Introduction to Calculus

Prison Mathematics Project

Introduction

Hello and welcome to the module on Introduction to Calculus! What follows is a module intended to support the reader in learning this fascinating topic. The Prison Mathematics Project (PMP) realises that you may be practising mathematics in an environment that is highly restrictive, so this text can both be used independently and does not require a calculator.

What is Calculus?

Calculus is a hugely important area of mathematics and the natural sciences. It was originally named "the calculus of infinitesimals", which is why some people refer to it as **the** calculus even today. There are two main subjects within this area: differential calculus and integral calculus. The two main pioneers of the subject are Isaac Newton and Gottfried Wilhelm Leibniz; they independently developed the subject throughout the 1600s. In particular, Newton was the first to apply calculus to physics and Leibniz formalised the topic and developed much of the notation we now use. Many more mathematicians have contributed to this rich topic.

Learning in this Module

The best way to learn mathematics is to do mathematics. Indeed, education isn't something that happens more than it is something we should all participate in. You will find various exercise questions and worked examples in these notes so that you may try to solve problems and deepen your understanding of this topic. Although the aim is for everything to only require the content of this module, you are encouraged to use any other sources you have at your disposal.

Acknowledgements

These notes are based on a lecture course by P. Walker at the University of Leeds.

Contents

1 Preliminaries 3

2 Basic Differentiation 4

3 Basic Integration 23

4 Multivariable Calculus 34

5 Matrix Calculus 49

6 Optimisation Problems 54

7 Exercise Solutions 55

1 Preliminaries

2 Basic Differentiation

We begin with an elementary introduction to differentiation, using the notion of a function, introduced in Section ??, and develop some intuition on limits. We hold off to Chapter ?? before introducing a rigorous definition of a limit of a function.

Limits and Continuity

Definition 2.1 Let $D \subseteq \mathbb{R}$ be some subset and $f : D \to \mathbb{R}$ be a function (recall we call D the domain of f). Suppose that $f(x)$ is defined for all $x \in D$ that is "*arbitrarily close*" to some number $a \in \mathbb{R}$; we do not require that $f(a)$ is defined. If f is "*arbitrarily close*" to some number $L \in \mathbb{R}$ whilst x is "*sufficiently close*" but **not** equal to a, then we call L the limit of f as x approaches a. We denote this by $\lim_{x\to a} f(x) = L$.

Note: In Definition 2.1, we used the following phrases: "*arbitrarily close*" and "*sufficiently close*". These are not at all mathematically rigorous; this is what we rectify in Chapter ??.

Example 2.2 Consider the function $f : [0, 2] \to \mathbb{R}$ defined by

$$f(x) = \begin{cases} 3 & \text{if } x = 1 \\ -5 & \text{if } x \neq 1 \end{cases}.$$

It is clear that $\lim_{x\to 1} f(x) = -5$. Indeed, according to Definition 2.1, we consider $x \in [0, 2]$ close to the value 1 but **not** equal to it, that is $x \neq 1$. But then f takes the value -5 for every $x \neq 1$, so this must be the limit.

Remark 2.3 The point of Example 2.2 is to show there is no expectation that $\lim_{x\to a} f(x) = f(a)$ in general. In fact, we shall soon define a *continuous* function to be precisely a function with the special property that this equation is always true (Chapter ?? uses this as a definition; Chapter ?? will prove this from a more rigorous setup).

Definition 2.4 Let $f : D \to \mathbb{R}$ be a function defined on a subset $D \subseteq \mathbb{R}$.

(i) A right-hand limit is a limit where $x \to a$ from above, denoted $\lim_{x\to a^+} f(x)$.

(ii) A left-hand limit is a limit where $x \to a$ from below, denoted $\lim_{x\to a^-} f(x)$.

Proposition 2.5 *For $f : D \to \mathbb{R}$ a function defined on $D \subseteq \mathbb{R}$, $\lim_{x\to a} f(x)$ exists if and only if the right-hand limit and left-hand limit exist where $\lim_{x\to a^+} f(x) = \lim_{x\to a^-} f(x)$.*

Proof: Deferred. □

Exercise 1 Determine if $\lim_{x\to 5} f(x)$ exists by considering the right-hand and left-hand limits of $f(x)$, where the function in question $f : \mathbb{R} \to \mathbb{R}$ is defined by

$$f(x) = \begin{cases} -1 & \text{if } x \leq 5 \\ 2 & \text{if } x > 5 \end{cases}.$$

Theorem 2.6 (Algebra of Limits) *Suppose f and g are functions where $\lim_{x\to a} f(x) = L$ and $\lim_{x\to a} g(x) = K$ exist. Then, the following are true:*

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

(ii) $\lim_{x\to a} \big(f(x)g(x)\big) = LK$.

(iii) $\lim_{x\to a} \big(f(x)/g(x)\big) = L/K$ **if** $g(x) \neq 0$ *for any* x *and* $K \neq 0$.

(iv) $\lim_{x\to a} \lambda f(x) = \lambda L$, *for any* $\lambda \in \mathbb{R}$.

(v) $\lim_{x\to a} f(x)^n = L^n$ *for any* $n \in \mathbb{N}$.

Proof: Deferred. □

Example 2.7 Let's use the Algebra of Limits to compute the following:

$$\lim_{x\to 3} f(x), \qquad f(x) = \frac{x^2 - 9}{x - 3}$$

First, the numerator is the difference of two squares, so it can be factorised as $(x - 3)(x + 3)$. Thankfully, we see that $f(x) = x + 3$; if we didn't realise this, on may begin to panic that the denominator was approaching zero. Therefore, using the Algebra of Limits(i), with $g(x) \equiv 3$ being the constant function, we see that $\lim_{x\to 3} f(x) = 3 + 3 = 6$.

Exercise 2 Compute the following by appealing to the Algebra of Limits:

(i) $\lim_{x \to 2} \frac{x^2 + 3x - 4}{x^3 - 7}$.

(ii) $\lim_{x \to 6} (x + 2)^4$.

(iii) $\lim_{x \to \frac{3}{4}} (x - \frac{3}{4}) \frac{\sin(x) + \tanh\big(x^2 + 3\big) \cos(6)}{x^{-8} + e^{x \tan(x)} - 2}$.

Definition 2.8 Let $D \subseteq \mathbb{R}$ be some subset and $f : D \to \mathbb{R}$ be a function. If f grows "*arbitrarily large*" as $x \to a \in \mathbb{R}$, then we say one of the following two things:

(i) f diverges to positive infinity if it grows large in the positive direction.

(ii) f diverges to negative infinity if it grows large in the negative direction.

These are denoted $\lim_{x \to a} f(x) = \infty$ and $\lim_{x \to a} f(x) = -\infty$, respectively. In either of these cases, we call the line $x = a$ a vertical asymptote of f.

Example 2.9 We will determine the vertical asymptotes of the function $f : \mathbb{R} \to \mathbb{R}$ given by

$$f(x) = \frac{x - 4}{x^2 + 5x}.$$

Indeed, we see that f diverges to infinity as the denominator grows "*arbitrarily small*", so the vertical asymptotes occur when $x^2 + 5x = 0$, that is at $x = 0$ and $x = -5$.

Definition 2.10 Let $f : D \to \mathbb{R}$ be a function defined on a subset $D \subseteq \mathbb{R}$. We say that f is continuous at $a \in D$ if $\lim_{x \to a} f(x) = f(a)$. We then call f continuous if it is continuous at every $a \in D$.

Example 2.11 Here are some examples of continuous functions.

(i) Every polynomial function is continuous.

(ii) Every rational function is continuous everywhere where the denominator is non-zero.

(iii) The function $f : \mathbb{R} \setminus \{0\} \to \mathbb{R}$ given by $f(x) = 1/x$ **is** continuous.

Note: One of the most important features of continuous functions is the intermediate value property; this essentially says that any continuous function defined on an interval can be drawn without taking your pen off the page.

Differentiable Functions

Definition 2.12 Let $f : D \to \mathbb{R}$ be a function defined on a subset $D \subseteq \mathbb{R}$. We say that f is differentiable at $a \in D$ if the following limit exists:

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

We then call f differentiable if the derivative exists at every $a \in D$.

Note: Equivalently, we can define the limit in Definition 2.12 to be

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

This is more practical to use when proving things in this particular section.

Example 2.13 We will prove that the function $g : \mathbb{R} \to \mathbb{R}$ given by $g(x) = x$ is differentiable. To that end, we will use the above note; let $a \in \mathbb{R}$. Consequently,

$$\begin{aligned} \lim_{h \to 0} \frac{g(a+h) - g(a)}{h} &= \lim_{h \to 0} \frac{a + h - a}{h} \\ &= \lim_{h \to 0} \frac{h}{h} \\ &= \lim_{h \to 0} 1 \\ &= 1. \end{aligned}$$

We have proven that $g'(a) = 1$ for any $a \in \mathbb{R}$.

Exercise 3 Prove (using the note) that $h : \mathbb{R} \to \mathbb{R}$ given by $h(x) = x^2$ is differentiable. [**Hint:** For any $a \in \mathbb{R}$, we should see that $h'(a) = 2a$.]

Proposition 2.14 *If a function f is differentiable at a, then is is continuous at a.*

Proof: We prove this rigorously later. However, we can give a proof now using (the non-rigorous) Definitions 2.10 and 2.12. By assumption, we know that the limit

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

exists. But now, we can write $f(x)$ somewhat perversely as

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

for all $x \neq a$. Taking the limit as $x \to a$, we see that

$$\begin{aligned}
\lim_{x \to a} f(x) &= \lim_{x \to a} \left(\frac{f(x) - f(a)}{x - a}(x - a) + f(a) \right) \\
&= \lim_{x \to a} \frac{f(x) - f(a)}{x - a} \lim_{x \to a} (x - a) + \lim_{x \to a} f(a) \\
&= f'(a) \cdot 0 + f(a) \\
&= f(a),
\end{aligned}$$

by the Algebra of Limits. Hence, since $\lim_{x \to a} f(x) = f(a)$, we see that f is continuous at a. □

Note: The converse to Proposition 2.14 is **not** true; continuity does **not** imply differentiability. For example, the function $f(x) = \sqrt{x^2}$ is continuous (everywhere, specifically at zero) but it is **not** differentiable at zero.

Exercise 4 Prove that $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = k$ for some fixed $k \in \mathbb{R}$ is differentiable.

We will now state a useful result which tells us how to apply differentiation to numerous differentiable functions that are collected together, either by addition or multiplication or composition.

Proposition 2.15 *Let f and g be differentiable functions. Then, the following are true:*

(i) $(f + g)'(x) = f'(x) + g'(x)$. (Sum Rule)

(ii) $(fg)'(x) = f'(x)g(x) + f(x)g'(x)$. (Product Rule)

(iii) $(f/g)'(x) = \left(f'(x)g(x) - f(x)g'(x)\right)/g(x)^2$ **if** $g(x) \neq 0$. (Quotient Rule)

(iv) $(f \circ g)'(x) = f'(g(x))g'(x)$. (Chain Rule)

Proof: Deferred. □

Remark 2.16 We can prove the Quotient Rule using the Product Rule and Chain Rule. Indeed, we can write $f/g = f \cdot 1/g$, so the problem of differentiating the left-hand side is reduced to differentiating the product of the functions f and $1/g$. Finally, we can differentiate $1/g$ by treating it as the composition of $1/x$ and g; this is where the Chain Rule applies.

> **Exercise 5** Prove that $f : \mathbb{R} \to \mathbb{R}$ where $f(x) = x^n$ is differentiable. You may use this:
>
> $$(a+b)^n = \sum_{k=0}^{n} \frac{n!}{k!(n-k)!} a^k b^{n-k},$$
>
> where $n! := n(n-1)(n-2)\cdots 1$ is the product of positive integers from 1 to n.
> [**Hint:** Proceed similarly to Exercise 3 and expand $f(x+h)$ using the above formula.]

Example 2.17 Here, we will use the rules of differentiation to compute some derivatives.

(i) Any polynomial is differentiable. Indeed, from Exercise 13, we know that powers of x are differentiable. Thus, the Sum Rule implies that sums of powers of x are differentiable; this is all a polynomial is. Explicitly, for a polynomial $p(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$, we know that its derivative is $p'(x) = na_n x^{n-1} + (n-1)a_{n-1}x^{n-2} + \cdots + a_1$.

(ii) Consider $f(x) = x\sin(x)$. Using $\sin'(x) = \cos(x)$ proven in Lemma 2.31 later, we can apply the Product Rule to conclude that $f'(x) = 1 \cdot \sin(x) + x \cdot \cos(x) = \sin(x) + x\cos(x)$.

(iii) Consider $g(x) = \cos\left(x^2\right)$. Using $\cos'(x) = -\sin(x)$ proven in Lemma 2.31 later, we can use the Chain Rule and Exercise 3 to conclude that $g'(x) = \cos'(x^2) \cdot 2x = -2x\sin\left(x^2\right)$.

> **Exercise 6** Find the derivative of $h(x) = x^2\cos\left(x^3\right)$ using Proposition 2.15.

All derivatives discussed thus far are so-called **first-order derivatives**, that is where we differentiate the function once. However, some classes of function (e.g. polynomials) are such that their derivatives are again differentiable, so we can take the second, third, fourth, etc. derivative.

> **Definition 2.18** Let f be a function where taking higher derivatives makes sense.
> (i) The second derivative of f is the function $f''(x) = (f')'(x)$.
> (ii) The third derivative of f is the function $f'''(x) = (f'')'(x)$.
> (iii) The k^{th} derivative of f is the function $f^{(k)}(x) = (f^{(k-1)})'(x)$.

Remark 2.19 Let $D \subseteq \mathbb{R}$ and $f : D \to \mathbb{R}$ be differentiable (and continuous, by Proposition 2.14). We call f *continuously differentiable* if its derivative $f' : D \to \mathbb{R}$ is continuous. Moreover, if the first k derivatives $f', f'', ..., f^{(k)}$ all exist and are continuous, we say that it is *k-times continuously differentiable*. Alternatively, we call f of *class* $C^k(D)$. If **all** derivatives exist, that is for every $k \in \mathbb{Z}^+$, then we call it *smooth*, or of *class* $C^\infty(D)$. We revisit this in Section **??**.

Note: The continuous functions on the domain $D \subseteq \mathbb{R}$ are said to be of *class* $C^0(D)$; this is such that it agrees with the notation introduced in Remark 2.19 above.

Exercise 7 Give one example of each of the following:

(i) A smooth function.

(ii) A class C^0 function which is **not** of class C^1 (or higher).

(iii) A class C^1 function which is **not** of class C^2 (or higher).

Definition 2.20 Let $D \subseteq \mathbb{R}$ and $f : D \to \mathbb{R}$ be differentiable. We call a a stationary point if $f'(a) = 0$. For a stationary point $a \in D$, we call it a local maximum if $f''(a) < 0$ and a local minimum if $f''(a) > 0$.

Example 2.21 Suppose we have the function $f(x) = 3x^2 - 2x + 74$. We can determine the stationary points by appealing to Example 2.17(i). Indeed, $f'(x) = 6x - 2$, from which we see that $f'(x) = 0$ if and only if $x = 1/3$. We have therefore found the only stationary point. To classify it, notice that $f''(x) = 6 > 0$, in particular $f''(1/3) > 0$, meaning it is a local minimum.

Note: There is a subtlety; if a is a stationary point of the function f and we see that $f''(a) = 0$, then we will have to do further work to determine what type of stationary point it is. Indeed, if $f''(a) = 0$ but $f'''(a) \neq 0$, then it is called an inflexion point.

We now state some theorems of differential calculus; we prove some using the non-rigorous definitions developed thus far but the true proofs will be left until Sections **??** and **??**.

Theorem 2.22 (Extreme Value Theorem) *Let $f : [a, b] \to \mathbb{R}$ be continuous. Then, there exist $c, d \in [a, b]$ such that $f(c) \leq f(x) \leq f(d)$ for all $x \in [a, b]$.*

Proof: Deferred. □

Remark 2.23 In words, Theorem 2.22 says that any continuous function whose domain is an interval, which we call $[a, b]$, is bounded and will always attain its maximum and minimum values. Indeed, we see it is bounded because the image $f(x)$ lies between the numbers $f(c)$ – the lower bound – and $f(d)$ – the upper bound. Moreover, it **attains** these bounds (the inequality in the statement is not strict) because the maximum and minimum can be outputted from the function by choosing suitable input, which we call d and c respectively.

Theorem 2.24 (Rolle's Theorem) *Let $f : [a, b] \to \mathbb{R}$ be continuous and differentiable on the open interval (a, b) with $f(a) = f(b)$. Then, there exists $c \in (a, b)$ such that $f'(c) = 0$.*

Proof: Because f is continuous on a closed interval, the Extreme Value Theorem applies, that is it attains its maximum and minimum values. There are now two cases to consider.

(i) If f attains its maximum and minimum values on the endpoints of $[a, b]$, that is $f(a)$ and $f(b)$, because we assume these are equal, it follows that $f'(x) = 0$ for every $x \in (a, b)$. Thus, choosing any element of this open interval will give us the necessary c.

(ii) If f does **not** attain its maximum and minimum values on the endpoints of $[a, b]$, then they will each occur at some *interior point*, that is in the open interval (a, b). Thus, f will have a so-called *local maximum* or *local minimum* at some $c \in (a, b)$, wherein $f'(c) = 0$. □

Example 2.25 Consider the function $f : [-r, r] \to \mathbb{R}$ given by $f(x) = \sqrt{r^2 - x^2}$, where $r > 0$. The graph of this function is the upper semi-circle centred at the origin. This is continuous on $[-r, r]$ and differentiable on $(-r, r)$; it is important this is open (excluding the endpoints) because it turns out f is **not** differentiable at $x = \pm r$. By Rolle's Theorem, there is a point $c \in (-r, r)$ such that $f'(c) = 0$. In fact, it turns out that $c = 0$.

Exercise 8 Explain if Rolle's Theorem applies to $f : [-2, 2] \to \mathbb{R}$ given by $f(x) = \sqrt{x^2}$. [**Hint:** Look at the note just above Theorem 2.22.]

Theorem 2.26 (Mean Value Theorem) *Let $f : [a, b] \to \mathbb{R}$ be continuous and differentiable on the open interval (a, b). Then, there exists $c \in (a, b)$ such that*

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

Proof: Define the function $g : [a, b] \to \mathbb{R}$ by the formula

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

We see that g is also continuous on $[a, b]$ and differentiable on (a, b). Moreover, $g(a) = f(a)$ and $g(b) = f(a)$, so the values of g at the endpoints coincides. Thus, we can apply Rolle's Theorem to g, giving the existence of $c \in (a, b)$ such that $g'(c) = 0$. However, we see that

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

Because $g'(c) = 0$, substituting $x = c$ into the above will yield

$$0 = f'(c) - \frac{f(b) - f(a)}{b - a},$$

which rearranges to give the result. □

Remark 2.27 Consider the graph $y = f(x)$ of a differentiable function $f : \mathbb{R} \to \mathbb{R}$. For each pair of numbers (a, b) where $a < b$, we can construct the chord (line segment) between the points $(a, f(a))$ and $(b, f(b))$. The Mean Value Theorem asserts that at some point $(c, f(c))$ on the graph between the previous two points, the tangent line to the graph is parallel to the chord. This is made clear by Figure 1 below.

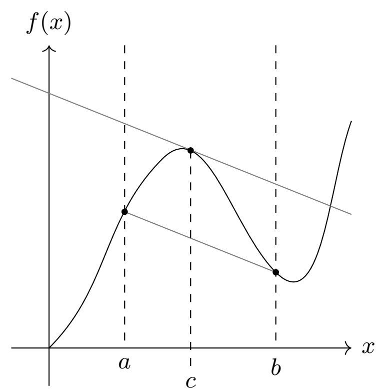


Figure 1: The geometric interpretation of the Mean Value Theorem.

Note: We used Rolle's Theorem to prove the Mean Value Theorem. More surprisingly, it turns out that Rolle's Theorem is a special case of the Mean Value Theorem, since if $f(a) = f(b)$, the fraction in Theorem 2.26 is zero.

Exercise 9 Use the Mean Value Theorem to prove the following: *if $I \subseteq \mathbb{R}$ is an interval and $f : I \to \mathbb{R}$ is differentiable with $f'(x) = 0$ for all $x \in [a, b]$, then f is constant.*
[**Hint:** Assume that f is not constant, so there exist $a, b \in I$ where $a < b$ and $f(a) \neq f(b)$.]

Notation 2.28 Throughout, we used *Lagrange's notation*, that is f', to denote a derivative. It is also useful to use *Leibniz's notation*, that is $\frac{\mathrm{d}f}{\mathrm{d}x}$, which is the derivative of f with respect to x.

Trigonometric Functions

We will now look at some special functions, starting with trigonometric functions, and get to grips with differentiating them. We recall that these functions were first introduced in Section

??. In particular, the note before Theorem ?? gives the so-called *compound-angle formulae* and Proposition ?? lists a number of properties of the functions $\sin(x)$ and $\cos(x)$.

Note: We do not introduce the trigonometric functions rigorously here; this requires *power series* from Chapter ??. However, we don't need to (for the non-rigorous setting here).

Recall from the aforementioned places that $\sin : \mathbb{R} \to [-1, 1]$ and $\cos : \mathbb{R} \to [-1, 1]$ satisfy these:

- $\sin^2(x) + \cos^2(x) = 1$.
- $\sin(x + y) = \sin(x)\cos(y) + \cos(x)\sin(y)$.
- $\cos(x + y) = \cos(x)\cos(y) - \sin(x)\sin(y)$.

Also, we define $\tan : \mathbb{R} \setminus \{\pi/2 + n\pi\} \to \mathbb{R}$ for all $n \in \mathbb{Z}$ by the formula $\tan(x) = \sin(x)/\cos(x)$.

Definition 2.29 We define the reciprocal functions cosecant, secant and cotangent as

$$\operatorname{cosec}(x) = \frac{1}{\sin(x)}, \qquad \sec(x) = \frac{1}{\cos(x)}, \qquad \cot(x) = \frac{1}{\tan(x)}.$$

Exercise 10 Prove the identities $1 + \cot^2(x) = \operatorname{cosec}^2(x)$ and $\tan^2(x) + 1 = \sec^2(x)$.
[**Hint:** Use the first identity $\sin^2(x) + \cos^2(x) = 1$ written above.]

Remark 2.30 We know also that sin and cos are 2π-periodic, meaning that $\sin(x + 2\pi) = \sin(x)$ and $\cos(x + 2\pi) = \cos(x)$; this is Proposition ??(i) and (ii). Additionally, tan is π-periodic. It turns out that by restricting the domains as follows, we can define inverse functions:

$$\begin{aligned} \sin &: [-\tfrac{\pi}{2}, \tfrac{\pi}{2}] \to [-1, 1], \\ \cos &: [0, \pi] \to [-1, 1], \\ \tan &: [-\tfrac{\pi}{2}, \tfrac{\pi}{2}] \to [-1, 1]. \end{aligned}$$

The inverses are arcsine, arccosine and arctangent, denoted $\sin^{-1}$, $\cos^{-1}$, $\tan^{-1}$ respectively.

Note: Be very aware to not get the inverse functions mixed up with the reciprocal functions.

Lemma 2.31 *The derivatives of the trigonometric functions are as follows:*

(i) $\sin'(x) = \cos(x)$.

(ii) $\cos'(x) = -\sin(x)$.

(iii) $\tan'(x) = \sec^2(x)$.

Proof: (i) We will use the definition (the note) directly, with a compound-angle formula. Indeed,

$$\begin{aligned}\sin'(x) &= \lim_{h\to 0} \frac{\sin(x+h)-\sin(x)}{h}\\ &= \lim_{h\to 0} \frac{\sin(x)\cos(h)+\cos(x)\sin(h)-\sin(x)}{h}\\ &= \lim_{h\to 0} \frac{\sin(x)(\cos h-1)+\cos(x)\sin(h)}{h}\\ &= \sin(x)\cdot 0+\cos(x)\cdot 1\\ &= \cos(x).\end{aligned}$$

(ii) Similarly, we can prove the second claim directly. Indeed,

$$\begin{aligned}\cos'(x) &= \lim_{h\to 0} \frac{\cos(x+h)-\cos(x)}{h}\\ &= \lim_{h\to 0} \frac{\cos(x)\cos(h)-\sin(x)\sin(h)-\cos(x)}{h}\\ &= \lim_{h\to 0} \frac{\cos(x)(\cos h-1)-\sin(x)\sin(h)}{h}\\ &= \cos(x)\cdot 0-\sin(x)\cdot 1\\ &= -\sin(x).\end{aligned}$$

(iii) We can apply the Quotient Rule to the definition of the tangent function, that is

$$\tan'(x) = \left(\frac{\sin}{\cos}\right)'(x) = \frac{\sin'(x)\cos(x)-\sin(x)\cos'(x)}{\cos^2(x)} = \frac{\sin^2(x)+\cos^2(x)}{\cos^2(x)} = \sec^2(x),$$

using (i) and (ii) just proved along with the identity $\sin^2(x)+\cos^2(x) = 1$. □

Lemma 2.32 *Let* $f : [-1,1] \to [-\frac{\pi}{2}, \frac{\pi}{2}]$ *be given by* $f(x) = \sin^{-1}(x)$. *Then,* $f'(x) = \frac{1}{\sqrt{1-x^2}}$.

Proof: Applying sin to both sides of $f(x) = \sin^{-1}(x)$ gives us $\sin\big(f(x)\big) = x$. Using Lemma 2.31(i) with the Chain Rule, differentiating both sides gives us $f'(x)\cos\big(f(x)\big) = 1$. Using the fact that $\sin^2(\theta) + \cos^2(\theta) = 1$ rearranges to $\cos(\theta) = \sqrt{1 - \sin^2(\theta)}$. Using $\theta = f(x)$, we obtain

$$\cos\big(f(x)\big) = \sqrt{1 - \sin^2(f(x))} = \sqrt{1 - x^2}.$$

Rearranging the equation we got after differentiating and combining everything together, we get

$$f'(x) = \frac{1}{\sqrt{1 - x^2}}. \qquad \square$$

Exercise 11 Compute the derivatives of the following functions:

(i) $f(x) = \cos^{-1}(x)$.

(ii) $f(x) = \tan^{-1}(x)$.

[**Hint:** Again, proceed as in the proof of Lemma 2.32, but use $\tan^2(\theta) + 1 = \sec^2(\theta)$.]

Exponential Function and Natural Logarithm

We will now motivate the introduction of another very useful function called the exponential function. Indeed, suppose we have a function $f : \mathbb{R} \to \mathbb{R}$ that satisfies $f'(x) = f(x)$ – this means that it is its own derivative. If we impose the condition that $f(0) = 1$, there is a unique solution to this *differential equation*; this unique solution is the exponential function $\exp(x)$.

Note: There are numerous ways to define the exponential function, and here are two more:

$$\exp(x) = \sum_{n=0}^{\infty} \frac{x^n}{n!} \qquad \text{and} \qquad \exp(x) = \lim_{n\to\infty} \left(1 + \frac{x}{n}\right)^n.$$

Remark 2.33 We now list some properties of the exponential function (without proof):

$$\exp(x + y) = \exp(x)\exp(y), \qquad \lim_{x\to\infty} \exp(x) = \infty, \qquad \lim_{x\to-\infty} \exp(x) = 0.$$

Moreover, we notice that $\exp : \mathbb{R} \to (0, \infty)$, so its range isn't the whole of the real numbers; it is a strictly positive function. It is also strictly increasing, i.e. $x > y$ implies that $\exp(x) > \exp(y)$.

Definition 2.34 The natural logarithm is the function $\log : (0, \infty) \to \mathbb{R}$ defined as the inverse of the exponential, that is $\log(x) = \exp^{-1}(x)$.

Lemma 2.35 *The natural logarithm satisfies the following properties:*

$$\log(xy) = \log(x) + \log(y), \qquad \lim_{x\to\infty} \log(x) = \infty, \qquad \lim_{x\to 0} \log(x) = -\infty.$$

Sketch of Proof: These follow from Remark 2.33, using the fact that $\log = \exp^{-1}$. □

Using Remark 2.33 inductively, we see that $\exp(x_1 + x_2 + \cdots + x_n) = \exp(x_1)\exp(x_2)\cdots\exp(x_n)$. Therefore, if we set each $x_i = \log(x)$, then we conclude that

$$\exp\big(n\log(x)\big) = \exp\big(\log(x)\big)^n = x^n,$$

so we can use the fact that exp and log are inverses to define integer powers of real numbers. Actually, this generalises to arbitrary powers of real numbers.

Definition 2.36 Let $x > 0$ and $k \in \mathbb{R}$. The k^{th} power of x is given by $x^k := \exp\big(k\log(x)\big)$.

Exercise 12 Using Definition 2.36, justify why $\log\left(a^b\right) = b\log(a)$ for all $a > 0$ and $b \in \mathbb{R}$.

Corollary 2.37 *Let $x, y > 0$. The natural logarithm satisfies* $\log\big(x/y\big) = \log(x) - \log(y)$.

Proof: Write $x/y = xy^{-1}$ and use the first property in Lemma 2.35 along with Exercise 12. □

Definition 2.38 The number $e \in \mathbb{R}$ is defined as $e := \exp(1)$.

Proposition 2.39 *We have that* $\log(e) = 1$. *Moreover, it follows that* $\exp(x) = e^x$.

Proof: By definition, we know that $\log(e) = \log\big(\exp(1)\big) = 1$, since log and exp are inverses of each other. Hence, using the formula in Definition 2.36, we have $e^x = \exp\big(x\log(e)\big) = \exp(x)$. □

Proposition 2.40 *We have the following derivatives:* $\exp'(x) = \exp(x)$ *and* $\log'(x) = 1/x$.

Proof: (i) The fact that $\exp'(x) = \exp(x)$ is a consequence of our definition of the exponential function; it was defined by it satisfying $f'(x) = f(x)$.

(ii) Let $f(x) = \log(x)$, which means that $\exp\big(f(x)\big) = x$. Differentiating both sides using the Chain Rule gives $f'(x)\exp'(f(x)) = 1$. As we established in (i), $\exp'(f(x)) = \exp\big(f(x)\big) = x$, so the expression rearranges to $f'(x) = 1/x$. □

Exercise 13 Reprove that the function $f(x) = x^n$ is differentiable with $f'(x) = nx^{n-1}$ by using Definition 2.36 to write $f(x) = \exp\big(n\log(x)\big)$ and applying the Chain Rule.

Hyperbolic Functions

Another group of functions behave similarly to the trigonometric functions; they are defined in terms of the exponential function. They are so named because one way to define them is in terms of a hyperbola (as opposed to a circle, which is how we can describe sin, cos, tan, and so forth).

Definition 2.41 The hyperbolic functions are defined as follows:

$$\begin{aligned} \sinh &: \mathbb{R} \to \mathbb{R}, & \sinh(x) &= \frac{e^x - e^{-x}}{2}, \\ \cosh &: \mathbb{R} \to [1, \infty), & \cosh(x) &= \frac{e^x + e^{-x}}{2}, \\ \tanh &: \mathbb{R} \to (-1, 1), & \tanh(x) &= \frac{e^{2x} - 1}{e^{2x} + 1}. \end{aligned}$$

Note: These are pronounced **shine**, **cosh** and **than** (but some prefer **sin-sh** and **tan-sh**).

Lemma 2.42 *For all $x \in \mathbb{R}$, we have* $\tanh(x) = \sinh(x)/\cosh(x)$.

Proof: See Exercise 14. □

Exercise 14 Prove Lemma 2.42.
[**Hint:** Use Definition 2.41 to re-write the right-hand side of the equation.]

In analogy to the trigonometric functions, there are a number of identities that the hyperbolic functions satisfy. Specifically, the three we next list are very similar to those for sin and cos written just before Definition 2.29.

Proposition 2.43 *The functions* sinh *and* cosh *satisfy the following for all* $x \in \mathbb{R}$*:*

(i) $\cosh^2(x) - \sinh^2(x) = 1$.

(ii) $\sinh(x+y) = \sinh(x)\cosh(y) + \cosh(x)\sinh(y)$.

(iii) $\cosh(x+y) = \cosh(x)\cosh(y) + \sinh(x)\sinh(y)$.

Sketch of Proof: One way to prove this is to again use Definition 2.41 directly. □

Definition 2.44 The reciprocal hyperbolic functions are defined as follows:

$$\begin{aligned} \operatorname{cosech} &: \mathbb{R}\setminus\{0\} \to \mathbb{R}\setminus\{0\}, & \operatorname{cosech}(x) &= \frac{1}{\sinh(x)}, \\ \operatorname{sech} &: \mathbb{R} \to (0,1], & \operatorname{sech}(x) &= \frac{1}{\cosh(x)}, \\ \coth &: \mathbb{R}\setminus\{0\} \to \mathbb{R}\setminus[-1,1], & \coth(x) &= \frac{1}{\tanh(x)}. \end{aligned}$$

Exercise 15 Prove the identities $1 - \tanh^2(x) = \operatorname{sech}^2(x)$ and $\coth^2(x) - 1 = \operatorname{cosech}^2(x)$. [**Hint:** Use the first identity $\cosh^2(x) - \sinh^2(x) = 1$ in Proposition 2.43.]

Remark 2.45 This is in clear analogy to Definition 2.29. Moreover, we again need to be careful not to mix up inverse functions and reciprocal functions. In the spirit of Remark 2.30, we can take inverses; we first restrict $\cosh : [0,\infty) \to [1,\infty)$. The inverses are $\sinh^{-1}$, $\cosh^{-1}$ and $\tanh^{-1}$.

Lemma 2.46 *The inverse hyperbolic functions are given by the following:*

$$\begin{aligned} \sinh^{-1} &: \mathbb{R} \to \mathbb{R}, & \sinh^{-1}(x) &= \log\Big(x + \sqrt{x^2-1}\Big), \\ \cosh^{-1} &: [0,\infty) \to [1,\infty), & \cosh^{-1}(x) &= \log\Big(x + \sqrt{x^2+1}\Big), \\ \tanh^{-1} &: (-1,1) \to \mathbb{R}, & \tanh^{-1}(x) &= \frac{1}{2}\log\left(\frac{1+x}{1-x}\right). \end{aligned}$$

Proof: ($\cosh^{-1}$) Suppose $f(x) = \cosh^{-1}(x)$, which means $x = \cosh\big(f(x)\big)$. By definition then,

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

We can rearrange this to get $e^{2f(x)} - 2xe^{f(x)} + 1 = 0$. This is just a quadratic in $e^{f(x)}$ (by which we mean if we replace all instances of $e^{f(x)}$ with another letter, y say, then we get $y^2 - 2xy + 1 = 0$

which is clearly quadratic). Therefore, we can use the quadratic formula to solve this:

$$e^{f(x)} = x \pm \sqrt{x^2-1} \qquad \Leftrightarrow \qquad f(x) = \log\Big(x \pm \sqrt{x^2-1}\Big).$$

We can use Exercise 12 along with the fact that $x - \sqrt{x^2-1} = (x+\sqrt{x^2-1})^{-1}$ to conclude that $\log\Big(x - \sqrt{x^2-1}\Big) = -\log\Big(x + \sqrt{x^2-1}\Big)$. Since the domain of cosh is non-negative, it follows that the range of $f(x) = \cosh^{-1}(x)$ is non-negative, so we need only take the positive sign. Hence,

$$\cosh^{-1}(x) = \log\Big(x + \sqrt{x^2-1}\Big).$$

($\sinh^{-1}$) Proceed similarly to $\cosh^{-1}$.

($\tanh^{-1}$) Proceed similarly to $\cosh^{-1}$ and $\sinh^{-1}$. □

Exercise 16 Prove the formula for $\sinh^{-1}$ given in Lemma 2.46.

We will now talk about the derivatives of the hyperbolic functions. Fortunately, a lot of the heavy-lifting was done when we discussed the exponential function and natural logarithm. We essentially get the next result for free (if we use the Product, Quotient and Chain Rules).

Theorem 2.47 *The hyperbolic functions are differentiable with these derivatives:*

(i) $\sinh'(x) = \cosh'(x)$.

(ii) $\cosh'(x) = \sinh(x)$.

(iii) $\tanh'(x) = \operatorname{sech}^2(x)$.

(iv) $\operatorname{cosech}'(x) = -\operatorname{cosech}(x)\coth(x)$.

(v) $\operatorname{sech}'(x) = -\operatorname{sech}(x)\tanh(x)$.

(vi) $\coth'(x) = -\operatorname{cosech}^2(x)$.

Sketch of Proof: Use the rules in Proposition 2.15 with Definitions 2.41 and 2.44. □

Proposition 2.48 *The inverse hyperbolic functions are differentiable with these derivatives:*

(i) $(\sinh^{-1})'(x) = \frac{1}{\sqrt{x^2+1}}$.

(ii) $(\cosh^{-1})'(x) = \frac{1}{\sqrt{x^2-1}}$.

(iii) $(\tanh^{-1})'(x) = \frac{1}{1-x^2}$.

Proof: ($\cosh^{-1}$) Let $f(x) = \cosh^{-1}(x)$, meaning that $x = \cosh\big(f(x)\big)$. Using Proposition 2.43(i),

$$\sinh\big(f(x)\big) = \sqrt{\cosh^2(f(x)) - 1} = \sqrt{x^2-1}.$$

The Chain Rule alongside Theorem 2.47 implies the result, namely

$$1 = f'(x)\sinh\big(f(x)\big) \qquad \Rightarrow \qquad f'(x) = \frac{1}{\sinh\big(f(x)\big)} = \frac{1}{\sqrt{x^2-1}}.$$

($\sinh^{-1}$) Proceed similarly to $\cosh^{-1}$; see Exercise 17.

($\tanh^{-1}$) Proceed similarly to $\cosh^{-1}$ and $\sinh^{-1}$, using $1-\tanh^2(x) = \operatorname{sech}^2(x)$. □

Exercise 17 Prove the formula for $(\sinh^{-1})'$ given in Proposition 2.48.

It may be that a function isn't given explicitly in terms of a single variable. For example, the equation $x^2+y^2=1$ describes the unit circle. We can think of this as the equation $f(x,y)=0$, where $f(x,y) = x^2+y^2-1$ is a function of two variables.

Definition 2.49 An implicit equation is an equation of the form $f(x_1,...,x_n)=0$, where f is a function of many variables (in this case, there are n variables).

Note: It is sometimes possible to rearrange an implicit equation to give it explicitly. For example, the equation $x-y+3=0$ can be rearranged to $y=x+3$; we can then think of $y=y(x)$, that is as a function of x, and proceed with differentiating (here, $y'(x)=1$). However, this is **not** guaranteed to happen for an implicit equation in general.

Example 2.50 Consider the implicit equation $x^2-xy+y^3=7$. We will think of $y=y(x)$ as a function of x; in this way, we will be able to get an expression for $y'(x)$ by applying the Product and Chain Rules. Indeed, differentiating this equation gives us

$$2x-y-xy'+3y^2y'=0.$$

Since we think of y as a function of x, we need to remember to differentiate it too. This becomes

$$(3y^2-x)y' = y-2x \qquad \Rightarrow \qquad y' = \frac{y-2x}{3y^2-x}.$$

Method – Implicit Differentiation: Consider an implicit equation in x and $y=y(x)$.

(i) Differentiate the equation as usual, but each time you differentiate y, put y' after it.

(ii) Rearrange for y'.

Exercise 18 Compute the implicit derivative of the equation $\sin(y) + x^2y^3 - \cos(x) = 2y$.

Remark 2.51 In fact, implicit differentiation can be made rigorous by the much more general Implicit Function Theorem (see Chapters **??** and **??**). However, the situation here includes only two variables (which we label x and y), so we can state this special case of the theorem once we have introduced *partial* derivatives in Section 4.

Let $f : \mathbb{R} \to \mathbb{R}$ be a function, and suppose we wish to approximate the value of f at a point **near** $a \in \mathbb{R}$. A first crude estimate would be $p_0(x) := f(a)$. However, if we know that f is differentiable at a, then we can make the so-called straight line estimate, that is

$$p_1(x) := f(a) + f'(a)(x - a).$$

We can continue this process, defining $p_n(x)$ for every $n \in \mathbb{Z}^+$; this motivates the next definition.

Definition 2.52 Let $f : \mathbb{R} \to \mathbb{R}$ be smooth and $a \in \mathbb{R}$. The Taylor series of f about a is

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

The n^{th} Taylor approximant of f at a is what we get by stopping the sum at $k = n$, i.e.

$$p_n(x) := \sum_{k=0}^{n} \frac{f^{(k)}(a)}{k!}(x - a)^k.$$

The moral of the story is this: given a one-time differentiable function $f : I \to \mathbb{R}$ defined on an open interval $I \subseteq \mathbb{R}$ and some number $a \in I$, then we can apply the Mean Value Theorem to find some $c \in I$ between a and any $x \in I \setminus \{a\}$ to conclude the following:

$$f(x) = f(a) + f'(c)(x - a) = p_0(x) + [\text{error term in } f'].$$

There is a generalisation: any n-times differentiable function defined on an open interval can be approximated by a polynomial of degree $n - 1$ plus an error in terms of the n^{th} derivative. Thus,

$$f(x) = p_{n-1}(x) + [\text{error term in } f^{(n)}].$$

Note: This result is Taylor's Theorem, stated and proved a bit later.

Example 2.53 We construct the Taylor series of $f(x) = x^3$ about $a = 2$. By definition, this is

$$\begin{aligned} f(x) &= f(2) + f'(2)(x-2) + \frac{1}{2!}f''(2)(x-2)^2 + \frac{1}{3!}f'''(2)(x-2)^3 + 0 + 0 + \cdots \\ &= 2^3 + 3(2^2)(x-2) + \frac{1}{2}6(2)(x-2)^2 + \frac{1}{6}6(x-2)^3 \\ &= 8 + 12(x-2) + 6(x-2)^3 + (x-2)^3. \end{aligned}$$

Note that $f^{(4)}(x) = 0$, so the fourth and all higher derivatives are zero

Exercise 19 Construct the second Taylor approximant of $f(x) = \sqrt{x}$ at $a = 4$.

Definition 2.54 The Maclaurin series of f is the Taylor series of f about $a = 0$.

Remark 2.55 It is debatable whether or not that Taylor series about 0 deserve their own name and definition. However, Taylor series in general (about zero or otherwise) are very important. We will see this time and again in Chapters **??** and **??**.

Proposition 2.56 *We have the following Maclaurin series:*

$$\begin{aligned} \sin(x) &= \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k+1)!} x^{2k+1}, \\ \cos(x) &= \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k)!} x^{2k}, \\ \exp(x) &= \sum_{k=0}^{\infty} \frac{1}{k!} x^k, \\ \log(1+x) &= \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k!} x^k. \end{aligned}$$

3 Basic Integration

We now look at the other side of a coin: is it possible to determine $f(x)$ if we only know $f'(x)$? The answer is a sort-of yes. As usual, we discuss functions of one variable throughout.

Integrals and Anti-Derivatives

Definition 3.1 Let $f : [a, b] \to \mathbb{R}$ be continuous. The integral of f over the interval $[a, b]$ is the limit of the area of rectangles that split the interval $[a, b]$ and fit just under/over the graph of the function f. We denote this area by the following expression:

$$\int_a^b f(x)\,\mathrm{d}x \in \mathbb{R}.$$

Remark 3.2 Consider a one-dimensional metal bar which is of length L. Hence, a position on the bar is just a point $x \in [0, L]$. Suppose that $\rho(x)$ is the density of the bar at x; this may change depending where on the bar we are. If we split the bar into n pieces of lengths $\delta x_1, ..., \delta x_n$, where each piece has a mass of $m_1, ..., m_n$, then we see for each $k = 1, ..., n$ that

$$\rho_k^{\min}\delta x_k \leq m_k \leq \rho_k^{\max}\delta x_k,$$

were $\rho_k^{\max}$ and $\rho_k^{\min}$ are the maximum and minimum densities of the k^{th} piece, respectively (recall that density $=$ mass$\boldsymbol{\nabla}\cdot$volume but we are only in one-dimension so volume is just length). Therefore, if the bar has mass M in total, we conclude that

$$\sum_{k=1}^{n} \rho_k^{\min}\delta x_k \leq M \leq \sum_{k=1}^{n} \rho_k^{\max}\delta x_k.$$

As we split the bar into more and more pieces, and we shrink the sizes of each of those pieces (meaning $n \to \infty$ and $\delta x_k \to 0$ for each k), then it follows that

$$\lim \rho_k^{\min} = \rho = \lim \rho_k^{\max}.$$

Thus, we get an expression of the mass (which we define as the integral):

$$M = \lim_{\delta x_k \to 0} \left(\sum_{k=1}^{n} \rho_k^{\max}\delta x_k \right) =: \int_0^L \rho(x)\,\mathrm{d}x.$$

Example 3.3 We can run through the construction in Remark 3.2 in the specific case of $\rho(x) = x$. Indeed, if the metal bar is of length L and it is split into n **equal** pieces, then we know that the length of the k^{th} piece is L/n (no matter what k is). In this case,

$$\rho_k^{\min} = \frac{(k-1)L}{n} \qquad \text{and} \qquad \rho_k^{\max} = \frac{kL}{n}.$$

As such, the so-called *lower and upper sums* are as follows:

$$\sum_{k=1}^{n} \frac{(k-1)L}{n}\frac{L}{n} = \frac{L^2}{n^2}\frac{n^2-n}{2} \qquad \text{and} \qquad \sum_{k=1}^{n} \frac{kL}{n}\frac{L}{n} = \frac{L^2}{n^2}\frac{n^2+n}{2}.$$

Above, we used the well-known formula for the sum of the first n positive integers (Exercise **??**). As $n \to \infty$, it is easy to see that **both** the lower and upper sums tend to $L^2/2$. Thus, we have

$$\int_0^L \rho(x)\,\mathrm{d}x = \int_0^L x\,\mathrm{d}x = \frac{L^2}{2}.$$

Exercise 20 Repeat Example 3.3 except with $\rho(x) = x^2$.
[**Hint:** When dealing with the lower and upper sums, use the formula in Example **??**.]

Note: For an integral $\int_a^b f(x)\,\mathrm{d}x$, we call f the integrand and $\mathrm{d}x$ the differential of x. We would then call f integrable on $[a, b]$. Moreover, a and b are the limits of integration.

We now state some useful integration properties. Again, the closest we have to motivating the notion of an integral is Remark 3.2, which is not rigorous. Thus, we do not give proper details of the proofs of these results until Section **??**.

Definition 3.4 Let f be integrable on $[a, b]$. Then, we define the following relation:

$$\int_a^b f(x)\,\mathrm{d}x = -\int_b^a f(x)\,\mathrm{d}x.$$

Proposition 3.5 *Let f and g be integrable over some intervals. The following are true:*

(i) $\int_a^b \alpha f(x) + \beta g(x)\,\mathrm{d}x = \alpha \int_a^b f(x)\,\mathrm{d}x + \beta \int_a^b g(x)\,\mathrm{d}x$ *for any* $\alpha, \beta \in \mathbb{R}$.

(ii) $\int_a^c f(x)\,\mathrm{d}x = \int_a^b f(x)\,\mathrm{d}x + \int_b^c f(x)\,\mathrm{d}x$.

Remark 3.6 The space of integrable functions on the interval $[a, b]$, denoted $L([a, b])$, is actually an $\mathbb{R}$-vector space in the sense of Definition **??**. This is clear from Proposition 3.5(i). The assignment $f \mapsto \int_a^b f(x)\,\mathrm{d}x$ is a linear map $L([a, b]) \to \mathbb{R}$ from this space to its ground field.

Definition 3.7 Let $f : [a, b] \to \mathbb{R}$ be continuous. A primitive (or anti-derivative) of f is a function $F : [a, b] \to \mathbb{R}$ such that $F'(x) = f(x)$ for all $x \in (a, b)$.

Lemma 3.8 *Let $f : [a, b] \to \mathbb{R}$ be continuous and F_1, F_2 be primitives of f. Then, $F_2 - F_1$ is constant.*

Proof: Simply apply the derivative to the difference $F_2(x) - F_1(x)$. Indeed, we get

$$(F_2 - F_1)'(x) = F_2'(x) - F_1'(x) = f(x) - f(x) = 0,$$

by the definition of a primitive. But a zero derivative implies it is constant, by Exercise 9. □

Lemma 3.9 *Let $f : [a, b] \to \mathbb{R}$ be continuous. Then, there exists a primitive F of f.*

Sketch of Proof: Define $I : [a, b] \to \mathbb{R}$ by $I(t) = \int_a^t f(x)\,\mathrm{d}x$; notice $I(t + h) - I(t) = \int_t^{t+h} f(x)\,\mathrm{d}x$ by using Proposition 3.5(ii). For 'small' values of h, we have $I(t + h) - I(t) \approx f(t)h$. Thus, the definition of the derivative implies

$$I'(t) = \lim_{h\to 0} \frac{I(t + h) - I(t)}{h} = \lim_{h\to 0} \frac{f(t)h}{h} = f(t).$$

Hence, I is a primitive of f. □

Theorem 3.10 (Fundamental Theorem of Calculus) *Let $f : [a, b] \to \mathbb{R}$ be continuous and F be a primitive of f. Then, the integral of f is the difference of the primitive at the endpoints, namely*

$$\int_a^b f(x)\,\mathrm{d}x = F(b) - F(a).$$

Sketch of Proof: Let $F : [a, b] \to \mathbb{R}$ be **any** primitive of f. Using I from Lemma 3.9, we know that $I - F$ is constant via Lemma 3.8. In particular, it is true that $I(a) - F(a) = I(b) - F(b)$. Rearranging this gives the result. □

Note: This is really the Fundamental Theorem of Calculus Version 2 (see Theorem **??**). The existence of I is the Fundamental Theorem of Calculus Version 1 (see Theorem **??**).

Exercise 21 Compute $\int_{-1}^3 x^2\,\mathrm{d}x$ by using the Fundamental Theorem of Calculus.

Thus far, the integrals we have discussed have been so-called *definite* integrals; this is where we integrate over an interval and the result is a number. However, we will develop further the notion of an indefinite integral and formulate a means to compute an anti-derivative of a given function (but, of course, there is no 'one-size-fits-all' method).

Note: We denote 'the' indefinite integral of f as $\int f(x)\,\mathrm{d}x$. It is unique up to an additive constant, meaning we have $\int f(x)\,\mathrm{d}x = F(x) + c$ for any primitive F of f and any $c \in \mathbb{R}$.

Example 3.11 Let $f(x) = x^{k+1}$ for some $k \in \mathbb{R} \setminus \{-1\}$. Recall its derivative from Exercises 5 and 13, namely $f'(x) = (k+1)x^k$. By the Fundamental Theorem of Calculus, this tells us

$$x^{k+1} = \int (k+1)x^k\,\mathrm{d}x \qquad \Rightarrow \qquad \int x^k\,\mathrm{d}x = \frac{1}{k+1}x^{k+1}.$$

Exercise 22 Find the indefinite integral of the function $f(x) = x^{-1}$.
[**Hint:** Use Proposition 2.40 and consider separately the cases $x > 0$ and $x < 0$.]

Lemma 3.12 *We have the following indefinite integrals:*
(i) $\int \sin(x)\,\mathrm{d}x = -\cos(x)$.
(ii) $\int \cos(x)\,\mathrm{d}x = \sin(x)$.
(iii) $\int \exp(x)\,\mathrm{d}x = \exp(x)$.

Integration by Substitution

We will now discuss the method of finding an indefinite integral by substitution. There are really two methods which feed into this, one being an *explicit* substitution and the other being a more relaxed *inspection* approach.

Method – Integration by Inspection: Let f be a continuous function with primitive F and g be any function with continuous derivative (i.e. class C^1). The Chain Rule implies

$$F'(g(x)) = F'(g(x))g'(x) = f(g(x))g'(x),$$

by definition of F being a primitive of f. The Fundamental Theorem of Calculus implies

$$F(g(x)) = \int F'(g(x))\,\mathrm{d}x = \int f(g(x))g'(x)\,\mathrm{d}x.$$

Example 3.13 Here are some examples of performing integration by inspection:

(i) $\int \cos(kx)\,\mathrm{d}x = \frac{1}{k}\sin(kx)$ for any $k \in \mathbb{R}$.

(ii) $\int \exp(kx)\,\mathrm{d}x = \frac{1}{k}\exp(kx)$ for any $k \in \mathbb{R}$.

(iii) $\int \frac{1}{ax+b} = \frac{1}{a}\log\big(|ax+b|\big)$, per Exercise 22.

Note: Let's give a commentary of Example 3.13(i) to get a better idea of what we are doing: the point is to notice that cos will integrate to sin. We then understand that the 'inside' will remain unchanged, so we expect the integral to include $\sin(kx)$. However, if we differentiate this using the Chain Rule, we actually get $k\cos(kx)$. Obviously, this is close to $\cos(kx)$ so we need only divide through by the factor we have at the start.

Exercise 23 Calculate the indefinite integral $\int \sin(-\omega x)$ where $\omega \in \mathbb{R}$.

Method – Integration by Explicit Substitution: Let $f(x)$ be a continuous function with primitive F. The idea is to express the variable x as a function of a new variable u, that is $x = x(u)$, such that the function $x(u)$ is either **strictly** increasing or **strictly** decreasing. If $x \in [a,b]$ is the original domain, then this corresponds to the new domain $u \in [\alpha,\beta]$. Now, the inverse of $x(u)$ is $u(x)$, that is u as a function of x. The Chain Rule implies

$$F'(x(u)) = F'(x(u))x'(u) = f(x(u))x'(u),$$

by definition of F being a primitive of f. The Fundamental Theorem of Calculus implies

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

The idea is that it is easy to compute the integral on the far left, rather than the far right.

Lemma 3.14 *We have the following substitutions for the integrand:*

(i) *If it includes* $\sqrt{c^2 - (ax+b)^2}$, *use the substitution* $ax + b = c\sin(u)$.

(ii) *If it includes* $\sqrt{(ax+b)^2 + c^2}$, *use the substitution* $ax + b = c\sinh(u)$.

(iii) *If it includes* $\sqrt{(ax+b)^2 - c^2}$, *use the substitution* $ax + b = c\cosh(u)$.

Example 3.15 We will use Method **??** to compute the following indefinite integral:

$$\int \frac{x}{\sqrt{4x^2 + 12x + 13}}\,\mathrm{d}x.$$

The denominator can be re-written $\sqrt{4x^2 + 12x + 13} = \sqrt{(2x+3)^2 + 2^2}$, so Lemma 3.14 suggests we use the substitution $2x+3 = 2\sinh(u)$, which is to say that $x = x(u) = \sinh(u) - 3/2$. From this, we can see that its derivative $x'(u) = \cosh(u)$ by using Theorem 2.47. We can now replace x in the integrand with its expression in terms of u, but we must then multiply everything by $x'(u)$; this is what the second integral says in Method **??**. Doing this gives us

$$\int \frac{\sinh(u) - 3/2}{2\cosh(u)} \cosh(u)\,\mathrm{d}u = \int \frac{1}{2}\sinh(u) - \frac{3}{4}\,\mathrm{d}u = \frac{1}{2}\cosh(u) - \frac{3}{4}u.$$

All that remains is to translate everything back in terms of x. Well, per the substitution we made, we see that $u = u(x) = \sinh^{-1}(x+3/2)$. Substituting this tells us that

$$\int \frac{x}{\sqrt{4x^2+12x+13}}\,\mathrm{d}x = \frac{1}{2}\cosh\left(\sinh^{-1}(x+3/2)\right) - \frac{3}{4}\sinh^{-1}(x+3/2).$$

Remark 3.16 The solution obtained in Example 3.15 isn't the prettiest, but we can be slightly cleverer to make the first term neater. Indeed, if we substitute $x = \sinh(u) - 3/2$ into the denominator of the integrand, we see that

$$\sqrt{(2x+3)^2+2^2} = \sqrt{4\sinh^2(u)+2^2} = 2\sqrt{\sinh^2(u)+1} = 2\cosh(u),$$

using the formula from Proposition 2.43(i). Consequently, we get the much more appealing

$$\frac{1}{2}\cosh(u) = \frac{1}{4}\sqrt{(2x+3)^2+2^2}.$$

Note: Since $\cosh^2(y) - \sinh^2(y) = 1$ for **any** y, this is true in particular for $y = \sinh^{-1}(x)$. Hence, we can conclude that $\cosh^2(\sinh^{-1} x) - x^2 = 1$ and a simple rearrangement gives

$$\cosh\left(\sinh^{-1} x\right) = \sqrt{1+x^2}.$$

Exercise 24 Compute the following definite integral:

$$\int_1^5 \frac{x^2-x}{\sqrt{x-1}}\,\mathrm{d}x.$$

[**Hint:** Use the substitution $u = \sqrt{x-1}$ and remember to change the limits of integration.]

Integration by Parts

Where integration by substitution is closely related to the Chain Rule, the next method is akin to the Product Rule. Indeed, this tells us how to integrate the product of two integrable functions.

Method – Integration by Parts: Let f and g be functions with continuous derivatives (i.e. class C^1). Recall that the Product Rule (Proposition 2.15(ii)) tells us the following:

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

The Fundamental Theorem of Calculus implies

$$f(x)g(x) = \int f'(x)g(x) + f(x)g'(x)\,\mathrm{d}x.$$

If we split the integral into two parts, then the above rearranges to the useful formula

$$\int f(x)g'(x)\,\mathrm{d}x = f(x)g(x) - \int f'(x)g(x)\,\mathrm{d}x.$$

The idea, when presented with an integral of the product of two functions, is to see if we already know the integral of one of them and the derivative of the other. This is the first step to being able to apply the integration by parts formula.

Example 3.17 We will use integration by parts to compute $\int \tan^{-1}(x)\,\mathrm{d}x$. The trick is to think of $\tan^{-1}(x) = \tan^{-1}(x) \cdot 1$ and use the formula in Method **??** where

$$\begin{aligned} f(x) &= \tan^{-1}(x) \quad &\Rightarrow \quad f'(x) &= \frac{1}{1+x^2}, \\ g'(x) &= 1 \quad &\Rightarrow \quad g(x) &= x. \end{aligned}$$

Recall that we know the derivative of f from Exercise 11(ii), and we know how to integrate constant functions. Consequently, the integration by parts formula gives precisely what we need:

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

Note: We made use of the following useful formula, which generalises Example 3.13(iii):

$$\int \frac{f'(x)}{f(x)}\,\mathrm{d}x = \log\left(|f(x)|\right).$$

Exercise 25 Compute the indefinite integral $\int \log(x)\,\mathrm{d}x$.
[**Hint:** Use the same trick as in Example 3.17 and consider $\log(x) = \log(x) \cdot 1$.]

It may be that in the process of doing integration by parts, we end up having to do an integral very similar to the one we are concerned with. We will explore an example of this shortly.

Definition 3.18 An integral recurrence relation is a relation relating an integral I_n that depends on $n \in \mathbb{N}$ to a lower-order integral I_k where $k < n$. We call I_0 the initial value of the recurrence relation.

Example 3.19 Suppose we wish to get an integral recurrence relation for $I_n := \int x^n \sin(x)\,\mathrm{d}x$. Using integration by parts, we can get a relation rather swiftly:

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

where the integral on the far right is precisely what we set out to solve (with $n-2$ instead of n).

Exercise 26 Find an explicit expression for $I_3 = \int x^3 \sin(x)\,\mathrm{d}x$.
[**Hint:** Use Example 3.19 to write I_3 in terms of I_1, and compute I_1 directly.]

Example 3.20 We will use Example 3.19 to compute $I_6 = \int x^6 \sin(x)\,\mathrm{d}x$. Indeed,

$$\begin{aligned} I_6 &= -x^6 \cos(x) + 6x^5 \sin(x) - 30 I_4 \\ &= -x^6 \cos(x) + 6x^5 \sin(x) - 30 \left(-x^4 \cos(x) + 4x^3 \sin(x) - 4(3) I_2\right) \\ &= -x^6 \cos(x) + 6x^5 \sin(x) + 30x^4 \cos(x) - 120x^3 \sin(x) + 360 I_2 \\ &= -x^6 \cos(x) + 6x^5 \sin(x) + 30x^4 \cos(x) - 120x^3 \sin(x) + 360 \left(-x^2 \cos(x) + 2x \sin(x) - 2I_0\right) \\ &= -x^6 \cos(x) + 6x^5 \sin(x) + 30x^4 \cos(x) - 120x^3 \sin(x) - 360x^2 \cos(x) + 720x \sin(x) - 720 I_0, \end{aligned}$$

but we can easily see the initial value is $I_0 = \int x^0 \sin(x)\,\mathrm{d}x - \int \sin(x)\,\mathrm{d}x = -\cos(x)$. Therefore,

$$I_6 = (-x^6 + 30x^4 - 360x^2 + 720) \cos(x) + (6x^5 - 120x^3 + 720x) \sin(x).$$

Note: We can get integral recurrence relations not only from integration by parts. Indeed,

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

is obtained by performing integration by explicit substitution with $u = \tan(x)$.

Integration using Partial Fractions

We will conclude this section by discussing the integrals of some rational functions of one variable. Indeed, we say that the integrand is an element of $\mathbb{R}(x)$, that is the set of quotients p/q where $p, q \in \mathbb{R}[x]$ are polynomials in one variable with coefficients in $\mathbb{R}$.

Definition 3.21 Let $f \in \mathbb{R}(x)$ be a rational function of one variable. The partial fraction expression for f has the following general form:

$$f(x) = p(x) + \sum_j \frac{a_j(x)}{b_j(x)},$$

where p is a polynomial and, for each j, the denominator b_j is a power of a polynomial which we can't factorise into polynomials of positive degree and the numerator a_j is a polynomial with lower degree than that of b_j.

Method – Partial Fractions: Let $f = p/q$ be a rational function of one variable.

(i) Fully factorise the denominator into a polynomial product: $q(x) = h_1^{k_1}(x)\cdots h_n^{k_n}(x)$.

(ii) Compare both sides of the following equation to solve for each numerator $A_{ij} \in \mathbb{R}[x]$, where the degree is one less than the denominator, i.e. $\deg(A_{ij}) = \deg(h_i^j) - 1$:

$$\frac{p(x)}{q(x)} = \left(\frac{A_{11}}{h_1(x)} + \frac{A_{12}}{h_1^2(x)} + \cdots + \frac{A_{1k_1}}{h_1^{k_1}(x)}\right) + \cdots + \left(\frac{A_{n1}}{h_n(x)} + \frac{A_{n2}}{h_n^2(x)} + \cdots + \frac{A_{nk_n}}{h_n^{k_n}(x)}\right).$$

Remark 3.22 What is stated in Method **??** works for the most general situations. That said, we will not really need all this detail (and the algebra looks rather scary when written out as it is there), so consider the next example to see what we do.

Example 3.23 Suppose we wish to express the following rational function using partial fractions:

$$\frac{x^2+5}{x^3-3x+2}.$$

According to Method **??**, the first job is to factorise the denominator as much as possible. Indeed, we can see that $x^3-3x+2=(x-1)^2(x+2)$. Now, the second part of the aforementioned method tells us that we need to consider sums of fractions where the denominators are each of the powers of the things that appear in what we just factorised. Hence, we need to look at

$$\frac{x^2+5}{x^3-3x+2}=\frac{A}{x-1}+\frac{B}{(x-1)^2}+\frac{C}{x+2}.$$

Note: We don't have many fractions; we use the simpler notation A, B, C over A_{11}, A_{12}, A_{21}.

In order to determine what A, B, C are, we can combine the right-hand side by finding a common denominator (of course, the common denominator will be $(x-1)^2(x+2)$ in this case). Hence,

$$\frac{x^2+5}{x^3-3x+2}=\frac{A(x-1)(x+2)}{x^3-3x+2}+\frac{B(x+2)}{x^3-3x+2}+\frac{C(x-1)^2}{x^3-3x+2}.$$

We can now choose specific values for x to determine what happens with the numerator. Well,

$$\begin{aligned} x=1 \quad &\Rightarrow \quad 1^2+5=0+B(1+2)+0 \quad \Rightarrow \quad B=2, \\ x=-2 \quad &\Rightarrow \quad (-2)^2+5=0+0+C(-2-1)^2 \quad \Rightarrow \quad C=1. \end{aligned}$$

It remains to compute the value of A. If we expand and simplify the numerators, we see that

$$\frac{x^2+5}{x^3-3x+2}=\frac{(A+C)x^2+(A+B-2C)x+(-2A+2B+C)}{x^3-3x+2}.$$

Comparing the x^2-coefficients, say, tells us $1=A+C$. Hence, $A=0$. We know therefore that

$$\frac{x^2+5}{x^3-3x+2}=\frac{2}{(x-1)^2}+\frac{1}{x+2}.$$

Exercise 27 Express the following rational function using partial fractions:

$$\frac{3x+7}{x^2+5x+6}.$$

Example 3.24 We will apply partial fraction decomposition to compute the following integral:

$$\int \frac{x^2-14x-5}{(x-4)(x^2+4x+13)}\,\mathrm{d}x.$$

First, notice that the denominator is fully factorised (since the quadratic factor cannot be further reduced using coefficients in $\mathbb{R}$). Hence, we expect the integrand to have the form

$$\frac{x^2 - 14x - 5}{(x-4)(x^2+4x+13)} = \frac{A}{x-4} + \frac{Bx+C}{x^2+4x+13},$$

where the numerator for the second fraction is one degree less than that of the denominator (the bottom is a quadratic so the top must be linear, per what Method **??**(ii) says). Proceeding as in Example 3.23 and Exercise 27, we see that $A = -1$, $B = 2$, $C = -2$. Thus, the integral is

$$\begin{aligned}
\int \frac{x^2 - 14x - 5}{(x-4)(x^2+4x+13)}\,\mathrm{d}x &= \int \frac{-1}{x-4} + \frac{2x-2}{x^2+4x+13}\,\mathrm{d}x \\
&= \int \frac{-1}{x-4} + \frac{2x+4}{x^2+4x+13} - \frac{6}{(x+2)^2+3^2}\,\mathrm{d}x \\
&= -\log\big(|x-4|\big) + \log\Big(|x^2+4x+13|\Big) + 2\tan^{-1}\left(\frac{x+2}{3}\right).
\end{aligned}$$

Here, we made use of the formula in the note just after Example 3.17 to deal with the first two integrands and we made the substitution $x + 2 = 3\tan(u)$ to deal with the third integrand (this is reminiscent of Lemma 3.14).

4 Multivariable Calculus

The next step on the journey through calculus is to look at functions of more than one variable. To this end, we can consider so-called *partial* derivatives and *partial* integrals; this is where we almost ignore all-but-one variable. However, there is a much more complicated theory beyond this; we will touch upon it but more is explained in Chapter **??**.

Partial Differentiation

Definition 4.1 A (real) multivariable function is a map of the form $f : \mathbb{R}^n \to \mathbb{R}$.

Example 4.2 Here are some examples of multivariable functions:

(i) Any map $f : \mathbb{R} \to \mathbb{R}$ is a 'multi'variable function (with $n = 1$).

(ii) The map $g : \mathbb{R}^2 \to \mathbb{R}$ given by $g(x, y) = x^7 + \cos(x + y)$ is a multivariable function.

(iii) The map $h : \mathbb{R}^3 \to \mathbb{R}$ given by $h(x_1, x_2, x_3) = x_1^2 + 4x_2^2 + x_3^2 - 9$ is a multivariable function.

Note: A two-variable function $f : \mathbb{R}^2 \to \mathbb{R}$ can be interpreted as defining a two-dimensional surface in $\mathbb{R}^3$. Such a surface will have so-called *local coordinates* $(x, y, f(x, y))$, that is where the final coordinate is precisely the output from the two-variable function of the previous coordinates. This is an extension of the notion of a graph (where we interpret $f : \mathbb{R} \to \mathbb{R}$ as defining a one-dimensional surface in $\mathbb{R}^2$ with local coordinates $(x, f(x))$).

Definition 4.3 Let $f : \mathbb{R}^n \to \mathbb{R}$ be a multivariable function in the variables $x_1, ..., x_n$. Then, we define partial derivative of f with respect to x_k for some $k = 1, ..., n$ as follows:

$$\frac{\partial f}{\partial x_k} := \lim_{h\to 0} \frac{f(x_1, ..., x_{k-1}, x_k + h, x_{k+1}, ..., x_n) - f(x_1, ..., x_n)}{h}.$$

Remark 4.4 The partial derivative of a multivariable function is simply the usual derivative as in Definition 2.12 **except** where we treat all variables as constant bar the one we are taking the derivative with respect to. In the formula above, we treat the variable x_k as a variable and all others $x_1, ..., x_{k-1}, x_{k+1}, ..., x_n$ simply as constants.

Note: We shall write both $\partial_{x_k} f$ and f_{x_k} for the partial derivative with respect to x_k.

Example 4.5 We will compute the partial derivatives of $f(x, y, z) = x^2 + yz$.

(i) The partial derivative $\frac{\partial f}{\partial x} = 2x$, since yz is being treated as constant, and we know that the derivative of a constant is zero.

(ii) The partial derivative $\frac{\partial f}{\partial y} = z$, since x^2 is being treated as a constant, and z is constant so we know that the derivative of a constant times a variable is just that constant.

(iii) The partial derivative $\frac{\partial f}{\partial z} = y$ for a near-identical reason.

Lemma 4.6 *Let $f : \mathbb{R}^n \to \mathbb{R}$. Then, f is constant if and only if $f_{x_k} = 0$ for* **all** *$k = 1, ..., n$.*

Proof: ($\Rightarrow$) This is trivial.

($\Leftarrow$) Suppose that $f_{x_1} = \cdots = f_{x_n} = 0$ and define the single-variable functions $g_k : \mathbb{R} \to \mathbb{R}$ by $g_k(t) = f(x_1, ..., x_{k-1}, t, x_{k+1}, ..., x_n)$, that is we replace x_k by the variable t. Because these are now functions of one variable, we can talk about their (usual) derivative. Indeed, we see that

$$g_i'(t) = \lim_{h \to 0} \frac{f(x_1, ..., x_{k-1}, t+h, x_{k+1}, ..., x_n) - f(x_1, ..., x_{k-1}, t, x_{k+1}, ..., x_n)}{h}.$$

Consequently, letting $t = x_k$, we conclude that $g_k'(x_k) = f_{x_k}(\mathbf{x}) = 0$, by assumption that every partial derivative of f is zero. Therefore, we know that g_k is actually constant (by Exercise 9). Doing this for all $k = 1, ..., n$ will mean that f is constant with respect to each variable and is therefore constant. □

Proposition 4.7 *Let $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}$. Then, the following are true:*

(i) $(f(\mathbf{x}) + g(\mathbf{x}))_{x_k} = f_{x_k}(\mathbf{x}) + g_{x_k}(\mathbf{x})$. (Sum Rule)

(ii) $(f(\mathbf{x})g(\mathbf{x}))_{x_k} = f_{x_k}(\mathbf{x})g(\mathbf{x}) + f(\mathbf{x})g_{x_k}(\mathbf{x})$. (Product Rule)

(iii) $(f(\mathbf{x})/g(\mathbf{x}))_{x_k} = (f_{x_k}(\mathbf{x})g(\mathbf{x}) - f(\mathbf{x})g_{x_k}(\mathbf{x}))/g(\mathbf{x})^2$ **if** $g(\mathbf{x}) \neq 0$. (Quotient Rule)

Proof: Omitted; this is similar to Proposition 2.15 (that of the usual derivative). □

We see that the story is almost identical to that of the usual derivative. However, we are missing a key ingredient: the Chain Rule. There is a (rather complicated) generalisation to multivariable functions but we begin with a simple case with one multivariable and one one-variable function.

Lemma 4.8 (Weak Chain Rule) *Let $f : \mathbb{R} \to \mathbb{R}$ be a function and $g : \mathbb{R}^n \to \mathbb{R}$. Then,*

$$\frac{\partial}{\partial x_k}(f \circ g)(\mathbf{x}) = f'(g(\mathbf{x}))\frac{\partial}{\partial x_k}g(\mathbf{x}).$$

Proof: Omitted; we will prove a stronger result later. □

Example 4.9 Suppose we wish to compute the partial derivatives of $\phi(x,y) = \sin(xy)$. Well,

$$\phi_x(x,y) = y\cos(xy) \qquad \text{and} \qquad \phi_y(x,y) = x\cos(xy),$$

since sin is a single-variable function (playing the role of f in the Weak Chain Rule) and xy is a two-variable function (playing the role of g in the Weak Chain Rule).

Exercise 28 Compute the partial derivatives f_x, f_y, f_z of the following three-variable function and support your answer with justifications of each step you make:

$$f(x,y,z) = xz + e^{yz} + \sin(xy).$$

We can adapt implicit differentiation to the partial derivative situation: for a multivariable function $f(x_1, ..., x_n)$, we can pick one variable to consider as a function of the rest.

Method – Implicit Partial Differentiation: Let $f(x_1, ..., x_n)$ be a multivariable function and consider the variable $x_i = x_i(x_1, ..., x_{i-1}, x_{i+1}, ...x_n)$ as a function of the other variables and suppose we wish to find an expression for the partial derivative $\frac{\partial x_i}{\partial x_k}$.

(i) Partial-differentiate the equation, but when you differentiate x_i, put $\frac{\partial x_i}{\partial x_k}$ after it.

(ii) Rearrange for $\frac{\partial x_i}{\partial x_k}$.

Example 4.10 Consider the equation $x^3z^2 - 5xy^5z = x^2 + y^3$ and treat $z = z(x,y)$ as a function of two variables. If we wish to determine an expression for z_x, say, then we appeal straight to Method **??**. Indeed, taking the partial derivative of the equation with respect to x yields

$$3x^2z^2 + 2x^3z\frac{\partial z}{\partial x} - 5y^5z - 5xy^5\frac{\partial z}{\partial x} = 2x.$$

Therefore, rearranging this gives us what we are after:

$$z_x = \frac{\partial z}{\partial x} = \frac{2x - 3x^2z^2 + 5y^5z}{2x^3z - 5xy^5}.$$

Note: This is, again, related to the Implicit Function Theorem (proven in Chapter **??**).

Exercise 29 Let $x^2y^2 + y^2z^2 + z^2x^2 = 7$ and find $\frac{\partial x}{\partial y}$ using implicit partial differentiation.

Definition 4.11 A higher-order partial derivative of the multivariable function $f : \mathbb{R}^n \to \mathbb{R}$ is one where we do successive partial differentiation. There are two main types to consider.

- There are pure higher-order partial derivatives of the form

$$\frac{\partial^m f}{\partial x_k^m}.$$

- There are mixed higher-order partial derivatives of the form

$$\frac{\partial^m f}{\partial x_{k_1}^{p_1} \partial x_{k_2}^{p_2} \cdots \partial x_{k_t}^{p_t}}, \qquad \text{where } p_1 + p_2 + \cdots + p_t = m.$$

Notation 4.12 As with usual partial derivatives, we can denote higher-order partial derivatives using subscript notation (i.e. analogous to how we write f_x to mean $\frac{\partial f}{\partial x}$). Indeed, we just write a list of subscripts representing the order in which to take the partial derivatives (e.g. $\frac{\partial^2 f}{\partial y \partial x}$ means $\frac{\partial}{\partial x}$ first and $\frac{\partial}{\partial y}$ second, so this is represented by f_{xy}).

Note: We were careful with the order of the mixed partial derivatives in Notation 4.12. Thankfully, there are some relatively-mild conditions where which we can disregard order.

Theorem 4.13 (Mixed Derivatives Theorem) *Let $D \subseteq \mathbb{R}^n$ and $f : D \to \mathbb{R}$ be a multivariable function with this property: $\mathbf{p} \in \mathbb{R}^n$ is a point and a neighbourhood of $\mathbf{p}$ is contained in D and f has continuous second-order partial derivatives at $\mathbf{p}$. Then, for all $i, j = 1, ..., n$,*

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{p}) = \frac{\partial^2 f}{\partial x_j \partial x_i}(\mathbf{p}).$$

Proof: We give an elementary proof for when $n = 2$. Indeed, let $D \subseteq \mathbb{R}^2$ be an open rectangle containing $\mathbf{p} = (a, b)$, that is $D = (a - \delta, a + \delta) \times (b - \varepsilon, b + \varepsilon)$ for some numbers $\delta, \varepsilon > 0$. Next, define the following multivariable functions for $0 < |h_1|, |h_2| < \varepsilon$:

$$\begin{aligned} u(h_1, h_2) &= f(a + h_1, b + h_2) - f(a + h_1, b), \\ v(h_1, h_2) &= f(a + h_1, b + h_2) - f(a, b + h_2), \\ w(h_1, h_2) &= f(a + h_1, b + h_2) - f(a + h_1, b) - f(a, b + h_2) + f(a, b). \end{aligned}$$

We apply the Mean Value Theorem to find $\gamma_1, \gamma_2, \delta_1, \delta_2 \in (0, 1)$ where the following are true:

$$\begin{aligned} w(h_1, h_2) &= u(h_1, h_2) - u(h_1, 0) \\ &= h_1 u_x(\gamma_1 h_1, h_2) \end{aligned}$$

$$
\begin{aligned}
&= h_1 \left(f_x(a + \gamma_1 h_1, b + h_2) - f_x(a + \gamma_1 h_1, b) \right) \\
&= h_1 h_2 f_{xy}(a + \gamma_1 h_1, b + \gamma_2 h_2)
\end{aligned}
$$

and

$$
\begin{aligned}
w(h_1, h_2) &= v(h_1, h_2) - v(h_1, 0) \\
&= h_2 v_y(h_1, \delta_2 h_2) \\
&= h_2 \left(f_y(a + h_1, b + \delta_2 h_2) - f_y(a, b + \delta_2 h_2) \right) \\
&= h_2 h_1 f_{yx}(a + \delta_1 h_1, b + \delta_2 h_2).
\end{aligned}
$$

Hence, we can equate the final line of each of the above groups of equations (since both are expression for the same function w). Since $h_1, h_2 \neq 0$, we can divide out the $h_1 h_2$ that appears in both, giving us the following equality:

$$f_{xy}(a + \gamma_1 h_1, b + \gamma_2 h_2) = f_{yx}(a + \delta_1 h_1, b + \delta_2 h_2).$$

If we take the limit as $h_1 \to 0$ and $h_2 \to 0$, then we obtain precisely the result we want, namely

$$\frac{\partial^2 f}{\partial x \partial y}(\mathbf{p}) = f_{yx}(a, b) = f_{xy}(a, b) = \frac{\partial^2 f}{\partial y \partial x}(\mathbf{p}). \qquad \square$$

Note: Theorem 4.13 also goes by the names Schwarz's Theorem and Clairaut's Theorem.

Exercise 30 Verify the Mixed Derivatives Theorem for $f(x, y) = xy^3 + x\sin(xy)$.

Definition 4.14 Let $f : \mathbb{R}^n \to \mathbb{R}$ be a multivariable function on the variables $x_1, ..., x_n$. The partial integral with respect to one of the variables is the usual integral where we treat all other variables as constant. This is denoted similarly to the integral of single-variable functions, i.e.

$$\int_a^b f(\mathbf{x}) \, \mathrm{d}x_k,$$

where the limits on the integral mean $x_k = a$ and $x_k = b$.

Example 4.15 The partial integral of the function $f(x, y) = yx^2$ with respect to x is

$$\int f(x, y) \, \mathrm{d}x = \frac{1}{3} yx^3 + \varphi(y),$$

where φ is **any** function of y. Indeed, as with usual integration, when we take an anti-derivative, they are the same up to constant. Here, if we partial-differentiate the right-hand side above with respect to x, we do indeed recover f. This is because we treat any y-variables as constant. Hence, our 'constant' here is a function in terms of the other variables.

Exercise 31 Compute the partial integral $\int_1^4 yx^2 \, \mathrm{d}x$.
[**Hint:** There will be no 'constant' $\varphi(y)$ here, like with usual integration over an interval.]

Geometry and Directional Derivatives

We will now begin to think more geometrically about partial derivatives. In doing so, using what we know about the usual one-variable case, we can build up some intuition on what these derivatives (and other differential operators) are telling us.

Example 4.16 Let $D \subseteq \mathbb{R}$ and $f : D \to \mathbb{R}$ be differentiable at $a \in D$. Recall in Theorem **??**(i) that a non-vertical line passing through the point (x_1, y_1) with gradient m can be given by $y - y_1 = m(x - x_1)$. We can interpret the number $f'(a)$ as the gradient of the straight line tangent to the graph $\big(x, f(x)\big)$ at the point $\big(a, f(a)\big)$. Substituting this point and gradient into the formula we just recalled gives us

$$f(x) - f(a) = f'(a)(x - a) \qquad \Leftrightarrow \qquad f(x) = f(a) + f'(a)(x - a).$$

Note: The expression $f(a) + f'(a)(x - a)$ at the end of Example 4.16 is well-defined for all values of $x \in D$, not just for $x = a$. However, when $x \neq a$, this formula is **not** equal to $f(x)$ in general. Hence, we call it the linear approximation of f at a and denote this by

$$f(x) \approx f(a) + f'(a)(x - a), \qquad x \in D.$$

We can generalise this concept to functions of many variables; we start with two variables.

Proposition 4.17 *Let $\Omega \subseteq \mathbb{R}^2$ and $f : \Omega \to \mathbb{R}$ be partially-differentiable at $(a, b) \in \Omega$. The* linear approximation of f at (a, b) *is $f(x, y) \approx f(a, b) + f_x(a, b)(x - a) + f_y(a, b)(y - b)$.*

Proof: Omitted; we prove a stronger result later. □

Remark 4.18 If we have a function $f(x)$ of one-variable, its *graph* is a subset of $\mathbb{R}^2$: we plot points $\big(x, f(x)\big)$ so it lives in a space of dimension one more than the number of variables $(1+1)$. Therefore, if we have now a function $f(x, y)$ of two variables, its *graph* is a subset of $\mathbb{R}^3$: we plot points $\big(x, y, f(x, y)\big)$. Hence, we call the graph of a one-variable function a curve and the graph of a two-variable function a surface. This gets much deeper in Chapter **??**. We provide a picture to bolster this intuition in Figure 2 below.

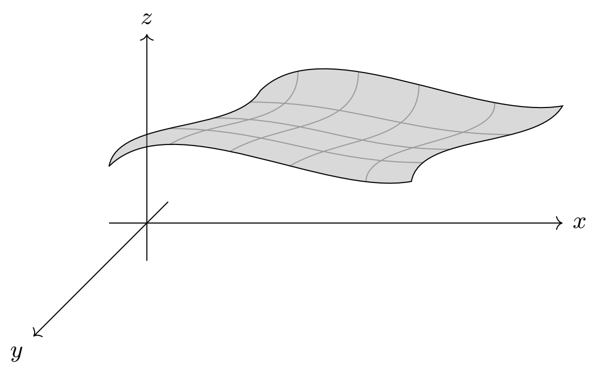


Figure 2: The graph of a two-variable function $f : \Omega \to \mathbb{R}$.

Exercise 32 Determine the linear approximation of $f(x, y) = x^2 + y^2$ at $(1, 2)$.

Using Remark 4.18, we can interpret Proposition 4.17 geometrically: when we take the partial derivative with respect to x, say, we are keeping y constant. In other words, we are looking at a 'slice' of our surface (the intersection of our surface with a plane in the y-axis, which is a curve). A similar story holds if we instead take the partial derivative with respect to y and keep x constant. Thus, we conclude the following important ideas:

- $f_x(a, b)$ can be treated as the tangent to the surface at the point (a, b) in the x-direction.
- $f_y(a, b)$ can be treated as the tangent to the surface at the point (a, b) in the y-direction.

Hence, this means the graph of the right-hand side $f(a, b) + f_x(a, b)(x - a) + f_y(a, b)(y - b)$ can be thought of as the **tangent plane** to the surface defined by $\big(x, y, f(x, y)\big)$.

Note: We have a slight problem; we can describe via partial derivatives the rates of change (tangents) in the x-direction and y-direction. But what about **any** old direction? This is what we concern ourselves with now.

Definition 4.19 Let $\Omega \subseteq \mathbb{R}^2$ and $f : \Omega \to \mathbb{R}$ have well-defined partial derivatives in each variable at $(a, b) \in \Omega$. The gradient of f at (a, b) is a function $\nabla f : \mathbb{R}^2 \to \mathbb{R}^2$ given by

$$\nabla f(a, b) = \big(f_x(a, b), f_y(a, b)\big) .$$

Remark 4.20 Suppose we have a function $f : X \to Y$. Then, the set of all functions of this form (with domain X and co-domain Y) we denote as $\mathcal{F}(X, Y)$. Using this, an equivalent way to describe the gradient of f at (a, b) is as a function $\nabla : \mathcal{F}(\Omega, \mathbb{R}) \times \mathbb{R}^2 \to \mathbb{R}^2$, that is we input into ∇ a function $\Omega \to \mathbb{R}$ and a point and we get from it another point.

Exercise 33 Find the gradient of $f(x, y) = x\sin(y) + x^7$ at the point $(2, \pi/2)$.

Suppose we are given a line in $\mathbb{R}^2$, that is in the xy-plane. From Definition **??**, if it passes through the point $\mathbf{a} = (a, b)$ in the direction $\mathbf{v} = (v, w)$, then it has the form $\gamma(t) = \mathbf{a} + t\mathbf{v}$. Note that this is a function $\gamma : \mathbb{R} \to \mathbb{R}^2$, because it takes one input, t, and outputs a vector with two components. Because $f : \Omega \to \mathbb{R}$ is a two-variable function (or, in other words, it takes as an input one vector with two components), we can substitute into it what we get out of γ:

$$\mathscr{F}(t) := f(\gamma(t)) = f(\mathbf{a} + t\mathbf{v}) = f(a + tv, b + tw).$$

Definition 4.21 Let $\Omega \subseteq \mathbb{R}^2$ and $f : \Omega \to \mathbb{R}$ have well-defined partial derivatives in each variable at $(a, b) \in \Omega$. The directional derivative of f at (a, b) along (v, w) is the derivative of the above function at $t = 0$, that is

$$\mathscr{F}'(0) = \lim_{h\to 0} \frac{\mathscr{F}(0 + h) - \mathscr{F}(0)}{h} = \lim_{h\to 0} \frac{f(a + hv, b + hw) - f(a, b)}{h}.$$

Example 4.22 Suppose we wish to find the directional derivative of $f(x, y) = x^2e^y$ at the point $(4, 0)$ along $(1, 1)$. First of all, we see that the line defined by this point and direction is $\gamma(t) = (4 + t, t)$. Hence, we can construct the one-variable function

$$\mathscr{F}(t) = f(\gamma(t)) = (4 + t)^2e^t.$$

According to Definition 4.21, we need to compute the derivative of this guy at $t = 0$. We can either do this (i) using the limit definition or (ii) using what we know about differentiating functions of one variable. We will do the second one, leaving the first for you to double-check. Indeed, we see from the Product and Chain Rules that

$$\mathscr{F}'(t) = 2(4 + t)e^t + (4 + t)^2e^t \qquad \Rightarrow \qquad \mathscr{F}'(0) = 8 + 16 = 24.$$

Lemma 4.23 *Let $\Omega \subseteq \mathbb{R}^2$ and $f : \Omega \to \mathbb{R}$ have well-defined partial derivatives in each variable at $(a, b) \in \Omega$. The directional derivative of f at (a, b) along (v, w) is*

$$\nabla f(a, b) \cdot (v, w).$$

Proof: We will use the linear approximation formula that we found in Proposition 4.17, namely $f(x,y) \approx f(a,b) + f_x(a,b)(x-a) + f_y(a,b)(y-b)$. However, this means precisely that

$$f(x,y) = f(a,b) + f_x(a,b)(x-a) + f_y(a,b)(y-b) + O(x^2+y^2),$$

where $O(x^2+y^2)$ are "error terms" in x^α and y^β for $\alpha, \beta \geq 2$. Fortunately, we do not need them for this construction. In Definition 4.21, we concern ourselves with the second limit. As such, we need an expression for $f(a+hv, b+hw)$. This is achieved by substituting $(x,y) = (a+hv, b+hw)$ into the above equation:

$$f(a+hu, b+hv) = f(a,b) + f_x(a,b)hv + f_y(a,b)hw + O(u^2+v^2).$$

Consequently, we conclude that the directional derivative is

$$\begin{aligned}
\mathscr{F}'(0) &= \lim_{h\to 0} \frac{f(a+hv, b+hw) - f(a,b)}{h} \\
&= \lim_{h\to 0} \frac{f_x(a,b)hv + f_y(a,b)hw + O(h^2)}{h} \\
&= \lim_{h\to 0} \big(f_x(a,b)v + f_y(a,b)w + O(h)\big) \\
&= f_x(a,b)v + f_y(a,b)w \\
&= \nabla f(a,b) \cdot (v,w).
\end{aligned}$$

□

Note: We used the notation $O(\cdot)$ in the above proof; this is big-O notation from analytic number theory and complexity theory. O stands for the German word *ordnung* (meaning "order"). The idea behind saying $f(x) = O(g(x))$ is that, as $x \to \infty$, the behaviours of the functions f and g are similar in this sense: there exists $M > 0$ and $k \in \mathbb{R}$ such that

$$\big|f(x)\big| \leq M\big|g(x)\big|, \qquad \text{when } x \geq k.$$

Exercise 34 Find the directional derivative of $f(x,y) = x\sin(y) + x^7$ at $(2, \pi/2)$ along $(3, \sqrt{5})$ by using Lemma 4.23. Check your answer by using Definition 4.21.

Remark 4.24 There is a very easy way to extend the definition of the directional derivative to not only functions defined with domains in $\mathbb{R}^2$, but whose domains live in a space $\mathbb{R}^n$ or arbitrary dimension. Indeed, let $\mathbf{v} \in \mathbb{R}^n$, $\Omega \subseteq \mathbb{R}^n$ and $f : \Omega \to \mathbb{R}$ be a function with well-defined partial derivatives in each variable at $\mathbf{a} \in \Omega$. The directional derivative of f at $\mathbf{a}$ along $\mathbf{v}$ is

$$\nabla f(\mathbf{a}) \cdot \mathbf{v} = f_{x_1}(\mathbf{a})v_1 + f_{x_2}(\mathbf{a})v_2 + \cdots + f_{x_n}(\mathbf{a})v_n.$$

Exercise 35 Find the directional derivative of $f(x, y, z) = x^2 + xy + z$ at $(1, 1, 1)$ along $(1, 2, 2)$.

Definition 4.25 Let $\mathbf{v} \in \mathbb{R}^n$, $\Omega \subseteq \mathbb{R}^n$ and $f : \Omega \to \mathbb{R}$ be a function with well-defined partial derivatives at $\mathbf{a} \in \Omega$. The directional derivative of f at $\mathbf{a}$ in the direction $\mathbf{u}$ is

$$\nabla f(\mathbf{a}) \cdot \frac{\mathbf{v}}{\|\mathbf{v}\|},$$

where $\|\mathbf{v}\| := \sqrt{v_1^2 + \cdots + v_n^2}$ is the magnitude of the vector as introduced in Chapter **??**.

Remark 4.26 The difference between Definition 4.25 and the definition in Remark 4.24 is subtle. In the former, we divide by the magnitude of the vector to get only the 'direction' that vector is pointing in. In the latter, we don't do this, and instead use the direction and length of the given vector. This is denoted by the use of terminology: *along* means we don't divide and *in the direction* means we do divide.

Exercise 36 Find the directional derivative of $f(x, y, z) = z^3 + 3x^2y^2 + \sin(z)$ at $(1, 4, 0)$ in the direction $(-2, 6, 3)$.

Proposition 4.27 *For $\Omega \subseteq \mathbb{R}^n$, suppose $f : \Omega \to \mathbb{R}$ is a function with well-defined partial derivatives at $\mathbf{a} \in \Omega$. If $\nabla f(\mathbf{a}) \neq 0$, then the gradient points in the direction that f changes most rapidly and its magnitude gives the maximal rate of change.*

Proof: Let $\theta \in [0, \pi]$ be the angle between $\nabla f(\mathbf{a})$ and suppose we take the directional derivative in the direction of a unit vector $\mathbf{u} \in \mathbb{R}^n$ (that is $\|\mathbf{u}\| = 1$). By (the generalisation of) Theorem **??**, we know that $\nabla f(\mathbf{a}) \cdot \mathbf{u} = \|\nabla f(\mathbf{a})\| \cos(\theta)$. Since $-1 \leq \cos(\theta) \leq 1$, this is maximal precisely when $\cos(\theta) = 1$, that is $\theta = 0$ and so $\nabla f(\mathbf{a})$ and $\mathbf{u}$ point in the same direction. □

Example 4.108 (Revisited) Recall we computed the directional derivative of $f(x, y) = x^2e^y$ at $(4, 0)$ along $(1, 1)$. We can now use Proposition 4.27 to determine the maximal rate of change of this function in this direction. Indeed, it is

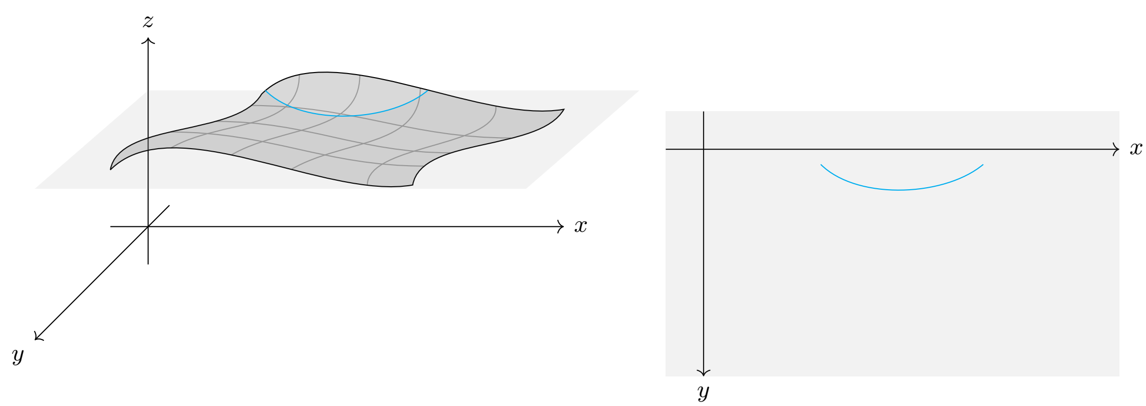
$$\|\nabla f(4, 0)\| = \|(8, 16)\| = 8\sqrt{5} \approx 17.888.$$

Definition 4.114 Let $\Omega \subseteq \mathbb{R}^n$, $f : \Omega \to \mathbb{R}$, $\mathbf{a} \in \Omega$ and $c \coloneqq f(\mathbf{a})$. The level set of f at $\mathbf{a}$ is

$$L_c(f) = \{\mathbf{x} \in \mathbb{R}^n : f(\mathbf{x}) = c\}.$$

Note: If $n = 2$, this is the contour of f at $\mathbf{a} = (a, b)$. A real-world application of this are the UK's Ordnance Survey maps; they show heights of terrain as inscribed smooth curves.

Remark 4.115 When $n = 2$, we know that the function $f : \Omega \to \mathbb{R}$ defines a surface (which we pictured in Figure 2). Although what we are about to say has an analogue in higher dimensions, it is easiest visualised and discussed when $n = 2$ (so the *graph* lives in three-dimensions). The level set of a two-variable function is the intersection of the surface in the xyz-space with the plane $z = c = f(\mathbf{a})$. This is pictured in Figure 3 below.



(a) The graph of $f : \Omega \to \mathbb{R}$. (b) The level set of $f : \Omega \to \mathbb{R}$.

Figure 3: A level set of a two-variable function $f : \Omega \to \mathbb{R}$.

Note: A level set is a special case of a fibre: for a map $f : X \to Y$ between sets and $y \in Y$, we say the fibre of y is the set $f^{-1}(y) \coloneqq \{x \in X : f(x) = y\}$ of elements that map to y.

From the proof of Proposition 4.27, we know that $\nabla f(\mathbf{a}) \cdot \mathbf{u} = \|\nabla f(\mathbf{a})\| \cos(\theta)$ for any unit vector $\mathbf{u} \in \mathbb{R}^n$. The gradient is orthogonal to the unit vector precisely when this formula is zero, i.e. when $\cos(\theta) = 0$. So, f is (locally) constant in the direction $\mathbf{u}$ which implies the next result.

Lemma 4.116 *For $\Omega \subseteq \mathbb{R}^n$, suppose $f : \Omega \to \mathbb{R}$ is a function with well-defined partial derivatives at $\mathbf{a} \in \Omega$. If $\nabla f(\mathbf{a}) \neq 0$, the gradient is orthogonal to the level set of f at $\mathbf{a}$.*

Proof: If $\mathbf{y} \in L_c(f)$ is a point in the level set of f at $c = f(\mathbf{a})$, choose a curve $\gamma : (-\varepsilon, \varepsilon) \to \mathbb{R}^n$

such that γ is contained in the level set $L_c(f)$ and $\gamma(0) = \mathbf{y}$. Then, $\mathscr{F}(t) = f(\gamma(t)) = c$, which implies that $\mathscr{F}'(0) = \nabla f(\mathbf{y}) \cdot \gamma'(0)0$. This tells us that the vectors $\gamma'(0)$ are orthogonal to the gradient of f at $\mathbf{y}$; they form the tangent plane to $L_c(f)$ at $\mathbf{y}$. □

Example 4.117 We will use a method based on Lemma 4.116 to find a unit vector perpendicular to the surface defined by $z = x^2 + y^2$ at the point $(1, 2, 5)$. Indeed, the defining equation can be re-written as $x^2 + y^2 - z = 0$. Therefore, this surface is actually the level set of the following function at zero:

$$f(x, y, z) \coloneqq x^2 + y^2 - z.$$

We quickly compute the gradient $\nabla f = (2x, 2y, -1)$, which is perpendicular to the surface. Hence, at $(1, 2, 5)$, the vector is $(2, 4, -1)$. It remains to divide through by its magnitude to get a vector of unit length, giving us the final answer

$$\frac{(2, 4, -1)}{\sqrt{21}}.$$

Exercise 37 Consider the function $f : \mathbb{R}^2 \to \mathbb{R}$ defined by $f(x, y) = x^2 + 3xy - x^2y + y^3$.

(i) Find the gradient vector ∇f.

(ii) Determine the rate of change of f at $(1, 1)$ in the direction $(3, -4)$.

(iii) At $(1, 1)$, find the maximum rate of increase of f and construct a unit vector pointing in the direction of this maximum rate of increase.

We conclude this subsection by making good on our promise in the proof, or lack thereof, of the Weak Chain Rule (Lemma 4.8).

Proposition 4.118 (Chain Rule for One Independent Variable) *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a function in the variables $x_1, ..., x_n$ and treat each $x_i : \mathbb{R} \to \mathbb{R}$ as differentiable functions of one independent variable t, say. Then, the derivative of $g(t) \coloneqq f(x_1(t), ..., x_n(t))$ is given by*

$$g'(t) = \frac{\partial f}{\partial x_1} x_1'(t) + \cdots + \frac{\partial f}{\partial x_n} x_n'(t).$$

Proof: We give an elementary proof for when $n = 2$ (the variables are x, y instead of x_1, x_2). Using the linear approximation of f at a point (a, b), we know that

$$f(a + \alpha, b + \beta) = f(a, b) + f_x(a, b)\alpha + f_y(a, b)\beta + O(\alpha^2 + \beta^2).$$

Using the Taylor series expressions after Definition 2.52, we now write write

$$x(t+h) = x(t) + x'(t)h + O(h^2) \qquad \text{and} \qquad y(t+h) = y(t) + y'(t)h + O(h^2).$$

For $a = x(t), \alpha = x'(t)h, b = y(t), \beta = y'(t)h$, we use limit definition of differentiation to see that

$$\begin{aligned} g'(t) &= \lim_{h\to 0} \frac{g(t+h) - g(t)}{h} \\ &= \lim_{h\to 0} \frac{f(x(t+h), y(t+h)) - f(x(t), y(t))}{h} \\ &= \lim_{h\to 0} \frac{f_x(x(t), y(t))x'(t)h + f_y(x(t), y(t))y'(t)h + O(h^2)}{h} \\ &= f_x(x(t), y(t))x'(t) + f_y(x(t), y(t))y'(t). \end{aligned}$$

□

Example 4.119 Consider the function $f : \mathbb{R}^3 \to \mathbb{R}$ given by $f(x, y, z) = \log(2x - 3y + 4z)$. We will apply Proposition 4.118 where we treat the variables as the following differentiable functions:

$$x = \exp(t), \qquad y = \log(t), \qquad z = \cosh(t).$$

We must first compute the partial derivatives of f with respect to each of the variables:

$$\frac{\partial f}{\partial x} = \frac{2}{2x - 3y + 4z}, \qquad \frac{\partial f}{\partial y} = \frac{-3}{2x - 3y + 4z}, \qquad \frac{\partial f}{\partial z} = \frac{4}{2x - 3y + 4z}.$$

It remains to compute the (usual) derivatives of the functions of t written at the top:

$$x'(t) = \exp(t), \qquad y'(t) = \frac{1}{t}, \qquad z'(t) = \sinh(t).$$

Substituting all this into the punchline of the previous proposition yields what we want, namely

$$f'(t) = \frac{2\exp(t) - 3\frac{1}{t} + 4\sinh(t)}{2x - 3y + 4z} = \frac{2\exp(t) - 3/t + 4\sinh(t)}{2\exp(t) - 3\log(t) + 4\cosh(t)}.$$

Exercise 38 Apply the Chain Rule for One Independent Variable to $f : \mathbb{R}^2 \to \mathbb{R}$ given by $f(x, y) = x^2 + y^2$, where we treat the variables as the following differentiable functions:

$$x(t) = \sin(t) \qquad \text{and} \qquad y(t) = t^3.$$

We have dealt with the situation where all our variables depend on one independent variable (parameter) t. Ultimately, we want to see what happens if all our variables depend on a collection of parameters $t_1, ..., t_k$. Before we make that jump, let's consider the situation where the dependence is on two independent variables s and t.

Proposition 4.120 (Chain Rule for Two Independent Variables) *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a function in the variables $x_1, ..., x_n$ and treat each $x_i : \mathbb{R}^2 \to \mathbb{R}$ as functions of two independent variables s and t, say, whose partial derivatives in each variable exist. Then, the partial derivatives of $g(s,t) := f(x_1(s,t), ..., x_n(s,t))$ are given by*

$$\frac{\partial g}{\partial s} = \frac{\partial f}{\partial x_1}\frac{\partial x_1}{\partial s} + \cdots + \frac{\partial f}{\partial x_n}\frac{\partial x_n}{\partial s} \qquad \text{and} \qquad \frac{\partial g}{\partial t} = \frac{\partial f}{\partial x_1}\frac{\partial x_1}{\partial t} + \cdots + \frac{\partial f}{\partial x_n}\frac{\partial x_n}{\partial t}.$$

Proof: Omitted. □

Note: A more general Chain Rule where $x_i : \mathbb{R}^k \to \mathbb{R}$ is given by the following formulae:

$$\frac{\partial g}{\partial t_j} = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}\frac{\partial x_i}{\partial t_j}, \qquad \text{where } j = 1, ..., k.$$

Remark 4.121 The note above provides us with the Chain Rule for k Independent Variables. However, there is a very slick way we can re-write the right-hand side of the expression here:

$$\frac{\partial g}{\partial t_j} = \nabla f \cdot \frac{\partial \mathbf{x}}{\partial t_j},$$

where $\mathbf{x} = (x_1, ..., x_n)$ and the notation $\partial \mathbf{x}/\partial t_j$ means that each entry of the vector $\mathbf{x}$ is partially-differentiated with respect to the same variable t_j. Similar notation appears in Section 5.

Exercise 39 Prove that for **any** function $f : \mathbb{R}^2 \to \mathbb{R}$ of two variables x and y where we treat the variables as $x(s,t) = \exp(s)\cos(t)$ and $y(s,t) = \exp(s)\sin(t)$, it satisfies

$$\sin(t)\frac{\partial f}{\partial s} + \cos(t)\frac{\partial f}{\partial t} = \exp(s)\frac{\partial f}{\partial y}.$$

[**Hint:** Use Proposition 4.120 to find $\partial f/\partial s$ and $\partial f/\partial t$; substitute into the left-hand side.]

Example 4.122 Let $f : \mathbb{R}^2 \to \mathbb{R}$ be **any** function of the variables z/x and x/y. We prove that

$$x\frac{\partial f}{\partial x} + y\frac{\partial f}{\partial y} + z\frac{\partial f}{\partial z} = 0.$$

The trick is to re-label the variables as $u := z/x$ and $v := x/y$. In reality, our function f is a function of three independent variables because u and v depend on x, y, z. Hence, we can apply the more general version of the Chain Rule discussed in the previous note to compute the derivatives we desire. We first determine all partial derivatives of $u = u(x, y, z)$ and $v = v(x, y, z)$:

$$\frac{\partial u}{\partial x} = -\frac{z}{x^2}, \qquad \frac{\partial u}{\partial y} = 0, \qquad \frac{\partial u}{\partial z} = \frac{1}{x},$$
$$\frac{\partial v}{\partial x} = \frac{1}{y}, \qquad \frac{\partial v}{\partial y} = -\frac{x}{y^2}, \qquad \frac{\partial v}{\partial z} = 0.$$

From the Chain Rule, we conclude the following:

$$\begin{aligned}
\frac{\partial f}{\partial x} &= \frac{\partial f}{\partial u}\frac{\partial u}{\partial x} + \frac{\partial f}{\partial v}\frac{\partial v}{\partial x} = -\frac{\partial f}{\partial u}\frac{z}{x^2} + \frac{\partial f}{\partial v}\frac{1}{y}, \\
\frac{\partial f}{\partial y} &= \frac{\partial f}{\partial u}\frac{\partial u}{\partial y} + \frac{\partial f}{\partial v}\frac{\partial v}{\partial y} = 0 - \frac{\partial f}{\partial v}\frac{x}{y^2}, \\
\frac{\partial f}{\partial z} &= \frac{\partial f}{\partial u}\frac{\partial u}{\partial z} + \frac{\partial f}{\partial v}\frac{\partial v}{\partial z} = \frac{\partial f}{\partial u}\frac{1}{x} + 0.
\end{aligned}$$

Substituting all this into the expression at the start of the example will imply the result:

$$x\frac{\partial f}{\partial x} + y\frac{\partial f}{\partial y} + z\frac{\partial f}{\partial z} = \frac{\partial f}{\partial v}\frac{x}{y} - \frac{\partial f}{\partial u}\frac{z}{x} - \frac{\partial f}{\partial v}\frac{x}{y} + \frac{\partial f}{\partial u}\frac{z}{x} = 0.$$

5 Matrix Calculus

Thus far, all our maps have had a co-domain of $\mathbb{R}$. Even as general as we were being at the end of the previous section (just above), we were still not hitting maximal generality. Truly, the most general situation one can apply the Chain Rule to (or calculus as a whole) is where we have $f : \mathbb{R}^n \to \mathbb{R}^m$ and $g : \mathbb{R}^k \to \mathbb{R}^n$ and we wish to differentiate the composition $f \circ g$.

The Jacobian Matrix

Definition 5.1 Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be a function in the variables $x_1, ..., x_n$, who themselves are functions of the variables $t_1, ..., t_m$ such that all relevant partial derivatives exist. The Jacobian matrix of the transformation $f(\mathbf{x}) = \mathbf{t}$ is the following $m \times n$ matrix:

$$J(x_1, ..., x_n) = \frac{\partial(t_1, ..., t_m)}{\partial(x_1, ..., x_n)} = \begin{pmatrix} \frac{\partial t_1}{\partial x_1} & \frac{\partial t_1}{\partial x_2} & \cdots & \frac{\partial t_1}{\partial x_n} \\ \frac{\partial t_2}{\partial x_1} & \frac{\partial t_2}{\partial x_2} & \cdots & \frac{\partial t_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial t_m}{\partial x_1} & \frac{\partial t_m}{\partial x_2} & \cdots & \frac{\partial t_m}{\partial x_n} \end{pmatrix}.$$

Notation 5.2 We abuse notation and instead write the Jacobian matrix as consisting of the partial derivatives $\partial f_i / \partial x_j$, where we think of the output of the function $f(\mathbf{x}) = (f_1(\mathbf{x}), ..., f_m(\mathbf{x}))$.

Note: The Jacobian determinant (or simply the Jacobian) is the determinant $\det(J)$ of this matrix. However, recall that from Chapter **??** that the determinant is only well-defined if the matrix is square (that is $m = n$). This will be important in Chapters **??** and **??**.

We can use the Jacobian matrix to capture the linear approximation information we have discussed earlier. Indeed, at least for a function of two variables $f : \mathbb{R}^2 \to \mathbb{R}$, recall that the linear approximation of f at (a, b) is given in Proposition 4.17 by

$$f(a, b) + f_x(a, b)(x - a) + f_y(a, b)(y - b).$$

Motivating the definition coming up next, we can re-write the above in vector notation as

$$f(a, b) + \nabla f(a, b) \begin{pmatrix} x - a \\ y - b \end{pmatrix}.$$

Definition 5.3 Let $f : \mathbb{R}^n \to \mathbb{R}^m$ and $\mathbf{a} \in \mathbb{R}^n$. The linear approximation of f at $\mathbf{a}$ is

$$f(\mathbf{x}) \approx f(\mathbf{a}) + J(\mathbf{a})(\mathbf{x} - \mathbf{a}).$$

Exercise 40 Determine the linear approximation of the following function at $\mathbf{a} = (2, 3)$:

$$f : \mathbb{R}^2 \to \mathbb{R}^3, \qquad f(x, y) = (x^2, 3y, x + y^3).$$

[**Hint:** Your final answer should be a (column) vector with three components.]

Changes of Variable

We can use the Jacobian matrix to determine or make changes of variables. In fact, we will see how the Jacobian (determinant) can help us with the task of changing from (x, y) to (s, t), say.

Example 5.4 Suppose we have two variables x, y given in terms of the variables s, t as follows:

$$x = s^2 t \qquad \text{and} \qquad y = t^2 s^{-1}.$$

Of course, we are assuming that $s \neq 0$ (otherwise s^{-1} is undefined). The task is to write s, t in terms of the variables x, y. To this end, we look for a combination of x and y that has no t-dependence, say; this will give us a way to write $s = s(x, y)$. Indeed, notice that

$$x = s^2 t \qquad \Rightarrow \qquad t = \frac{x}{s^2} \qquad \Rightarrow \qquad y = \frac{(x/s^2)^2}{s} = \frac{x^2}{s^5} \qquad \Rightarrow \qquad s = \frac{x^{2/5}}{y^{1/5}}.$$

We must work under the assumption that $y \neq 0$ (and therefore $t \neq 0$ and $x \neq 0$). Substituting this into the first equation will give us the rest of what we are after, namely

$$x = s^2 t \qquad \Rightarrow \qquad x = \frac{x^{4/5}}{y^{2/5}} t \qquad \Rightarrow \qquad t = x^{1/5} y^{2/5}.$$

Note: We can't always swap between pairs of variables via elementary means like above.

Theorem 5.5 *Let $f : \mathbb{R}^2 \to \mathbb{R}^2$ be given in the variables x and y, which are functions $x, y : \mathbb{R}^2 \to \mathbb{R}$ of s and t. Suppose all partial derivatives exist and are continuous.*

(i) *If the Jacobian $\det\big(J(a,b)\big) \neq 0$ for some $(a,b) \in \mathbb{R}^2$, then we can write s and t as functions of x and y with continuous partial derivatives, for all (s,t) "sufficiently close" to (a,b).*

(ii) *Let S be an open subset (i.e. no boundary) of the st-plane, where s and t are given in terms of x and y inside the set S and where $\det\big(J(a,b)\big) \neq 0$ for all $(a,b) \in S$. Then, the corresponding subset X in the xy-plane is also open. Moreover, if s and t are injective (functions of x and y) on S, then the inverse parametrisation from X to S (functions x and y of s and t) has continuous partial derivatives.*

Proof: Deferred; this is a special case of the Inverse Function Theorem. □

The point of Theorem 5.5 is this: if we wish to make a change of variables, we should **start** by restricting ourselves to points where the Jacobian is non-zero since, otherwise, it is hopeless.

Example 5.6 Consider the polar coordinates which we expressed in Lemma **??**:

$$x = r\cos(\theta) \qquad \text{and} \qquad y = r\sin(\theta),$$

where $r \geq 0$ is a non-negative real number and $\theta \in (-\pi, \pi]$. We again strive to write r and θ in terms of x and y. We know that this can be done (up to certain conditions) because we proved it back in Lemma **??**. However, the best we can do is to get $\theta : \mathbb{R}^2 \setminus \{(x,0) : x \leq 0\} \to (-\pi, \pi)$. This is a function which doesn't allow us to put in a half-line in the xy-plane and which misses π from its output. If we try to fix this, we will break the continuity of this function.

Exercise 41 Compute the Jacobian of the transformation in Example 5.6.

Note: In Example 5.6, we can compute $x^2 + y^2$ to eliminate θ and get an expression for r, but we must have that $r \neq 0$ (otherwise when we eliminate to get an expression for θ, we will have an undefined quantity). This is captured in the answer to Exercise 41.

From our discussions in Section 4, we can write the Chain Rule (for Two Independent Variables) as follows (compare this to the note after Proposition 4.120):

$$\begin{pmatrix} f_s \\ f_t \end{pmatrix} = \begin{pmatrix} x_s & y_s \\ x_t & y_t \end{pmatrix} \begin{pmatrix} f_x \\ f_y \end{pmatrix}.$$

The matrix above is simply the transpose of the Jacobian matrix. This equation therefore tells us that we can write f_x and f_y in terms of f_s and f_t **if** we can invert the (transpose of the) Jacobian matrix. This is why its determinant is so important!

Remark 5.7 In general, we can capture the Chain Rule from the previous section as follows:

$$\frac{\partial f}{\partial \mathbf{t}} = J(\mathbf{t})^T \frac{\partial f}{\partial \mathbf{x}}.$$

Proposition 5.8 *Let $x_1, ..., x_n : \mathbb{R}^n \to \mathbb{R}$ be functions in the variables $t_1, ..., t_n$ where the partial derivatives with respect to each variable exists. If the corresponding Jacobian matrix $J(\mathbf{x})$ is invertible, the the inverse is given by*

$$J(\mathbf{x})^{-1} = \left(\frac{\partial(t_1, ..., t_n)}{\partial(x_1, ..., x_n)} \right)^{-1} = \frac{\partial(x_1, ..., x_n)}{\partial(t_1, ..., t_n)} = J(\mathbf{t}).$$

Proof: Consider a map $f : \mathbb{R}^n \to \mathbb{R}^n$ in the variables $x_1, ..., x_n$ (which we already assume are parametrised in terms of $t_1, ..., t_n$). Looking at the Chain Rule expression in Remark 5.7, we apply it to the following n cases: $f(\mathbf{t}) = t_i$ for all $i = 1, ..., n$. In each case, we get something like the following, where the 1 is in the i^{th} entry in the left-hand vector:

$$\begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} = J(\mathbf{t})^T \begin{pmatrix} \partial t_i / \partial x_1 \\ \vdots \\ \partial t_i / \partial x_n \end{pmatrix}.$$

Combining all of these into a single equation (and using a property in Lemma **??**) tells us that

$$\mathbb{1}_n = J(\mathbf{t})^T J(\mathbf{x})^T = J(\mathbf{x}) J(\mathbf{t}).$$

Thus, the Jacobian matrices on the right-hand side are inverses to each other, by definition. $\square$

Method – Change of Variables: Let $x_1, ..., x_n$ be each given in terms of $t_1, ..., t_n$.

(i) Construct the Jacobian matrix $J(x_1, ..., x_n)$.

(ii) Find the Jacobian $\det(J)$ using the methods of Section **??**.

- If it is non-zero, move to Step (iii).
- If it is zero, we are done and we cannot change variables.

(iii) Compute the inverse matrix to determine $t_1, ..., t_n$ each in terms of $x_1, ..., x_n$.

Example 5.9 We will apply Method **??** to $x = s^2 + t^2$ and $y = st$ in order to write s and t in terms of the variables x and y. The first task is to construct the Jacobian matrix, and it is

$$J(x, y) = \begin{pmatrix} \dfrac{\partial x}{\partial s} & \dfrac{\partial x}{\partial t} \\ \dfrac{\partial y}{\partial s} & \dfrac{\partial y}{\partial t} \end{pmatrix} = \begin{pmatrix} 2s & 2t \\ t & s \end{pmatrix}.$$

Note that the Jacobian is $2s^2 - 2t^2$, which is non-zero if and only if $s \neq \pm t$. Now, the pairs $(-s, -t)$ and (t, s) each give the same x and y values as the pair (s, t). Therefore, the transformation is **not** injective, and it has no inverse. That being said, we can restrict to a domain on which the transformation is injective, e.g. we will assume that $s > \pm t$. Now, we see that

$$x + 2y = (s + t)^2 \qquad \text{and} \qquad x - 2y = (s - t)^2.$$

As such, we can take square-roots and solve the system simultaneously for s and t:

$$s = \frac{1}{2}\left(\sqrt{x + 2y} + \sqrt{x - 2y}\right) \qquad \text{and} \qquad t = \frac{1}{2}\left(\sqrt{x + 2y} - \sqrt{x - 2y}\right).$$

Exercise 42 Verify Proposition 5.8 in the situation where $x = s^2 + t^2$ and $y = st$.
[**Hint:** Use Example 5.9 to get $\frac{\partial s}{\partial x}, \frac{\partial s}{\partial y}, \frac{\partial t}{\partial x}, \frac{\partial t}{\partial y}$ and form $J(s, t)$; compare it to $J(x, y)^{-1}$.]

6 Optimisation Problems

7 Exercise Solutions

We provide detailed solutions to the exercises interwoven within each section of the module. Hopefully you have given these questions a try whilst on your learning journey with the module. But mathematics is difficult, so don't feel disheartened if you had to look up an answer before you knew where to begin (we have all done it)!

Solutions to Exercises in Section 2