, continuity at a point a consists of three separate conditions:

(i) $f(a)$ is defined (that is, a lies in the domain of f)

(ii) the limit $\lim_{x \to a} f(x)$ exists as a real number

(iii) the value of this limit equals the function value: $\lim_{x\to a} f(x) = f(a)$

Using the precise ε-δ notion of limit, continuity at a can be written as follows (see Proposition 3.1):

Proposition 3.1. A function f is continuous at a if and only if for every $\varepsilon > 0$ there exists $\delta > 0$ such that

$$|x-a| < \delta \quad \Rightarrow \quad |f(x) - f(a)| < \varepsilon.$$

Note that the condition $0 < |x-a|$ is no longer necessary in the definition of continuity in Proposition 3.1 because we also require the inequality at $x = a$, where both sides are equal.

3.2 Types of Discontinuities

When a function fails to be continuous at a point, it is useful to classify the type of discontinuity. We begin by discussing what a removable discontinuity is with an example. Let

$$f(x) = \frac{x^2 - x - 2}{x - 2} \quad \text{where } x \neq 2.$$

We can factor the numerator as $x^2 - x - 2 = (x-2)(x+1)$, so for $x \neq 2$, we have $f(x) = x + 1$. So,

$$\lim_{x\to 2} f(x) = \lim_{x\to 2} (x+1) = 3,$$

but $f(2)$ is not defined since $x = 2$ is excluded from the domain. As such, f is discontinuous at 2, even though the limit exists. This leads to the following notion (Definition 3.2):

Definition 3.2 (removable discontinuity and continuous extension). Let f be a function and $a \in \mathbb{R}$ such that the limit $\lim_{x\to a} f(x)$ exists. If either

- $f(a)$ is not defined, or
- $f(a)$ is defined but $f(a) \neq \lim_{x\to a} f(x)$,

then we say that f has a removable discontinuity at a. Next, define a new function f_1 as follows:

$$f_1(x) = \begin{cases} f(x) & x \neq a; \\ \lim_{x\to a} f(x) & x = a. \end{cases}$$

Then, f_1 is continuous at a and is called the continuous extension of f at a.

Definition 3.3 (infinite discontinuity). Let f be a function and $a \in \mathbb{R}$. Suppose that at least one of the one-sided limits is infinite, i.e.

$$\lim_{x \to a^+} f(x) = \pm\infty \quad \text{or} \quad \lim_{x \to a^-} f(x) = \pm\infty.$$

Then the vertical line $x = a$ is called a vertical asymptote of the graph of $y = f(x)$, and we say that f has an infinite discontinuity at a.

For rational functions, vertical asymptotes occur exactly at the zeros of the denominator (after cancellation of common factors).

Proposition 3.2. Let $f(x) = \frac{P(x)}{Q(x)}$ be a rational function, where P and Q are polynomials with no common factor of positive degree. Then f has an infinite discontinuity at a if and only if $Q(a) = 0$.

Another important type of discontinuity occurs when left and right limits are finite but unequal. We call these jump discontinuities (Definition 3.4).

Definition 3.4 (jump discontinuity). Let f be a function and $a \in \mathbb{R}$. Suppose that both one-sided limits $\lim_{x \to a^-} f(x)$ and $\lim_{x \to a^+} f(x)$ exist as real numbers, but

$$\lim_{x \to a^-} f(x) \neq \lim_{x \to a^+} f(x).$$

Then we say that f has a jump discontinuity at a.

Example 3.1 (floor function). For each $x \in \mathbb{R}$, there exists a unique integer n such that $n \leq x < n+1$. This integer is denoted by $\lfloor x \rfloor$ and called the *floor* (or greatest integer) of x (Figure 3.1). For every $n \in \mathbb{Z}$, we have

$$\lim_{x \to n^+} \lfloor x \rfloor = n \quad \text{and} \quad \lim_{x \to n^-} \lfloor x \rfloor = n - 1,$$

so these two one-sided limits are different and the two-sided limit does not exist. Thus, $\lfloor x \rfloor$ has a jump discontinuity at each $n \in \mathbb{Z}$.

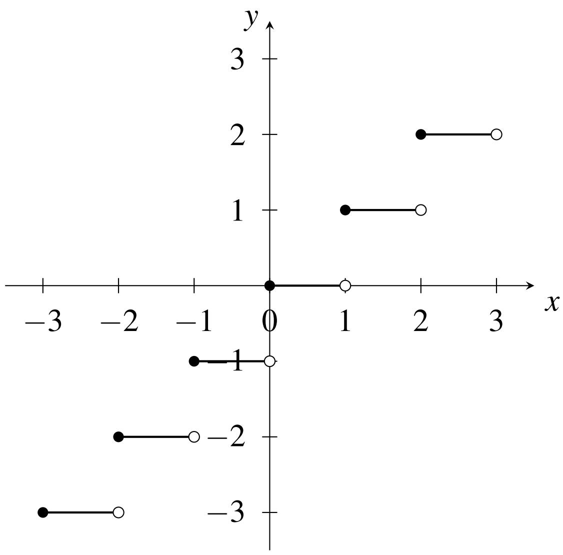


Figure 3.1: Graph of the floor function $y = \lfloor x \rfloor$

Definition 3.5 (one-sided continuity). Let f be a function and $a \in \mathbb{R}$.

(i) We say that f is continuous from the right at a if

$$\lim_{x \to a^+} f(x) = f(a)$$

(ii) We say that f is continuous from the left at a if

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

The relation between one-sided and two-sided continuity is as follows:

Theorem 3.2. A function f is continuous at a if and only if it is continuous from both the left and the right at a. That is,

$$\lim_{x \to a} f(x) = f(a) \quad \Leftrightarrow \quad \lim_{x \to a^-} f(x) = f(a) \text{ and } \lim_{x \to a^+} f(x) = f(a).$$

Definition 3.6 (continuity on intervals). Let f be a function and let I be an interval.

(i) f is continuous on (a,b) if f is continuous at every point $x \in (a,b)$

(ii) f is continuous on $[a,b]$ if f is continuous at every $x \in (a,b)$, f is continuous from the right at a, and f is continuous from the left at b

(iii) f is continuous on $[a,b)$ if f is continuous at every $x \in (a,b)$ and f is continuous from the right at a

(iv) f is continuous on $(a,b]$ if f is continuous at every $x \in (a,b)$ and f is continuous from the left at b

3.3 Algebra of Continuous Functions

We next show that continuity is preserved under the usual algebraic operations.

Theorem 3.3 (algebra of continuous functions). Let f and g be functions that are continuous at a, and let $c \in \mathbb{R}$ be a constant. Then, the following hold:

- **(i)** the function $x \mapsto cf(x)$ is continuous at a
- **(ii)** the sum $f+g$ is continuous at a
- **(iii)** the difference $f-g$ is continuous at a
- **(iv)** the product fg is continuous at a
- **(v)** the quotient f/g is continuous at a whenever $g(a) \neq 0$

Proof. Since f and g are continuous at a, then

$$\lim_{x \to a} f(x) = f(a) \quad \text{and} \quad \lim_{x \to a} g(x) = g(a).$$

The basic limit laws in Propositions 2.1 and 2.2 then give

$$\lim_{x \to a} cf(x) = cf(a) \quad \text{and} \quad \lim_{x \to a} \big(f(x) + g(x)\big) = f(a) + g(a),$$

and similarly for the difference, product, and (when $g(a) \neq 0$) quotient. In each case, the limit equals the value of the corresponding function at $x = a$, so the resulting function is continuous at a. □

Proposition 3.3. The constant function $f(x) = c$ is continuous on $\mathbb{R}$.

Proof. Let $a \in \mathbb{R}$ and $\varepsilon > 0$ be arbitrary. Take any $\delta > 0$, for instance $\delta = 1$. Then,

$$|x-a| < \delta \quad \Rightarrow \quad |f(x) - f(a)| = |c-c| = 0 < \varepsilon.$$

Hence, $\lim_{x \to a} f(x) = f(a) = c$. □

Proposition 3.4. The identity function $f(x) = x$ is continuous on $\mathbb{R}$.

Proof. Let $a \in \mathbb{R}$ and $\varepsilon > 0$ be arbitrary. Choose $\delta = \varepsilon$. If $|x-a| < \delta$, then

$$|f(x) - f(a)| = |x-a| < \delta = \varepsilon,$$

and the result follows. □

We then discuss the continuity of power functions. Let $n \in \mathbb{N}$. A power function is of the form $x \mapsto x^n$.

Proposition 3.5. Each power function x^n with $n \in \mathbb{N}$ is continuous on $\mathbb{R}$.

Proof. We can prove this by induction on n. The case $n=1$ is the identity function which has already been treated in Proposition 3.4. If x^n is continuous and the identity function is continuous, then the product $x^{n+1} = x^n \cdot x$ is continuous by **(iv)** of Theorem 3.3. Thus, all powers are continuous. □

Next, recall that a monomial has the form cx^n with $c \in \mathbb{R}$. Since it is the product of the constant function c and the continuous power function x^n, every monomial is continuous on $\mathbb{R}$. A polynomial is a finite sum of monomials. That is,

$$P(x) = c_n x^n + \cdots + c_1 x + c_0$$

so by repeated use of Theorem 3.3, we obtain the following result (Theorem 3.4):

Theorem 3.4. Every polynomial is continuous on $\mathbb{R}$.

Recall that a rational function is a quotient $R(x) = \frac{P(x)}{Q(x)}$, where P and Q are polynomials and Q is not the zero polynomial. The domain of R is $\{x \in \mathbb{R} : Q(x) \neq 0\}$. Using the quotient rule for continuous functions (**(v)** of Theorem 3.3), we obtain the following result (Theorem 3.5):

Theorem 3.5. Every rational function is continuous on its domain.

3.4 Substitution in Limits

We now consider how limits behave under change of variables.

Theorem 3.6 (substitution in limits). Let f and g be functions. Suppose that

1. $\lim\limits_{x \to a} f(x) = b$
2. $\lim\limits_{y \to b} g(y) = c$
3. $f(x) \neq b$ for all x in some open interval around a, except possibly at $x = a$

Then,

$$\lim_{x \to a} g(f(x)) = c.$$

Proof. Let $\varepsilon > 0$ be arbitrary. Because $\lim\limits_{y \to b} g(y) = c$, there exists $\delta_2 > 0$ such that

$$0 < |y - b| < \delta_2 \quad \Rightarrow \quad |g(y) - c| < \varepsilon.$$

Because $\lim\limits_{x \to a} f(x) = b$, there exists $\delta > 0$ such that

$$0 < |x - a| < \delta \quad \Rightarrow \quad 0 < |f(x) - b| < \delta_2.$$

Now for such x,

$$0 < |f(x) - b| < \delta_2 \quad \Rightarrow \quad |g(f(x)) - c| < \varepsilon.$$

Hence,

$$0 < |x-a| < \delta \quad \Longrightarrow \quad |g(f(x)) - c| < \varepsilon,$$

which proves the claim. □

A particularly important special case is the substitution $x = a+h$. If we let $h = x-a$, then $x \to a$ with $x \neq a$ corresponds to $h \to 0$ with $h \neq 0$. This yields the following result (Theorem 3.7):

Theorem 3.7. For any function f and any $a \in \mathbb{R}$,

$$\lim_{x\to a} f(x) = \lim_{h\to 0} f(a+h),$$

provided at least one of these limits exists.

Consequently, continuity at a can also be characterised by a limit in the increment h. As such, a function f is continuous at a if and only if

$$\lim_{h\to 0} f(a+h) = f(a).$$

The substitution theorem in Theorem 3.7 can be simplified considerably when the outer function is continuous.

Theorem 3.8 (limit of a composite function). Suppose that

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

and that g is continuous at b. Then,

$$\lim_{x\to a} g(f(x)) = g(b).$$

Proof. Since g is continuous at b, for every $\varepsilon > 0$, there exists $\delta_2 > 0$ such that

$$|y-b| < \delta_2 \quad \Rightarrow \quad |g(y) - g(b)| < \varepsilon.$$

Because $\lim_{x\to a} f(x) = b$, there exists $\delta > 0$ such that

$$0 < |x-a| < \delta \quad \Rightarrow \quad |f(x) - b| < \delta_2.$$

Combining the two implications (with $y = f(x)$) yields

$$0 < |x-a| < \delta \quad \Longrightarrow \quad |g(f(x)) - g(b)| < \varepsilon,$$

which proves the desired limit. □

Because the limit of $f(x)$ as $x \to a$ is $f(a)$ whenever f is continuous at a, we also obtain a simple continuity rule for composites (Theorem 3.9).

Theorem 3.9 (continuity of a composite function). If f is continuous at a and g is continuous at $f(a)$, then the composite function $g \circ f$ defined by

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

is continuous at a.

Proof. We have

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

Thus, $\lim_{x \to a} (g \circ f)(x) = (g \circ f)(a)$ and the result follows. □

3.5 Root Functions and Rational Powers

Let $n \in \mathbb{N}$. The n^{th} root function is $x \mapsto x^{1/n} = \sqrt[n]{x}$. When n is odd, $\sqrt[n]{x}$ is defined for all $x \in \mathbb{R}$ as the unique real number y such that $y^n = x$. When n is even, $\sqrt[n]{x}$ is defined for $x \geq 0$ as the unique number $y \geq 0$ such that $y^n = x$.[1]

Theorem 3.10. Let $n \in \mathbb{N}$. Then the function $x \mapsto \sqrt[n]{x}$ is continuous on $\mathbb{R}$ if n is odd, and continuous on $[0, \infty)$ if n is even.

A complete proof uses the fact that the function $x \mapsto x^n$ is continuous and strictly monotone on its domain, and therefore has a continuous inverse. Instead, we give an explicit ε-δ proof in the important case $x \mapsto \sqrt{x}$ at a positive point in Theorem 3.11.

Theorem 3.11. The square-root function $x \mapsto \sqrt{x}$ is continuous at every $a > 0$.

Proof. Fix $a > 0$ and let $\varepsilon > 0$ be arbitrary. We must find $\delta > 0$ such that

$$0 < |x - a| < \delta \quad \Rightarrow \quad \left|\sqrt{x} - \sqrt{a}\right| < \varepsilon.$$

First, since $\sqrt{x}$ is defined only for $x \geq 0$, we must ensure that our δ is small enough so that $x \geq 0$ whenever $|x - a| < \delta$. It suffices to require $\delta \leq a$. Then,

$$0 < |x - a| < \delta \leq a \quad \Rightarrow \quad x > a - \delta \geq 0. \tag{3.1}$$

For such x, we have

$$\left|\sqrt{x} - \sqrt{a}\right| = \frac{|x - a|}{\sqrt{x} + \sqrt{a}} < \frac{|x - a|}{\sqrt{a}} \leq \frac{\delta}{\sqrt{a}},$$

because $\sqrt{x} + \sqrt{a} \geq \sqrt{a}$. Thus it is enough to choose δ such that $\delta/\sqrt{a} \leq \varepsilon$, i.e. $\delta \leq \varepsilon\sqrt{a}$. Combining this with the earlier restriction $\delta \leq a$ (3.1), we set $\delta = \min\{a, \varepsilon\sqrt{a}\}$. Then,

$$0 < |x - a| < \delta \quad \Rightarrow \quad \left|\sqrt{x} - \sqrt{a}\right| < \frac{\delta}{\sqrt{a}} \leq \varepsilon,$$

which proves continuity at a. □

[1] For readers interested in the *proof* of this result, please refer to Theorem 1.21 of W. Rudin's classic '*Principles of Mathematical Analysis*' [4]. You have been warned.

We then discuss continuity for rational powers of x. Every rational number $r \in \mathbb{Q}$ can be written uniquely as $r = m/n$, where $m \in \mathbb{Z}$, $n \in \mathbb{N}$ and $\gcd(m,n) = 1$. If n is odd, we define

$$x^{m/n} = \left(\sqrt[n]{x}\right)^m \quad \text{for all } x \in \mathbb{R}.$$

On the other hand, if n is even, we define

$$x^{m/n} = \left(\sqrt[n]{x}\right)^m \quad \text{for } x \geq 0.$$

Theorem 3.12. For every rational $r \in \mathbb{Q}$, the power function $x \mapsto x^r$ is continuous on its natural domain.

We give a rough sketch of the proof. Write $r = m/n$ with m and n as per the above discussion. Then, $x^r = (\sqrt[n]{x})^m$ is a composition and product of continuous functions, hence is continuous on the domain where it is defined.

3.6 Trigonometric Functions and their Continuity

A geometric argument (Figure 3.2) shows that for $0 < x < \pi/2$, we have

$$0 < \sin x < x.$$

To see why, we consider the unit circle centred at the origin O. Let $P = (1,0)$ be the point where the circle intersects the positive x-axis, and let x be an angle satisfying $0 < x < \frac{\pi}{2}$. We draw the radius from O making an angle x with the positive x-axis; its endpoint on the unit circle is denoted by A. Since $0 < x < \frac{\pi}{2}$, the point A lies in the first quadrant and has coordinates $A = (\cos x, \sin x)$. In particular, $\sin x > 0$.

A basic geometric fact about circles is that between two distinct points on a circle, the straight line segment (the chord) joining them is always shorter than the corresponding circular arc. In our situation, the straight line segment $\overline{PA}$ (chord) is shorter than the arc $\widehat{PA}$, so $|PA| < x$. Moreover, the vertical segment $\overline{BA}$ is one side of the right-angled triangle $\triangle OBA$ and is certainly shorter than the hypotenuse $\overline{PA}$. It follows that $|BA| < |PA| < x$.

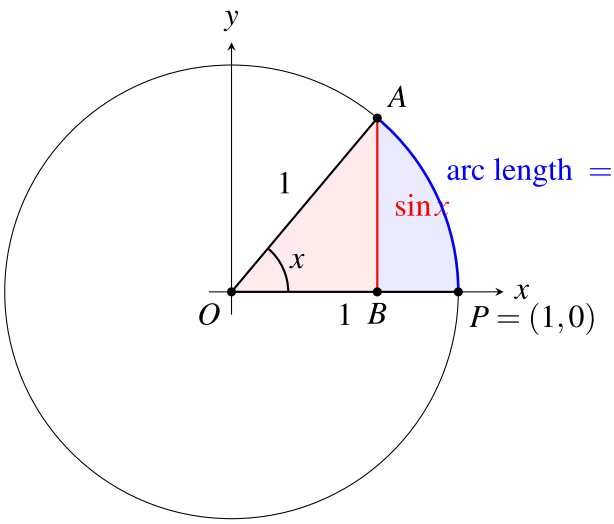


Figure 3.2: Geometric illustration of $0 < \sin x < x$ for $0 < x < \pi/2$

Therefore, as $x \to 0^+$,

$$0 \leq \sin x \leq x \quad \Rightarrow \quad \lim_{x \to 0^+} \sin x = 0$$

by the squeeze theorem. For negative x, the identity $\sin(-x) = -\sin x$ implies that for $-\pi/2 < x < 0$, we have $x < \sin x < 0$. Again, squeezing between x and 0 shows that

$$\lim_{x \to 0^-} \sin x = 0.$$

Hence, the two-sided limit exists and equals 0, which is also $\sin 0$. Therefore by Theorem 2.2, $\sin x$ is continuous at 0.

For the cosine function, one can show that for $|x| < \pi/2$, we have

$$1 - x^2 \leq \cos x \leq 1.$$

As $x \to 0$, both bounding functions tend to 1, so the squeeze theorem yields

$$\lim_{x \to 0} \cos x = 1 = \cos 0.$$

Thus, $\cos x$ is continuous at 0.

Recall that the addition formulae for sine and cosine are

$$\sin(a+b) = \sin a \cos b + \cos a \sin b$$
$$\cos(a+b) = \cos a \cos b - \sin a \sin b$$

These can be proved geometrically, but we omit the proof. Using these formulae and the continuity at 0, we can show continuity at any $a \in \mathbb{R}$ (Theorem 3.13).

Theorem 3.13. For every $a \in \mathbb{R}$,

$$\lim_{x \to a} \sin x = \sin a, \quad \text{and} \quad \lim_{x \to a} \cos x = \cos a.$$

Hence, $\sin x$ and $\cos x$ are continuous on $\mathbb{R}$.

Proof. Let $x = a + h$. Then $h \to 0$ as $x \to a$. Using the addition formulae, we have

$$\sin(a+h) = \sin a \cos h + \cos a \sin h$$
$$\cos(a+h) = \cos a \cos h - \sin a \sin h$$

Taking limits as $h \to 0$ and using $\lim_{h \to 0} \sin h = 0$ and $\lim_{h \to 0} \cos h = 1$, we obtain

$$\lim_{x \to a} \sin x = \sin a \cdot 1 + \cos a \cdot 0 = \sin a$$
$$\lim_{x \to a} \cos x = \cos a \cdot 1 - \sin a \cdot 0 = \cos a$$

Thus the result follows. $\square$

The remaining trigonometric functions are defined in terms of $\sin x$ and $\cos x$ by

$$\tan x = \frac{\sin x}{\cos x} \quad \cot x = \frac{\cos x}{\sin x} \quad \sec x = \frac{1}{\cos x} \quad \csc x = \frac{1}{\sin x}.$$

Since $\sin x$ and $\cos x$ are continuous everywhere, and quotients of continuous functions are continuous wherever the denominator is non-zero, we obtain the following theorem (Theorem 3.14):

Theorem 3.14. The functions $\sin x$ and $\cos x$ are continuous on $\mathbb{R}$. The functions $\tan x$ and $\sec x$ are continuous on

$$\mathbb{R} \setminus \left\{ \frac{\pi}{2} + k\pi \mid k \in \mathbb{Z} \right\} \quad \text{where } \cos x = 0,$$

and the functions $\cot x$ and $\csc x$ are continuous on

$$\mathbb{R} \setminus \{k\pi \mid k \in \mathbb{Z}\} \quad \text{where } \sin x = 0.$$

3.7 Intermediate Value Theorem

One of the most fundamental consequences of continuity is the intermediate value theorem, which formalizes the intuitive idea that the graph of a continuous function has no *jumps*.

Theorem 3.15 (intermediate value theorem, simple version). Let f be continuous on a closed interval $[a,b]$. Suppose that $f(a) < 0 < f(b)$ or $f(a) > 0 > f(b)$. Then, there exists a point $c \in (a,b)$ such that

$$f(c) = 0.$$

See Figure 3.3 for a geometric interpretation of Theorem 3.15.

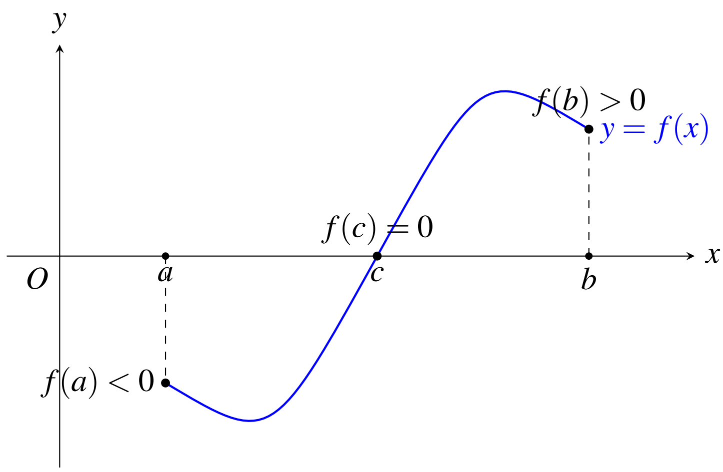


Figure 3.3: Illustration of the intermediate value theorem

Theorem 3.15 asserts the existence of at least one root in (a,b) but it does not provide its exact location or the number of roots. A more general formulation is as follows (Theorem 3.16):

Theorem 3.16 (intermediate value theorem, general version). Let f be continuous on $[a,b]$ and suppose that $f(a) \neq f(b)$. If N is any number between $f(a)$ and $f(b)$ (i.e. either $f(a) < N < f(b)$ or $f(b) < N < f(a)$), then there exists $c \in (a,b)$ such that

$$f(c) = N.$$

The simple version (Theorem 3.15) is recovered by setting $N = 0$. Also, Figure 3.4 gives a geometric interpretation of Theorem 3.16.

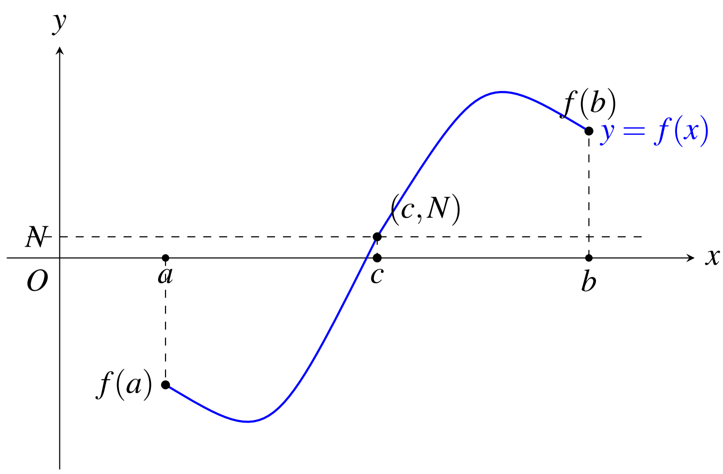


Figure 3.4: Illustration of the intermediate value theorem

Example 3.2. Show that the equation

$$4x^3 - 6x^2 + 3x - 2 = 0 \tag{3.2}$$

has at least one real root.

Solution. Let $f(x) = 4x^3 - 6x^2 + 3x - 2$. Since f is a polynomial, it is continuous on $\mathbb{R}$ by Theorem 3.4. We see that $f(0) = -2 < 0$ and $f(2) = 12 > 0$. By the intermediate value theorem (Theorem 3.15), there exists $c \in (0,2)$ such that $f(c) = 0$. Thus the equation (3.2) has at least one real solution in that interval. $\square$

Chapter 4

Differentiation and Applications

4.1 Definition of the Derivative

Consider the curve given by $y = f(x) = x^2$. We would like to find the equation of the tangent line at the point $P = \left(a, a^2\right)$. For a nearby point

$$Q = (a+h, f(a+h)) = \left(a+h, (a+h)^2\right),$$

the slope of the secant line PQ is

$$m_{PQ} = \frac{f(a+h) - f(a)}{h} = \frac{(a+h)^2 - a^2}{h}.$$

If this slope tends to a limit as $h \to 0$, we interpret the limit as the slope of the tangent line at P.

Definition 4.1 (derivative using first principles). Let f be a function and let a be in its domain. We say that f is differentiable at a if the limit

$$\lim_{h \to 0} \frac{f(a+h) - f(a)}{h}$$

exists as a finite real number. In that case the limit is called the derivative of f at a and is denoted by $f'(a)$:

$$f'(a) = \lim_{h \to 0} \frac{f(a+h) - f(a)}{h} \tag{4.1}$$

Equivalently, using the variable x approaching a, we say that f is differentiable at a with derivative $f'(a)$ if and only if

$$f'(a) = \lim_{x \to a} \frac{f(x) - f(a)}{x - a} \tag{4.2}$$

To see why, set $h = x - a$ so that $x = a + h$ and $x \to a$ is equivalent to $h \to 0$. Then,

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

and the two limits (4.1) and (4.2) coincide.

Definition 4.2 (tangent line). If f is differentiable at a, the tangent line to the curve $y = f(x)$ at the point $(a, f(a))$ is the line with equation

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

Example 4.1. Let $f(x) = x^2 - 8x + 9$. Compute $f'(3)$.

$$f'(3) = \lim_{h\to 0} \frac{f(3+h) - f(3)}{h}.$$

We have $f(3+h) = -6 - 2h + h^2$ and $f(3) = -6$. So,

$$\frac{f(3+h)-f(3)}{h} = \frac{\left(-6-2h+h^2\right) - (-6)}{h} = \frac{-2h+h^2}{h} = -2+h.$$

Taking $h \to 0$ gives

$$f'(3) = \lim_{h\to 0}(-2+h) = -2.$$

Hence, f is differentiable at $x = 3$ with derivative $f'(3) = -2$, and the tangent line at $x = 3$ is given by the equation $y = -2x$.

Example 4.2 (a function defined piecewise). Define

$$f(x) = \begin{cases} x^2 \sin\left(\frac{1}{x}\right) & \text{if } x \neq 0; \\ 0; & \text{if } x = 0. \end{cases}$$

We compute $f'(0)$ as follows:

$$f'(0) = \lim_{h\to 0} \frac{f(h) - f(0)}{h} = \lim_{h\to 0} \frac{h^2 \sin\left(\frac{1}{h}\right)}{h} = \lim_{h\to 0} h \sin\left(\frac{1}{h}\right).$$

For $h \neq 0$, we have $-1 \leq \sin\left(\frac{1}{h}\right) \leq 1$ so $\left|\sin\left(\frac{1}{h}\right)\right| \leq 1$. Thus,

$$\left|h \sin\left(\frac{1}{h}\right)\right| \leq |h| \quad \text{so} \quad -|h| \leq h\sin\left(\frac{1}{h}\right) \leq |h|.$$

Since

$$\lim_{h\to 0}(-|h|) = 0 \quad \text{and} \quad \lim_{h\to 0}|h| = 0,$$

the sqeueze theorem (Theorem 2.3) yields

$$\lim_{h\to 0} h \sin\left(\frac{1}{h}\right) = 0.$$

Therefore, $f'(0) = 0$. The tangent line at $x = 0$ is the x–axis.

Example 4.3 (falling body). Let $s = s(t)$ describe the position of a particle moving along a straight line at time t. The instantaneous velocity at time a is defined to be the derivative

$$v(a) = s'(a).$$

The speed at $t = a$ is the absolute value

$$|v(a)| = \left|s'(a)\right|.$$

Suppose a ball is dropped from rest near the Earth's surface, so that its height is

$$s(t) = \frac{1}{2}gt^2 \quad \text{where } t \geq 0$$

and $g \approx 9.8\,\mathrm{m/s^2}$ is the gravitational acceleration. Then,

$$s'(t) = \lim_{h\to 0} \frac{\frac{1}{2}g(t+h)^2 - \frac{1}{2}gt^2}{h} = \lim_{h\to 0} \frac{\frac{1}{2}g\left(2th+h^2\right)}{h} = \lim_{h\to 0} \frac{1}{2}g(2t+h) = gt.$$

Thus, the instantaneous velocity after 5 seconds is $s'(5) = 5g \approx 49\,\mathrm{m/s}$.

4.2 Derivative as a Function

Definition 4.1 gives, for each fixed a, a number $f'(a)$ whenever the limit exists. In many cases, we can express the derivative for all x at once. Let f be a function. The derivative of f is the function f' defined (by its values at points where the limit exists) by

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

Equivalently,

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

The process of computing f' is called differentiation.

Example 4.4. Let $f(x) = \frac{1}{x}$, where $x \neq 0$. Fix $a \neq 0$. Then

$$f'(a) = \lim_{x\to a} \frac{f(x) - f(a)}{x-a} = \lim_{x\to a} \frac{\frac{1}{x} - \frac{1}{a}}{x-a} = \lim_{x\to a} \frac{a-x}{ax(x-a)} = \lim_{x\to a} \frac{-1}{ax} = -\frac{1}{a^2}.$$

Thus, the derivative function is

$$f'(x) = -\frac{1}{x^2} \quad \text{where } x \neq 0.$$

Definition 4.3. A function f is said to be differentiable on an open interval I if it is differentiable at each point $a \in I$.

Theorem 4.1. If f is differentiable at a, then f is continuous at a.

Proof. Assume f is differentiable at a, so by (4.2),

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

exists. We want to show

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

For $x \neq a$, the trick is to write

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

Hence,

$$f(x) = \frac{f(x) - f(a)}{x - a} \cdot (x - a) + f(a).$$

Taking limits as $x \to a$ and using the algebra of limits,

$$\lim_{x \to a} f(x) = \left(\lim_{x \to a} \frac{f(x) - f(a)}{x - a} \right) \left(\lim_{x \to a} (x - a) \right) + \lim_{x \to a} f(a) = f'(a) \cdot 0 + f(a) = f(a).$$

Thus, f is continuous at a. $\square$

Note that the converse of Theorem 4.1 is false. That is, continuity does not imply differentiability. A standard example is $f(x) = |x|$, which is continuous everywhere but not differentiable at 0.

Example 4.5 (MA2002 AY21/22 Sem 1). Let

$$f(x) = \begin{cases} e^{-1/x^2} & \text{if } x \neq 0; \\ 0 & \text{if } x = 0. \end{cases}$$

Show that f is twice differentiable and find $f''(x)$.

Solution. For $x \neq 0$, we have $f(x) = e^{-1/x^2}$. Then,

$$f'(x) = e^{-1/x^2} \cdot \frac{2}{x^3} = \frac{2}{x^3} e^{-1/x^2} \quad \text{where } x \neq 0.$$

Differentiating again using the product rule (Theorem 4.2) yields

$$f''(x) = \frac{4 - 6x^2}{x^6} e^{-1/x^2} \quad \text{where } x \neq 0.$$

We have found an expression for $f(x)$. Now, we shall prove that f is twice differentiable on $\mathbb{R}$ by first proving that f' exists at 0 (i.e. prove that $f'(0)$ exists and we will see that it is equal to 0) and then proving that f'' exists at 0 (i.e. prove that $f''(0)$ exists and we will see that it is equal to 0). By (4.2), we have

$$f'(0) = \lim_{h \to 0} \frac{f(h) - f(0)}{h} = \lim_{h \to 0} \frac{e^{-1/h^2}}{h} = 0.$$

Here, we used L'Hôpital's rule (Theorem 4.20), which will be covered in due course. Hence, f' exists at 0 and $f'(0) = 0$.

Next, for $h \neq 0$, by (4.2) again, we have

$$\frac{f'(h) - f'(0)}{h} = \frac{f'(h)}{h} = \frac{2}{h^4} e^{-1/h^2}.$$

Again, by a repeated application of L'Hôpital's rule (Theorem 4.20), we have

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

Therefore, f is twice differentiable on $\mathbb{R}$. To conclude,

$$f''(x) = \begin{cases} \frac{4 - 6x^2}{x^6} e^{-1/x^2} & \text{if } x \neq 0; \\ 0 & \text{if } x = 0. \end{cases}$$

$\square$

4.3 Differentiation Rules

We now develop general rules that allow us to differentiate combinations of functions more easily.

Proposition 4.1. Let $c \in \mathbb{R}$ be a constant and f be differentiable at x. Then,

$$\frac{d}{dx}(c) = 0 \quad \text{and} \quad \frac{d}{dx}(cf(x)) = cf'(x).$$

Proof. For the constant function $f(x) = c$,

$$\frac{d}{dx}(c) = \lim_{h\to 0} \frac{c-c}{h} = \lim_{h\to 0} 0 = 0.$$

For the constant multiple, we have

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

□

Proposition 4.2. If f and g are differentiable at x, then so are $f+g$ and $f-g$, and

$$(f+g)'(x) = f'(x) + g'(x) \quad \text{and} \quad (f-g)'(x) = f'(x) - g'(x).$$

Proof. For the sum,

$$\begin{aligned}
(f+g)'(x) &= \lim_{h\to 0} \frac{(f(x+h)+g(x+h)) - (f(x)+g(x))}{h} \\
&= \lim_{h\to 0} \frac{f(x+h)-f(x)}{h} + \lim_{h\to 0} \frac{g(x+h)-g(x)}{h} \\
&= f'(x) + g'(x).
\end{aligned}$$

The difference rule follows by writing $f - g = f + (-1)g$ and using the constant multiple rule (Proposition 4.1). □

We then introduce the product and quotient rules (Theorems 4.2 and 4.3) but omit their proofs.

Theorem 4.2 (product rule). If f and g are differentiable at x, then fg is differentiable at x and

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

Theorem 4.3 (quotient rule). Let f and g be differentiable at x and assume $g(x) \neq 0$. Then

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

Theorem 4.4 (power rule for integer exponents). Let $n \in \mathbb{Z}$. Then the function $x \mapsto x^n$ is differentiable on its domain and

$$\frac{d}{dx}(x^n) = nx^{n-1}, \tag{4.3}$$

for all $x \in \mathbb{R}$ if $n \geq 0$, and for all $x \neq 0$ if $n < 0$.

Proof. First suppose $n \in \mathbb{N}$. By the binomial theorem,

$$(x+h)^n = x^n + \binom{n}{1} x^{n-1} h + \binom{n}{2} x^{n-2} h^2 + \cdots + h^n.$$

Hence,

$$\frac{(x+h)^n - x^n}{h} = \binom{n}{1} x^{n-1} + \binom{n}{2} x^{n-2} h + \cdots + h^{n-1}.$$

Taking $h \to 0$ yields

$$\frac{d}{dx}(x^n) = \binom{n}{1} x^{n-1} = nx^{n-1}.$$

For $n = 0$, we have $x^0 = 1$ and the derivative is $0 = 0 \cdot x^{-1}$, which is consistent with (4.3).

For $n \in \mathbb{Z}_-$, write $n = -m$ with $m \in \mathbb{N}$. Then $x^n = \frac{1}{x^m}$. Using the quotient rule (Theorem 4.3) together with the case for $m > 0$, one finds

$$\frac{d}{dx}(x^n) = -\frac{mx^{m-1}}{x^{2m}} = -mx^{-m-1} = nx^{n-1},$$

for $x \neq 0$. $\square$

Corollary 4.1. Every polynomial

$$P(x) = a_n x^n + \cdots + a_1 x + a_0$$

is differentiable on $\mathbb{R}$, and

$$P'(x) = na_n x^{n-1} + \cdots + 2a_2 x + a_1.$$

Proof. A polynomial is a finite sum of constant multiples of integer power functions. By the power rule and the sum and constant multiple rules (Theorem 4.4 and Propositions 4.2 and 4.1 respectively), the derivative is as stated. $\square$

As a result, every rational function $R(x) = \frac{P(x)}{Q(x)}$, where P and Q are polynomials and $Q(x) \neq 0$, is differentiable on its domain, with derivative given by the quotient rule.

Next, we discuss derivatives of trigonometric functions. We recall the classical geometric inequalities for $0 < h < \frac{\pi}{2}$ as follows:

$$\sin h < h < \tan h = \frac{\sin h}{\cos h}.$$

Recall that this can be proven by considering Figure 3.2. From this, we get

$$\cos h < \frac{\sin h}{h} < 1 \quad \text{where } 0 < h < \frac{\pi}{2}.$$

A similar argument for negative h shows that for $-\frac{\pi}{2} < h < 0$, we have

$$\cos h < \frac{\sin h}{h} < 1.$$

Hence, for all h with $0 < |h| < \frac{\pi}{2}$, we have

$$\cos h < \frac{\sin h}{h} < 1.$$

Letting $h \to 0$, we get

$$\lim_{h\to 0} \cos h = 1 \quad \text{and} \quad \lim_{h\to 0} 1 = 1,$$

and by the squeeze theorem (Theorem 2.3), we have

$$\lim_{h\to 0} \frac{\sin h}{h} = 1. \tag{4.4}$$

To obtain $\frac{\cos h - 1}{h}$, for $0 < h < \frac{\pi}{2}$ we have

$$\cos h - 1 > \cos^2 h - 1 = -\sin^2 h,$$

and $\cos h - 1 < 0$, so

$$-\frac{\sin^2 h}{h} < \frac{\cos h - 1}{h} < 0.$$

As $h \to 0^+$, we have

$$\lim_{h\to 0^+} \left(-\frac{\sin^2 h}{h}\right) = \lim_{h\to 0^+} \left(\frac{\sin h}{h} \cdot (-\sin h)\right) = 1 \cdot 0 = 0.$$

Thus by the squeeze theorem (Theorem 2.3),

$$\lim_{h\to 0^+} \frac{\cos h - 1}{h} = 0.$$

A similar argument for $h \to 0^-$ shows

$$\lim_{h\to 0} \frac{\cos h - 1}{h} = 0. \tag{4.5}$$

Theorem 4.5. For all $x \in \mathbb{R}$, we have

$$\frac{d}{dx}(\sin x) = \cos x \quad \text{and} \quad \frac{d}{dx}(\cos x) = -\sin x. \tag{4.6}$$

Proof. For $\sin x$, we have

$$\begin{aligned}
\frac{d}{dx}(\sin x) &= \lim_{h\to 0} \frac{\sin(x+h) - \sin x}{h} \\
&= \lim_{h\to 0} \frac{\sin x \cos h + \cos x \sin h - \sin x}{h} \\
&= \lim_{h\to 0} \left(\sin x \cdot \frac{\cos h - 1}{h} + \cos x \cdot \frac{\sin h}{h}\right) \\
&= \sin x \cdot 0 + \cos x \cdot 1 \quad \text{by (4.4) and 4.5)}
\end{aligned}$$

which is equal to $\cos x$.

For $\cos x$,

$$\begin{aligned}\frac{d}{dx}(\cos x) &= \lim_{h\to 0}\frac{\cos(x+h)-\cos x}{h}\\ &= \lim_{h\to 0}\frac{\cos x\cos h-\sin x\sin h-\cos x}{h}\\ &= \lim_{h\to 0}\left(\cos x\cdot\frac{\cos h-1}{h}-\sin x\cdot\frac{\sin h}{h}\right)\\ &= \cos x\cdot 0-\sin x\cdot 1 \quad \text{by (4.4) and 4.5)}\end{aligned}$$

which is equal to $-\sin x$. □

Using the quotient rule (Theorem 4.3) and the derivatives in (4.6). For all x in the respective domains, one can deduce that

$$\begin{aligned}\frac{d}{dx}(\tan x) &= \sec^2 x\\ \frac{d}{dx}(\cot x) &= -\csc^2 x\\ \frac{d}{dx}(\sec x) &= \sec x\tan x\\ \frac{d}{dx}(\csc x) &= -\csc x\cot x\end{aligned}$$

For example, one can see that

$$\frac{d}{dx}(\tan x)=\frac{d}{dx}\left(\frac{\sin x}{\cos x}\right)=\frac{\cos x\cdot\cos x-\sin x\cdot(-\sin x)}{\cos^2 x}=\frac{\cos^2 x+\sin^2 x}{\cos^2 x}=\frac{1}{\cos^2 x}=\sec^2 x.$$

Theorem 4.6 (chain rule). Let f be differentiable at x and g be differentiable at $f(x)$. Then the composition $g\circ f$ is differentiable at x and

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

Equivalently, if $y=f(x)$ and $z=g(y)$, then

$$\frac{dz}{dx}=\frac{dz}{dy}\cdot\frac{dy}{dx}.$$

We now discuss higher derivatives. Let f be a function. The zeroth derivative of f is f itself and is denoted by $f^{(0)}$. For $n\in\mathbb{N}$, the n^{th} derivative of f is defined recursively by

$$f^{(n)}=\left(f^{(n-1)}\right)'.$$

We say f is n times differentiable if $f^{(n)}$ exists on the domain of interest. If $y=f(x)$, the n^{th} derivative is denoted by

$$f^{(n)}(x)=\frac{d^n y}{dx^n}.$$

Let f be a differentiable function. If f' is differentiable, then the second derivative of f is the derivative of f':

$$f'' = (f')'.$$

If $y = f(x)$, we also write

$$f''(x) = \frac{d^2y}{dx^2}.$$

Intuitively, $f'(x)$ measures the rate of change of $f(x)$ with respect to x, and $f''(x)$ measures the rate of change of $f'(x)$, giving information about the curvature of the graph.

Example 4.6 (position, velocity and acceleration). Let $s = s(t)$ be the position of a particle moving along a line. The velocity is $v(t) = s'(t)$ and the acceleration is $a(t) = v'(t) = s''(t)$.

For instance, if $s(t) = t^3 - 6t^2 + 9t$, then

$$v(t) = s'(t) = 3t^2 - 12t + 9 \quad \text{and} \quad a(t) = s''(t) = 6t - 12.$$

4.4 Implicit Differentiation

An equation of the form

$$F(x,y) = 0$$

is said to define y as an implicit function of x near a point (x_0, y_0) on the curve if, in some neighbourhood of x_0, there exists a differentiable function $y = y(x)$ such that $F(x, y(x)) = 0$ and $y(x_0) = y_0$.

Recall that the unit circle is given by the equation $x^2 + y^2 = 1$. For a point (x_0, y_0) on the circle with $y_0 > 0$, we can solve explicitly to obtain $y = \sqrt{1-x^2}$. So, near such a point y is an explicit function of x. For $y_0 < 0$, we can write $y = -\sqrt{1-x^2}$. However, near the points $(\pm 1, 0)$ the graph fails the vertical line test, so y is not a single-valued function of x there.

We now discuss the method of implicit differentiation. The idea is as follows: if $F(x,y) = 0$ and $y = y(x)$ is a differentiable implicit function, then differentiating both sides with respect to x (using the chain rule for y) gives an equation involving dy/dx, from which dy/dx can be solved. In using implicit differentiation, we assume differentiability of the implicit function; this method does not by itself *prove* differentiability.

Example 4.7 (unit circle). Consider $x^2 + y^2 = 1$. Differentiating both sides with respect to x gives

$$\frac{d}{dx}\left(x^2\right) + \frac{d}{dx}\left(y^2\right) = \frac{d}{dx}(1) \quad \Rightarrow \quad 2x + 2y\frac{dy}{dx} = 0.$$

Solving for $\frac{dy}{dx}$ gives $\frac{dy}{dx} = -\frac{x}{y}$. At a point $P = (x_0, y_0)$ with $y_0 \neq 0$, this gives the slope of the tangent line at P, which is $-\frac{x_0}{y_0}$.

Example 4.8 (power rule extends to rational exponents). Assume $x > 0$, and let $r = \frac{m}{n}$ with $m \in \mathbb{Z}$ and $n \in \mathbb{N}$, $\gcd(m,n) = 1$. Define

$$y = x^r = x^{m/n}.$$

Then, $x^m = y^n$. Differentiating both sides with respect to x yields

$$\frac{d}{dx}(x^m) = \frac{d}{dx}(y^n) \quad \Rightarrow \quad mx^{m-1} = ny^{n-1}\frac{dy}{dx}.$$

Thus,

$$\frac{dy}{dx} = \frac{m}{n}\frac{x^{m-1}}{y^{n-1}}.$$

Since $y^{n-1} = \left(x^{m/n}\right)^{n-1} = x^{m(n-1)/n}$, we get

$$\frac{dy}{dx} = \frac{m}{n}x^{m-1-\frac{m(n-1)}{n}} = \frac{m}{n}x^{\frac{m}{n}-1} = rx^{r-1}.$$

Thus, the power rule extends to rational exponents on $(0,\infty)$.

Example 4.9. Consider the curve

$$x^3 + y^3 = 3xy.$$

This is known as the *folium of Descartes* (Figure 4.1). It is a rational algebraic curve with a characteristic loop in the first quadrant and a node at the origin. In the special case shown here (with parameter $a = 1$), the folium has a notable point at $\left(\frac{3}{2}, \frac{3}{2}\right)$ and an oblique asymptote given by the line $x + y + 1 = 0$.

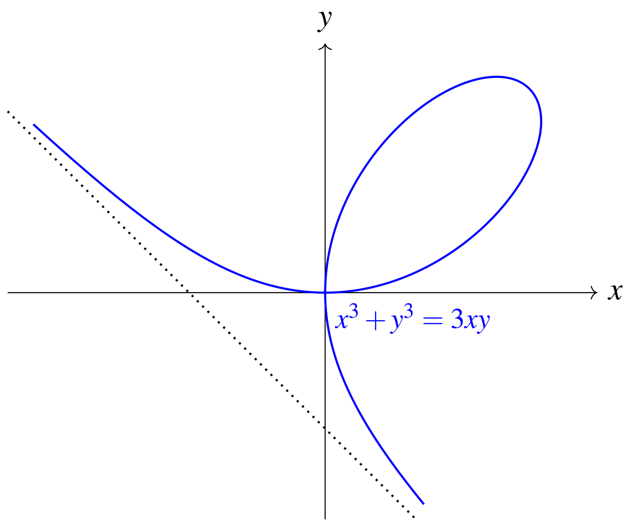


Figure 4.1: The folium of Descartes $x^3 + y^3 = 3xy$

Differentiating both sides with respect to x, we have

$$3x^2 + 3y^2\frac{dy}{dx} = 3y + 3x\frac{dy}{dx}.$$

Rearranging,

$$3y^2\frac{dy}{dx} - 3x\frac{dy}{dx} = 3y - 3x^2 \quad \text{so} \quad \left(3y^2 - 3x\right)\frac{dy}{dx} = 3\left(y - x^2\right).$$

Thus,

$$\frac{dy}{dx} = \frac{y - x^2}{y^2 - x}.$$

4.5 Extreme Values

Definition 4.4 (absolute maxima and minima). Let f be a real-valued function with domain $D \subseteq \mathbb{R}$.

(i) We say that f has an absolute maximum value at $c \in D$ if

$$f(c) \geq f(x) \quad \text{for every } x \in D.$$

Then $f(c)$ is called the absolute maximum (or global maximum) of f on D.

(ii) We say that f has an absolute minimum value at $c \in D$ if

$$f(c) \leq f(x) \quad \text{for every } x \in D.$$

Then $f(c)$ is called the absolute minimum (or global minimum) of f on D.

The absolute maximum and minimum are collectively called the extreme values of f on D.

Example 4.10. Let

$$f(x) = 3x^4 - 16x^3 + 18x^2 \quad \text{on the closed interval } [-1, 3.5].$$

One can check that $f(-1) = 37$ and $f(3) = -27$. On the domain $[-1, 3.5]$, the largest value is 37, attained at $x = -1$, and the smallest value is -27, attained at $x = 3$. Thus,

$$\max_{[-1,3.5]} f = 37 \quad \text{and} \quad \min_{[-1,3.5]} f = -27.$$

Definition 4.5 (local maxima and minima). Let f be a real-valued function with domain $D \subseteq \mathbb{R}$ and let $c \in D$.

(i) We say that f has a local maximum value at c if there exists an open interval I with $c \in I$ such that

$$f(c) \geq f(x) \quad \text{for all } x \in I \cap D.$$

(ii) We say that f has a local minimum value at c if there exists an open interval I with $c \in I$ such that

$$f(c) \leq f(x) \quad \text{for all } x \in I \cap D.$$

Local maxima and minima are also called relative extrema, and together they are called local extreme values.

Note that local extreme values are always attained at interior points of the domain: if f has a local extremum at c, then c is not an endpoint of D (unless D has isolated points).

Theorem 4.7 (extreme value theorem). Let f be continuous on a finite closed interval $[a,b]$. Then, there exist points $c,d \in [a,b]$ such that

$$f(c) \leq f(x) \leq f(d) \quad \text{for every } x \in [a,b].$$

In particular, f attains both an absolute minimum and an absolute maximum on $[a,b]$.

The extreme value theorem (Theorem 4.7) guarantees the existence but not uniqueness of the extreme values. The following conditions are essential:

(i) If f is not continuous on $[a,b]$, it may fail to attain its extrema

(ii) If the interval is not closed or not bounded, f may fail to attain a maximum or minimum

Definition 4.6 (critical point). Let f be a function and c an interior point of its domain. We call c a critical point of f if either of the following holds:

(i) $f'(c)$ exists and $f'(c) = 0$ (a stationary point), or

(ii) $f'(c)$ does not exist

Theorem 4.8 (Fermat's theorem). Let f be a function and c an interior point of its domain. If f has a local maximum or local minimum at c and f is differentiable at c, then

$$f'(c) = 0.$$

Proof. Assume f has a local minimum at c and is differentiable at c. Then there exists $\delta > 0$ such that

$$f(c) \leq f(x) \quad \text{for all } x \in (c-\delta, c+\delta).$$

For $x > c$ with x close to c we have

$$\frac{f(x) - f(c)}{x - c} \geq 0$$

because the numerator is non-negative and the denominator is positive. Hence

$$\lim_{x \to c^+} \frac{f(x) - f(c)}{x - c} \geq 0. \tag{4.7}$$

Similarly, for $x < c$ close to c we have $f(x) \geq f(c)$ and $x - c < 0$, so

$$\frac{f(x) - f(c)}{x - c} \leq 0,$$

and therefore,

$$\lim_{x \to c^-} \frac{f(x) - f(c)}{x - c} \leq 0. \tag{4.8}$$

Since f is differentiable at c, these one-sided limits both exist and are equal to $f'(c)$, so

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

Combining the inequalities (4.7) and (4.8) respectively, we obtain $f'(c) \geq 0$ and $f'(c) \leq 0$, hence $f'(c) = 0$. The case of a local maximum is similar. □

Note that if f has a local extreme value at c, then c must be a critical point. The converse is false: critical points need not correspond to local extrema.

Combining the extreme value theorem (Theorem 4.7 with Fermat's theorem (Theorem 4.8) gives a practical way to find global extrema on a closed interval. We call this the closed interval method (Theorem 4.9).

Theorem 4.9 (closed interval method). Let f be continuous on $[a, b]$. Then, the absolute maximum and minimum values of f on $[a, b]$ are attained either at the endpoints a or b, or at critical points $c \in (a, b)$ of f.

To find the extreme values of f on $[a, b]$, we proceed as follows:

(i) Compute $f(a)$ and $f(b)$

(ii) Find all critical points $c \in (a, b)$, and compute $f(c)$ at each such point

(iii) The largest of these values is the absolute maximum; the smallest is the absolute minimum

Example 4.11. Let

$$f(x) = x^3 - 3x^2 + 1 \quad \text{where } x \in \left[-\frac{1}{2}, 4\right].$$

Then,

$$f'(x) = 3x^2 - 6x = 3x(x - 2),$$

so the critical points are $x = 0$ and $x = 2$, both interior to the interval. We have $f\left(-\frac{1}{2}\right) = \frac{1}{8}$, $f(0) = 1$, $f(2) = -3$, and $f(4) = 17$. Hence, the absolute minimum is -3 at $x = 2$, and the absolute maximum is 17 at $x = 4$.

Theorem 4.10 (Rolle's theorem). Let f be a function such that the following hold:

(i) f is continuous on $[a, b]$

(ii) f is differentiable on (a, b) and

(iii) $f(a) = f(b)$

Then, there exists at least one point $c \in (a,b)$ such that $f'(c) = 0$.

Try to see how Rolle's theorem applies to the graph in Figure 4.2.

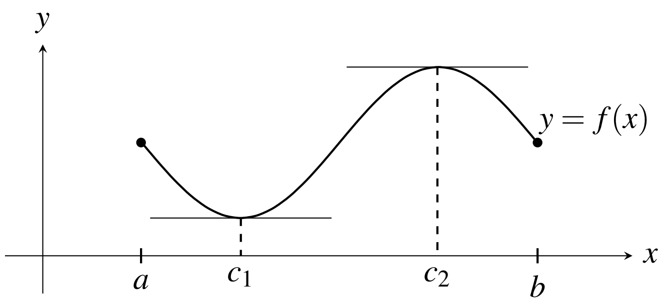


Figure 4.2: Example of a graph that satisfies Rolle's theorem

Proof. If f is constant on $[a,b]$, then $f'(x) = 0$ for all $x \in (a,b)$ and the result is trivial.

Otherwise, f is not constant. By the extreme value theorem (Theorem 4.7), f attains an absolute maximum and an absolute minimum on $[a,b]$. Since $f(a) = f(b)$ and f is not constant, at least one of these extreme values must occur at some interior point $c \in (a,b)$. At such a point c, f has a local extremum, so by Fermat's theorem (Theorem 4.8), $f'(c) = 0$. □

Theorem 4.11 (mean value theorem). Let f be a continuous function on $[a,b]$ and differentiable on (a,b). Then, there exists $c \in (a,b)$ such that

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

See Figure 4.3 for a geometric interpretation of the mean value theorem.

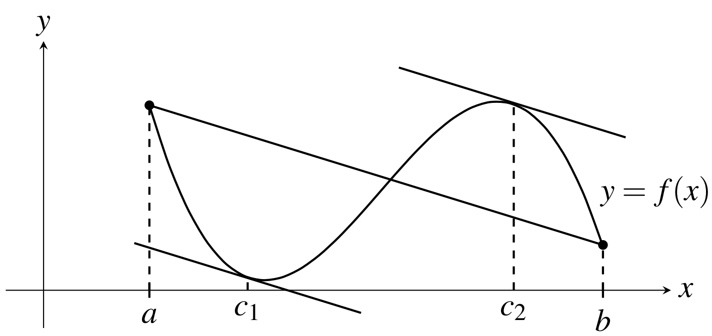


Figure 4.3: Geometric interpretation of the mean value theorem

Proof. Consider the auxiliary function

$$h(x) = f(x) - \left[\frac{f(b) - f(a)}{b - a}(x - a) + f(a)\right].$$

This is the difference between f and the straight line joining the points $(a, f(a))$ and $(b, f(b))$. The function h is continuous on $[a,b]$ and differentiable on (a,b). Furthermore, $h(a) = 0$ and $h(b) = 0$. By Rolle's theorem (Theorem 4.10), there exists $c \in (a,b)$ such that $h'(c) = 0$. However,

$$h'(x) = f'(x) - \frac{f(b) - f(a)}{b - a} \quad \text{so} \quad 0 = h'(c) = f'(c) - \frac{f(b) - f(a)}{b - a}$$

which is the desired identity. □

We then discuss some consequences of the mean value theorem.

Theorem 4.12. Let f be continuous on an interval I and differentiable on the interior of I.

(i) If $f'(x) = 0$ for every interior point x of I, then there exists a constant C such that $f(x) = C$ for all $x \in I$

(ii) If $f'(x) = g'(x)$ for every interior point x of I, where f and g are continuous on I and differentiable on its interior, then there exists a constant C such that

$$f(x) = g(x) + C \quad \text{for all } x \in I$$

Proof. We first prove **(i)**. Take any $a, b \in I$ with $a < b$. The restriction of f to $[a,b]$ is continuous and differentiable on (a,b). By the mean value theorem (Theorem 4.11), there exists $c \in (a,b)$ with

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

But $f'(c) = 0$ by assumption, hence $f(b) = f(a)$. Therefore f is constant on I.

For **(ii)**, the trick is to define $h(x) = f(x) - g(x)$. Then, h is continuous on I, differentiable on the interior of I, and $h'(x) = 0$ there. By **(i)** there is a constant C with $h(x) = C$ on I, i.e. $f(x) = g(x) + C$. □

Example 4.12. Using Theorem 4.12, one can prove the Pythagorean identity

$$\sin^2 x + \cos^2 x = 1 \quad \text{for all } x \in \mathbb{R}.$$

Set $f(x) = \sin^2 x + \cos^2 x$. Then, one can deduce that $f'(x) = 0$ so f is constant. Evaluating at $x = 0$ gives $f(0) = 1$, hence $f(x) = 1$.

4.6 Monotonicity and Concavity

Definition 4.7 (increasing and decreasing functions). Let $I \subseteq \mathbb{R}$ be an interval.

(i) A function f is increasing on I if for all $a, b \in I$ with $a < b$ we have

$$f(a) < f(b).$$

(ii) A function f is decreasing on I if for all $a, b \in I$ with $a < b$ we have

$$f(a) > f(b).$$

Theorem 4.13 (increasing/decreasing test). Let f be continuous on an interval I and differentiable on the interior of I.

(i) If $f'(x) > 0$ for every interior point x of I, then f is increasing on I

(ii) If $f'(x) < 0$ for every interior point x of I, then f is decreasing on I

Proof. We will only prove **(i)** as **(ii)** is analogous. Take $a, b \in I$ with $a < b$. By the mean value theorem (Theorem 4.11), there exists $c \in (a, b)$ with

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

Since $f'(c) > 0$ and $b - a > 0$, it follows that $f(b) - f(a) > 0$, i.e. $f(a) < f(b)$. □

Theorem 4.14 (partial converse). Suppose that f is differentiable on an open interval I.

(i) If f is increasing on I, then $f'(x) \geq 0$ for all $x \in I$

(ii) If f is decreasing on I, then $f'(x) \leq 0$ for all $x \in I$

Example 4.13 (MA2002 AY21/22 Sem 1). Let $f(x) = x(x^2 - 1)^{2/3}$. Find the open intervals on which f is increasing and decreasing.

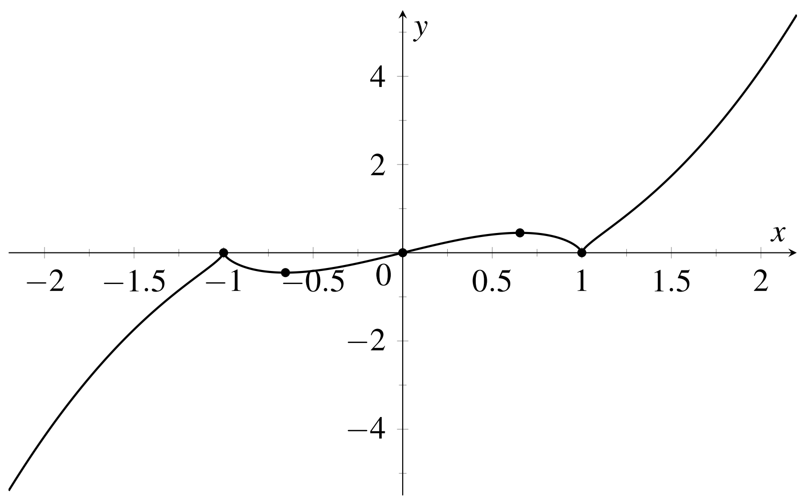


Figure 4.4: Graph of $y = x\left(x^2 - 1\right)^{2/3}$

Solution. We have

$$\begin{aligned} f'(x) &= x \cdot \frac{2}{3}(x^2-1)^{-1/3}(2x) + (x^2-1)^{2/3} \\ &= \frac{4}{3}x^2(x^2-1)^{-1/3} + (x^2-1)^{2/3} \\ &= (x^2-1)^{-1/3}\left(\frac{7}{3}x^2 - 1\right) \end{aligned}$$

Note that $x = -1$ and $x = 1$ are asymptotes of $y = f'(x)$. Setting $f'(x) = 0$, we obtain $x = \pm\sqrt{3/7}$. As such, the interval on which f is increasing is $(-\infty, -1) \cup \left(-\sqrt{3/7}, \sqrt{3/7}\right) \cup (1, \infty)$, and the interval on which f is decreasing is $\left(-1, -\sqrt{3/7}\right) \cup \left(1, \sqrt{3/7}\right)$. ☐

We then introduce the first derivative test for local extrema (Theorem 4.15).

Theorem 4.15 (first derivative test). Let f be continuous at a critical point c and differentiable on an open interval containing c, except possibly at c itself.

- **(i)** If $f'(x)$ changes sign from positive to negative as x increases through c, then f has a local maximum at c
- **(ii)** If $f'(x)$ changes sign from negative to positive as x increases through c, then f has a local minimum at c
- **(iii)** If $f'(x)$ does not change sign at c (e.g. positive on both sides, or negative on both sides), then f has no local extremum at c

We will only give a proof of **(ii)** as the other cases are analogous.

Proof. Assume that $f'(x) < 0$ for x in some interval (a, c) and $f'(x) > 0$ for x in some interval (c, b), with $a < c < b$. Then, f is decreasing on $(a, c]$ and increasing on $[c, b)$ by Theorem 4.13. Hence, for every $x \in (a, b)$, we have $f(c) \leq f(x)$, so f has a local minimum at c. ☐

Recall from our discussion in Chapter 4.3 that if f' exists on an interval and is differentiable, then the second derivative of f is defined to be $f''(x) = (f')'(x)$. Before we introduce the second derivative test, we need a lemma.

Lemma 4.1. If $\lim_{x \to a} g(x)$ exists and is positive, then there exists $\delta > 0$ such that $g(x) > 0$ for all x with $0 < |x - a| < \delta$.

Proof. Let $L = \lim_{x \to a} g(x) > 0$. Take $\varepsilon = L > 0$. By Definition 2.4, there exists $\delta > 0$ such that

$$0 < |x - a| < \delta \quad \Rightarrow \quad |g(x) - L| < \varepsilon = L.$$

Then, $g(x) > L - \varepsilon = 0$ for all such x. ☐

Theorem 4.16 (second derivative test). Let f be twice differentiable at c with $f'(c) = 0$.

(i) If $f''(c) > 0$, then f has a local minimum at c

(ii) If $f''(c) < 0$, then f has a local maximum at c

Proof. We will only prove **(i)**. Assume $f'(c) = 0$ and $f''(c) > 0$. By first principles (Definition 4.1), we have

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

Define

$$g(x) = \frac{f'(x)}{x - c} \text{ for } x \neq c \quad \text{so} \quad \lim_{x \to c} g(x) = f''(c) > 0.$$

By Lemma 4.1, there exists $\delta > 0$ such that $g(x) > 0$ whenever $0 < |x - c| < \delta$. Thus,

$$x < c \quad \Rightarrow \quad x - c < 0 \text{ and } g(x) > 0 \quad \Rightarrow \quad f'(x) < 0$$

and

$$x > c \quad \Rightarrow \quad x - c > 0 \text{ and } g(x) > 0 \quad \Rightarrow \quad f'(x) > 0.$$

So f' is negative to the left of c and positive to the right of c, and the first derivative test implies that f has a local minimum at c. The case $f''(c) < 0$ is analogous. □

Note that if $f'(c) = 0$ and $f''(c) = 0$, the second derivative test is inconclusive: f may have a local maximum, a local minimum, or neither. For example, say $f(x) = x^3$. Then, $f'(0) = f''(0) = 0$ but f has no local extremum. On the other hand, if $g(x) = x^4$, then $f'(0) = f''(0) = 0$ and it has a local minimum at $x = 0$.

Example 4.14 (MA2002 AY21/22 Sem 1). Continuing from Example 4.16, let $f(x) = x(x^2 - 1)^{2/3}$. Find the x-coordinates of the local maximum and minimum points of f.

Solution. We first set $f'(x) = 0$. Then, we see that $x = \pm\sqrt{3/7}$, for which the local maximum occurs at $x = \sqrt{3/7}$ and the local minimum occurs at $x = -\sqrt{3/7}$.

Earlier, we showed that

$$f'(x) = (x^2 - 1)^{-1/3} \left(\frac{7}{3} x^2 - 1 \right)$$

so $f'(x) \to -\infty$ as $x \to -1$ or $x \to 1$. Consider an ε-neighbourhood at the point $x = -1$, where $\varepsilon > 0$ is arbitrarily small. Note that $f(-1) = 0$, but $f(-1 - \varepsilon) < 0$ and $f(-1 + \varepsilon) < 0$, which asserts that a local maximum occurs at $x = -1$. In a similar fashion, $f(1) = 0$, but $f(1 - \varepsilon) > 0$ and $f(1 + \varepsilon) > 0$, which asserts that a local minimum occurs at $x = 1$.

To summarise, the local minima are at $x = -\sqrt{3/7}$ and $x = 1$, and the local maxima are at $x = \sqrt{3/7}$ and $x = -1$. □

Definition 4.8 (concavity). Let f be differentiable on an open interval I.

(i) We say that f is concave up on I if for any $a,b \in I$ with $a \neq b$,

$$f(b) - f(a) > f'(a)(b-a). \tag{4.9}$$

Geometrically, the graph of f lies above all its tangent lines on I.

(ii) We say that f is concave down on I if for any $a,b \in I$ with $a \neq b$,

$$f(b) - f(a) < f'(a)(b-a).$$

Geometrically, the graph of f lies below all its tangent lines on I.

Theorem 4.17. Let f be differentiable on an open interval I.

(i) If f is concave up on I, then f' is increasing on I

(ii) If f is concave down on I, then f' is decreasing on I

Proof. We will only prove **(i)**. Assume f is concave up. Take $a < b$ in I. From the concavity inequalities (4.9) applied to pairs (a,b) and (b,a), and using the mean value theorem (Theorem 4.11), one can show that

$$f'(a) < \frac{f(b) - f(a)}{b-a} < f'(b),$$

so $f'(a) < f'(b)$ for all $a < b$, which means f' is increasing. The concave down case is similar. $\square$

Theorem 4.18 (concavity test). Let f be twice differentiable on an open interval I.

(i) If $f''(x) > 0$ for all $x \in I$, then f is concave up on I

(ii) If $f''(x) < 0$ for all $x \in I$, then f is concave down on I

We only prove **(i)** as **(ii)** is analogous.

Proof. Note that $f'' = (f')'$. If $f''(x) > 0$ on I, then f' is increasing on I, so f is concave up. $\square$

Example 4.15 (MA2002 AY21/22 Sem 1). Continuing from Example 4.16, let $f(x) = x(x^2-1)^{2/3}$. Find the open intervals on which f is concave up and concave down.

Solution. As

$$f''(x) = \frac{4}{9}x\left(7x^2-9\right)\left(x^2-1\right)^{-4/3},$$

f is concave up when $f'' > 0$. That is, $\left(-3/\sqrt{7}, -1\right) \cup (-1,0) \cup \left(3/\sqrt{7}, \infty\right)$.

On the other hand, f is concave down when $f'' < 0$. That is, $\left(-\infty, -3/\sqrt{7}\right) \cup (0,1) \cup (1, 3/\sqrt{7})$. $\square$

Definition 4.9 (inflection point). Let f be continuous at c. We say that f has an inflection point at c if the concavity of f changes at c; that is, f is concave up on one side of c and concave down on the other side.

Theorem 4.19. If f has an inflection point at c and f is twice differentiable at c, then $f''(c) = 0$.

Proof. Without loss of generality, suppose f changes from concave down to concave up at c. Then, f' is decreasing on (a,c) and increasing on (c,b) for some $a < c < b$. Hence f' has a local minimum at c as shown in Figure 4.5.

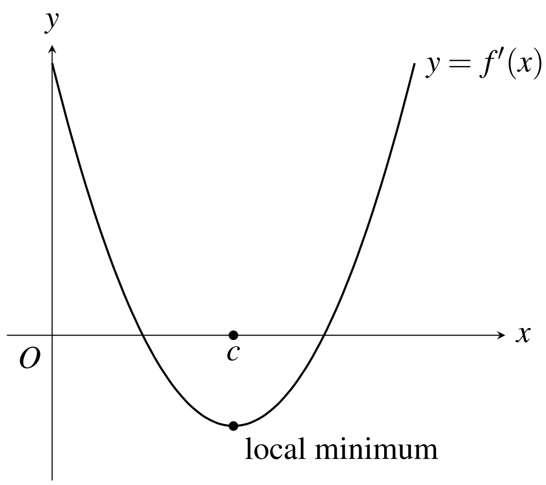


Figure 4.5: The derivative f' is decreasing on (a,c), increasing on (c,b), so f' has a local minimum at c

Since f' is differentiable at c, Fermat's theorem (Theorem 4.8) applied to f' yields $(f')'(c) = 0$, i.e. $f''(c) = 0$. □

Example 4.16 (MA2002 AY21/22 Sem 1). Continuing from Example 4.16, let $f(x) = x(x^2-1)^{2/3}$. Find the x-coordinates of the inflection points of f.

Solution. $x = 0$, $x = -3/\sqrt{7}$ and $x = 3/\sqrt{7}$ since the concavity of f change here. □

To sketch the graph of a reasonably nice function f, one systematically uses the information given by f', f''. The general strategy is as follows:

(i) Determine the domain of f, and identify intercepts and any asymptotes

(ii) Compute $f'(x)$ and find all critical points in the domain

(iii) Use the sign of f' to determine intervals where f is increasing or decreasing, and to locate local maxima and minima (first derivative test)

(iv) Compute $f''(x)$, study its sign to determine intervals where f is concave up or concave down, and find possible inflection points

(v) Combine all this information to draw a qualitative graph

4.7 Optimisation Problems

Many applied problems ask us to maximise or minimise a quantity (area, volume, cost, distance, time, etc.) subject to some constraints. The general strategy is as follows:

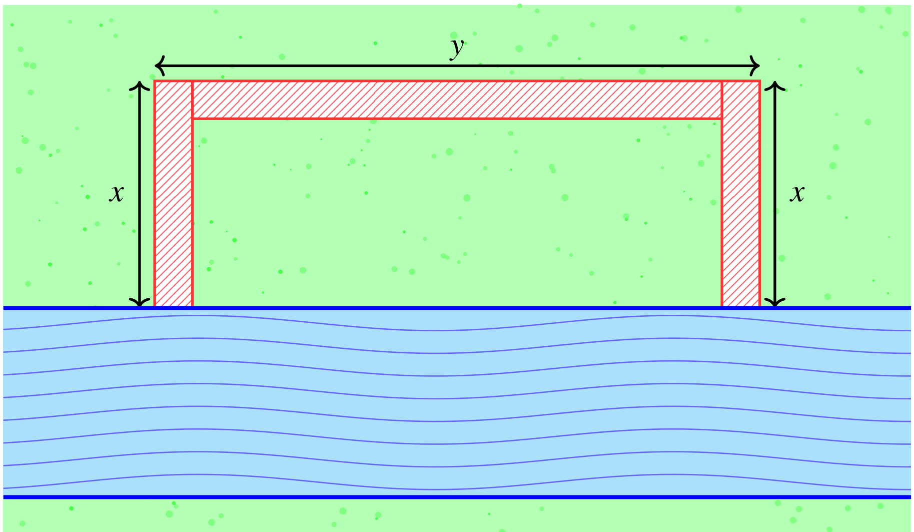
(i) Read the problem carefully and understand what is being optimized

(ii) Introduce variables and express the quantity to be optimised as a function $F(x)$ of a single variable by eliminating other variables using the constraints

(iii) Determine the domain of $F(x)$ from the context (often an interval of possible values)

(iv) Find critical points of F and, together with endpoints, determine which give the desired absolute maximum or minimum (using the closed interval method or the increasing/decreasing test)

Example 4.17. Suppose a farmer has 3000 metres of fencing and wishes to fence off a rectangular field that borders a river. Suppose he does not need fencing along the riverbank. What are the dimensions of the field that has the largest area?



Solution. Let x denote the length of each side perpendicular to the river, and let y denote the length of the side parallel to the river. Since no fencing is needed along the riverbank, the total length of fencing used is $2x + y = 3000$. The area A of the rectangular field is $A = xy$. Using the constraint, we express y in terms of x so

$$y = 3000 - 2x.$$

Thus, the area as a function of x is

$$A(x) = x(3000 - 2x) = 3000x - 2x^2.$$

To maximise the area, we differentiate $A(x)$ with respect to x to obtain

$$A'(x) = \frac{d}{dx}\left(3000x - 2x^2\right) = 3000 - 4x.$$

We find the critical point by setting $A'(x) = 0$ so $x = 750$. Next, we check that this critical point gives a maximum. The second derivative is

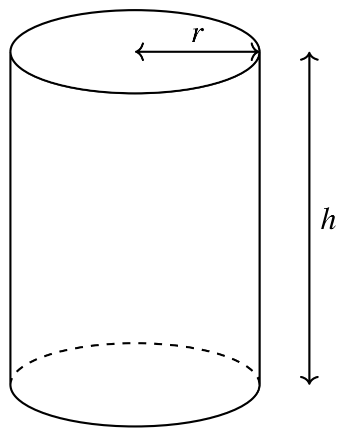
$$A''(x) = \frac{d}{dx}(3000 - 4x) = -4 < 0,$$

so $A(x)$ is concave down and $x = 750$ gives a maximum area. Substituting $x = 750$ into the constraint equation $2x + y = 3000$, we find y, which is $y = 1500$. Therefore, the rectangular field with the largest area has dimensions

$$x = 750 \text{ metres (perpendicular to the river)} \quad \text{and} \quad y = 1500 \text{ metres (along the river).}$$

□

Example 4.18. A cylindrical can must hold volume $V = 1\ \text{m}^3$. Find the dimensions that minimize the surface area (and hence the cost of metal).



Solution. Let r be the radius and h the height. Then,

$$V = \pi r^2 h = 1 \quad \text{so} \quad h = \frac{1}{\pi r^2},$$

and the total surface area is

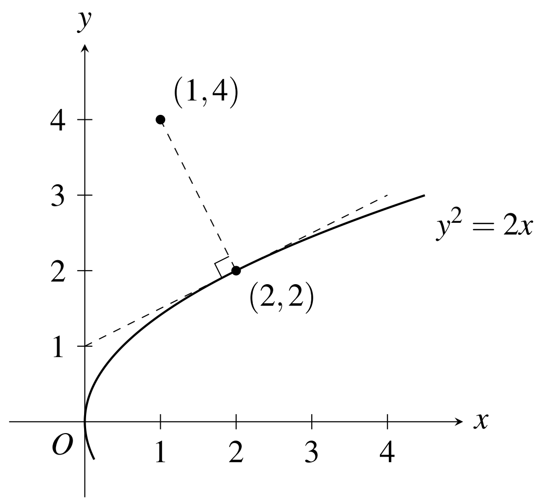
$$A(r) = 2\pi r^2 + 2\pi r h = 2\pi r^2 + 2\pi r \cdot \frac{1}{\pi r^2} = 2\pi r^2 + \frac{2}{r}, \quad \text{where } r > 0.$$

As such,

$$A'(r) = 4\pi r - \frac{2}{r^2} = \frac{4\pi r^3 - 2}{r^2}.$$

Setting $A'(r) = 0$ gives $4\pi r^3 - 2 = 0$, so $r = \left(\frac{1}{2\pi}\right)^{1/3}$. One can check that A' is negative for small r and positive for large r, hence this gives the unique minimum. Using $h = \frac{1}{\pi r^2}$, we obtain the corresponding optimal height. □

Example 4.19. Find the point on the parabola $y^2 = 2x$ that is closest to the point $(1,4)$.



Solution. A point on the parabola has coordinates (x,y) with $y^2 = 2x$. The squared distance to $(1,4)$ is

$$D^2 = (x-1)^2 + (y-4)^2 .$$

Using $x = \frac{y^2}{2}$, write

$$D^2(y) = \left(\frac{y^2}{2} - 1\right)^2 + (y-4)^2 .$$

It is enough to minimise $D^2(y)$ for $y \in \mathbb{R}$. After expanding and simplifying, one obtains

$$D^2(y) = \frac{1}{4}y^4 - 8y + 17 \quad \text{so} \quad \left(D^2\right)'(y) = y^3 - 8$$

so the critical points satisfy $y^3 = 8$, i.e. $y = 2$. Checking the sign of the derivative shows that this gives the minimum. The corresponding x is $x = \frac{y^2}{2} = \frac{4}{2} = 2$. Hence, the point $(2,2)$ on the parabola is closest to $(1,4)$. $\square$

Example 4.20 (Snell's law). Consider two homogeneous optical media separated by a planar interface. Let A be a point in medium 1 where the speed of light is v_1 and B be a point in medium 2 where the speed of light is v_2. Suppose a ray of light travels from A to B as shown in Figure 4.6.

Let θ_1 denote the angle of incidence (angle between the incident ray and the normal to the interface) and θ_2 denote the angle of refraction (angle between the refracted ray and the normal to the interface). This gives us the standard setup where Snell's law states that

$$\frac{\sin\theta}{v_1} = \frac{\sin\theta}{v_2},$$

or equivalently, $n_1 \sin\theta_1 = n_2 \sin\theta_2$, where $n_i = c/v_i$ is the refractive index in medium i. Here, c denotes the speed of light in vacuum, which is approximately 3×10^8 m/s.

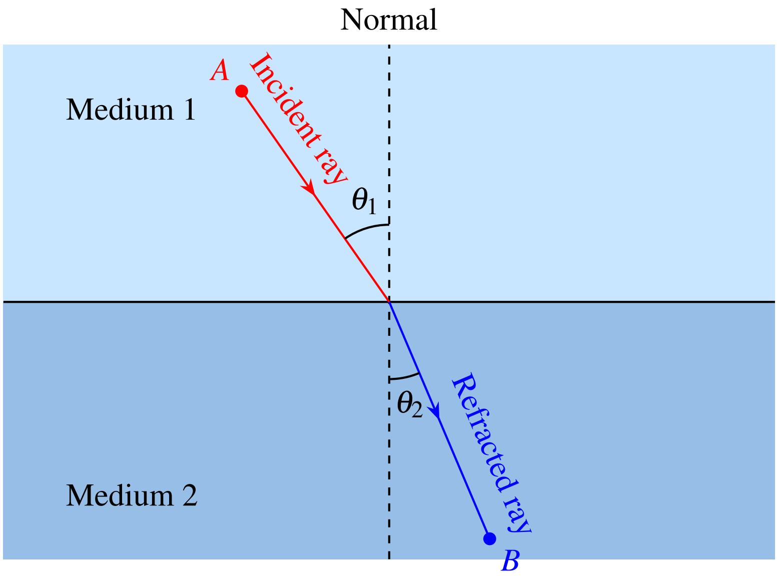


Figure 4.6: Ray of light travelling from A to B

From the point of view of optimisation, Snell's law can be derived from Fermat's principle of least time: among all possible paths that a ray of light could take from A to B, the actual path is the one that minimises the travel time. To make this precise, place a coordinate system so that the interface is the x–axis and medium 1 lies above the interface while medium 2 lies below it (Figure 4.6). Let

$$A = (-a, h_1) \in \text{medium 1} \quad \text{and} \quad B = (b, -h_2) \in \text{medium 2},$$

where $a, b, h_1, h_2 > 0$ are fixed. Any broken path from A to B that consists of a straight segment in medium 1 meeting a straight segment in medium 2 at some point P on the interface can be described by $P = (x, 0)$. We treat x as the variable to be chosen so that the travel time is minimised.

The distance from A to P is

$$AP = \sqrt{(x+a)^2 + h_1^2},$$

and the distance from P to B is

$$PB = \sqrt{(b-x)^2 + h_2^2}.$$

Since the speed of light in medium i is v_i, the time taken to traverse each segment is distance divided by speed. Thus, the total travel time as a function of x is

$$T(x) = \frac{AP}{v_1} + \frac{PB}{v_2} = \frac{1}{v_1}\sqrt{(x+a)^2 + h_1^2} + \frac{1}{v_2}\sqrt{(b-x)^2 + h_2^2}. \tag{4.10}$$

We now have an optimisation problem, which is to find $x \in \mathbb{R}$ that minimises $T(x)$. Assuming T is differentiable and the minimiser occurs at an interior point, a necessary condition is $T'(x) = 0$. Differentiating both sides of (4.10), we obtain

$$T'(x) = \frac{x+a}{v_1\sqrt{(x+a)^2 + h_1^2}} - \frac{b-x}{v_2\sqrt{(b-x)^2 + h_2^2}}.$$

Setting $T'(x) = 0$ gives

$$\frac{x+a}{v_1\sqrt{(x+a)^2+h_1^2}} = \frac{b-x}{v_2\sqrt{(b-x)^2+h_2^2}}.$$

Next, we interpret the fractions in terms of the angles of incidence and refraction. As mentioned, θ_1 is the angle between the incident ray AP and the normal, and θ_2 is the angle between the refracted ray PB and the normal. Because the normal is vertical, the horizontal component of AP is $|x+a|$ and the length of AP is $\sqrt{(x+a)^2+h_1^2}$, so

$$\sin\theta_1 = \frac{\text{horizontal component of } AP}{|AP|} = \frac{|x+a|}{\sqrt{(x+a)^2+h_1^2}}.$$

Similarly,

$$\sin\theta_2 = \frac{|b-x|}{\sqrt{(b-x)^2+h_2^2}}.$$

Ignoring the absolute values (which only encode the orientation along the interface) and comparing with the equation obtained from $T'(x) = 0$, we see that

$$\frac{\sin\theta_1}{v_1} = \frac{\sin\theta_2}{v_2}.$$

Equivalently, in terms of refractive indices $n_i = c/v_i$, this is $n_1\sin\theta_1 = n_2\sin\theta_2$ which is precisely Snell's law!

Example 4.21 (MA2002 AY21/22 Sem 1). A triangle is bounded by the tangent line to $y = e^x$ (where $x < 1$) and the axes. Find the coordinates of the tangent point so that the triangle attains its largest area. Justify your answer.

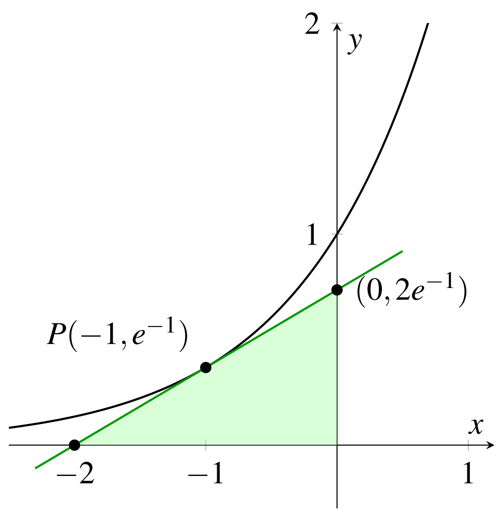


Figure 4.7: The graph of $y = e^x$ and a tangent line

Solution. We refer to Figure 4.7. Consider a point $P(a,e^a)$. The equation of the tangent to P is

$$y - e^a = e^a(x-a) \quad \text{so} \quad y = e^a x - ae^a + e^a.$$

To find the y-intercept, set $x=0$, so $y=e^a(1-a)$. To find the x-intercept, set $y=0$, so $x=a-1$.

Thus, the tangent intersects the y-axis and x-axis at $(0,e^a(1-a))$ and $(a-1,0)$ respectively. As $a-1<0$, the area of the triangle formed is

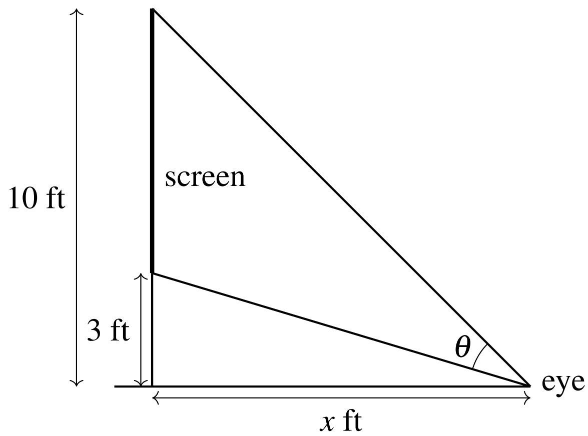
$$\frac{1}{2}\cdot e^a(1-a)\cdot(1-a)=-\frac{e^a}{2}\cdot(1-a)^2.$$

Now,

$$f'(a)=-\frac{e^a}{2}\cdot(a^2-1),$$

so $f'(a)=0$ implies $a=-1$ since $a<1$. One can use the first derivative test (the second derivative test fails here) to verify that at point P with coordinates $(-1,e^{-1})$, the area of the triangle formed is a maximum. $\square$

Example 4.22 (MA2002 AY23/24 Sem 2). An auditorium with a flat floor has a large screen on one wall. The lower edge of the screen is 3ft above eye level and the upper edge of the screen is 10ft above the eye level. How far from the screen should you stand to maximise your viewing angle θ? Give your answer in exact value.



Solution. By applying Pythagoras' theorem twice, the lengths of the bottom slant line and the top slant line are $\sqrt{x^2+9}$ and $\sqrt{x^2+100}$ respectively.

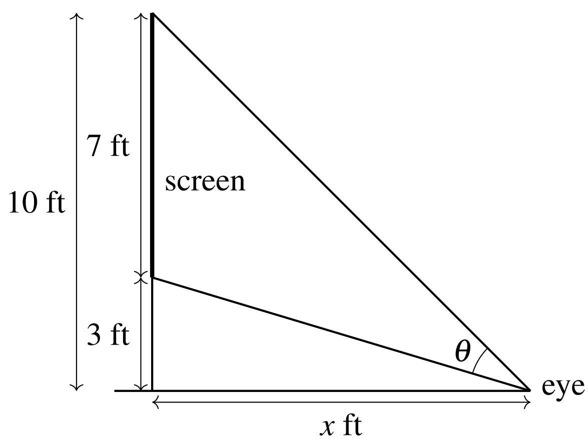


Figure 4.8: Annotated figure

By applying the cosine rule on the top triangle in Figure 4.8, we have

$$7^2 = \left(\sqrt{x^2+9}\right)^2 + \left(\sqrt{x^2+100}\right)^2 - 2\sqrt{x^2+9}\sqrt{x^2+100}\cos\theta.$$

So,

$$\cos\theta = \frac{30+x^2}{\sqrt{(x^2+9)(x^2+100)}}.$$

By some tedious implicit differentiation, we have

$$-\sin\theta\frac{d\theta}{dx} = \frac{49x\left(x^2-30\right)}{\left[(x^2+9)(x^2+100)\right]^{3/2}}.$$

Setting $\frac{d\theta}{dx} = 0$, we have $x = 0$ or $x = \pm\sqrt{30}$. Since $x > 0$, we only accept $x = \sqrt{30}$. Subsequently, one can use the first derivative test to deduce that when $x = \sqrt{30}$, θ is at a maximum. □

4.8 L'Hôpital's Rule

When evaluating limits of the form

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

it may happen that both numerator and denominator tend to zero or both tend to $\pm\infty$. In these situations, the quotient is said to be in an indeterminate form of type

$$\frac{0}{0} \quad \text{or} \quad \frac{\infty}{\infty}.$$

In such cases one can sometimes simplify the limit by applying H'Hôpital's rule.

Theorem 4.20 (L'Hôpital's rule: 0/0 case). Let f and g be differentiable on an open interval I containing a (except possibly at a) and suppose that

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

and that $g'(x) \neq 0$ for all $x \in I\setminus\{a\}$. Assume that the limit

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

exists (finite or infinite). Then,

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

Theorem 4.21 (L'Hôpital's rule: ∞/∞ case). Let f and g be differentiable on an open interval I containing a (possibly $a = \pm\infty$). Suppose that

$$\lim_{x\to a} |f(x)| = \infty \quad \text{and} \quad \lim_{x\to a} |g(x)| = \infty,$$

and that $g'(x) \neq 0$ for all x sufficiently close to a. If the limit

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

exists (finite or infinite), then

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

Note that L'Hôpital's rule may be applied repeatedly if the resulting limit is still of indeterminate type. The hypotheses (differentiability, type of indeterminate form, existence of the derivative quotient limit) in Theorems 4.20 and 4.21 are essential; the rule must not be used blindly.

Also, based on my prior experience with this course, evaluating limits using series expansion, i.e. $\sin x \approx x - \frac{x^3}{3!}$ are not permitted.

Example 4.23. Evaluate

$$\lim_{x\to 0} \frac{\sin x}{x}.$$

Solution. As $x \to 0$ both numerator and denominator tend to 0, so we have a $0/0$ indeterminate form. Applying L'Hôpital's rule (Theorem 4.20), we have

$$\lim_{x\to 0} \frac{\sin x}{x} = \lim_{x\to 0} \frac{\cos x}{1} = \cos 0 = 1.$$

Recall that this limit can also be proved using geometry and the squeeze theorem (Theorem 2.3) as per our discussion in (4.4). □

Example 4.24. Evaluate

$$\lim_{x\to\infty} \frac{\ln x}{x}.$$

Solution. Both $\ln x$ and x tend to ∞ as $x \to \infty$, so we have an ∞/∞ form. Applying L'Hôpital's rule (Theorem 4.21), we have

$$\lim_{x\to\infty} \frac{\ln x}{x} = \lim_{x\to\infty} \frac{1/x}{1} = \lim_{x\to\infty} \frac{1}{x} = 0.$$

□

Example 4.25. Evaluate

$$\lim_{x\to 0^+} x \ln x.$$

Solution. This is not in quotient form, but the trick is to rewrite the expression as

$$x \ln x = \frac{\ln x}{1/x}.$$

As $x \to 0^+$, $\ln x \to -\infty$ and $1/x \to \infty$, so we have an ∞/∞ form. Applying l'Hôpital's rule (Theorem 4.21), we have

$$\lim_{x \to 0^+} x \ln x = \lim_{x \to 0^+} \frac{1/x}{-1/x^2} = \lim_{x \to 0^+} (-x) = 0.$$

$\square$

Example 4.26 (MA2002 AY21/22 Sem 1). Let f be an increasing continuous function on $\mathbb{R}$ such that

$$\lim_{x \to \infty} \frac{1}{x} \int_0^x f(t)\, dt = 1.$$

Prove that $\lim_{x \to \infty} f(x)$ exists and equals 1.[1]

Solution. We proceed by contradiction. Suppose on the contrary that the limit of $f(x)$ as x tends to infinity does not exist. Since f is an increasing function, by Definition 4.7, as x increases, f will always increase and never level off from some value of x onwards. In other words,

$$\lim_{x \to \infty} f(x) = \infty \quad \text{so} \quad \int_0^x f(t)\, dt = \infty$$

However, by L'Hôpital's rule (Theorem 4.21), we have

$$\lim_{x \to \infty} \frac{1}{x} \int_0^x f(t)\, dt = \lim_{x \to \infty} f(x) = \infty$$

which is a contradiction. Thus, the limit $\lim_{x \to \infty} f(x)$ exists.

Now, we prove that the desired limit is equal to 1. Again, we proceed with contradiction. First, suppose on the contrary that $\lim_{x \to \infty} f(x) < 1$. This means that for all real values of x, we have $f(x) < 1$ because f is an increasing function. However, this implies

$$\lim_{x \to \infty} \frac{1}{x} \int_0^x f(t)\, dt < \lim_{x \to \infty} \frac{1}{x} \int_0^x dt = \lim_{x \to \infty} \frac{x}{x} = 1$$

which is a contradiction.

Next, suppose on the contrary that $\lim_{x \to \infty} f(x) > 1$. Then, in a similar notion to Definition 2.7 (but not exactly), for all $x > M$, there exists $M \geq 0$ such such that $f(x) > 1$, where we again used the fact that f is increasing. The trick now is to define a constant

$$K = \int_0^M f(t)\, dt.$$

Thus,

$$\lim_{x \to \infty} \int_0^x f(t)\, dt = \lim_{x \to \infty} \left(K + \int_M^x f(t)\, dt \right) = \infty$$

because for $x > M$, we have

$$\int_M^x f(t)\, dt > \int_M^x dt = x - M \quad \text{so} \quad \lim_{x \to \infty} \int_M^x f(t)\, dt > \lim_{x \to \infty} (x - M) = \infty.$$

By L'Hôpital's rule (Theorem 4.21), we have

$$\lim_{x \to \infty} \frac{1}{x} \int_0^x f(t)\, dt = \lim_{x \to \infty} f(x) > 1$$

which is a contradiction. We conclude that the limit exists, and it must be equal to 1. $\square$

[1] Interested readers should look up Cesàro mean.

4.9 Injective Functions and Inverses

Definition 4.10 (injective function). Let f be a function with domain $D \subseteq \mathbb{R}$. We say that f is one-to-one (or injective) on D if for all $a, b \in D$,

$$a \neq b \Rightarrow f(a) \neq f(b) \quad \text{or equivalently} \quad f(a) = f(b) \Rightarrow a = b.$$

The two formulations in Definition 4.10 are logically equivalent — they are contrapositives of each other.

Example 4.27. We give some simple examples of injective functions.

(i) Let $f(x) = x^3$ with domain $\mathbb{R}$. If $f(a) = f(b)$, then $a^3 = b^3$, so $a = b$.[2] Hence, f is one-to-one on $\mathbb{R}$.

(ii) Let $g(x) = \frac{1}{x}$ with domain $\mathbb{R} \setminus \{0\}$. If $g(a) = g(b)$, then $\frac{1}{a} = \frac{1}{b}$ so $a = b$. As such, g is one-to-one on $\mathbb{R} \setminus \{0\}$.

Definition 4.11 (inverse function). Let $f : A \to B$ be a one-to-one function with domain A and range B. For each $y \in B$ there exists a unique $x \in A$ such that

$$y = f(x).$$

The inverse function of f is the function $f^{-1} : B \to A$ defined by

$$f^{-1}(y) = x \Leftrightarrow y = f(x) \quad \text{for all } x \in A, y \in B.$$

Thus B is the domain of f^{-1} and A is the range of f^{-1}.

We now describe a method to find the inverse of a function algebraically. Suppose f is one-to-one. Then, write $y = f(x)$. Solve the equation $y = f(x)$ for x in terms of y: $x = f^{-1}(y)$. We then interchange the roles of x and y to present the inverse function as $y = f^{-1}(x)$.

On the Cartesian plane, interchanging x and y is equivalent to reflecting across the line $y = x$. Thus the graph of f^{-1} is the reflection of the graph of f across $y = x$.

Definition 4.12 (monotonic function). Let I be an interval. For all $a, b \in I$, a function $f : I \to \mathbb{R}$ is said to be

(i) increasing on I if $a < b \Rightarrow f(a) < f(b)$

(ii) decreasing on I if $a < b \Rightarrow f(a) > f(b)$

Either case is called monotonic.

[2] Use the difference of cubes identity $a^3 - b^3 = (a-b)(a^2 + ab + b^2)$ and argue why we should not take into consideration $a^2 + ab + b^2 = 0$.

Proposition 4.3. If f is increasing on I, then f is one-to-one on I. The same holds if f is decreasing.

Proof. Suppose f is increasing. Let $a,b \in I$ and assume $f(a) = f(b)$. If $a < b$, then because f is increasing, then $f(a) < f(b)$ which is a contradiction. If $a > b$, then $f(a) > f(b)$, which is again a contradiction. Hence, neither $a < b$ nor $a > b$ is possible, so $a = b$. Thus, f is one-to-one. The decreasing case is similar. $\square$

Theorem 4.22. Let f be continuous and one-to-one on an interval I. Then f is either strictly increasing on I or strictly decreasing on I.

See the footnote[3] for a rough sketch of the proof.

Theorem 4.23. Suppose f is one-to-one and continuous on an interval I. Then the inverse f^{-1} is continuous on its domain.

Proof. Assume f is increasing (the decreasing case is analogous). Then, f^{-1} is increasing on its domain. Let b be a point in the domain of f^{-1}, and set $a = f^{-1}(b) \in I$.

To show right-continuity at b, let $\varepsilon > 0$ and choose $\varepsilon_1 \in (0, \varepsilon]$ such that $[a, a+\varepsilon_1] \subseteq I$. Since f is increasing and continuous, we can set

$$\delta = f(a+\varepsilon_1) - b > 0.$$

Then for $0 < y - b < \delta$, we have

$$b < y < f(a+\varepsilon_1) \Rightarrow a < f^{-1}(y) < a+\varepsilon_1 \leq a+\varepsilon.$$

Thus, $\left|f^{-1}(y) - a\right| < \varepsilon$ for y sufficiently close to b from the right. A similar argument gives left-continuity, and the endpoint behaviour is handled one-sidedly. Hence, f^{-1} is continuous. $\square$

Theorem 4.24 (derivative of inverse function). Let f be one-to-one and continuous on an interval I, and let a be an interior point of I such that f is differentiable at a with $f'(a) \neq 0$. Let $b = f(a)$. Then f^{-1} is differentiable at b and

$$\left(f^{-1}\right)'(b) = \frac{1}{f'(a)}.$$

---

[3] Assume f is not monotone. Then there exist $\alpha, \beta \in I$ with $\alpha < \beta$ and $f(\alpha) > f(\beta)$, and also some $a, b \in I$ with $a < b$ and $f(a) < f(b)$. One constructs a continuous function

$$g(x) = f(a + x(\alpha - a)) - f(b + x(\beta - b)) \quad \text{where } 0 \leq x \leq 1,$$

and uses the intermediate value theorem (Theorem 3.16) to show that f must take the same value at two different points, contradicting injectivity.

Equivalently, if $y = f(x)$, then

$$\frac{dy}{dx} = f'(x) \neq 0 \quad \Rightarrow \quad \frac{dx}{dy} = \frac{1}{f'(x)}.$$

Proof. Let $y = f(x)$, so $x = f^{-1}(y)$. Then by Definition 4.1, we have

$$\left(f^{-1}\right)'(b) = \lim_{y\to b} \frac{f^{-1}(y) - f^{-1}(b)}{y-b} = \lim_{y\to b} \frac{x-a}{f(x)-f(a)}.$$

By continuity of f^{-1} at b, we have $y \to b$ implies $x \to a$, and by injectivity, $y \neq b$ implies $x \neq a$. Thus

$$\left(f^{-1}\right)'(b) = \lim_{x\to a} \frac{x-a}{f(x)-f(a)} = \frac{1}{\lim\limits_{x\to a} \dfrac{f(x)-f(a)}{x-a}} = \frac{1}{f'(a)}.$$

The result follows. □

4.10 Inverse Trigonometric Functions

The function $\sin x$ is not one-to-one on $\mathbb{R}$, but it is strictly increasing on $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$ with range $[-1,1]$.

Definition 4.13 (arcsine). For $x \in [-1,1]$, the unique $y \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$ such that $\sin y = x$ is denoted by $y = \arcsin x$ or $y = \sin^{-1} x$. Thus,

$$\arcsin : [-1,1] \to \left[-\frac{\pi}{2}, \frac{\pi}{2}\right] \quad \text{is the inverse of } \sin x \text{ restricted to } \left[-\frac{\pi}{2}, \frac{\pi}{2}\right].$$

Note that the function arcsin is continuous on $[-1,1]$ and differentiable on $(-1,1)$ with

$$\frac{d}{dx}\arcsin x = \frac{1}{\sqrt{1-x^2}} \quad \text{where } -1 < x < 1.$$

The function $\cos x$ is not one-to-one on $\mathbb{R}$, but it is strictly decreasing on $[0,\pi]$ with range $[-1,1]$.

Definition 4.14 (arccosine). For $x \in [-1,1]$, the unique $y \in [0,\pi]$ such that $\cos y = x$ is denoted by $y = \arccos x$ or $y = \cos^{-1} x$. Then,

$$\arccos : [-1,1] \to [0,\pi] \quad \text{is the inverse of } \cos x \text{ restricted to } [0,\pi].$$

Again, note that the function arccos is continuous on $[-1,1]$ and differentiable on $(-1,1)$ with

$$\frac{d}{dx}\arccos x = -\frac{1}{\sqrt{1-x^2}} \quad \text{where } -1 < x < 1.$$

Moreover,

$$\arcsin x + \arccos x = \frac{\pi}{2} \quad \text{where } -1 \leq x \leq 1.$$

This identity follows from the fact that on the chosen principal branches $\sin\left(\frac{\pi}{2}-\theta\right) = \cos\theta$ and uniqueness of inverses.

The tangent function $\tan x$ is strictly increasing and continuous on $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$ with range $\mathbb{R}$.

Definition 4.15 (arctangent). For $x \in \mathbb{R}$, the unique $y \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$ such that $\tan y = x$ is denoted by $y = \arctan x$ or $y = \tan^{-1} x$. Then,

$$\arctan : \mathbb{R} \to \left(-\frac{\pi}{2}, \frac{\pi}{2}\right) \quad \text{is the inverse tangent function.}$$

4.11 Hyperbolic Functions

Definition 4.16. The hyperbolic sine and hyperbolic cosine functions are defined for $x \in \mathbb{R}$ by

$$\sinh x = \frac{e^x - e^{-x}}{2} \quad \text{and} \quad \cosh x = \frac{e^x + e^{-x}}{2}.$$

Note that the hyperbolic functions satisfy the identity

$$\cosh^2 x - \sinh^2 x = 1 \quad \text{for all } x \in \mathbb{R}.$$

Thus, the parametric curve parametrised by $x = \cosh t$ and $y = \sinh t$ lies on the hyperbola $x^2 - y^2 = 1$ (the right branch, since $\cosh(t) > 0$). Also, the derivatives are

$$\frac{d}{dx}\sinh x = \cosh x \quad \text{and} \quad \frac{d}{dx}\cosh x = \sinh x.$$

Moreover, for all $x \in \mathbb{R}$, we have $\cosh x \geq 1$ and $\sinh x$ is strictly increasing with range $\mathbb{R}$.

CHAPTER 5

Integration and Applications

5.1 Integration as Area under a Curve

We begin with the intuitive problem of finding the area of a plane region. Suppose $f : [a,b] \to \mathbb{R}$ is continuous and non–negative, and consider the region

$$R = \left\{(x,y) \in \mathbb{R}^2 : a \le x \le b, 0 \le y \le f(x)\right\}.$$

We want to define the area $A(R)$ in a rigorous way. The basic idea is to approximate R by rectangles whose total area can be computed easily, and then pass to the limit as the mesh of the partition goes to 0. We say that a partition P of the interval $[a,b]$ is a finite increasing sequence

$$P = \{x_0, x_1, \dots, x_n\} \quad \text{such that } a = x_0 < x_1 < \cdots < x_n = b$$

The norm of the partition P is

$$\|P\| = \max_{1 \le i \le n} (x_i - x_{i-1}).$$

For each subinterval $[x_{i-1}, x_i]$, we choose a sample point $x_i^* \in [x_{i-1}, x_i]$. The corresponding Riemann sum of f with respect to P and the choice of sample points $\{x_i^*\}$ is

$$S(f; P, \{x_i^*\}) = \sum_{i=1}^{n} f(x_i^*)\,\Delta x_i \quad \text{where } \Delta x_i = x_i - x_{i-1}.$$

Geometrically, $f(x_i^*)\,\Delta x_i$ is the area of a rectangle with base $[x_{i-1}, x_i]$ and height $f(x_i^*)$ (see red rectangle in Figure 5.1), so the Riemann sum is the total area of these rectangles, approximating $A(R)$.

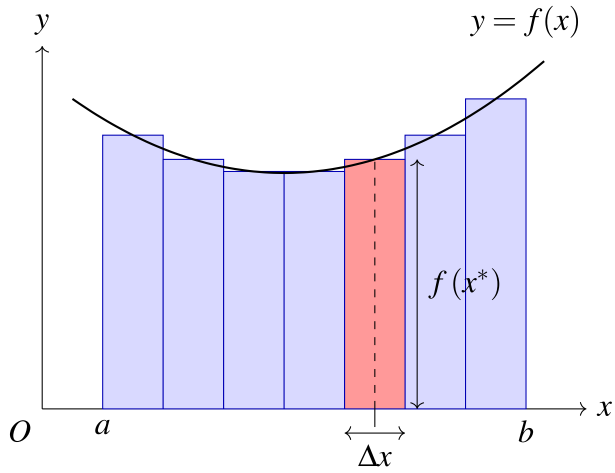


Figure 5.1: Integration as area under a curve

Definition 5.1 (Riemann integrability criterion). Let $f : [a,b] \to \mathbb{R}$ be a bounded function. We say that f is Riemann integrable on $[a,b]$ if there exists a real number I with the property that for every $\varepsilon > 0$ there exists $\delta > 0$ such that

$$\|P\| < \delta \quad \Rightarrow \quad |S(f;P,\{x_i^*\}) - I| < \varepsilon$$

for every partition P of $[a,b]$ and every choice of sample points $\{x_i^*\}$. In that case we write

$$I = \int_a^b f(x)\ dx \quad \text{and call this the definite integral of } f \text{ from } a \text{ to } b.$$

Thus, the integral $\int_a^b f(x)\ dx$ is the common limit of all Riemann sums for f over partitions with norms tending to 0.

Theorem 5.1. If $f : [a,b] \to \mathbb{R}$ is continuous, then f is Riemann integrable on $[a,b]$.

We omit the proof; it relies on the fact that a continuous function on a closed interval is *uniformly continuous* and bounded. This is definitely out of scope of the course — you would need knowledge from MA3210 Mathematical Analysis II. See Theorem 51.10 of [3] if you are interested in the proof. When we write

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

we say that f is the integrand, x is the dummy variable of integration, a and b are the lower and upper limits, and the symbol $\int$ is the integral sign. The value of the integral does not depend on the letter used for the dummy variable. That is to say,

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

Example 5.1. Let $c \in \mathbb{R}$ be constant and consider $f(x) = c$ on $[a,b]$. For any partition P and any sample points, we obtain

$$S(f;P,\{x_i^*\}) = \sum_{i=1}^{n} c\,\Delta x_i = c\sum_{i=1}^{n}\Delta x_i = c\,(b-a)\,.$$

Therefore,

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

This agrees with the elementary area formula for a rectangle of base $(b-a)$ and height c.

Let $f:[a,b]\to\mathbb{R}$ be continuous and non–negative. Then, f is Riemann integrable on $[a,b]$ and

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

is equal to the ordinary Euclidean area of the region between the graph of f, the x-axis and the vertical lines $x=a$ and $x=b$. This follows by comparing the area of the region with the areas of lower and upper step–function approximations built from Riemann sums and letting the norm go to 0.

Proposition 5.1. Let $f:[0,1]\to\mathbb{R}$ be Riemann integrable on $[0,1]$. Then,

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n} f\left(\frac{k}{n}\right) = \int_0^1 f(x)\;dx.$$

Proof. Since f is Riemann integrable on $[0,1]$, for any partition $P=\{x_0,x_1,\ldots,x_n\}$ of $[0,1]$, the Riemann integral is defined as:

$$\int_0^1 f(x)\,dx = \lim_{\|P\|\to 0}\sum_{i=1}^{n} f(c_i)\Delta x_i$$

where $c_i \in [x_{i-1},x_i]$ and $\|P\| = \max\limits_{1\le i\le n}(x_i - x_{i-1})$. The trick is to consider the uniform partition with n subintervals

$$x_k = \frac{k}{n} \quad \text{where } k = 0,1,2,\ldots,n.$$

For this partition, we have $\Delta x_k = x_k - x_{k-1} = \frac{k}{n} - \frac{k-1}{n} = \frac{1}{n}$ for all k, and $\|P\| = \frac{1}{n}\to 0$ as $n\to\infty$. Say we choose the right endpoint as our sample point, i.e. $c_k = x_k = \frac{k}{n}$. Then, the Riemann sum becomes:

$$\sum_{k=1}^{n} f(c_k)\Delta x_k = \sum_{k=1}^{n} f\left(\frac{k}{n}\right)\cdot\frac{1}{n} = \frac{1}{n}\sum_{k=1}^{n} f\left(\frac{k}{n}\right)$$

Since f is Riemann integrable and $\|P\| = \frac{1}{n}\to 0$ as $n\to\infty$, we have

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n} f\left(\frac{k}{n}\right) = \lim_{\|P\|\to 0}\sum_{k=1}^{n} f(c_k)\Delta x_k = \int_0^1 f(x)\,dx$$

and the result follows. ∎

Example 5.2. Use Riemann sum to find

$$\lim_{n\to\infty}\sum_{k=1}^{n}\ln\sqrt[n]{1+\frac{k}{n}}.$$

Solution. We have

$$\lim_{n\to\infty}\sum_{k=1}^{n}\ln\sqrt[n]{1+\frac{k}{n}}=\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}\ln\left(1+\frac{k}{n}\right)$$

Using Proposition 5.1, the limit becomes

$$\int_0^1 \ln(1+x)\ dx.$$

Using integration by parts (Theorem 5.18 though this will be formally discussed in Chapter 5.6), we have

$$\int_0^1 \ln(1+x)\ dx=[x\ln(1+x)]_0^1-\int_0^1\frac{x}{x+1}\ dx=\ln 2-\int_0^1 1-\frac{1}{x+1}\ dx$$

which evaluates to $2\ln 2-1$. □

Let f and g be Riemann integrable on $[a,b]$ and $c\in\mathbb{R}$ is fixed. First, note that if $a=b$ and f is defined at a, then

$$\int_a^a f(x)\ dx=0.$$

If $a<b$, then

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

These follow directly from the definition in terms of Riemann sums, by reversing the orientation of the interval.

Theorem 5.2 (additivity). Let $a,b,c\in\mathbb{R}$ with $a\le c\le b$, and assume f is integrable on $[a,b]$. Then,

$$\int_a^c f(x)\ dx+\int_c^b f(x)\ dx=\int_a^b f(x)\ dx.$$

Theorem 5.2 expresses the idea that the total signed area on $[a,b]$ is the sum of the signed areas on $[a,c]$ and $[c,b]$.

Theorem 5.3 (linearity). Let f,g be integrable on $[a,b]$ and $c\in\mathbb{R}$. Then,

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

$$\int_a^b cf(x)\ dx=c\int_a^b f(x)\ dx$$

Theorem 5.4 (order). If $f(x) \geq g(x)$ for all $x \in [a,b]$, then

$$\int_a^b f(x)\, dx \geq \int_a^b g(x)\, dx.$$

Corollary 5.1 (integral bounds). Let m and M be the minimum and maximum of the continuous function $f\colon [a,b] \to \mathbb{R}$. Then

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

Proof. We have $m \leq f(x) \leq M$ for all x, hence $m \leq f(x)$ and $f(x) \leq M$. Integrating and using the linearity and constant-function formula (Theorem 5.3) yields the desired inequalities. □

The Fundamental Theorem of Calculus (FTC) establishes the precise connection between derivatives and integrals. It has two parts (Theorems 5.5 and 5.7).

Theorem 5.5 (Fundamental Theorem of Calculus Part I). Let $f : [a,b] \to \mathbb{R}$ be continuous. Define

$$F(x) = \int_a^x f(t)\, dt \quad \text{where } x \in [a,b].$$

Then, F is differentiable on (a,b) and

$$F'(x) = f(x) \quad \text{for all } x \in (a,b).$$

Proof. Fix $x \in (a,b)$. Let $h \neq 0$ be arbitrarily small enough such that $x+h \in [a,b]$, we have

$$\frac{F(x+h)-F(x)}{h} = \frac{1}{h}\left(\int_a^{x+h} f(t)\, dt - \int_a^x f(t)\, dt\right) = \frac{1}{h}\int_x^{x+h} f(t)\, dt.$$

If $h > 0$, then this quantity equals

$$\frac{1}{h}\int_x^{x+h} f(t)\, dt,$$

while if $h < 0$ we can rewrite it as

$$\frac{1}{h}\int_{x+h}^{x} f(t)\, dt.$$

In both cases, it is the average value of f over an interval shrinking to the fractional part of x, denoted by $\{x\}$. Since f is continuous at x, by Proposition 3.1, for every $\varepsilon > 0$ there exists $\delta > 0$ such that $|t-x| < \delta$ implies $|f(t) - f(x)| < \varepsilon$. If $|h| < \delta$ we have, for $h > 0$,

$$\left|\frac{1}{h}\int_x^{x+h} f(t)\, dt - f(x)\right| = \left|\frac{1}{h}\int_x^{x+h} (f(t)-f(x))\, dt\right| \leq \frac{1}{h}\int_x^{x+h} |f(t)-f(x)|\, dt \leq \varepsilon.$$

A similar estimate holds for $h < 0$. Hence,

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

and F is differentiable at x with $F'(x) = f(x)$. □

Thus the operation 'integrate from a to x' is an antiderivative operation: it produces a function whose derivative is the original integrand. Speaking of antiderivatives, we shall define what they are here. Let $f : I \to \mathbb{R}$ be a function on an interval I. A function $F : I \to \mathbb{R}$ is called an antiderivative of f if

$$F'(x) = f(x) \quad \text{for all } x \in I$$

where F' is defined.

Theorem 5.6 (uniqueness up to a constant). If F_1 and F_2 are antiderivatives of the same function f on an interval I, then there exists a constant $C \in \mathbb{R}$ such that

$$F_2(x) = F_1(x) + C \quad \text{for all } x \in I.$$

Proof. Consider $G(x) = F_2(x) - F_1(x)$. Then

$$G'(x) = F_2'(x) - F_1'(x) = f(x) - f(x) = 0 \quad \text{on } I.$$

By **(i)** of Theorem 4.12, G is constant on I, so $F_2 = F_1 + C$ for some C. □

Theorem 5.7 (Fundamental Theorem of Calculus Part II). Let $f : [a,b] \to \mathbb{R}$ be continuous and let F be any antiderivative of f on $[a,b]$. Then,

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

Proof. Define

$$G(x) = \int_a^x f(t)\ dt.$$

By the first part of the Fundamental Theorem of Calculus (Theorem 5.5), G is an antiderivative of f on $[a,b]$. Thus, G and F are antiderivatives of the same function on the interval, so by Theorem 5.6, there exists C with

$$G(x) = F(x) + C.$$

Taking $x = a$ gives $G(a) = \int_a^a f(t)\ dt = 0$, so $0 = F(a) + C$ and $C = -F(a)$. Therefore, $G(x) = F(x) - F(a)$. Setting $x = b$ yields

$$\int_a^b f(t)\ dt = G(b) = F(b) - F(a)$$

as claimed. □

The Fundamental Theorems of Calculus (Theorems 5.5 and 5.7) suggest the following notation. Let f be a function on an interval I. An indefinite integral of f is the family of all antiderivatives of f. That is,

$$\int f(x)\ dx\ =\ F(x)+C,$$

where F is any fixed antiderivative of f on I and $C \in \mathbb{R}$ is an arbitrary constant. By Theorem 5.7, we can compute definite integrals by first finding an indefinite integral, which is

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

whenever $F'(x) = f(x)$. As expected, linearity carries over to indefinite integrals. That is,

$$\int (f(x)+g(x))\ dx = \int f(x)\ dx + \int g(x)\ dx \quad \text{and} \quad \int cf(x)\ dx = c\int f(x)\ dx$$

up to additive constants.

5.2 The Natural Logarithm

Definition 5.2. The natural logarithm function $\ln x$ is defined for $x > 0$ by the integral

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

Note that $\ln x$ is continuous and differentiable on $(0,\infty)$ with the following properties:

$$\ln 1 = 0 \quad \text{and} \quad \frac{d}{dx}\ln x = \frac{1}{x} \quad \text{and} \quad \frac{d^2}{dx^2}\ln x = -\frac{1}{x^2}. \tag{5.1}$$

Consequently, $\ln x$ is strictly increasing and concave down on $(0,\infty)$. Next, from the integral definition and properties of improper integrals, one can show

$$\lim_{x\to 0^+} \ln(x) = -\infty, \qquad \lim_{x\to\infty} \ln(x) = \infty,$$

so the range of $\ln(x)$ is all of $\mathbb{R}$.

Theorem 5.8 (logarithm of a product). For $a > 0$ and $x > 0$,

$$\ln(ax) = \ln a + \ln x.$$

Proof. Fix $a > 0$ and define $f(x) = \ln(ax) - \ln x$ for $x > 0$. Then,

$$f'(x) = \frac{d}{dx}\ln(ax) - \frac{d}{dx}\ln x = \frac{a}{ax} - \frac{1}{x} = 0.$$

Hence, f is constant on $(0,\infty)$. Evaluating at $x = 1$ gives

$$f(1) = \ln(a) - \ln(1) = \ln(a),$$

so $f(x) = \ln(a)$ for all $x > 0$, i.e. $\ln(ax) = \ln(a) + \ln(x)$. □

Theorem 5.9 (logarithm of a rational power). Let $x > 0$ and $r \in \mathbb{Q}$. Then,

$$\ln\left(x^r\right) = r \ln x.$$

Proof. Fix $r \in \mathbb{Q}$ and define $g(x) = \ln\left(x^r\right) - r \ln x$ for $x > 0$. Then,

$$g'(x) = \frac{d}{dx} \ln\left(x^r\right) - r \frac{d}{dx} \ln x = \frac{r x^{r-1}}{x^r} - \frac{r}{x} = 0.$$

Thus g is constant. Evaluating at $x = 1$ yields $g(1) = 0$, so $g(x) = 0$ and hence, we conclude that $\ln\left(x^r\right) = r \ln(x)$. □

For $x < 0$, we have $-x > 0$ and

$$\frac{d}{dx} \ln(-x) = -1 \cdot \frac{1}{-x} = \frac{1}{x}.$$

Also, for $x \neq 0$, we have

$$\frac{d}{dx} \ln|x| = \frac{1}{x} \quad \text{and} \quad \int \frac{1}{x}\, dx = \ln|x| + C.$$

Moreover, if a, b have the same sign, then

$$\int_a^b \frac{1}{x}\, dx = \ln|b| - \ln|a|.$$

Definition 5.3 (logarithmic differentiation). Let

$$y = \left[f_1(x)\right]^{r_1} \cdots \left[f_n(x)\right]^{r_n} \quad \text{where } r_1, \ldots, r_n \in \mathbb{Q}$$

and each f_k is a non-zero differentiable function. The method of logarithmic differentiation consists of the following steps:

(i) taking absolute values, so

$$|y| = \prod_{k=1}^{n} \left|f_k(x)\right|^{r_k}$$

(ii) taking natural logarithms, i.e.

$$\ln|y| = \sum_{k=1}^{n} r_k \ln\left|f_k(x)\right|$$

(iii) differentiating both sides with respect to x to obtain

$$\frac{1}{y}\frac{dy}{dx} = \sum_{k=1}^{n} r_k \frac{f_k'(x)}{f_k(x)}.$$

Example 5.3. We give an example of the method of logarithmic differentiation. Let

$$y = \frac{\left(x^2+1\right)\sqrt{x+3}}{x-1} \quad \text{where } x > 1.$$

Then,

$$\ln y = \ln\left(x^2+1\right) + \frac{1}{2}\ln(x+3) - \ln(x-1).$$

Differentiating both sides yields

$$\frac{1}{y}\frac{dy}{dx} = \frac{2x}{x^2+1} + \frac{1}{2}\cdot\frac{1}{x+3} - \frac{1}{x-1}$$

so

$$\frac{dy}{dx} = \left[\frac{2x}{x^2+1} + \frac{1}{2(x+3)} - \frac{1}{x-1}\right]\frac{\left(x^2+1\right)\sqrt{x+3}}{x-1}.$$

5.3 The Exponential Function

We know from (5.1) that $\ln x$ is strictly increasing and continuous on $(0,\infty)$ with range $\mathbb{R}$. Hence, it is bijective, and has an inverse $\exp : \mathbb{R} \to (0,\infty)$.

Definition 5.4. There is a unique real number $e > 0$ such that $\ln e = 1$. This number is called Euler's number. Numerically,

$$e \approx 2.718281828459045\cdots.$$

Definition 5.5. The exponential function (base e) is

$$\exp(x) = e^x,$$

defined as the inverse of $\ln(x)$ and it satisfies the following properties:

$$\ln(e^x) = x \text{ where } x \in \mathbb{R} \quad \text{and} \quad e^{\ln y} = y \text{ where } y > 0.$$

Theorem 5.10. The exponential function e^x is differentiable on $\mathbb{R}$ and

$$\frac{d}{dx}e^x = e^x.$$

Proof. Let $y = e^x$, so $x = \ln y$. Then, we have

$$\frac{dy}{dx} = \frac{1}{\frac{dx}{dy}} = \frac{1}{\frac{d}{dy}\ln(y)} = \frac{1}{1/y} = y = e^x.$$

$\square$

For a rational exponent $r = \frac{m}{n}$, we can define $e^r = \left(\sqrt[n]{e}\right)^m$, and for real x extend via continuity, i.e. $e^x = \exp(x)$.

Definition 5.6. Let $a > 0$. For $x \in \mathbb{R}$, define the exponential function of base a by

$$a^x = \exp(x\ln(a)) = e^{x\ln(a)}.$$

By Definition 5.6, we have

$$\ln(a^x) = x\ln(a) \quad \text{for all } a > 0 \text{ and } x \in \mathbb{R},$$

consistent with the rational-exponent case.

Theorem 5.11. Let $a > 0$ and $x, y \in \mathbb{R}$. Then, the following hold:

(i) $a^x a^y = a^{x+y}$

(ii) $a^{-x} = \frac{1}{a^x}$

(iii) $(a^x)^y = a^{xy}$

(iv) a^x is differentiable on $\mathbb{R}$ and

$$\frac{d}{dx}a^x = a^x \ln a$$

Proof. **(i)-(iii)** follow from the corresponding identities for e^x and properties of ln. For **(iv)**, write $a^x = e^{x\ln(a)}$ and apply the chain rule (Theorem 4.6) to obtain

$$\frac{d}{dx}a^x = \frac{d}{dx}e^{x\ln(a)} = \ln(a)\, e^{x\ln(a)} = a^x \ln(a).$$

□

Definition 5.7. Let $a \in \mathbb{R}$. For $x > 0$, define

$$x^a = e^{a\ln x}.$$

Theorem 5.12. For any $a \in \mathbb{R}$, the function $f(x) = x^a$ is differentiable on $(0,\infty)$ with

$$\frac{d}{dx}x^a = ax^{a-1} \quad \text{for all } x > 0.$$

Moreover,

$$\int x^a\, dx = \begin{cases} \ln|x| + C & \text{if } a = -1; \\ \frac{x^{a+1}}{a+1} + C & \text{if } a \neq -1. \end{cases}$$

Proof. Write $x^a = e^{a\ln(x)}$ for $x > 0$. Let $u = a\ln(x)$. Then

$$\frac{d}{dx}x^a = \frac{d}{dx}e^u = \frac{du}{dx}e^u = \frac{a}{x}e^{a\ln(x)} = \frac{a}{x}x^a = ax^{a-1}.$$

The antiderivatives follow by reversing the differentiation. □

Example 5.4. Find $\frac{d}{dx}(x^x)$ for $x > 0$.

Solution. Let $y = x^x$. Taking logarithms, $\ln y = \ln(x^x) = x\ln x$. Differentiating both sides yields

$$\frac{1}{y}\frac{dy}{dx} = \ln(x) + 1 \quad \text{so} \quad \frac{dy}{dx} = (\ln x + 1)x^x.$$

□

Theorem 5.13.

$$e = \lim_{x\to 0}(1+x)^{1/x}.$$

Proof. Write

$$(1+x)^{1/x} = \exp\left(\frac{1}{x}\ln(1+x)\right).$$

Using L'Hôpital's rule (Theorem 4.20) on $\frac{\ln(1+x)}{x}$ as $x\to 0$, we have

$$\lim_{x\to 0}\frac{\ln(1+x)}{x} = \lim_{x\to 0}\frac{\frac{1}{1+x}}{1} = 1.$$

Hence,

$$\lim_{x\to 0}(1+x)^{1/x} = 1 = e.$$

□

From Theorem 5.13, we see that a useful strategy for limits involving positive functions is as follows. We first express $f(x)^{g(x)}$ as $\exp(g(x)\ln(f(x)))$. We then evaluate the limit inside the exponential (often via l'Hôpital's rule) and then apply continuity of the exponential function.

Example 5.5. Evaluate

$$\lim_{x\to 0^+} x^x.$$

Solution. We have $x^x = \exp(x\ln(x))$. As $x\to 0^+$, $x\ln(x)\to 0$, which can be computed by writing

$$x\ln(x) = \frac{\ln(x)}{1/x}$$

and applying l'Hôpital's rule (Theorem 4.21). Thus, we see that the desired limit is equal to $\exp 0 = 1$. □

Example 5.6 (MA2002 AY21/22 Sem 1). Let n be a fixed positive integer. Find the following limit and simplify the answer.

$$\lim_{x\to 0}\left(\frac{e^x+e^{2x}+\ldots+e^{nx}}{n}\right)^{1/x}$$

Solution. The trick[1] is to first consider the natural logarithm of the function and note that $1 = \frac{n}{n}$. Then, we have

$$\ln\left(\frac{e^x+e^{2x}+\ldots+e^{nx}}{n}\right)^{1/x} = \frac{1}{x}\ln\left(1+\frac{e^x+e^{2x}+\ldots+e^{nx}-n}{n}\right)$$

[1] It might be difficult to think of the trick (other than the natural logarithm). Some may attempt this question using the squeeze theorem or using the geometric series at the start but would realise that their attempts are futile.

As $x \to 0$ on the right side, we observe that the expression is in indeterminate form. Using L'Hôpital's rule (Theorem 4.20), the limit as $x \to 0$ becomes

$$\lim_{x\to 0} \frac{e^x + 2e^{2x} + \ldots + ne^{nx}}{e^x + e^{2x} + \ldots + e^{nx}} = \lim_{x\to 0} \frac{1+2+\ldots+n}{n}.$$

The numerator on the right side is an arithmetic series, for which the sum is given by $\frac{1}{2}n(n+1)$. Reverse engineering, our original limit is $e^{\frac{1}{2}(n+1)}$. □

As mentioned, integration can be viewed geometrically as computing the area under a curve. Here, we compare the area under $y = \ln x$ with simple shapes whose areas are easy to compute, and this translates into sharp estimates for $\ln(n!)$ (since $\ln(n!)$ is closely tied to sums of $\ln k$). The outcome is a *Stirling-type bound* for the factorial, showing that $n!$ is well-approximated by $(n/e)^n \sqrt{n}$ up to an explicit constant factor (as in the inequalities proved in parts **(iii)** and **(iv)** of Example 5.7). Let us discuss Example 5.7 in detail.

Example 5.7 (MA2002 AY21/22 Sem 1).

(i) For $x \geq 1$, let $n = \lfloor x \rfloor$ and define

$$f(x) = (n+1-x)\ln n + (x-n)\ln(n+1).$$

Show that for all $x \geq 1$, $f(x) \leq \ln x$ and that for all $n \in \mathbb{Z}^+$,

$$\int_1^n f(x)\, dx = \ln(n!) - \frac{1}{2}\ln n. \tag{5.2}$$

(ii) For $x \geq 1$, let n be the unique integer such that $n - 1/2 < x < n + 1/2$ and define

$$g(x) = \frac{x}{n} - 1 + \ln n.$$

Show that for all $x \geq 1$, $g(x) \geq \ln x$ and that for all $n \in \mathbb{Z}^+$,

$$\int_1^n g(x)\, dx = \int_1^n f(x)\, dx + \frac{1}{8}\left(1 - \frac{1}{n}\right). \tag{5.3}$$

(iii) Use the results in **(i)** and **(ii)** to conclude that for all $n \in \mathbb{Z}^+$,

$$\frac{7}{8} \leq \ln(n!) - \left(n + \frac{1}{2}\right)\ln n + n \leq 1.$$

(iv) Use the results in **(iii)** to conclude that

$$e^{7/8} \leq \frac{n!}{(n/e)^n\sqrt{n}} \leq e.$$

Solution.

(i) We first prove that for all $x \geq 1$, we have $f(x) \leq \ln x$. Equivalently, we can define

$$g(x) = (n+1-x)\ln n + (x-n)\ln(n+1) - \ln x$$

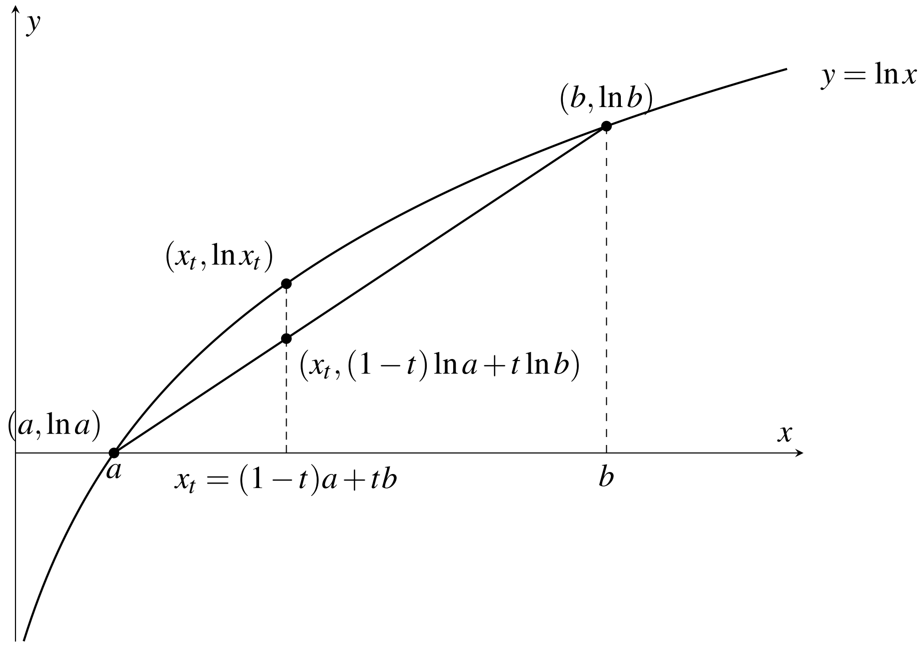
so it suffices to prove that for all $x \geq 1$, we have $g(x) \leq 0$. The trick is to first note that $x-n \in [0,1]$ so we can let $t = x-n$. As such,

$$g(x) = (1-t)\ln n + t\ln(n+1) - \ln x.$$

Since ln is concave down on $(0,\infty)$, for all $a,b>0$ and $t \in [0,1]$, we have

$$\ln[(1-t)a+tb] \geq (1-t)\ln a + t\ln b. \tag{5.4}$$

Here is a geometric interpretation.



Applying (5.4) with $a = n$ and $b = n+1$ yields

$$(1-t)\ln n + t\ln(n+1) \leq \ln x$$

so it follows that for all $x \geq 1$, we have $g(x) \leq 0$.

We then prove that (5.2) holds. Note that $n+1-x$ and $x-n$ are periodic functions with discontinuities at integer points. Next, $\ln n$ and $\ln(n+1)$ also have discontinuities at integer points. One can sketch the graph of $f(x)$ and see that for $x \in [1,2]$, we obtain a right-angled triangle, whereas for $x \in [2,3], x \in [3,4], \ldots$, we obtain trapeziums. As such,

$$\int_1^n f(x)\,dx = \int_1^2 f(x)\,dx + \int_2^3 f(x)\,dx + \ldots + \int_{n-1}^n f(x)\,dx$$

which is equal to

$$\frac{1}{2}\cdot 1\cdot \ln 2 + \frac{1}{2}\cdot 1\cdot(\ln 2+\ln 3) + \frac{1}{2}\cdot 1\cdot(\ln 3+\ln 4) + \ldots + \frac{1}{2}\cdot 1\cdot(\ln(n-1)+\ln n).$$

Upon simplification, we obtain

$$\ln 2 + \ln 3 + \ldots + \ln n - \frac{1}{2}\ln n = \ln(n!) - \frac{1}{2}\ln n.$$

(ii) Let

$$h(x) = \frac{x}{n} - 1 + \ln n - \ln x$$

so it suffices to prove that $h(x) \geq 0$ for all $x \geq 1$. We have

$$h(x) = \frac{x}{n} - 1 - \ln\left(\frac{x}{n}\right).$$

Let $u = \frac{x}{n}$ which is > 0. Then, $h(x) = \phi(u)$ where $\phi(u) = u - 1 - \ln u$. It suffices to show that $\phi(u) \geq 0$ for all $u > 0$. One sees that

$$\phi'(u) = 1 - \frac{1}{u} \quad \text{and} \quad \phi''(u) = \frac{1}{u^2} > 0$$

so ϕ is convex on $(0, \infty)$, and its only critical point is $u = 1$. Since $\phi'' > 0$, this critical point is a global minimum. We see that $\phi(1) = 0$ so $\phi(u) \geq \phi(1) = 0$ for all $u > 0$. As such, the first claim holds.

We then prove that (5.3) holds. Since n is the unique integer such that $n - \frac{1}{2} < x < n + \frac{1}{2}$, the trick is to consider $n < x + \frac{1}{2} < n + 1$ so we can define $n = \lfloor x + \frac{1}{2} \rfloor$. So,

$$g(x) = \frac{x}{\lfloor x + 1/2 \rfloor} - 1 + \ln\left\lfloor x + \frac{1}{2} \right\rfloor.$$

Again, we see that the region between $y = g(x)$ and the x-axis from $x = 1$ to $x = n$ comprises a triangle between $x = 1$ and $x = \frac{3}{2}$ and a trapezium between $x = \frac{3}{2}$ and $x = \frac{5}{2}$, as well as between $x = \frac{5}{2}$ and $x = \frac{7}{2}$, and so on.

The area of the triangle is

$$\frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}.$$

The area of the trapezium bounded by $y = g(x)$, the x-axis, and the ordinates $x = \frac{3}{2}$ and $x = \frac{5}{2}$ is

$$\frac{1}{2} \cdot 1 \cdot \left(\frac{3}{4} - 1 + \ln(2) + \frac{1}{4} + \ln 2\right) = \ln 2.$$

Subsequently, the area of the trapezium bounded by $y = g(x)$, the x-axis, and the ordinates $x = \frac{5}{2}$ and $x = \frac{7}{2}$ is $\ln 3$. However, the upper limit of the integral of g is n as mentioned in (5.3), which is $\in \mathbb{Z}$. Based on the pattern established earlier,

$$\int_{n-1/2}^{n+1/2} g(x)\, dx = \ln n,$$

but the integral from n to $n + 1/2$ should be omitted. This is the area of another trapezium, which is

$$\frac{1}{2} \underbrace{\left(n + \frac{1}{2} - n\right)}_{\text{width}} \underbrace{\left[g(n) + g\left(n + \frac{1}{2}\right)\right]}_{\text{sum of parallel sides}} = \frac{1}{4}\left[g(n) + g\left(n + \frac{1}{2}\right)\right]$$

which simplifies to

$$\frac{1}{4}\left(\frac{n}{n} - 1 + \ln n + \frac{n + 1/2}{n} - 1 + \ln n\right) = \frac{1}{4}\left(2\ln n + \frac{1}{2n}\right).$$

In a similar fashion as compared to **(i)**,

$$\begin{aligned}\int_1^n g(x)\ dx &= \frac{1}{8} + \ln 2 + \ln 3 + \ldots + \ln n - \frac{1}{4}\left(2\ln n + \frac{1}{2n}\right) \\ &= \frac{1}{8} + \ln 1 + \ln 2 + \ln 3 + \ldots + \ln n - \frac{1}{4}\left(2\ln n + \frac{1}{2n}\right) \\ &= \frac{1}{8} + \ln(n!) - \frac{1}{2}\ln n - \frac{1}{8n} \\ &= \ln(n!) - \frac{1}{2}\ln n + \frac{1}{8}\left(1 - \frac{1}{n}\right)\end{aligned}$$

so the result follows.

(iii) First, set

$$q = \ln(n!) - \left(n + \frac{1}{2}\right)\ln n + n.$$

Note that

$$q = \ln(n!) - \frac{1}{2}\ln n + n(1 - \ln n) = \int_1^n f(x)\ dx + n(1 - \ln n).$$

Since $f(x) \le \ln x$ from **(i)**, then

$$\int_1^n f(x)\ dx \le \int_1^n \ln x\ dx = n\ln n - n + 1.$$

It is thus clear that an upper bound for q is obtained, and that is 1.

Using **(ii)**, since $g(x) \ge \ln x$, then

$$\int_1^n g(x)\ dx \ge \int_1^n \ln x\ dx = n\ln n - n + 1.$$

Next, using (5.3), we have

$$\int_1^n f(x)\ dx \ge \frac{1}{8}\left(\frac{1}{n} - 1\right) + n\ln n - n + 1.$$

Hence,

$$q \ge n - n\ln n + \frac{1}{8n} - \frac{1}{8} + n\ln n - n + 1 = \frac{7}{8} + \frac{1}{8n} \ge \frac{7}{8},$$

so a lower bound for q is obtained, which is $7/8$.

(iv) Since q is bounded between $7/8$ and 1, we have

$$e^{7/8} \le \frac{e^n n!}{n^{(2n+1)/2}} \le e.$$

The term that is sandwiched here is indeed the one given at the start of the question, so we are done. □

5.4 Improper Integrals

The definite integral as defined so far only applies to bounded functions on finite closed intervals. Improper integrals extend the concept to allow unbounded intervals or integrands.

Definition 5.8 (infinite intervals). Let f be continuous on $[a,\infty)$. If the limit

$$\lim_{b\to\infty}\int_a^b f(x)\,dx$$

exists (as a finite real number), we define the improper integral

$$\int_a^\infty f(x)\,dx = \lim_{b\to\infty}\int_a^b f(x)\,dx$$

and say that it converges. Otherwise the improper integral is said to diverge.

Similarly, if f is continuous on $(-\infty,b]$, we define

$$\int_{-\infty}^b f(x)\,dx = \lim_{a\to-\infty}\int_a^b f(x)\,dx,$$

whenever the limit exists.

For an integral over $(-\infty,\infty)$ we set

$$\int_{-\infty}^\infty f(x)\,dx = \int_{-\infty}^c f(x)\,dx + \int_c^\infty f(x)\,dx$$

for some (and hence any) c, provided both integrals on the right converge.

Example 5.8 (The p–series test on $[1,\infty)$). In MA2108 Mathematical Analysis I, we will encounter the p-series test. It says that

$$\sum_{n=1}^\infty \frac{1}{n^p} \text{ converges} \quad \text{if and only if } p>1.$$

As such, the mentioned series diverges for $0<p\le 1$. Now, for $p>0$, consider

$$\int_1^\infty \frac{1}{x^p}\,dx = \lim_{b\to\infty}\int_1^b x^{-p}\,dx. \tag{5.5}$$

For $p\neq 1$, we obtain

$$\int_1^b x^{-p}\,dx = \frac{b^{1-p}-1}{1-p},$$

so

$$\int_1^\infty \frac{1}{x^p}\,dx = \begin{cases} \frac{1}{p-1} & \text{if } p>1; \\ \text{diverges} & \text{if } 0<p\le 1. \end{cases}$$

For $p=1$, we have

$$\int_1^b \frac{1}{x}\,dx = \ln b \to \infty,$$

so the improper integral diverges. Comparing with the p-series test, this is precisely the same threshold $p=1$ that we saw in the improper integral in (5.5). For readers interested in a deeper discussion (related to the integral test for convergence and to some, the Cauchy condensation test) which involves concepts from MA2108 Mathematical Analysis I, please check out Chapter 59 of [3].

Definition 5.9 (unbounded integrands). Let f be continuous on $(a,b]$ but unbounded near a. If the limit

$$\lim_{\varepsilon\to 0^+}\int_{a+\varepsilon}^{b} f(x)\,dx$$

exists (as a finite real number), we define

$$\int_a^b f(x)\,dx = \lim_{\varepsilon\to 0^+}\int_{a+\varepsilon}^{b} f(x)\,dx.$$

Similar definitions apply when f is unbounded near b, or when there is a singularity at a point c inside (a,b); in the latter case one splits the integral at c.

Example 5.9. Consider

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

The integrand is unbounded at $x=0$, but

$$\int_\varepsilon^1 x^{-1/2}\,dx = 2-2\sqrt{\varepsilon}$$

which tends to 2. Thus

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

In contrast,

$$\int_0^1 \frac{1}{x}\,dx = \lim_{\varepsilon\to 0^+}\int_\varepsilon^1 \frac{1}{x}\,dx = \lim_{\varepsilon\to 0^+}(\ln 1-\ln\varepsilon) = \infty,$$

so the latter improper integral diverges.

Theorem 5.14 (comparison test). Let f and g be continuous on $[a,\infty)$ with $0\le f(x)\le g(x)$ for all $x\ge a$.

(i) If $\int_a^\infty g(x)\,dx$ converges, then $\int_a^\infty f(x)\,dx$ converges

(ii) If $\int_a^\infty f(x)\,dx$ diverges, then $\int_a^\infty g(x)\,dx$ diverges

The proof of Theorem 5.14 uses the monotonicity of the integral and properties of limits of increasing sequences of partial integrals. We omit it here, but anyway some readers will encounter it again in MA2108 Mathematical Analysis I.

5.5 Integration by Substitution

Integration by substitution is the analogue of the chain rule in differentiation.

Theorem 5.15 (integration by substitution). Let g be differentiable on an interval

I and let f be continuous on an interval containing $g(I)$. Then

$$\int f(g(x))\, g'(x)\ dx = \int f(u)\ du \quad \text{where } u = g(x).$$

In practice, one sets $u = g(x)$, computes $du = g'(x)\, dx$ and rewrites the integral entirely in terms of u.

Proof. Consider the composite function

$$F(x) = \int f(g(x))\, g'(x)\ dx.$$

Let H be an antiderivative of f, so $H'(u) = f(u)$. By the chain rule (Theorem 4.6),

$$\frac{d}{dx} H(g(x)) = H'(g(x))\, g'(x) = f(g(x))\, g'(x).$$

Thus $H(g(x))$ is an antiderivative of the integrand, and

$$\int f(g(x))\, g'(x)\ dx = H(g(x)) + C.$$

On the other hand,

$$\int f(u)\ du = H(u) + C,$$

so substituting $u = g(x)$ gives the same expression. □

Theorem 5.16 (integration by substitution). Let $g \colon [\alpha, \beta] \to \mathbb{R}$ be differentiable with continuous derivative, and let f be continuous on an interval containing $g([\alpha, \beta])$. Then,

$$\int_{\alpha}^{\beta} f(g(x))\, g'(x)\ dx = \int_{g(\alpha)}^{g(\beta)} f(u)\ du.$$

Proof. Let

$$F(u) = \int f(u)\ du$$

be an antiderivative of f. Then, $F'(u) = f(u)$. By the chain rule (Theorem 4.6),

$$\frac{d}{dx} F(g(x)) = F'(g(x))\, g'(x) = f(g(x))\, g'(x).$$

Thus, by the Fundamental Theorem of Calculus (Theorem 5.7), we have

$$\int_{\alpha}^{\beta} f(g(x))\, g'(x)\ dx = F(g(\beta)) - F(g(\alpha)).$$

The right side equals

$$\int_{g(\alpha)}^{g(\beta)} f(u)\ du$$

again by the Fundamental Theorem of Calculus (Theorem 5.7), since F is an antiderivative of f. □

When performing integration by substitution, we replace an integral in the variable x by one in a new variable t, chosen so that the integrand simplifies. Sometimes, it is convenient to run this procedure backwards, starting from a substitution that converts a given integrand into another type of integrand (for example a rational function in t). We begin with a motivating example. Say we wish to evaluate

$$\int \frac{dx}{1+\sqrt{x}}.$$

If we naively try the substitution $t=\sqrt{x}$, or equivalently $x=t^2$, where $t \geq 0$, then $dx = 2t\ dt$. The integral becomes

$$\int \frac{dx}{1+\sqrt{x}} = \int \frac{2t}{1+t}\ dt = \int \left(2-\frac{2}{1+t}\right) dt = 2t - 2\ln(1+t) + C.$$

Replacing t with $\sqrt{x}$ yields

$$\int \frac{dx}{1+\sqrt{x}} = 2\sqrt{x} - 2\ln\left(1+\sqrt{x}\right) + C.$$

A direct differentiation check shows that the derivative of the right side is indeed $\frac{1}{1+\sqrt{x}}$. We now introduce the inverse substitution rule (Theorem 5.17), which can be seen as a result that is equivalent to integration by substitution (Theorem 5.15).

Theorem 5.17 (inverse substitution rule). Let f be continuous on an interval and let $x=g(t)$ be a one-to-one differentiable function with continuous derivative g'. Then,

$$\int f(x)\ dx = \int f(g(t))\, g'(t)\ dt,$$

where the integral on the right is taken with respect to t. After integrating in t, one then replaces t by the inverse function $g^{-1}(x)$ to express the result in terms of x.

Note that the usual integration by substitution method (Theorem 5.15) is often used to simplify the integrand. On the other hand, the inverse substitution rule (Theorem 5.17) is often used to change the type of integrand (for instance, to turn an integral involving $\sqrt{x}$ into a rational function of t). The requirement that g be one-to-one ensures that an inverse function g^{-1} is well-defined on the relevant interval.

Example 5.10. Evaluate

$$\int \frac{dx}{x(1+x^4)}.$$

Solution. Set $t=\frac{1}{x}$ so $x=\frac{1}{t}$. As such,

$$\frac{dx}{dt} = -\frac{1}{t^2} \quad \text{so} \quad dx = -\frac{1}{t^2}\ dt.$$

The integral becomes

$$\int \frac{dx}{x(1+x^4)} = -\int \frac{t^3}{1+t^4}\ dt.$$

Now, let $u = 1+t^4$, so $\frac{du}{dt} = 4t^3$, which implies $du = 4t^3\,dt$. Then

$$-\int \frac{t^3}{1+t^4}\,dt = -\frac{1}{4}\int \frac{1}{u}\,du = -\frac{1}{4}\ln|u| + C = -\frac{1}{4}\ln\left(1+t^4\right) + C.$$

Finally, $t = \frac{1}{x}$ gives

$$\int \frac{dx}{x\left(1+x^4\right)} = -\frac{1}{4}\ln\left(1+\frac{1}{x^4}\right) + C.$$

□

Example 5.11. Evaluate

$$\int \frac{dx}{(1+x^2)^n} \quad \text{where } n \in \mathbb{N}.$$

Solution. Note that

$$\frac{d}{dx}\tan^{-1}(x) = \frac{1}{1+x^2} \quad \Rightarrow \quad \int \frac{dx}{1+x^2} = \tan^{-1}x + C.$$

For general n, let $x = \tan t$, where $t \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$ so $\frac{dx}{dt} = \sec^2 t$. Then, Then,

$$1+x^2 = 1+\tan^2 t = \sec^2 t,$$

so

$$\int \frac{dx}{(1+x^2)^n} = \int \frac{\sec^2 t}{(\sec^2 t)^n}\,dt = \int \frac{1}{(\sec t)^{2n-2}}\,dt = \int \cos^{2n-2} t\,dt.$$

Thus, these integrals reduce to integrals of even powers of cosine, which can be handled using trigonometric identities (such as half-angle formulas) and recursion. See Example 5.20. □

Example 5.12. We see that

$$\int \tan x\,dx = \int \frac{\sin x}{\cos x}\,dx.$$

Let $u = \cos x$ so $\frac{du}{dx} = -\sin x$. As such,

$$\int \tan x\,dx = \int -\frac{1}{u}\,du = -\ln|u| + C = -\ln|\cos x| + C.$$

Example 5.13. Note that

$$\int \sec x\,dx = \int \frac{\sec(x)(\sec(x)+\tan(x))}{\sec(x)+\tan(x)}\,dx.$$

Let $u = \sec x + \tan x$. Then, $\frac{du}{dx} = \sec x \tan x + \sec^2 x$. After some algebraic manipulation, one obtains

$$\int \sec x\,dx = \ln|\sec x + \tan x| + C.$$

Example 5.14 (MA2002 AY23/24 Sem 2). Let f be any positive continuous function on $[0, \pi/2]$. Find

$$\int_0^{\pi/2} \frac{f(\cos x)}{f(\cos x) + f(\sin x)}\,dx.$$

Hint: Use the identity $\cos(\pi/2 - x) = \sin x$.

Solution. Let the original integral be I. Using the substitution $u = \pi/2 - x$, we have $du = -dx$, so

$$\begin{aligned} I &= -\int_{\pi/2}^{0} \frac{f(\cos(\pi/2 - u))}{f(\cos(\pi/2 - u)) + f(\sin(\pi/2 - u))}\, du \\ &= \int_0^{\pi/2} \frac{f(\sin u)}{f(\sin u) + f(\cos u)}\, du \\ &= \int_0^{\pi/2} \frac{f(\sin x)}{f(\sin x) + f(\cos x)}\, dx \end{aligned}$$

So,

$$I + I = \int_0^{\pi/2} \frac{f(\cos x)}{f(\cos x) + f(\sin x)}\, dx + \int_0^{\pi/2} \frac{f(\sin x)}{f(\sin x) + f(\cos x)}\, dx = \int_0^{\pi/2} 1\, dx = \frac{\pi}{2}.$$

Hence, $I = \pi/4$. □

5.6 Integration by Parts

Let $u = u(x)$ and $v = v(x)$ be differentiable functions with continuous derivatives on an interval. Recall that the product rule (Theorem 4.2) states that

$$\frac{d}{dx}(u(x)v(x)) = u'(x)v(x) + u(x)v'(x).$$

Integrating both sides with respect to x,

$$\int \left(u'(x)v(x) + u(x)v'(x)\right)\, dx = u(x)v(x) + C.$$

Rearranging gives the integration by parts formula (Theorem 5.18). The art in using integration by parts is to choose u and dv so that the resulting integral $\int v\, du$ is simpler than the original $\int u\, dv$.

Theorem 5.18 (integration by parts). If $u = u(x)$ and $v = v(x)$ are differentiable with continuous derivatives, then

$$\int u(x)v'(x)\, dx = u(x)v(x) - \int u'(x)v(x)\, dx.$$

In differential notation (with $du = u'(x)\,dx$, $dv = v'(x)\,dx$), we have

$$\int u\, dv = uv - \int v\, du.$$

Example 5.15. Compute

$$\int \ln x\, dx \quad \text{where } x > 0.$$

Solution. Take $u = \ln x$ and $dv = dx$ so $du = \frac{1}{x}\, dx$ and $v = x$. As such,

$$\int \ln x\, dx = x\ln x - \int x \cdot \frac{1}{x}\, dx = x\ln x - \int 1\, dx = x\ln x - x + C.$$

□

Example 5.16. Compute

$$\int x \sin x \, dx.$$

Solution. Take $u = x^2$ and $dv = e^x \, dx$ so $du = 2x \, dx$ and $v = e^x$. Then,

$$\int x \sin x \, dx = -x \cos x + \int \cos x \, dx = -x \cos x + \sin x + C.$$

□

Example 5.17. Compute

$$\int x^2 e^x \, dx.$$

Solution. First, choose $u = x^2$ and $dv = e^x \, dx$, so $du = 2x \, dx$ and $v = e^x$. As such,

$$\int x^2 e^x \, dx = x^2 e^x - 2 \int x e^x \, dx.$$

For $\int x e^x \, dx$, integrate by parts again with $u = x$ and $dv = e^x \, dx$, so $du = dx$ and $v = e^x$. This yields

$$\int x e^x \, dx = x e^x - \int e^x \, dx = x e^x - e^x + C.$$

As such,

$$\int x^2 e^x \, dx = x^2 e^x - 2\left(x e^x - e^x\right) + C = \left(x^2 - 2x + 2\right) e^x + C.$$

□

Example 5.18. Compute

$$\int \sin^{-1} x \, dx \quad \text{where } -1 < x < 1.$$

Solution. Take $u = \sin^{-1} x$ and $dv = dx$, so $du = \frac{1}{\sqrt{1-x^2}} \, dx$ and $v = x$. As such,

$$\int \sin^{-1} x \, dx = x \sin^{-1} x - \int \frac{x}{\sqrt{1-x^2}} \, dx.$$

For the remaining integral, let $w = 1 - x^2$, so $\frac{dw}{dx} = -2x$. Then,

$$\int \frac{x}{\sqrt{1-x^2}} \, dx = -\frac{1}{2} \int \frac{1}{\sqrt{w}} \, dw = -\sqrt{w} + C = -\sqrt{1-x^2} + C.$$

Hence,

$$\int \sin^{-1} x \, dx = x \sin^{-1} x + \sqrt{1-x^2} + C.$$

□

Sometimes, integration by parts leads to an equation involving the original integral, which can then be solved algebraically.

Example 5.19. Compute

$$\int e^x \sin x \, dx.$$

Solution. First, set $u = e^x$ and $dv = \sin x\, dx$, so $du = e^x\, dx$ and $v = -\cos x$. As such,

$$\int e^x \sin x\, dx = -e^x \cos x + \int e^x \cos x\, dx.$$

Now, we integrate the second integral by parts. So, choose $u = e^x$ and $dv = \cos x\, dx$, so $du = e^x\, dx$ and $v = \sin x$. As such,

$$\int e^x \cos x\, dx = e^x \sin x - \int e^x \sin x\, dx.$$

Hence,

$$\int e^x \sin x\, dx = -e^x \cos x + e^x \sin x - \int e^x \sin x\, dx.$$

Let

$$I = \int e^x \sin x\, dx.$$

Then,

$$I = -e^x \cos x + e^x \sin x - I \quad \text{so} \quad 2I = e^x (\sin x - \cos x).$$

Hence,

$$\int e^x \sin x\, dx = \frac{1}{2} e^x (\sin x - \cos x) + C.$$

□

Example 5.20 (reduction formula). For any non-zero integer n, prove that

$$\int \cos^n x\, dx = \frac{1}{n} \cos^{n-1} x \sin x + \frac{n-1}{n} \int \cos^{n-2} x\, dx.$$

Solution. The trick is to write

$$\int \cos^n x\, dx = \int \cos^{n-1} x \cos x\, dx.$$

Let $u = \cos^{n-1} x$ and $dv = \cos x\, dx$, so $du = -(n-1)\cos^{n-2} x \sin x\, dx$ and $v = \sin x$. As such,

$$\int \cos^n x\, dx = \cos^{n-1} x \sin x + (n-1) \int \cos^{n-2} x \sin^2 x\, dx.$$

Using $\sin^2 x = 1 - \cos^2 x$,

$$\int \cos^n x\, dx = \cos^{n-1} x \sin x + (n-1) \int \cos^{n-2} x\, dx - (n-1) \int \cos^n x\, dx.$$

Collect the integrals of $\cos^n x$ on the left so

$$n \int \cos^n x\, dx = \cos^{n-1} x \sin x + (n-1) \int \cos^{n-2} x\, dx.$$

Dividing by n gives the claimed formula. □

Example 5.21 (MA2002 AY22/23 Sem 1). Suppose f is a continuously differentiable function. That is, the derivative of f is continuous. Prove that

$$\lim_{n\to\infty} \frac{1}{n^2} \sum_{k=1}^{n} \left[k f'\left(\frac{k}{n}\right) + n f\left(\frac{k}{n}\right)\right] = f(1).$$

Solution. One should be well-versed with the contents from Chapter 5.1. We see that

$$\lim_{n\to\infty}\frac{1}{n^2}\sum_{k=1}^{n}kf'\left(\frac{k}{n}\right)=\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}\frac{k}{n}f'\left(\frac{k}{n}\right)=\int_0^1 xf'(x)\ dx$$

and

$$\lim_{n\to\infty}\frac{1}{n^2}\sum_{k=1}^{n}nf\left(\frac{k}{n}\right)=\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}f\left(\frac{k}{n}\right)=\int_0^1 f(x)\ dx.$$

As such,

$$\lim_{n\to\infty}\frac{1}{n^2}\sum_{k=1}^{n}\left[kf'\left(\frac{k}{n}\right)+nf\left(\frac{k}{n}\right)\right]=\int_0^1 xf'(x)+f(x)\ dx.$$

Performing integration by parts on the first integrand yields

$$\int_0^1 xf'(x)\ dx=[xf(x)]_0^1-\int_0^1 f(x)\ dx=f(1)-\int_0^1 f(x)\ dx$$

and the result follows from here. □

5.7 Trigonometric Substitution

When the integrand contains square roots of quadratic expressions in x, completing the square often leads to one of the standard forms

$$\sqrt{a^2-x^2}\quad\text{or}\quad\sqrt{a^2+x^2}\quad\text{or}\quad\sqrt{x^2-a^2},$$

where $a>0$, which can be handled by appropriate trigonometric substitutions based on Pythagorean identities.

(i) If $\sqrt{a^2-x^2}$ appears, use $x=a\sin t$ where $t\in\left[-\frac{\pi}{2},\frac{\pi}{2}\right]$. Then,

$$\sqrt{a^2-x^2}=a\cos t.$$

(ii) If $\sqrt{a^2+x^2}$ appears, use $x=a\tan t$ where $t\in\left(-\frac{\pi}{2},\frac{\pi}{2}\right)$. Then,

$$\sqrt{a^2+x^2}=a\sec t.$$

(iii) If $\sqrt{x^2-a^2}$ appears, use $x=a\sec t$ where $t\in\left[0,\frac{\pi}{2}\right)\cup\left[\pi,\frac{3\pi}{2}\right)$. Then,

$$\sqrt{x^2-a^2}=a\tan t.$$

Example 5.22. Compute

$$\int\sqrt{1-x^2}\ dx.$$

Solution. Let $x=\sin t$ where $t\in\left[-\frac{\pi}{2},\frac{\pi}{2}\right]$. As such, $\sqrt{1-x^2}=\cos t$ and $dx=\cos t\ dt$. Then,

$$\int\sqrt{1-x^2}\ dx=\int\cos^2 t\ dt=\int\frac{1+\cos 2t}{2}\ dt=\frac{t}{2}+\frac{1}{4}\sin 2t+C.$$

Since $\sin 2t=2\sin t\cos t=2x\sqrt{1-x^2}$ and $t=\sin^{-1}x$, we obtain

$$\int\sqrt{1-x^2}\ dx=\frac{1}{2}\sin^{-1}x+\frac{1}{2}x\sqrt{1-x^2}+C.$$

□

Example 5.23. Compute
$$\int \sqrt{1+x^2}\,dx.$$

Solution. Let $x = \tan t$ where $t \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$. So, $\sqrt{1+x^2} = \sec t$ and $dx = \sec^2 t\,dt$. Then, Then,
$$\int \sqrt{1+x^2}\,dx = \int \sec t \sec^2 t\,dt = \int \sec^3 t\,dt.$$
Using the known formula
$$\int \sec^3 t\,dt = \frac{1}{2}\sec t\tan t + \frac{1}{2}\ln|\sec t + \tan t| + C,$$
we get
$$\int \sqrt{1+x^2}\,dx = \frac{1}{2}x\sqrt{1+x^2} + \frac{1}{2}\ln\left(\sqrt{1+x^2}+x\right) + C.$$
□

Example 5.24. Compute
$$\int \frac{x}{\sqrt{5+4x-x^2}}\,dx.$$

Solution. First complete the square in the denominator to obtain
$$5+4x-x^2 = 3^2-(x-2)^2.$$
Let $x-2 = 3\sin t$, where $t \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$. So, $\sqrt{5+4x-x^2} = 3\cos t$ and $x = 2+3\sin t$. As such, $dx = 3\cos t\,dt$. Thus,
$$\int \frac{x}{\sqrt{5+4x-x^2}}\,dx = \int \frac{2+3\sin t}{3\cos t}\cdot 3\cos t\,dt = \int (2+3\sin t)\,dt$$
which is equal to $2t-3\cos t+C$. We then rewrite this expression in terms of x. Since $x-2 = 3\sin t$, we have
$$\sin t = \frac{x-2}{3} \quad \text{and} \quad \cos t = \frac{\sqrt{5+4x-x^2}}{3} \quad \text{and} \quad t = \sin^{-1}\left(\frac{x-2}{3}\right).$$
Hence,
$$\int \frac{x}{\sqrt{5+4x-x^2}}\,dx = 2\sin^{-1}\left(\frac{x-2}{3}\right) - \sqrt{5+4x-x^2} + C,$$
□

Example 5.25 (MA2002 AY21/22 Sem 1). Evaluate the following definite integral:
$$\int_1^2 \frac{1}{x^2(x^2+4)^{3/2}}\,dx$$

Solution. It hints to us to consider the Pythagorean identity $\tan^2\theta + 1 = \sec^2\theta$, so the substitution required is $x^2 = 4\tan^2\theta$, or rather, $x = 2\tan\theta$. This yields $dx = 2\sec^2\theta\,d\theta$. Then, the integral becomes
$$\begin{aligned}\int_{\arctan(1/2)}^{\pi/4} \frac{2\sec^2\theta}{4\tan^2\theta\cdot 8\sec^3\theta}\,d\theta &= \frac{1}{16}\int_{\arctan(1/2)}^{\pi/4} \frac{1}{\tan^2\theta\sec\theta}\,d\theta \\ &= \frac{1}{16}\int_{\arctan(1/2)}^{\pi/4} \cos\theta\cot^2\theta\,d\theta\end{aligned}$$

We then use the Pythagorean identity $\cot^2\theta = \csc^2\theta - 1$ and the fact that $-\csc\theta$ is an antiderivative of $\cot\theta\csc\theta$. As such, the integral becomes

$$\frac{1}{16}\int_{\arctan(1/2)}^{\pi/4} \cot\theta\csc\theta - \frac{1}{16}\int_{\arctan(1/2)}^{\pi/4} \cos\theta\, d\theta = -\frac{1}{16}\left[\sin\theta + \csc\theta\right]_{\arctan(1/2)}^{\pi/4}.$$

Given that $\theta = \arctan(1/2)$, we wish to find expressions for $\sin\theta$ and $\csc\theta$. As $\tan\theta = 1/2$, we can construct a right triangle with legs 1 and 2 and hypotenuse $\sqrt{5}$, where the angle θ is opposite the leg with length 1. As such, $\sin\theta = 1/\sqrt{5}$ and $\csc\theta = \sqrt{5}$. To conclude, the answer is

$$-\frac{1}{16}\left(\frac{1}{\sqrt{2}} + \sqrt{2} - \frac{1}{\sqrt{5}} - \sqrt{5}\right).$$

□

5.8 Integration of Rational Functions

Some simple rational integrals include the following:

$$\int \frac{dx}{x^n} = \begin{cases} \ln|x| + C & \text{if } n = 1; \\ \frac{x^{1-n}}{1-n} + C & \text{if } n \geq 2, \end{cases}$$

and, via the substitution $x = \tan t$,

$$\int \frac{dx}{(1+x^2)^n} = \int \cos^{2n-2} t\, dt,$$

for integers $n \geq 1$, which we have already discussed in Example 5.11.

We now discuss the factorisation of polynomials over the real numbers $\mathbb{R}$.

Theorem 5.19. Every non-constant polynomial with real coefficients can be uniquely written as a product of real linear factors and real irreducible quadratic factors.

We give a more precise formulation of Theorem 5.19. If $P(x)$ is a real polynomial of degree ≥ 1, then we can write

$$P(x) = \prod_j (x + a_j)^{r_j} \prod_k (x^2 + b_k x + c_k)^{s_k},$$

where $r_j, s_k \in \mathbb{N}$, $b_k, c_k \in \mathbb{R}$ and $b_k^2 < 4c_k$, and the factorisation is unique up to ordering. Recall that a rational function

$$f(x) = \frac{A(x)}{B(x)}$$

is proper if $\deg A < \deg B$. Otherwise it is called improper.

Proposition 5.2 (division algorithm). Given any rational function $f(x) = \frac{A(x)}{B(x)}$, with polynomials A, B and $\deg B \geq 1$, there exist unique polynomials $Q(x)$, $R(x)$ such that

$$A(x) = B(x)Q(x) + R(x) \quad \text{where } \deg R < \deg B.$$

Thus,

$$f(x) = Q(x) + \frac{R(x)}{B(x)},$$

so integration reduces to integrating a polynomial and a proper rational function.

We now discuss the method of partial fraction decomposition. Let

$$f(x) = \frac{A(x)}{B(x)}$$

be a proper rational function, and we wish to factor $B(x)$ as a product of linear and irreducible quadratic factors. Suppose

$$B(x) = \prod_j (x+a_j)^{r_j} \prod_k (x^2 + b_k x + c_k)^{s_k}$$

is the factorisation of $B(x)$ over $\mathbb{R}$ into linear and irreducible quadratic factors. Then, the proper rational function $\dfrac{A(x)}{B(x)}$ can be expressed uniquely as a sum of partial fractions of the form

$$\sum_j \left(\frac{A_{j,1}}{x+a_j} + \frac{A_{j,2}}{(x+a_j)^2} + \cdots + \frac{A_{j,r_j}}{(x+a_j)^{r_j}} \right) + \sum_k \left(\frac{B_{k,1}x + C_{k,1}}{x^2 + b_k x + c_k} + \cdots + \frac{B_{k,s_k}x + C_{k,s_k}}{(x^2 + b_k x + c_k)^{s_k}} \right),$$

where the coefficients $A_{j,\ell}$, $B_{k,m}$, $C_{k,m} \in \mathbb{R}$ are uniquely determined. Note that the total number of unknown coefficients in the partial fraction decomposition equals $\deg B(x)$.

For example, for linear factors, we have

$$\int \frac{dx}{(x+a)^k} = \begin{cases} \ln|x+a| + C, & \text{if } k = 1; \\ \frac{(x+a)^{1-k}}{1-k} + C & \text{if } k \geq 2. \end{cases}$$

For quadratic factors, we can use the identity

$$x^2 + bx + c = \left(x + \frac{b}{2}\right)^2 + \alpha^2 \quad \text{where } \alpha^2 = c - \frac{b^2}{4} > 0.$$

Example 5.26. Decompose

$$\frac{4x}{x^3 - x^2 - x + 1}$$

into partial fractions and integrate.

Solution. By partial fraction decomposition, one can see that

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

Hence,

$$\int \frac{4x}{x^3-x^2-x+1}\,dx = -\ln|x+1| + \ln|x-1| - \frac{2}{x-1} + C.$$

□

Example 5.27 (MA2002 AY23/24 Sem 2). Find

$$\int \frac{7x^2-13x+13}{(x-2)(x^2-2x+3)}\,dx.$$

Solution. Note that $x^2-2x+3=(x-1)^2+2$, which cannot be split into linear factors with real coefficients. In fact, recall from our earlier discussion that such polynomials are said to be irreducible over $\mathbb{R}$. So, there exist $A,B,C\in\mathbb{R}$ such that

$$\frac{7x^2-13x+13}{(x-2)(x^2-2x+3)} = \frac{A}{x-2} + \frac{Bx+C}{(x-1)^2+2}.$$

Hence,

$$\begin{aligned} 7x^2-13x+13 &= A\left[(x-1)^2+2\right] + (Bx+C)(x-2) \\ &= (A+B)x^2 + (-2A-2B+C)x + 3A-2C \end{aligned}$$

This implies $A+B=7$, $-2A-2B+C=-13$ and $3A-2C=13$. Substituting $B=7-A$ and $C=\frac{1}{2}(3A-13)$ into the second equation, we have

$$-2A-2(7-A)+\frac{1}{2}(3A-13) = -13 \quad \text{so } A=5.$$

Consequently, $B=2$ and $C=1$. Hence, the partial fraction decomposition of the integrand is

$$\frac{5}{x-2} + \frac{2x+1}{(x-1)^2+2}.$$

Integrating, we obtain

$$\int \frac{5}{x-2}\,dx + \int \frac{2x+1}{(x-1)^2+2}\,dx = 5\ln|x-2| + \int \frac{2x-2}{(x-1)^2+2}\,dx + \int \frac{3}{(x-1)^2+2}\,dx$$

which is equal to

$$5\ln|x-2| + \ln\left[(x-1)^2+2\right] + \frac{3}{\sqrt{2}}\tan^{-1}\left(\frac{x-1}{\sqrt{2}}\right) + c.$$

□

Example 5.28 (MA2002 AY23/24 Sem 2). For what value(s) of a does

$$\int_1^{\infty} \frac{ax}{x^2+1} - \frac{1}{2x}\,dx$$

converge? Evaluate the corresponding integral(s).

Solution. We have

$$\begin{aligned}\int_1^\infty \frac{ax}{x^2+1} - \frac{1}{2x}\,dx &= \frac{1}{2}\left[a\ln\left(x^2+1\right) - \ln x\right]_1^\infty \\ &= \frac{1}{2}\lim_{R\to\infty}\ln\left[\frac{\left(R^2+1\right)^a}{R}\right] - \frac{a}{2}\ln 2\end{aligned}$$

Note that the second expression $a\ln 2/2$ is not affected by R, so it suffices to consider the first expression. Hence, we must have

$$\lim_{R\to\infty}\ln\left[\frac{\left(R^2+1\right)^a}{R}\right] = 0 \quad \text{so} \quad \lim_{R\to\infty}\frac{\left(R^2+1\right)^a}{R} = 1.$$

By L'Hôpital's rule (Theorem 4.21), one can deduce that $a = 1/2$. □

Example 5.29 (bounding π). On the interval $[0,1]$ we have

$$\frac{1}{2} \le \frac{1}{1+x^2} \le 1.$$

Multiplying by $x^4(1-x)^4 \ge 0$, integrating on $[0,1]$, and using explicit computations yields

$$\frac{1}{1260} \le \frac{22}{7} - \pi \le \frac{1}{630}.$$

This provides a classical rational approximation $\frac{22}{7}$ to π.

5.9 Universal Trigonometric Substitution

For integrals of the form

$$\int f(\sin(x), \cos(x))\,dx,$$

where f is a rational expression in two variables, there is a powerful substitution that converts everything into a rational function of a single variable. Let

$$t = \tan\left(\frac{x}{2}\right) \text{ where } -\pi < x < \pi \quad \text{so that} \quad x = 2\tan^{-1} t.$$

We recall the identities

$$\sin x = \frac{2\tan\left(\frac{x}{2}\right)}{1+\tan^2\left(\frac{x}{2}\right)} \quad \text{and} \quad \cos x = \frac{1-\tan^2\left(\frac{x}{2}\right)}{1+\tan^2\left(\frac{x}{2}\right)}.$$

Hence,

$$\sin x = \frac{2t}{1+t^2} \quad \text{and} \quad \cos x = \frac{1-t^2}{1+t^2}.$$

Furthermore, differentiating $x = 2\tan^{-1} t$ gives

$$\frac{dx}{dt} = \frac{2}{1+t^2} \quad \Rightarrow \quad dx = \frac{2}{1+t^2}\,dt.$$

Theorem 5.20 (universal trigonometric substitution). For $-\pi < x < \pi$, we have

$$\int f(\sin(x), \cos(x))\, dx = \int f\left(\frac{2t}{1+t^2}, \frac{1-t^2}{1+t^2}\right) \frac{2}{1+t^2}\, dt,$$

where $t = \tan\left(\frac{x}{2}\right)$. The integrand on the right is a rational function of t.

Example 5.30. Compute

$$\int \sec(x)\, dx \quad \text{for } -\pi < x < \pi \text{ and } x \neq \pm\frac{\pi}{2}.$$

Solution. Let $t = \tan\left(\frac{x}{2}\right)$. Then,

$$\cos x = \frac{1-t^2}{1+t^2} \quad \text{so} \quad \sec x = \frac{1+t^2}{1-t^2}$$

and

$$dx = \frac{2}{1+t^2}\, dt.$$

As such,

$$\int \sec x\, dx = \int \frac{1+t^2}{1-t^2} \cdot \frac{2}{1+t^2}\, dt = \int \frac{2}{1-t^2}\, dt.$$

By partial fraction decomposition, one can write

$$\frac{2}{1-t^2} = \frac{1}{1+t} + \frac{1}{1-t},$$

so that

$$\int \sec x\, dx = \ln|1+t| - \ln|1-t| + C = \ln\left|\frac{1+t}{1-t}\right| + C.$$

Using $t = \tan\left(\frac{x}{2}\right)$ and the identity

$$\sec x + \tan x = \frac{1+t}{1-t},$$

we obtain the well-known formula

$$\int \sec x\, dx = \ln|\sec x + \tan x| + C,$$

which is valid on any interval where $\sec(x)$ is continuous. □

Example 5.31. Find

$$\int \frac{dx}{\sin x + 2\cos x + 3} \quad \text{where } -\pi < x < \pi.$$

Solution. With $t = \tan\left(\frac{x}{2}\right)$, we have

$$\sin x = \frac{2t}{1+t^2} \quad \text{and} \quad \cos x = \frac{1-t^2}{1+t^2} \quad \text{and} \quad dx = \frac{2}{1+t^2}\, dt.$$

As such,

$$\sin x + 2\cos x + 3 = \frac{2t}{1+t^2} + 2 \cdot \frac{1-t^2}{1+t^2} + 3 = \frac{t^2+2t+5}{1+t^2}.$$

Hence,

$$\int \frac{dx}{\sin x + 2\cos x + 3} = \int \frac{\frac{2}{1+t^2}\, dt}{\frac{t^2+2t+5}{1+t^2}} = \int \frac{2}{t^2+2t+5}\, dt.$$

Completing the square yields $t^2+2t+5 = (t+1)^2+2^2$. One can then use the substitution $u = \frac{1}{2}(t+1)$ to obtain

$$\int \frac{2}{t^2+2t+5}\, dt = \int \frac{2}{4(u^2+1)} \cdot 2\, du = \int \frac{1}{u^2+1}\, du = \tan^{-1}(u) + C = \tan^{-1}\left(\frac{t+1}{2}\right) + C.$$

Thus,

$$\int \frac{dx}{\sin x + 2\cos x + 3} = \tan^{-1}\left(\frac{1}{2}\tan\left(\frac{x}{2}\right) + \frac{1}{2}\right) + C \quad \text{where } -\pi < x < \pi.$$

□

As an extension of Example 5.31, on $[-\pi, \pi]$, one can extend this antiderivative to a continuous function $F_1(x)$, and use it to evaluate definite integrals such as

$$\int_0^{\pi} \frac{dx}{\sin x + 2\cos x + 3} = F_1(\pi) - F_1(0) = \frac{\pi}{2} - \tan^{-1}\left(\frac{1}{2}\right).$$

5.10 Area of Plane Regions

Let $f : [a,b] \to \mathbb{R}$ be a continuous and non-negative function. Consider the region in the plane bounded above by the graph $y = f(x)$, below by the x-axis, and between the vertical lines $x = a$ and $x = b$. As shown in Figure 5.2, the area of this region is defined to be

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

This agrees with the intuitive picture in Figure 5.1 obtained by approximating the region by rectangles. Here is a heuristic justification. We partition $[a,b]$ into subintervals of equal length $\Delta x = \frac{b-a}{n}$. On each subinterval, choose a sample point x_k^*. The area of the thin rectangle above x_k^* is approximately $f\left(x_k^*\right)\Delta x$. Summing and taking limits,

$$A \approx \sum_{k=1}^{n} f(x_k^*)\Delta x \quad \text{which tends to} \quad \int_a^b f(x)\ dx$$

as $\Delta x \to 0$. This is precisely the Riemann integral.

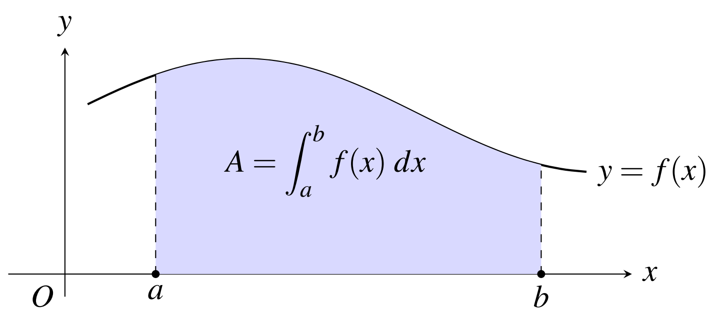


Figure 5.2: Area below a curve $y = f(x)$

We then discuss the area between two curves. Let $f, g : [a,b] \to \mathbb{R}$ be continuous functions with $f(x) \geq g(x)$ for all $x \in [a,b]$. Consider the region bounded above by $y = f(x)$, below by $y = g(x)$, and between $x = a$ and $x = b$. As shown in Figure 5.3, the area of this region is

$$A = \int_a^b (f(x) - g(x))\, dx.$$

Again, we give a heuristic justification. At a point $x \in [a,b]$, consider the vertical line segment joining the lower curve $y = g(x)$ to the upper curve $y = f(x)$. Its length is

$$\ell(x) = f(x) - g(x) \geq 0.$$

Approximating the region by thin vertical strips of width Δx, we have

$$A \approx \sum \ell(x^*)\, \Delta x = \sum (f(x^*) - g(x^*))\, \Delta x.$$

Taking the limit as $\Delta x \to 0$ gives the stated integral.

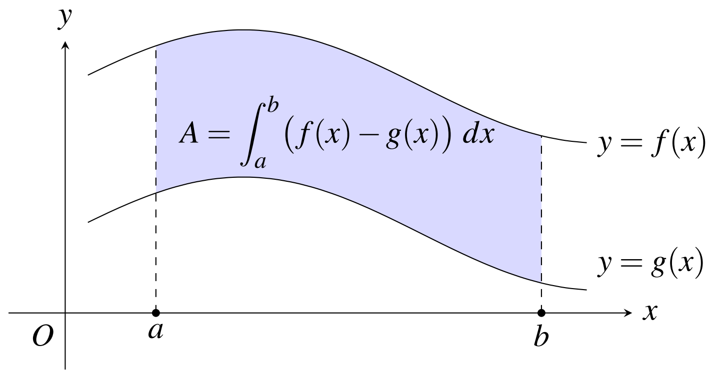
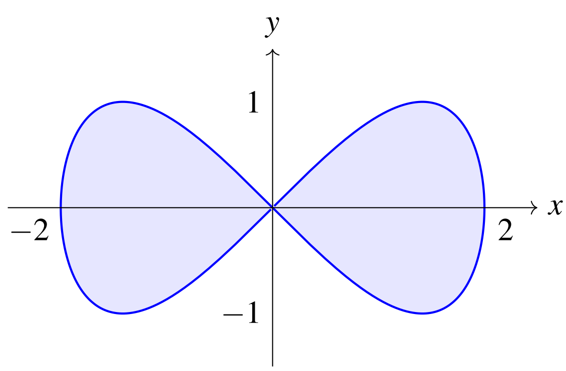


Figure 5.3: Area between two curves $y = f(x)$ and $y = g(x)$

Example 5.32 (MA2002 AY21/22 Sem 1). Let R be the region enclosed by the curve defined by $x^4 = 4(x^2 - y^2)$. Find the area of R.



Solution. Setting $y = 0$, we see that the curve intersects the axes at $(-2,0), (0,0)$ and $(2,0)$. With some simple algebraic manipulation, the top half of the curve can be represented by

$$y = \sqrt{x^2 - \frac{1}{4}x^4}.$$

As such, due to symmetry, the area of R is four times the integral over the top half of the curve from 0 to 2. That is,

$$\text{Area} = 4\int_0^2 \sqrt{x^2 - \frac{1}{4}x^4}\, dx = 2\int_0^2 \sqrt{4x^2 - x^4}\, dx = 2\int_0^2 x\sqrt{4-x^2}\, dx.$$

To finish off, one can use the substitution $u = x^2$. We omit the remaining details. The reader can check that the area is equal to $\frac{16}{3}$ units^2. □

Example 5.33 (MA2002 AY23/24 Sem 2). Let $f(x) = -x^2 + 4x + 2$ and $g(x) = x^2 - 6x + 10$. Find the area of the shaded region R bounded by f and g.

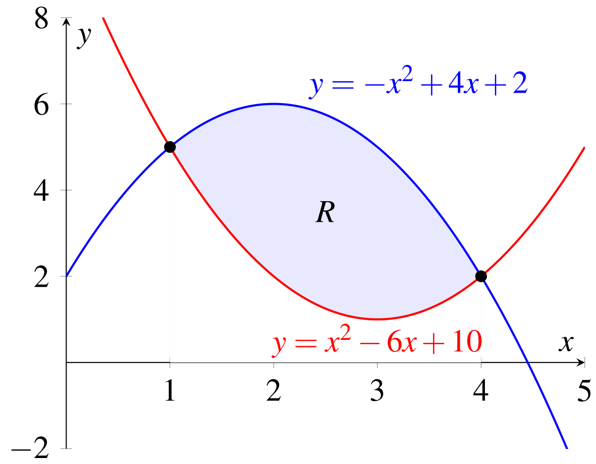


Figure 5.4: The graphs of $y = f(x)$ and $y = g(x)$ with shaded region R

Solution. We first find the intersection points. Setting $f(x) = g(x)$, we have $2x^2 - 10x + 8 = 0$, so $x^2 - 5x + 4 = 0$. The roots are $x = 1$ and $x = 4$. Since $f \geq g$ on $[1, 4]$, then the area of the shaded region R is

$$\int_1^4 f(x) - g(x)\ dx = \int_1^4 -2x^2 + 10x - 8\ dx = 9.$$

□

Sometimes it is more convenient to slice the region horizontally. Suppose a plane region R lies between the horizontal lines $y = c$ and $y = d$, and for each $y \in [c, d]$ the intersection of R with the horizontal line is a segment from $x = x_{\text{left}}(y)$ to $x = x_{\text{right}}(y)$, where both functions are continuous and $x_{\text{right}}(y) \geq x_{\text{left}}(y)$. Then, the area of R is

$$A = \int_c^d \left(x_{\text{right}}(y) - x_{\text{left}}(y)\right)\ dy.$$

To see why, at a fixed height y, the horizontal slice has length

$$L(y) = x_{\text{right}}(y) - x_{\text{left}}(y).$$

Approximating by thin horizontal strips of height Δy,

$$A \approx \sum L(y^*)\,\Delta y = \sum \left(x_{\text{right}}(y^*) - x_{\text{left}}(y^*)\right)\Delta y.$$

Passing to the limit yields the formula.

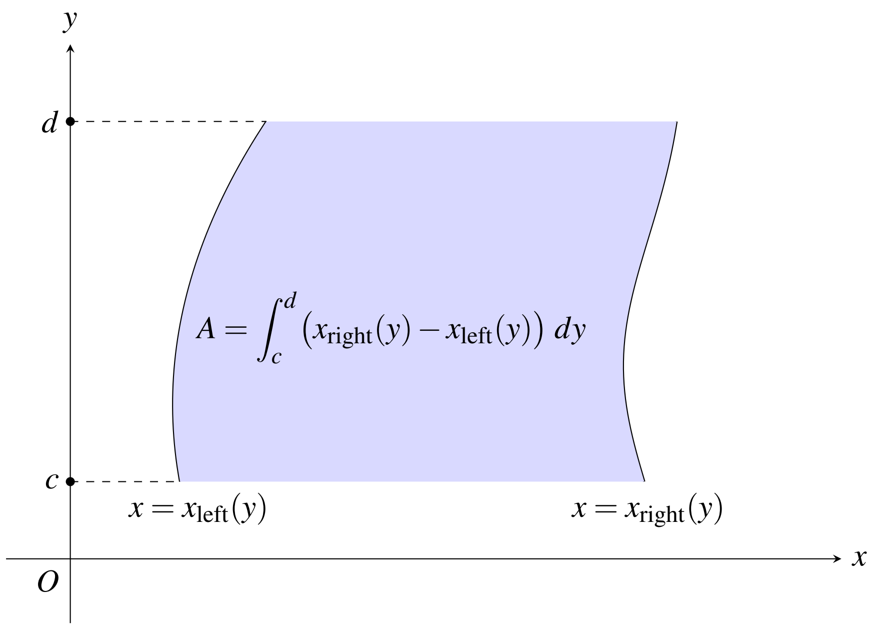


Figure 5.5: Area between two curves $x = x_{\text{right}}(y)$ and $x = x_{\text{left}}(y)$

Example 5.34. Consider the region enclosed between the curves $y = x^2$ and $y = 2x - x^2$.

Using vertical slices, the intersection points satisfy $x^2 = 2x - x^2$ so $2x^2 - 2x = 0$. For $x \in [0,1]$, we have $2x - x^2 \geq x^2$ so the length of the vertical slice is $\ell(x) = 2x - 2x^2$. Thus,

$$A = \int_0^1 \left(2x - 2x^2\right)\,dx = \frac{1}{3}.$$

Using horizontal slices, for $y \in [0,1]$, we have $y = x^2$ and $y = 2x - x^2$ which imply $x = \sqrt{y}$ and $x = 1 - \sqrt{1-y}$ respectively. The left boundary is $x = 1 - \sqrt{1-y}$ and the right boundary is $x = \sqrt{y}$, so the slice length is $L(y) = \sqrt{y} + \sqrt{1-y} - 1$. Again, one checks that

$$A = \int_0^1 L(y)\,dy = \frac{1}{3},$$

which agrees with the previous computation.

5.11 Volume of Revolution (Disc Method)

A solid is a three-dimensional region in space. Suppose the solid lies between the planes $x = a$ and $x = b$ along the x-axis, and let $A(x)$ denote the area of the cross-section of the solid cut by the plane perpendicular to the x-axis at position x, assuming $A(x)$ is continuous on $[a,b]$. Under these assumptions, the volume of the solid is

$$V = \int_a^b A(x)\,dx.$$

To see why, we partition $[a,b]$ into subintervals of length Δx. For a sample point x_k^* in the k^{th} subinterval, approximate the slice of the solid between x_k and x_{k+1} by a cylinder of base area $A\left(x_k^*\right)$ and thickness Δx, so its volume is $A\left(x_k^*\right)\Delta x$. Summing and taking limits,

$$V \approx \sum A\left(x_k^*\right)\Delta x \quad \longrightarrow \quad \int_a^b A(x)\ dx.$$

An entirely analogous formula holds if the solid is described along the y-axis. That is, if $A(y)$ is the cross-sectional area of the plane perpendicular to the y-axis, then

$$V = \int_c^d A(y)\ dy.$$

Example 5.35 (cone). For example, one can recover the classic formula for the volume of a cone of base area A and height h. Place the cone with its vertex at the origin and its base in the plane $z = H$. Let $A(h)$ be the area of the cross-section at height h, where $0 \le h \le H$. By the similarity of triangles, the cone of height h is a scaled copy of the whole cone, with linear scaling factor $\frac{h}{H}$. Hence, the areas scale by the square of this factor. That is to say,

$$\frac{A(h)}{A} = \left(\frac{h}{H}\right)^2 \quad \Rightarrow \quad A(h) = A\left(\frac{h}{H}\right)^2 = \frac{A}{H^2}h^2.$$

Thus,

$$V = \int_0^H A(h)\ dh = \frac{A}{H^2}\int_0^H h^2\ dh = \frac{A}{H^2}\cdot\frac{H^3}{3} = \frac{1}{3}AH.$$

In particular, for a circular base of radius r, we have $A = \pi r^2$, so $V = \frac{1}{3}\pi r^2 H$.

Example 5.36 (sphere). We can also recover the formula for the volume of a unit sphere. For a more satisfying discussion involving spherical coordinates, check out MA2104 Multivariable Calculus. Anyway, we consider the unit sphere

$$x^2 + y^2 + z^2 = 1.$$

Place the sphere so that it lies between the planes $x = -1$ and $x = 1$. For each $x \in [-1, 1]$, the cross-section perpendicular to the x-axis is a disk of radius $\sqrt{1 - x^2}$, hence the area is

$$A(x) = \pi\left(1 - x^2\right).$$

Therefore,

$$V = \int_{-1}^1 \pi\left(1 - x^2\right)\ dx = \frac{4}{3}\pi.$$

We now discuss the disc method for finding the volume of the solid of revolution.

Theorem 5.21 (disc method for solids of revolution about x-axis). Let $f : [a,b] \to \mathbb{R}$ be a continuous function. Consider the region bounded by the graph $y = f(x)$, the x-axis, and the vertical lines $x = a$, $x = b$. Suppose $f(x) \ge 0$ on $[a,b]$. Rotate this region about the x-axis to form a solid of revolution. At each x, the cross-section perpendicular to the x-axis is a disk of radius $f(x)$ and area

$$A(x) = \pi\left(f(x)\right)^2.$$

The volume of the solid is

$$V = \int_a^b \pi (f(x))^2 \, dx.$$

Example 5.37. Let $f(x) = \sqrt{x}$ on $[0,1]$. Rotating the region under f about the x-axis (Figure 5.6) produces a solid whose volume is

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

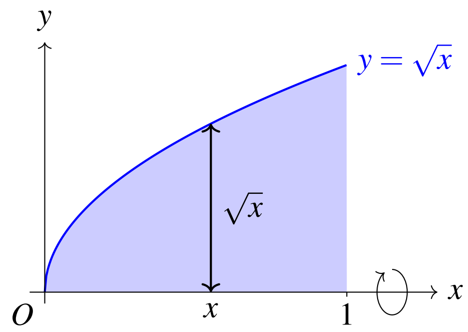


Figure 5.6: Volume of solid of revolution of $y = \sqrt{x}$

Theorem 5.22 (disc method for solids of revolution about x-axis). Now suppose $f, g\colon [a,b] \to \mathbb{R}$ are continuous with $f(x) \geq g(x) \geq 0$. Consider the region between the two curves and rotate it about the x-axis. For a fixed x, the cross-section is an annulus: the outer radius is $f(x)$, the inner radius is $g(x)$, and the area of the cross-section is

$$A(x) = \pi (f(x))^2 - \pi (g(x))^2.$$

The volume of the solid obtained by rotating the region between $y = f(x)$ and $y = g(x)$ about the x-axis is

$$V = \int_a^b \pi \left[(f(x))^2 - (g(x))^2 \right] dx.$$

Theorem 5.23 (disc method for solids of revolution about y-axis). If a region bounded by $x = g(y)$, $x = h(y)$, $y = c$, $y = d$ is rotated about the y-axis, then similar formulas hold with integration in y.

Example 5.38. If the region between $x = 0$ and $x = \sqrt[3]{y}$ for $0 \leq y \leq 8$ is rotated about the y-axis, each cross-section perpendicular to the y-axis is a disk of radius $\sqrt[3]{y}$, so

$$V = \int_0^8 \pi \left(\sqrt[3]{y}\right)^2 dy = \frac{96}{5}\pi.$$

Example 5.39. Let the region be the disk enclosed by the circle

$$x^2 + (y-2)^2 = 1.$$

Rotating this disk about the x-axis produces a solid. To compute its volume, we use the disc method. For a fixed $x \in [-1,1]$, the circle gives $y = 2 \pm \sqrt{1-x^2}$. Thus, the outer radius is $2+\sqrt{1-x^2}$, the inner radius is $2-\sqrt{1-x^2}$, and

$$A(x) = \pi\left[\left(2+\sqrt{1-x^2}\right)^2 - \left(2-\sqrt{1-x^2}\right)^2\right] = 8\pi\sqrt{1-x^2}.$$

Hence,

$$V = \int_{-1}^{1} 8\pi\sqrt{1-x^2}\,dx = 8\pi\cdot\frac{\pi}{2} = 4\pi^2.$$

5.12 Volume of Revolution (Shell Method)

The disc method in Chapter 5.11 uses cross-sections perpendicular to the axis of rotation. Sometimes it is more convenient to use cross-sections parallel to the axis, resulting in cylindrical shells.

Theorem 5.24 (cylindrical shell method). The volume of the solid formed by rotating the region under $y = f(x)$ on $[a,b]$, $a \geq 0$, about the y-axis is

$$V = \int_a^b 2\pi x f(x)\,dx. \tag{5.6}$$

Example 5.40. Let $f(x) = 2x^2 - x^3$ on $[0,2]$. Consider the region between the graph of f and the x-axis on this interval and rotate it about the y-axis. The radius of the shell at x is $r = x$, the height is $h = f(x)$, so the shell area is

$$A(x) = 2\pi x\left(2x^2 - x^3\right) = 2\pi\left(2x^3 - x^4\right).$$

Thus,

$$V = \int_0^2 2\pi x\left(2x^2 - x^3\right)\,dx = \frac{16}{5}\pi.$$

More generally, if a region between two curves $y = u(x)$ and $y = v(x)$ with $u(x) \geq v(x)$ is rotated about the y-axis, then each vertical segment of height $u(x) - v(x)$ gives a shell of area

$$A(x) = 2\pi x\left(u(x) - v(x)\right),$$

and thus,

$$V = \int_a^b 2\pi x\left(u(x) - v(x)\right)\,dx.$$

We shall see why (5.6) holds. Let $f : [a,b] \to \mathbb{R}$ be continuous and non-negative with $a \geq 0$. Consider the region under $y = f(x)$ above the x-axis on $[a,b]$. Rotate this region about the y-axis to form a solid. The region can be viewed as the union of vertical segments from $y = 0$ to $y = f(x)$. Fix x and rotate the vertical line segment at x about the y-axis. The result is the lateral surface of a thin right circular cylinder (a cylindrical shell) of radius $r = x$ and height $h = f(x)$. If the thickness is Δx, then the lateral surface area is approximately

$$A(x)\Delta x \approx 2\pi r h \Delta x = 2\pi x f(x)\Delta x,$$

and the shell volume is this area times the thickness Δx. Thus the total volume is approximated by

$$\sum 2\pi x^* f(x^*)\,\Delta x.$$

Example 5.41. Consider the region bounded between $y = x^2$ and $y = x$ on $[0,1]$.

When rotated about the y-axis, using the shell method, for each $x \in [0,1]$, the vertical segment has height $x - x^2$, so

$$A(x) = 2\pi x\left(x - x^2\right) = 2\pi\left(x^2 - x^3\right).$$

As such,

$$V = \int_0^1 2\pi\left(x^2 - x^3\right)\,dx = \frac{\pi}{6}.$$

On the other hand, using the disc method with respect to y, solving $y = x$ gives $x = y$, and $y = x^2$ gives $x = \sqrt{y}$. The outer radius is $\sqrt{y}$, the inner radius is y, so

$$A(y) = \pi\left((\sqrt{y})^2 - y^2\right) = \pi\left(y - y^2\right),$$

hence,

$$V = \int_0^1 \pi\left(y - y^2\right)\,dy = \frac{\pi}{6},$$

confirming the shell-method result.

Example 5.42 (MA2002 AY21/22 Sem 1). Let R be the region enclosed by the curve defined by $x^4 = 4\left(x^2 - y^2\right)$. Find the volume of the solid formed by rotating R completely about the y-axis. Refer to Example 5.32 for a sketch of the graph.

Solution. By the shell method, the volume is

$$2\int_0^2 2\pi xy\,dx = 4\pi\int_0^2 x\sqrt{x^2 - \frac{1}{4}x^4}\,dx = 2\pi\int_0^2 x^2\sqrt{4 - x^2}\,dx.$$

Using the substitution $x = 2\sin\theta$, we have $dx = 2\cos\theta\,d\theta$, so the integral becomes

$$2\pi\int_0^{\pi/2} 4\sin^2\theta\cdot 2\cos\theta\cdot 2\cos\theta\,d\theta = 32\pi\int_0^{\pi/2}\sin^2\theta\cos^2\theta\,d\theta = 8\pi\int_0^{\pi/2}\sin^2 2\theta\,d\theta.$$

Using the identity $\cos 4\theta = 1 - 2\sin^2 2\theta$, it is easy to show that the volume is $2\pi^2$ units^3. We leave this to the reader. □

Example 5.43 (MA2002 AY23/24 Sem 2). Continuing from Example 5.33, let $f(x) = -x^2 + 4x + 2$ and $g(x) = x^2 - 6x + 10$. Find the volume of the solid generated when R is revolved about the y-axis. See Figure 5.4 for a sketch.

Solution. Using the shell method, the volume is equal to

$$\int_1^4 2\pi x\left(f(x) - g(x)\right)\,dx = 2\pi\int_1^4 x\left(-2x^2 + 10x - 8\right)\,dx$$

which is equal to 45π. □

5.13 Arc Length of a Curve

Let f be a continuous function on $[a,b]$ that is differentiable with continuous derivative f' on (a,b). Consider the graph of $y = f(x)$ from $x = a$ to $x = b$. We shall approximate the curve by a polygonal path. If Δx is small and the tangent is not vertical, then the change in y is approximately

$$\Delta y \approx f'(x)\,\Delta x.$$

The length of a small segment is

$$\Delta L \approx \sqrt{(\Delta x)^2 + (\Delta y)^2} \approx \sqrt{1 + (f'(x))^2}\,\Delta x.$$

Adding and taking limits gives the formula for arc length (Theorem 5.25).

Theorem 5.25 (arc length). Let $f : [a,b] \to \mathbb{R}$ be differentiable with continuous derivative on $[a,b]$. The length of the curve $y = f(x)$, where $a \le x \le b$, is

$$L = \int_a^b \sqrt{1 + (f'(x))^2}\,dx.$$

If a curve is given as $x = g(y)$, with g differentiable and g' continuous on $[c,d]$, then the length of the curve is

$$L = \int_c^d \sqrt{1 + (g'(y))^2}\,dy.$$

Example 5.44. Find the arc length of

$$y = \frac{4\sqrt{2}}{3}x^{3/2} - 1 \quad \text{where } 0 \le x \le 1.$$

Solution. We have

$$\frac{dy}{dx} = 2\sqrt{2}x^{1/2} \quad \text{so} \quad 1 + \left(\frac{dy}{dx}\right)^2 = 1 + 8x,$$

so

$$L = \int_0^1 \sqrt{1+8x}\,dx.$$

Let $u = 1 + 8x$, so $du = 8\,dx$. As such, $dx = \frac{1}{8}\,du$. Then,

$$L = \frac{1}{8}\int_1^9 \sqrt{u}\,du = \frac{13}{6}.$$

□

Example 5.45 (circumference of the unit circle). The unit circle $x^2 + y^2 = 1$ has circumference 2π. We may compute this by considering the upper semicircle

$$y = \sqrt{1 - x^2} \quad \text{where } 0 \le x \le 1,$$

and then multiplying its length by 4 (since the full circle consists of four quarter arcs). We have

$$\frac{dy}{dx} = \frac{-x}{\sqrt{1-x^2}} \quad \text{so} \quad \sqrt{1+\left(\frac{dy}{dx}\right)^2} = \frac{1}{\sqrt{1-x^2}}.$$

The arc length of the quarter circle is

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

Hence, the entire circumference is 2π.

Example 5.46. Find the length of the curve $y = \sqrt{x}$ from $(0,0)$ to $(1,1)$.

Solution. Directly using $y = f(x)$ gives

$$\frac{dy}{dx} = \frac{1}{2\sqrt{x}}$$

which is not continuous at $x = 0$. Instead, we shall parametrise the same curve as $x = g(y) = y^2$ for $0 \leq y \leq 1$. Now,

$$\frac{dx}{dy} = 2y \quad \text{so} \quad 1+\left(\frac{dx}{dy}\right)^2 = 1+4y^2.$$

Thus,

$$L = \int_0^1 \sqrt{1+4y^2}\,dy.$$

Using the substitution $u = 2y$, we have

$$L = \frac{1}{2}\int_0^2 \sqrt{1+u^2}\,du.$$

Using the standard antiderivative

$$\int \sqrt{1+u^2}\,du = \frac{1}{2}u\sqrt{1+u^2} + \frac{1}{2}\ln\left(u+\sqrt{1+u^2}\right) + C,$$

we obtain

$$L = \frac{1}{2}\sqrt{5} + \frac{1}{4}\ln\left(\sqrt{5}+2\right).$$

□

5.14 Surface Area of Revolution

Before deriving the general formula, we recall two geometric facts. First, consider a right circular cone of base radius r and slant height ℓ. Then, it has a lateral surface area of

$$A = \pi r\ell.$$

To see why, if we cut along a line from the vertex to the base and flatten the surface, we obtain a sector of a circle of radius ℓ. The circumference of the base $(2\pi r)$ corresponds to the arc length of the sector, while the full circle of radius ℓ has circumference $2\pi\ell$. The area of the sector is to $\pi\ell^2$ as its arc length is to $2\pi\ell$, hence,

$$\frac{A}{\pi\ell^2} = \frac{2\pi r}{2\pi\ell} = \frac{r}{\ell}$$

so $A = \pi r\ell$. Next, a frustum of a right circular cone with slant height ℓ and base radii $r_1 < r_2$ has lateral surface area

$$A = \pi (r_1 + r_2)\,\ell.$$

Again to see why, the frustum can be obtained by removing a smaller cone from a larger similar cone. If the larger cone has slant height ℓ_2 and base radius r_2, and the smaller cone has slant height ℓ_1 and base radius r_1, then similarity gives

$$\frac{r_1}{\ell_1} = \frac{r_2}{\ell_2},$$

and $\ell = \ell_2 - \ell_1$. Subtract the two lateral surface areas, we have $\pi r_2\ell_2 - \pi r_1\ell_1$, and then we rearrange to obtain $A = \pi (r_1 + r_2)\,\ell$.

We now derive the surface area of revolution formula (Theorem 5.26). Let $f : [a,b] \to \mathbb{R}$ be continuous and non-negative, differentiable with continuous derivative on $[a,b]$. Consider the curve $y = f(x)$, $a \le x \le b$, and rotate it about the x-axis to form a surface of revolution. Approximate the curve by a polygonal path. For a small increment Δx, the curve segment from $(x, f(x))$ to $(x+\Delta x, f(x+\Delta x))$ is approximated by a straight line of length ΔL. Rotating this segment about the x-axis produces a conical frustum whose slant height is ΔL and whose radii are approximately $f(x)$ and $f(x+\Delta x)$. Its lateral surface area is approximately

$$\Delta A \approx \pi\left(f(x) + f(x+\Delta x)\right)\Delta L.$$

Thus,

$$\frac{\Delta A}{\Delta x} \approx \pi\left(f(x) + f(x+\Delta x)\right)\frac{\Delta L}{\Delta x}.$$

Taking the limit as $\Delta x \to 0$, and using the arc-length formula (Theorem 5.25)

$$\frac{dL}{dx} = \sqrt{1 + (f'(x))^2},$$

we obtain

$$\frac{dA}{dx} = 2\pi f(x)\sqrt{1 + (f'(x))^2}.$$

Theorem 5.26 (surface area of revolution). Let f be nonnegative and differentiable with continuous derivative on $[a,b]$. The area of the surface formed by rotating the curve $y = f(x)$, $a \le x \le b$, about the x-axis is

$$S = \int_a^b 2\pi f(x)\sqrt{1 + (f'(x))^2}\,dx. \tag{5.7}$$

In terms of the variable y, if a curve is given by $x = g(y)$ and is rotated about the

y-axis, then the corresponding formula is

$$S = \int_c^d 2\pi g(y) \sqrt{1 + (g'(y))^2}\, dy.$$

Example 5.47 (surface area of the unit sphere). The upper semicircle of the unit circle is given by

$$y = \sqrt{1 - x^2} \quad \text{where } -1 \leq x \leq 1.$$

Rotating this curve about the x-axis produces the unit sphere. In Example 5.45, we already computed

$$\frac{dy}{dx} = \frac{-x}{\sqrt{1-x^2}} \quad \text{so} \quad \sqrt{1 + \left(\frac{dy}{dx}\right)^2} = \frac{1}{\sqrt{1-x^2}}.$$

Thus,

$$2\pi y \sqrt{1 + \left(\frac{dy}{dx}\right)^2} = 2\pi \sqrt{1-x^2} \cdot \frac{1}{\sqrt{1-x^2}} = 2\pi.$$

The surface area is

$$S = \int_{-1}^{1} 2\pi\, dx = 2\pi \cdot 2 = 4\pi.$$

Example 5.48. Consider the curve $y = x^2$ from $(0,0)$ to $(1,1)$ rotated about the y-axis. It is more convenient to write the curve as $x = \sqrt{y}$, where $0 \leq y \leq 1$. Then,

$$\frac{dx}{dy} = \frac{1}{2\sqrt{y}} \quad \text{so} \quad \sqrt{1 + \left(\frac{dx}{dy}\right)^2} = \sqrt{1 + \frac{1}{4y}} = \sqrt{\frac{4y+1}{4y}}.$$

The radius of rotation is $x = \sqrt{y}$, so

$$2\pi x \sqrt{1 + \left(\frac{dx}{dy}\right)^2} = 2\pi \sqrt{y} \cdot \sqrt{\frac{4y+1}{4y}} = \pi \sqrt{4y+1}.$$

Therefore,

$$S = \int_0^1 \pi \sqrt{4y+1}\, dy.$$

Using the substitution $u = 4y + 1$, we obtain

$$S = \frac{\pi}{4} \int_1^5 \sqrt{u}\, du = \frac{\pi}{6}\left(5\sqrt{5} - 1\right).$$

Example 5.49 (MA2002 AY21/22 Sem 1). Let f be a function that is non-negative, increasing and continuous on $[0,2]$, and differentiable on $(0,2)$. A surface S is formed by rotating the curve $y = f(x)$ completely about the x-axis. Suppose that the area of the portion of S on any interval $[a,b] \subseteq [0,2]$ is always $10\pi(b-a)$. If $f(0) = 3$, find the expression for $f(x)$.

Though this is ideally a problem on ordinary differential equations (Chapter 6), it can be solved as we just need to work with (5.7).

Solution. Using (5.7), we have

$$\int_0^x 2\pi f(t)\sqrt{1+(f'(t))^2}\,dt = 10\pi x, \tag{5.8}$$

where $a=0$ and $b=x$ and $[a,b]\subseteq[0,2]$. By the Fundamental Theorem of Calculus (Theorem 5.5), differentiating both sides of (5.8), we have

$$2\pi f(x)\sqrt{1+(f'(x))^2} = 10\pi.$$

The differential equation becomes[2]

$$f(x)\sqrt{1+(f'(x))^2} = 5 \quad \text{so} \quad (f(x))^2+\left(f(x)f'(x)\right)^2 = 25.$$

Using the substitution $u(x)=(f(x))^2$, we have $u'(x)=2f(x)f'(x)$, so the differential equation now becomes

$$u(x)+\frac{1}{4}(u'(x))^2 = 25.$$

Differentiating both sides, we have

$$u'(x)\left(1+\frac{1}{2}u''(x)\right) = 0.$$

As such, either $u'(x)=0$ or $u''(x)=-2$. The former yields $u(x)=c_1$, so $f(x)=\pm\sqrt{c_1}$. Since $f(0)=3$, this solution yields $f(x)=3$ for all $0\le x\le 2$. However, substituting this into (5.7), for $a\le x\le b$, we have

$$\int_a^b 2\pi\cdot 3\,dx = 6\pi(b-a)\neq 10\pi(b-a).$$

As such, we consider the latter of the two solutions, for which it yields $u(x)=-x^2+c_2x+c_3$, so $f(x)=\pm\sqrt{-x^2+c_2x+c_3}$, but since f is non-negative, then $f(x)=\sqrt{-x^2+c_2x+c_3}$. Substituting $f(0)=3$ gives $c_3=9$. Substituting $f(x)=\sqrt{-x^2+c_2x+9}$ into (5.7), for $a\le x\le b$, we have

$$\int_a^b 2\pi\sqrt{-x^2+c_2x+9}\cdot\sqrt{1+\left(\frac{c_2-2x}{2\sqrt{-x^2+c_2x+9}}\right)^2} = 10\pi(b-a).$$

Thus,

$$\begin{aligned}\int_a^b \sqrt{-4x^2+4c_2x+36+c_2^2+4x^2-4c_2x}\,dx &= 10(b-a)\\ \int_a^b \sqrt{36+c_2^2}\,dx &= 10(b-a)\end{aligned}$$

so $c_2=8$. We reject $c_2=-8$. To see why, if $f(x)=\sqrt{-x^2-8x+9}$, then the domain of f does not include values of x for $x>1$. Hence, $f(x)=\sqrt{-x^2+8x+9}$. □

---

[2]There is some semblance to solving this differential equation as compared to Clairaut's equation, for which the latter is given by

$$y(x) = x\frac{dy}{dx}+f\left(\frac{dy}{dx}\right).$$

CHAPTER 6

Ordinary Differential Equations

6.1 Introduction

Let $I \subseteq \mathbb{R}$ be an interval and let $y : I \to \mathbb{R}$ be an unknown function. An ordinary differential equation (ODE) of order n for y is an equation of the form

$$F\left(x, y(x), y'(x), y''(x), \ldots, y^{(n)}(x)\right) = 0,$$

where F is a given function of $n+2$ variables, and $y^{(k)}$ denotes the k–th derivative of y with respect to x. The largest order of derivative of y that appears in the ODE is called the *order* of the ODE. In particular, a *first-order ODE* involves y and y' but no higher derivatives, and can usually be written in the form

$$\frac{dy}{dx} = F(x, y),$$

for some function F of two variables.

Next, let $J \subseteq I$ be an interval. A function $y : J \to \mathbb{R}$ is a *solution* of the ODE on J if y is sufficiently differentiable on J and the ODE is satisfied for every $x \in J$.

A family of solutions depending on one or more arbitrary constants is called a general solution. A solution that cannot be obtained from the general solution by specialising the constants is sometimes called a singular solution.

Example 6.1. Consider the first-order ODE

$$\frac{dy}{dx} = 1 - \sqrt{x} \quad \text{where } x \geq 0.$$

Integrating both sides with respect to x gives

$$y(x) = \int 1 - \sqrt{x}\, dx = x - \frac{2}{3}x^{3/2} + c, \tag{6.1}$$

where $c \in \mathbb{R}$ is an arbitrary constant. This family is the general solution of the ODE on any interval where the right side of (6.1) is defined.

The simplest first-order ODEs are those of the form

$$\frac{dy}{dx} = f(x),$$

where f is a function of x alone. In this case, any antiderivative of f gives a solution. Suppose $f : I \to \mathbb{R}$ is a continuous function. Then, the general solution of

$$\frac{dy}{dx} = f(x)$$

on I is

$$y(x) = \int f(x)\,dx = F(x) + C,$$

where F is any fixed antiderivative of f and $C \in \mathbb{R}$. Similarly, when

$$\frac{dy}{dx} = g(y)$$

depends only on y, we can treat x as a function of y. Say g is continuous and non-zero on an interval J and consider

$$\frac{dy}{dx} = g(y).$$

Then, on any subinterval where $g(y) \neq 0$, the solutions are implicitly given by

$$x = \int \frac{1}{g(y)} dy = G(y) + C.$$

Note that if $y = C_0$ is a constant solution of the algebraic equation $g(y) = 0$, then $y(x) = C_0$ is a solution of $\frac{dy}{dx} = g(y)$ since the derivative vanishes identically. Such constant solutions are usually singular with respect to the family obtained by integrating $\frac{1}{g(y)}$.

Definition 6.1 (initial value problem). A first-order initial value problem (IVP) consists of a first-order ODE

$$\frac{dy}{dx} = F(x,y) \quad \text{together with a condition } y(a) = b,$$

where (a,b) is a specified point in the plane. A function y is a solution of the IVP if it solves the ODE on some interval containing a and satisfies the initial condition.

In many situations, once the general solution $y(x)$ depending on a constant C is known, the initial condition determines the value of C and hence a unique particular solution.

Example 6.2. Consider

$$\frac{dy}{dx} = y \quad \text{where } y(0) = 1.$$

The general solution of $\frac{dy}{dx} = y$ is $y(x) = Ce^x$.[1] Imposing $y(0) = 1$ gives $1 = Ce^0 = C$, so $C = 1$. Hence the unique solution to the IVP is $y(x) = e^x$.

6.2 Separable First-Order Equations

[1] We will discuss methods to solving first-order differential equations in due course.

Definition 6.2 (separable ODE). A first-order ODE

$$\frac{dy}{dx} = F(x,y)$$

is called separable if F can be factorised as $F(x,y) = f(x)\,g(y)$ for suitable single–variable functions f and g.

Assume $g(y) \neq 0$ in a region of interest. Then, the equation

$$\frac{dy}{dx} = f(x)\,g(y)$$

can be rewritten formally as

$$\frac{1}{g(y)}\frac{dy}{dx} = f(x) \quad \text{so} \quad \frac{1}{g(y)}dy = f(x)\,dx.$$

Integrating both sides, we obtain

$$\int \frac{1}{g(y)}dy = \int f(x)\,dx$$

which provides an implicit relation between x and y.

Theorem 6.1 (general solution of separable equations). Let f and g be continuous on intervals, and consider the separable ODE

$$\frac{dy}{dx} = f(x)\,g(y).$$

(i) Any constant $y = C$ satisfying $g(C) = 0$ gives a constant (singular) solution.

(ii) On any region where $g(y) \neq 0$, the general solution is obtained by integrating

$$\int \frac{1}{g(y)}\, dy = \int f(x)\, dx.$$

Example 6.3. Solve

$$2\sqrt{xy}\frac{dy}{dx} = 1 \quad \text{where } x,y > 0.$$

Solution. We have $\sqrt{y}\,dy = \frac{1}{2}\sqrt{x}\,dx$. Integrating both sides yields

$$\int \sqrt{y}\,dy = \int \frac{1}{2}\sqrt{x}\,dx$$

so $\frac{2}{3}y^{3/2} = \frac{1}{3}x^{3/2} + C$. We can make y the subject of the equation. Anyway, this implicitly defines the general solution on the region $x,y > 0$. □

6.3 Homogeneous First-Order Equations

Definition 6.3 (homogeneous function). Let $F : \mathbb{R}^m \to \mathbb{R}$ be defined on a cone (for example $\mathbb{R}^m \setminus \{0\}$). We say that F is homogeneous of degree n if

$$F(tx_1, \ldots, tx_m) = t^n F(x_1, \ldots, x_m)$$

for every $t \in \mathbb{R} \setminus \{0\}$ for which both sides are defined.

Example 6.4. We give some examples of homogeneous functions.

(i) For any non-negative integer n, each monomial $x^{n-i}y^i$ is homogeneous of degree n, and so is any homogeneous polynomial

$$\sum_{i=0}^{n} a_i x^{n-i} y^i.$$

(ii) Any linear function $F(x_1, \ldots, x_m) = a_1x_1 + \cdots + a_mx_m$ is homogeneous of degree 1.

(iii) A function F is homogeneous of degree 0 if $F(tx_1, \ldots, tx_m) = F(x_1, \ldots, x_m)$; in two variables this means that $F(x,y)$ depends only on the ratio y/x.

Definition 6.4 (homogeneous first-order ODE). A first-order ODE

$$\frac{dy}{dx} = F(x,y) \tag{6.2}$$

is called homogeneous if F is homogeneous of degree 0, that is,

$$F(tx,ty) = F(x,y) \quad \text{for all } t \neq 0.$$

Equivalently, $F(x,y)$ can be expressed in the form $F(x,y) = \Phi\left(\frac{y}{x}\right)$ for some single–variable function Φ wherever $x \neq 0$.

For equations of the form (6.2), the substitution $z(x) = \frac{y(x)}{x}$ reduces the ODE to a separable equation in x and z. Indeed, writing $y(x) = xz(x)$, we obtain

$$\frac{dy}{dx} = z(x) + x\frac{dz}{dx},$$

and since $F(x,y) = \Phi\left(\frac{y}{x}\right)$, we have

$$\frac{dy}{dx} = F(x,y) = \Phi\left(\frac{y}{x}\right) = \Phi(z).$$

Hence

$$z(x) + x\frac{dz}{dx} = \Phi(z(x)) \quad \text{or equivalently} \quad x\frac{dz}{dx} = \Phi(z) - z.$$

Provided that $\Phi(z) - z \neq 0$, this differential equation is separable because we can write it as

$$\frac{dz}{\Phi(z) - z} = \frac{dx}{x}.$$

Example 6.5. Solve the homogeneous equation

$$x^2\frac{dy}{dx} = y^2 + 2xy \quad \text{where } x \neq 0.$$

Solution. We write the differential equation as

$$\frac{dy}{dx} = \left(\frac{y}{x}\right)^2 + 2\left(\frac{y}{x}\right).$$

Set $z = y/x$, so $y = xz$ and

$$\frac{dy}{dx} = z + x\frac{dz}{dx} \quad \text{so} \quad z + x\frac{dz}{dx} = z^2 + 2z.$$

Hence,

$$x\frac{dz}{dx} = z^2 + z = z(z+1).$$

Assuming $z \neq 0$ and $z \neq -1$ (these will correspond to singular solutions), we have

$$\frac{dz}{z(z+1)} = \frac{dx}{x}.$$

By partial fraction decomposition, as

$$\frac{1}{z(z+1)} = \frac{1}{z} - \frac{1}{z+1},$$

we have

$$\int\left(\frac{1}{z} - \frac{1}{z+1}\right)dz = \int\frac{dx}{x} \quad \text{so} \quad \ln\left|\frac{z}{z+1}\right| = \ln|x| + C.$$

Exponentiating, we have

$$\left|\frac{z}{z+1}\right| = C|x|.$$

We replace z with y/x and absorb $\pm$ into C. After some algebraic manipulation, we see that the solution to the differential equation can be rewritten in the form

$$y = \frac{x^2}{C-x} \quad \text{with } C \neq 0$$

which is the general solution on suitable intervals. The excluded cases $z \equiv 0$ and $z \equiv -1$ correspond to the singular solutions $y \equiv 0$ and $y \equiv -x$. □

6.4 First-Order Linear Equations

A first-order ODE of the form

$$\frac{dy}{dx} = F(x,y)$$

is called linear if $F(x,y)$ is an affine (hence linear) function of y. That is,

$$\frac{dy}{dx} + p(x)y = q(x),$$

where p and q are given functions of x. When $q \equiv 0$, the differential equation is homogeneous. That is,

$$\frac{dy}{dx} + p(x)\,y = 0.$$

This equation is separable since

$$\frac{1}{y}\frac{dy}{dx} = -p(x),$$

so

$$\int \frac{1}{y}\,dy = -\int p(x)\ dx \quad \Rightarrow \quad \ln|y| = -P(x) + C,$$

where P is an antiderivative of p. Hence, $y(x) = Ce^{-P(x)}$.

For the general linear equation

$$\frac{dy}{dx} + p(x)\,y = q(x), \tag{6.3}$$

we seek a non-zero function $v(x)$ such that the left side becomes the derivative of $v(x)\,y(x)$. That is to say,

$$\frac{d}{dx}(v(x)\,y(x)) = v(x)\frac{dy}{dx} + v'(x)\,y(x).$$

Comparing with

$$v(x)\left(\frac{dy}{dx} + p(x)\,y\right) = v(x)\frac{dy}{dx} + v(x)\,p(x)\,y,$$

we want $v'(x) = v(x)\,p(x)$, i.e. v must solve the homogeneous linear equation. As before, this gives $v(x) = e^{P(x)}$ where $P'(x) = p(x)$. The function v is called an integrating factor. Multiplying the original equation (6.3) by v, we obtain

$$\frac{d}{dx}(v(x)\,y(x)) = v(x)\,q(x).$$

Integrating,

$$v(x)\,y(x) = \int v(x)\,(x)\,dx + C,$$

and thus

$$y(x) = \frac{1}{v(x)}\left(\int v(x)\,q(x)\,dx + C\right).$$

If we chose a different antiderivative $P_1(x) = P(x) + k$, then $v_1(x) = e^{P_1(x)} = e^k v(x)$ is a constant multiple of v. Multiplying the entire equation by a non-zero constant does not change the set of solutions, so any integrating factor of this form gives the same general solution.

Example 6.6. Solve

$$x\frac{dy}{dx} = x^2 + 3y \quad \text{where } x > 0. \tag{6.4}$$

Solution. Rewriting in standard form,

$$\frac{dy}{dx} - \frac{3}{x}y = x.$$

Here $p(x) = -\frac{3}{x}$, so

$$P(x) = \int p(x)\, dx = \int -\frac{3}{x} dx = -3\ln x,$$

and an integrating factor is

$$v(x) = e^{P(x)} = e^{-3\ln x} = x^{-3}.$$

Multiplying (6.4) throughout by x^{-3} gives

$$x^{-3}\frac{dy}{dx} - 3x^{-4}y = x^{-2} \quad \text{so} \quad \frac{d}{dx}\left(x^{-3}y\right) = x^{-2}.$$

Integrating both sides, we have

$$x^{-3}y = \int x^{-2} dx = -x^{-1} + C,$$

so

$$y(x) = x^3\left(-\frac{1}{x} + C\right) = -x^2 + Cx^3.$$

□

Example 6.7 (MA2002 AY21/22 Sem 1). Consider the following initial value problem:

$$\frac{dy}{dx} = 2e^{-x}y^2 + 2y - 3e^x \quad \text{where } y = 0 \text{ at } x = 0$$

(i) Use the substitution $y = e^x + \frac{1}{z}$ to convert the differential equation of the given initial value problem into a first order linear equation in x and z.

(ii) Solve the differential equation obtained in **(i)**. Hence, solve the initial value problem. Express the answer as $y = f(x)$.

Solution.

(i) The substitution yields

$$\frac{dy}{dx} = e^x - z^{-2}\frac{dz}{dx}$$

so the differential equation becomes

$$2e^{-x}\left(e^x + \frac{1}{z}\right)^2 + 2\left(e^x + \frac{1}{z}\right) - 3e^x = e^x - z^{-2}\frac{dz}{dx}$$
$$2e^{-x}\left(e^{2x} + \frac{2e^x}{z} + z^{-2}\right) + 2e^x + 2z^{-1} = 4e^x - z^{-2}\frac{dz}{dx}$$

As such,

$$\frac{dz}{dx} + 6z = -2e^{-x}. \tag{6.5}$$

(ii) The integrating factor is $e^{\int 6\,dx} = e^{6x}$. Hence, multiplying both sides of (6.5) by the integrating factor, we have

$$\frac{d}{dx}\left(ze^{6x}\right) = -2e^{5x} \quad \text{so} \quad ze^{6x} = -\frac{2}{5}e^{5x} + c.$$

When $x = y = 0$, then $z = -1$, so $c = 0.6$. Hence,

$$y = -\frac{3e^{x}(e^{5x}-1)}{2e^{5x}+3}$$

is the solution to the initial value problem. ☐

Sometimes, an equation is not linear in y, but is linear in x when we view x as a function of y. For example, suppose we can write the equation in the form

$$\frac{dx}{dy} + \widetilde{p}(y)\,x = \widetilde{q}(y)\,.$$

Then, the usual integrating factor method (now with independent variable y) applies. This is especially useful when the original equation is of the form

$$H(x,y)\,\frac{dy}{dx} = G(x,y)\,,$$

and H and G are such that solving for $\frac{dx}{dy}$ yields a linear equation in x.

6.5 Bernoulli's Equation

Definition 6.5 (Bernoulli equation). A Bernoulli equation is a first-order ODE of the form

$$\frac{dy}{dx} + p(x)\,y = q(x)\,y^{n}, \tag{6.6}$$

where p and q are given functions of x, and $n \in \mathbb{R}$ is a given constant.

If $n = 0$ or $n = 1$, the equation is already linear or at least separable (see our discussions in Chapters 6.3 and 6.4). The interesting case is when $n \neq 0, 1$. Assume $n \neq 0, 1$ and $y(x) \neq 0$ on an interval. Introduce the change of variables

$$z(x) = y(x)^{1-n}\,.$$

Then, by the chain rule (Theorem 4.6), we have

$$\frac{dz}{dx} = (1-n)\,y^{-n}\frac{dy}{dx}.$$

Multiplying the Bernoulli equation (6.6) by $(1-n)\,y^{-n}$, we obtain

$$(1-n)\,y^{-n}\frac{dy}{dx} + (1-n)\,p(x)\,y^{1-n} = (1-n)\,q(x)\,,$$

i.e.

$$\frac{dz}{dx} + (1-n)\,p(x)\,z = (1-n)\,q(x)\,.$$

This is a first-order linear ODE in z. After solving for z, we recover y from $y^{1-n} = z$.

Example 6.8. Solve
$$x\frac{dy}{dx}+y=x^4y^3 \quad \text{where } x>0.$$
Rewriting,
$$\frac{dy}{dx}+\frac{1}{x}y=x^3y^3.$$
This is a Bernoulli equation with $n=3$. Let
$$z=y^{1-3}=y^{-2} \quad \text{so} \quad \frac{dz}{dx}=-2y^{-3}\frac{dy}{dx}.$$
Multiplying the original equation by $-2y^{-3}$ yields
$$-2y^{-3}\frac{dy}{dx}-\frac{2}{x}y^{-2}=-2x^3 \quad \text{so} \quad \frac{dz}{dx}-\frac{2}{x}z=-2x^3.$$
This becomes a linear equation in z. Here $p(x)=-\frac{2}{x}$, so
$$P(x)=\int -\frac{2}{x}dx=-2\ln x \quad \text{and} \quad v(x)=e^{P(x)}=x^{-2}.$$
Then
$$\frac{d}{dx}\left(x^{-2}z\right)=x^{-2}\left(-2x^3\right)=-2x.$$
Integrating both sides yields $x^{-2}z=-x^2+C$ so $z(x)=-x^4+Cx^2$. Recall that $z=y^{-2}$, hence
$$y^{-2}=Cx^2-x^4.$$
On intervals where the right–hand side is non-zero, this determines $y(x)$.

6.6 Modelling with First-Order ODEs

We now give several important applications of first-order ODEs in modelling time–dependent phenomena. We begin with discussing exponential growth and decay. Suppose $y(t)$ denotes the size of a quantity at time t (for instance, a population, the amount of substance, or the value of an investment). A basic assumption is that the rate of change is proportional to the current size. That is,
$$\frac{dy}{dt}=ky,$$
where k is a constant. When $k>0$ we have natural growth, when $k<0$ we have natural decay. The ODE is separable, so
$$\frac{1}{y}\frac{dy}{dt}=k \quad \Rightarrow \quad \int \frac{1}{y}dy=\int k\,dt,$$
so $\ln|y|=kt+C$ and hence, $y(t)=Ce^{kt}$. If $y(0)=y_0$ is prescribed, then
$$y(t)=y_0e^{kt}.$$
We then discuss compound interest. Say an amount $A_0>0$ is invested at an interest rate $r>0$ per year. If interest is compounded once per year, after t years, the value is
$$A(t)=A_0(1+r)^t.$$

If interest is compounded n times per year, then the interest rate per compounding period is r/n, and there are nt periods in t years; thus

$$A_n(t) = A_0\left(1+\frac{r}{n}\right)^{nt}.$$

The interest is said to be compounded continuously if we define

$$A(t) = \lim_{n\to\infty} A_0\left(1+\frac{r}{n}\right)^{nt},$$

provided the limit exists. Using the standard limit

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

as in Theorem 5.13, we obtain $A(t) = A_0 e^{rt}$. Thus, continuous compounding leads to the same exponential law as the ODE

$$\frac{dA}{dt} = rA.$$

We then discuss radioactive decay and half-life. In radioactive decay, the rate at which a substance decays is proportional to the amount remaining. Let $m(t)$ denote the mass of the radioactive substance at time t. Then,

$$\frac{dm}{dt} = km,$$

with $k < 0$. The solution is

$$m(t) = m(0)\,e^{kt}.$$

The half-life $t_{1/2}$ of the substance is the time required for half of the initial mass to decay. That is, $m\left(t_{1/2}\right) = \frac{1}{2}m(0)$. Substituting this into $m(t) = m(0)\,e^{kt}$ gives

$$\frac{1}{2}m(0) = m(0)\,e^{kt_{1/2}} \quad \text{so} \quad e^{kt_{1/2}} = \frac{1}{2} \quad \text{so} \quad k = -\frac{\ln 2}{t_{1/2}}.$$

Thus,

$$m(t) = m(0)\exp\left(-\frac{\ln 2}{t_{1/2}}t\right).$$

Example 6.9 (radiocarbon dating). Carbon-14 has half–life approximately 5730 years. A sample in which 10% of the original carbon-14 has decayed still contains 90% of the original mass, so $m(t) = 0.9m(0)$. Writing

$$m(t) = m(0)\exp\left(-\frac{\ln 2}{5730}t\right),$$

we solve

$$0.9 = \exp\left(-\frac{\ln 2}{5730}t\right)$$

and obtain $t \approx 871$ years.

We now discuss logistic population growth. First, note that the pure exponential model, also known as the Malthusian model, denoted by

$$\frac{dP}{dt} = kP,$$

assumes unlimited resources, which is unrealistic for large populations. A more realistic model, known as Verhulst's model, assumes a carrying capacity $M > 0$ and that the growth rate decreases linearly as P approaches M. Let $P(t)$ be the population at time t. Suppose

$$\frac{dP}{dt} = r(M-P)P \quad \text{where } r > 0,$$

so that when P is small (relative to M), the equation behaves like $\frac{dP}{dt} \approx rMP$ (exponential growth), while as $P \to M$ the growth rate tends to 0. This is the logistic equation. It is an example of a Bernoulli equation because it can be written as

$$\frac{dP}{dt} - rMP = -rP^2.$$

Let $z = P^{-1}$. Then,

$$\frac{dz}{dt} = -P^{-2}\frac{dP}{dt}.$$

Multiplying the logistic equation by $-P^{-2}$ gives

$$-P^{-2}\frac{dP}{dt} + rMP^{-1} = r \quad \text{so} \quad \frac{dz}{dt} + rMz = r.$$

This equation is linear in z. The integrating factor is e^{rMt}, so

$$\frac{d}{dt}\left(e^{rMt}z\right) = re^{rMt},$$

and therefore,

$$e^{rMt}z = \int re^{rMt}\,dt + C = \frac{r}{rM}e^{rMt} + C = \frac{1}{M}e^{rMt} + C.$$

Thus,

$$z(t) = \frac{1}{M} + Ce^{-rMt},$$

and hence,

$$P(t) = \frac{1}{z(t)} = \frac{M}{1+CMe^{-rMt}}.$$

Renaming the constant CM as C, we usually write

$$P(t) = \frac{M}{1+Ce^{-rMt}}.$$

We observe the following properties. If $P(0) = P_0$, then $C = \frac{M}{P_0} - 1$. Next, as $t \to \infty$, $P(t) \to M$; as $t \to -\infty$, $P(t) \to 0$. Lastly, if $0 < P_0 < M$, then $C > 0$, P is strictly increasing, and the graph has an inflection point when $P(t) = M/2$.

We then introduce Newton's law of cooling, which states that the rate of change of the temperature of an object is proportional to the difference between its temperature and that of the surrounding medium. Let $T(t)$ be the temperature of the object at time t, and let T_S be the constant surrounding temperature. Then,

$$\frac{dT}{dt} = -r(T - T_S) \quad \text{where } r > 0.$$

Set $A(t) = T(t) - T_S$. Then,

$$\frac{dA}{dt} = -rA \quad \text{so} \quad A(t) = A(0)e^{-rt}.$$

Thus,

$$T(t) = T_S + (T(0) - T_S)e^{-rt}.$$

Example 6.10. A boiled egg at 98°C is placed in water at 18°C. After 5 minutes, its temperature is 38°C. Assuming the water temperature remains constant at 18°C, how long in total will it take for the egg to cool to 20°C?

Solution. Here $T_S = 18$ and $T(0) = 98$. Hence,

$$T(t) = 18 + 80e^{-rt}.$$

The condition $T(5) = 38$ gives $38 = 18 + 80e^{-5r}$ so $r = \frac{1}{5}\ln 4$. To find the time t such that $T(t) = 20$, we solve the equation $20 = 18 + 80e^{-rt}$ so $t \approx 13$ minutes. Since 5 minutes have already elapsed, it takes a further $13 - 5 \approx 8$ minutes for the egg to reach 20°C. □

Lastly, we discuss the draining tank problem and a variant of it known as the mixing problem. First, consider a cylindrical tank with cross–sectional area A_T (constant) and water depth $h(t)$ at time t. Suppose there is a sharp-edged hole of area A_H at the bottom. According to Torricelli's law, the speed v of the water exiting the hole is

$$v = \sqrt{2gh},$$

where g is the acceleration due to gravity. The outflow volume rate is

$$\frac{dV}{dt} = -A_H v = -A_H\sqrt{2gh}.$$

On the other hand, $V(t) = A_T h(t)$, so

$$\frac{dV}{dt} = A_T\frac{dh}{dt}.$$

Equating the two expressions for $\frac{dV}{dt}$ gives

$$A_T\frac{dh}{dt} = -A_H\sqrt{2gh},$$

or

$$\frac{dh}{dt} = -k\sqrt{h} \quad \text{so} \quad k = \frac{A_H\sqrt{2g}}{A_T} > 0.$$

This is a separable differential equation, for which we obtain $2\sqrt{h} = -kt + C$. If $h(0) = h_0$, then $2\sqrt{h_0} = C$, so

$$h(t) = \left(\sqrt{h_0} - \frac{k}{2}t\right)^2.$$

The tank empties when $h(t) = 0$, that is, when $t = \frac{2\sqrt{h_0}}{k}$.

Example 6.11. A right circular cylindrical tank of radius 1m and height 4m is full. The water drains through a small hole at the bottom at a rate

$$\frac{dV}{dt} = -0.1\sqrt{h}\,\mathrm{m}^3/\mathrm{min},$$

where h is measured in metres. The volume is $V = \pi h$ since the radius is 1, so

$$\pi\frac{dh}{dt} = -0.1\sqrt{h},$$

that is,

$$\frac{dh}{dt} = -\frac{0.1}{\pi}\sqrt{h} = -\frac{1}{10\pi}\sqrt{h}.$$

Separating variables,

$$\frac{dh}{\sqrt{h}} = -\frac{1}{10\pi}dt.$$

Integrating,

$$2\sqrt{h} = -\frac{t}{10\pi} + C.$$

If $h(0) = 4$, then $2\sqrt{4} = 4 = C$. Hence

$$h(t) = \left(2 - \frac{t}{20\pi}\right)^2.$$

The tank is empty when $h(t) = 0$, i.e. $t = 40\pi$ minutes.

Mixing (tank) problems model how the amount of a dissolved substance (e.g. salt) changes in a well-stirred tank while fluid flows in and out. First, we make some standard assumptions as follows. First, the contents of the tank are perfectly mixed at every time t. Hence, the concentration in the outflow equals the concentration in the tank. Next, the inflow and outflow volumetric rates are known functions of time. Lastly, the inflow solute concentration is known.

Let $m(t)$ denote the mass of solute in the tank at time t and $V(t)$ denote the volume of solution in the tank at time t. Next, let $r_{\text{in}}(t)$ (volume/time) and $r_{\text{out}}(t)$ (volume/time) denote the inflow and outflow rates respectively, and $c_{\text{in}}(t)$ (mass/volume) denote the solute concentration in the inflow. Then, the fundamental modelling principle is

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

If the inflow concentration is $c_{\text{in}}(t)$, then the solute enters at

$$\text{rate in}(t) = r_{\text{in}}(t)\,c_{\text{in}}(t).$$

At time t, the concentration in the tank is

$$c_{\text{tank}}(t) = \frac{m(t)}{V(t)}.$$

Because the tank is well-stirred, the outflow concentration equals $c_{\text{tank}}(t)$. Hence,

$$\text{rate out}(t) = r_{\text{out}}(t)\cdot\frac{m(t)}{V(t)}.$$

Substituting the expressions for the two rates gives the standard mixing ordinary differential equation, which states that

$$m'(t) = r_{\text{in}}(t)\,c_{\text{in}}(t) - r_{\text{out}}(t)\,\frac{m(t)}{V(t)}.$$

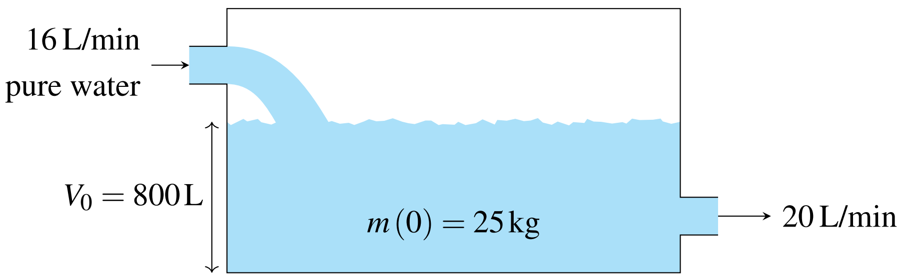
Note that the volume changes according to net flow. That is,

$$V'(t) = r_{\text{in}}(t) - r_{\text{out}}(t) \quad \text{with initial condition } V(0) = V_0.$$

In particular, if r_{in} and r_{out} are constants, then

$$V(t) = V_0 + (r_{\text{in}} - r_{\text{out}})\,t.$$

Example 6.12 (MA2002 AY23/24 Sem 2). Pure water flows into a tank at rate of 16 L/min, and the stirred mixture flows out of the tank at a rate of 20 L/min. The tank initially holds 800 litres of solution containing 25 kg of salt. When will the tank have exactly 5 kg of salt? Give your answer to the nearest integer in the unit of minute.



Solution. Let the mass of salt at time t be $m(t)$. Then,

$$m'(t) = -20 \cdot \frac{m(t)}{800-4t} \quad \text{so} \quad \frac{m'(t)}{m(t)} = -\frac{20}{800-4t}.$$

Integrating both sides yields

$$\ln m = 5\ln|t-200| + c.$$

When $t=0$, we have $m=25$, so $c = \ln 25 - 5\ln 200$. As such, when $m=5$, we have

$$\ln 5 = 5\ln|t-200| + \ln 25 - 5\ln 200.$$

Solving and taking the minimum of the t values, we have $t \approx 55$ minutes. □

Bibliography

[1] Stewart, J. (2016). *Calculus*. 8th ed. Cengage Learning. ISBN: 9781285740621.

[2] Munkres, J. *Topology*. 2nd ed. Pearson, Upper Saddle River, New Jersey, 2000. ISBN: 9780131816299.

[3] Johnsonbaugh, R. and Pfaffenberger, W. E. *Foundations of Mathematical Analysis*. Dover Publications, Mineola, New York, 2010. ISBN: 9780486477664.

[4] Rudin, W. *Principles of Mathematical Analysis*. 3rd ed. McGraw-Hill, New York, 1976. ISBN: 9780070542358.