MSIN0011 Calculus and Modelling

Matthew M. Jones

March 24, 2022

Contents

Week 1: Foundations

Sets

A **set** is a collection of **elements**. To describe a set we either list all its element (using **ellipses** for infinite sets), or use a **predicate**. For example:

- $C = \{\text{red}, \text{orange}, \text{yellow}\}$
- $\mathbb{N} = \{1, 2, 3, 4, \ldots\}$
- $P = \{p \colon p \text{ is a prime number}\}$

To indicate an element is contained in a set we use the symbol $\in$. For example,

$$\text{orange} \in C.$$

Similarly if an item is **not** in a set we use the symbol $\notin$. For example,

$$\text{blue} \notin C.$$

Subsets

If all the elements of a set are also elements of another set then we say the first set is a **subset** of the second. This is written, for example,

$$P \subset \mathbb{N}.$$

Unions and intersections

The **union** of two sets A and B is the set

$$A \cup B = \{x \colon x \in A \text{ or } x \in B\}.$$

The **intersection** of two sets A and B is the set

$$A \cap B = \{x \colon x \in A \text{ and } x \in B\}.$$

The empty set

The set that contains **no element** is called the **empty set** and is denoted by the special symbol

$$\emptyset$$

The empty set is a subset for every set. In notation, for any set A,

$$\emptyset \subset A.$$

Intervals

Intervals are especially useful sets of real numbers that lie between two values. The notation for intervals is the following:

$$\begin{aligned}[a,b] &= \{x\colon a \le x \le b\}\\ (a,b) &= \{x\colon a < x < b\},\\ (a,b] &= \{x\colon a < x \le b\},\\ [a,b) &= \{x\colon a \le x < b\}.\end{aligned}$$

- The values a and b are called the **endpoints**
- Intervals of the form $[a,b]$ are called **closed intervals**
- Intervals of the form (a,b) are called **open intervals**

Note: the values a and b must satisfy $a \le b$ otherwise the interval is the empty set. Also a or b or both can be $\pm\infty$.

For example,

$$\mathbb{R} = (-\infty, \infty)$$

Special sets

There are a number of special sets that we will routinely use so they have their own special characters:

- $\mathbb{N} = \{1, 2, 3, 4, \dots\}$, the **natural numbers**;
- $\mathbb{Z} = \{\dots, -2, -1, 0, 1, 2, \dots\}$, the **integers**;
- $\mathbb{Q} = \{p/q\colon p, q \in \mathbb{Z}, q \neq 0\}$, the **rationals**; and
- $\mathbb{R}$ the **real numbers**.

You might also have encountered

- $\mathbb{C} = \{a + \mathrm{i}b\colon a, b \in \mathbb{R}\}$, the **complex numbers**.

Note: we have the following chain of subsets,

$$\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R} \subset \mathbb{C}.$$

Note: there is also a bigger set of numbers called Hamiltonians that include the complex numbers as a subset, they're useful when manipulating 3d vectors in computer graphics. They are denoted $\mathbb{H}$.

Tuples

A **tuple** is an **ordered pair of items**, for example $(1, 2)$. If A and B are sets then the set of tuples (a, b) where $a \in A$ and $b \in B$ is denoted

$$A \times B.$$

Hence,

$$A \times B = \{(a, b) \colon a \in A, \text{and } b \in B\}.$$

For example,

$$[0, 1] \times [-1, 1] = \{(x, y) \colon 0 \leq x \leq 1, -1 \leq y \leq 1\}.$$

As a special case we write $A \times A = A^2$. For example,

$$\mathbb{R}^2 = \{(x, y) \colon x, y \in \mathbb{R}\}.$$

Tuples can be extended to an arbittary number of items, for example,

$$(3, -2, 1)$$

is a 3-tuple and in general an ordered list of n items is called a n-tuple.

The definitions above extend to these higher order tuples. For example,

$$[0, 2] \times [-1, 1] \times [2, 3] = \{(x, y, z) \colon x \in [0, 2], y \in [-1, 1], z \in [2, 3]\}$$

Hence $(1, 0, 2) \in [0, 2] \times [-1, 1] \times [2, 3]$, for example.

The set $\mathbb{R}^n$ is an important case, it consists of all the n-tuples of the form

$$(x_1, x_2, \dots, x_n)$$

where each x_i is a real number.

Functions

A **function** f is a rule that associates to each element of a set A - called the **domain** - an element of a second set B - called the co-domain. We write this as

$$f\colon A \to B$$

The range of the function f is the set of all elements of B that are mapped by f from some element of A. We write this as

$$f(A).$$

For example suppose $f\colon \mathbb{R} \to \mathbb{R}$ is defined by

$$f(x) = x^2.$$

Then $f(\mathbb{R}) = [0, \infty)$.

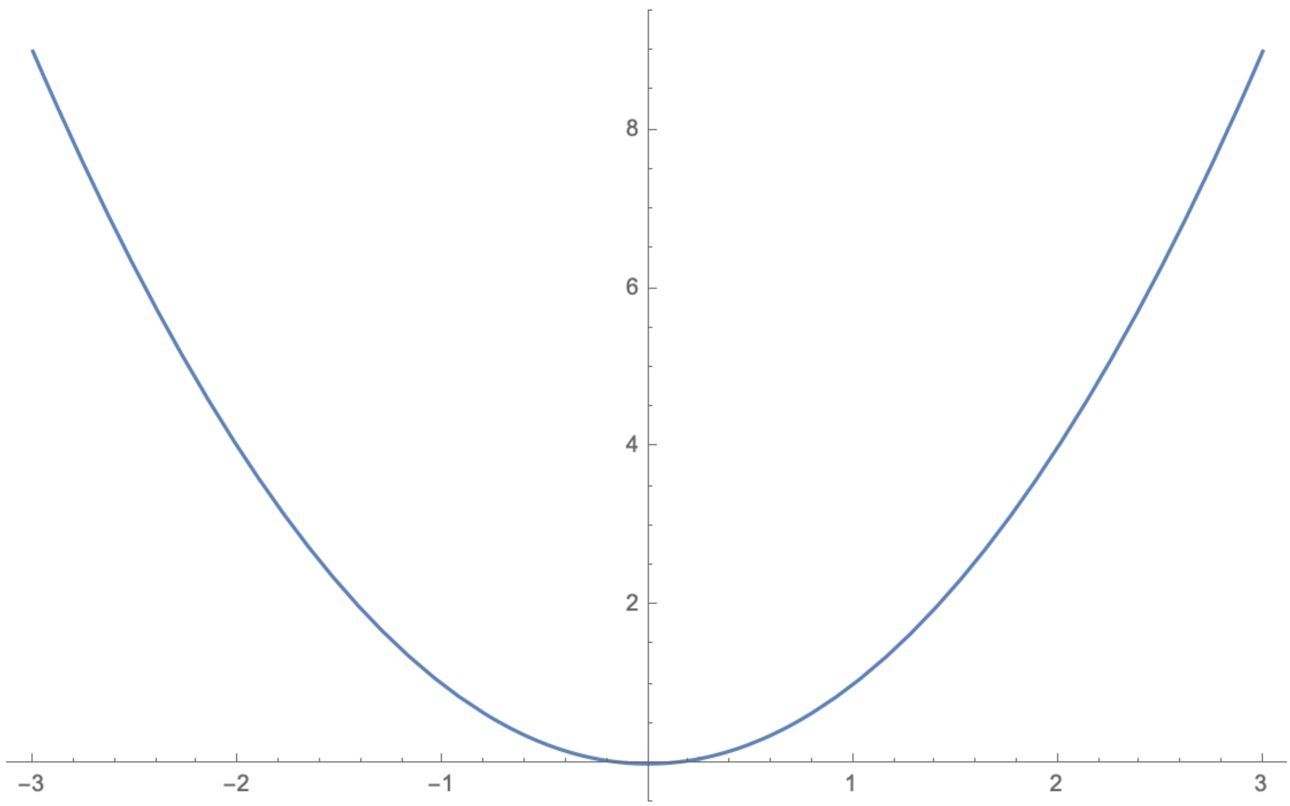


Figure 1: The graph of $f(x) = x^2$

The graph of this function is given in Figure 1.

Note: the graph of a function $f\colon A \to B$ is a subset of $A \times B$.

We often define a function using the following two-line format, where the special arrow $\mapsto$ is used to show that a variable is **mapped to** and expression.

$$\begin{aligned} f\colon A &\to B \\ x &\mapsto f(x) \end{aligned}$$

For example,

$$\begin{aligned} f\colon \mathbb{N} &\to \mathbb{N} \\ x &\mapsto 3x+1 \end{aligned}$$

Sometimes a function can be defined with different expressions in different subsets of its domain. In this case we use the following format.

$$\begin{aligned} f\colon \mathbb{R} &\to \mathbb{R} \\ x &\mapsto \begin{cases} 1, & \text{if } x < 1; \\ x^2, & \text{if } 1 \leq x \leq 2 \\ 4, & \text{if } x > 2. \end{cases} \end{aligned}$$

This is called a **piecewise** function. It graph is given in Figure 2.

Note: the range of this function is $f(\mathbb{R}) = [1,4]$.

The domain and range of a function need not be one of the **nice** sets we've seen so far. Let $X = (-1,1) \cup (2\pi, \infty)$. Then we may define,

$$\begin{aligned} g\colon X &\to \mathbb{R} \\ x &\mapsto \begin{cases} x^2, & \text{if } x \in (-1,1); \\ \sin x, & \text{if } x \in (2\pi, \infty) \end{cases} \end{aligned}$$

This has graph given below in Figure 3

Here, the range is $f(X) = [-1,1]$.

Function composition

We can compose two funcions f and g to produce a third, $f \circ g$ which is defined to be

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

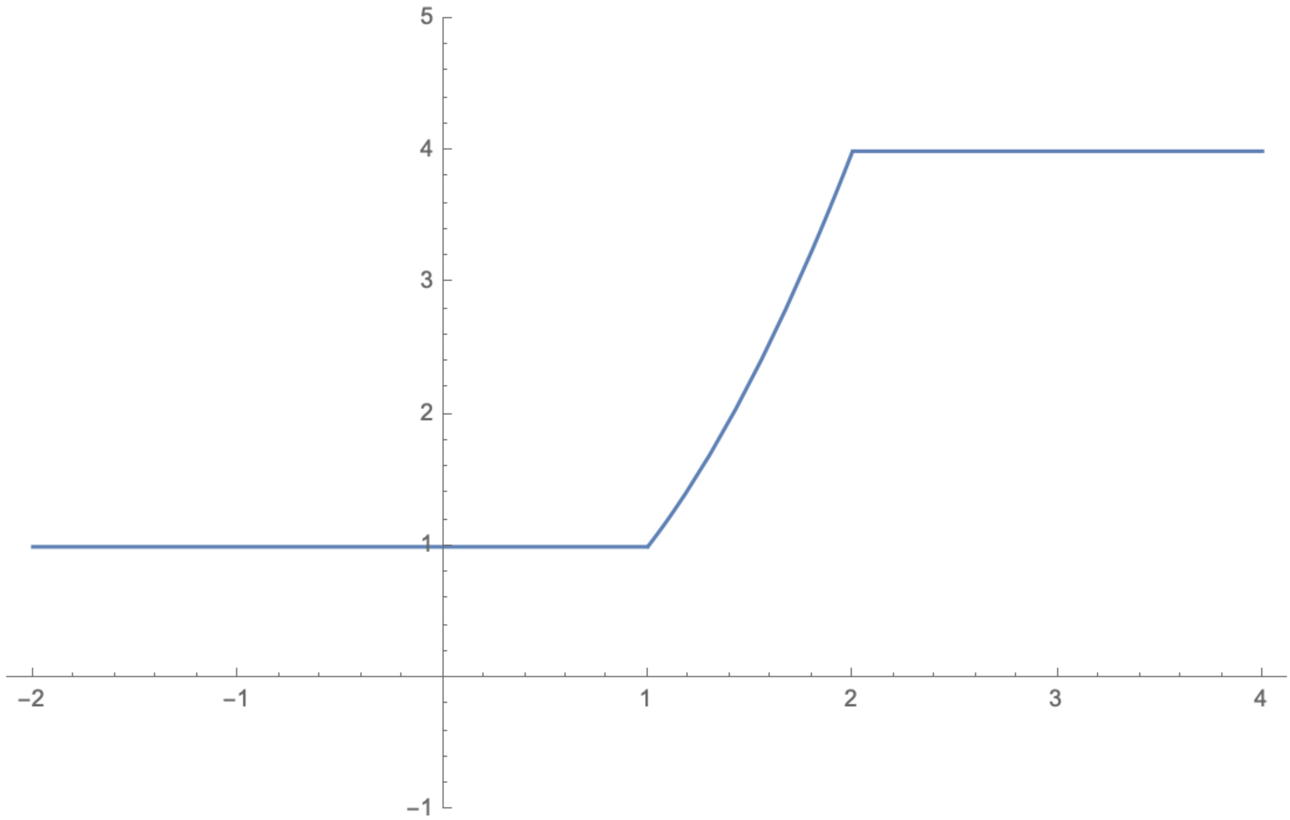


Figure 2: The graph of a piecewise function

for all x. But we do need to be careful. This is only possible if the range of g is a subset of the domain of f.

For example, suppose

$$f\colon [0,\infty) \to \mathbb{R}$$
$$x \mapsto \sqrt{x}$$

and

$$g\colon \mathbb{R} \to \mathbb{R}$$
$$x \mapsto \mathrm{e}^x.$$

Then $f \circ g$ is defined for all $x \in \mathbb{R}$ since the range of g satisfies

$$g(\mathbb{R}) = (0,\infty) \subset [0,\infty)$$

which is the domain of f.

On the other hand if we define,

$$h\colon \mathbb{R} \to \mathbb{R}$$
$$x \mapsto 1 - x^2.$$

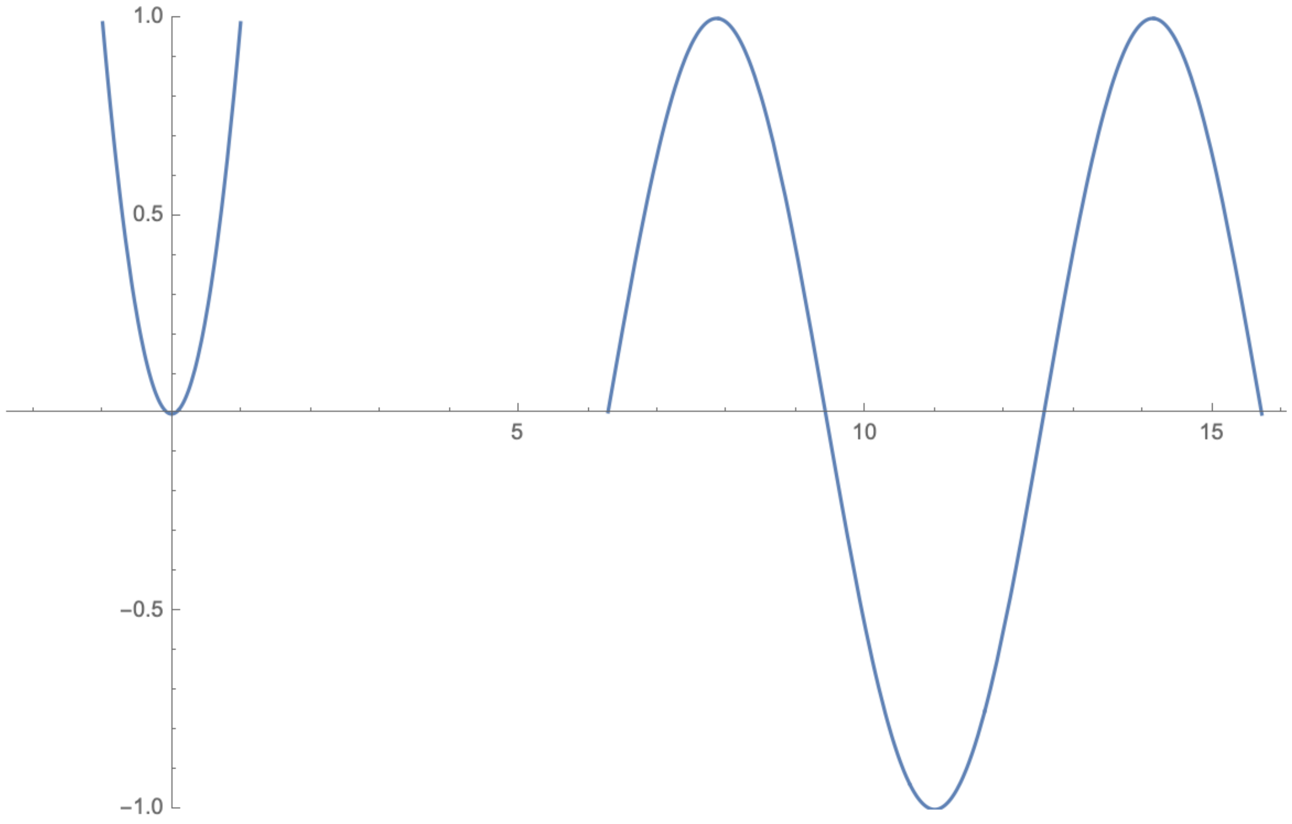


Figure 3: A domain that isn't a single interval

then $f \circ h$ cannot be defined because the range of h is

$$h(\mathbb{R}) = (-\infty, 1] \not\subset [0, \infty)$$

Injectivity

A function f is said to be **injective** or **one-to-one** if whenever $x \neq y$, $f(x) \neq f(y)$.

The following useful graphical technique can be used to determine if a function is injective:

- f is injective if, and only if **every** horizontal line intersects the graph of f at at most one point (i.e. either one point or no points)

For example, the function

$$\begin{aligned} f \colon \mathbb{R} &\to \mathbb{R} \\ x &\mapsto 1 - x - 2x^3 \end{aligned}$$

is injective. Its graph is in Figure 4.

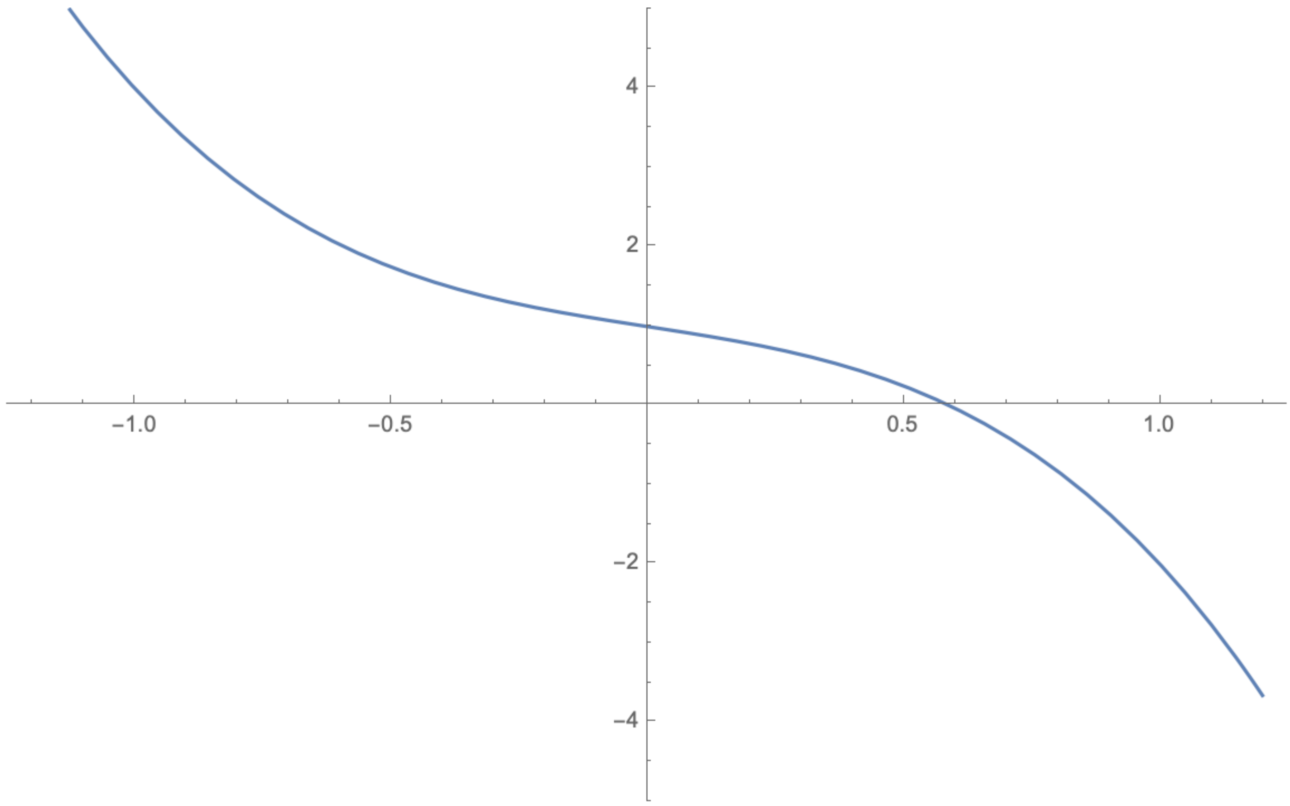


Figure 4: An injective function

Whereas the function

$$\begin{aligned} f\colon \mathbb{R} &\to \mathbb{R} \\ x &\mapsto 1 + x - 2x^3 \end{aligned}$$

is not injective. Its graph is given here, where the horizontal line drawn is an example of a line that intersects the graph at three points.

Surjectivity

A function $f\colon A \to B$ is said to be **surjective** or **onto** if

$$f(A) = B.$$

In other words if **every** element of B is mapped by some element of A.

For example,

$$\begin{aligned} f\colon \mathbb{R} &\to \mathbb{R} \\ x &\mapsto x^2 \end{aligned}$$

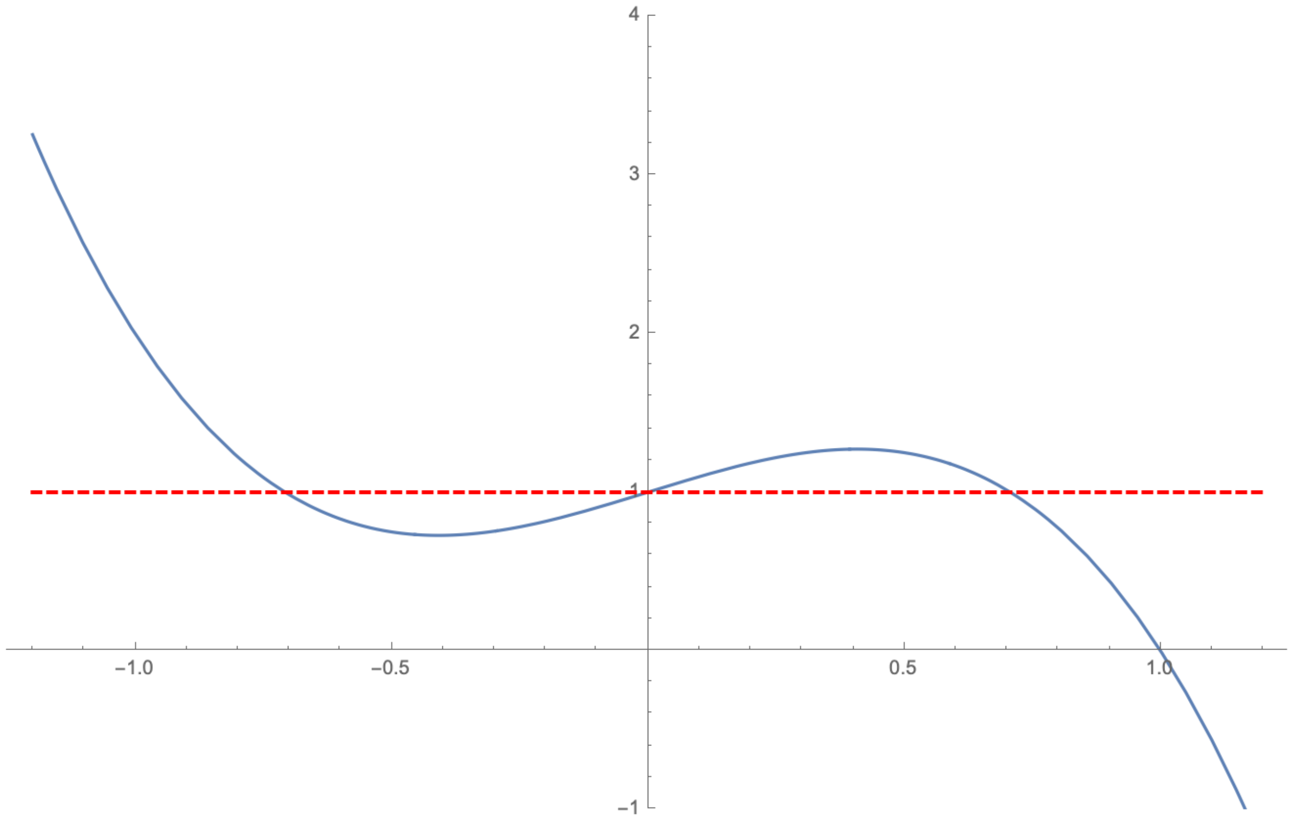


Figure 5: A non-injective function

is **not** surjective since $f(A) = [0, \infty) \neq \mathbb{R}$.

Note:

- surjectivity and injectivity have no relation, a function can be neither, either, or both.
- We always have the option of **making** a function $f\colon A \to B$ surjective by simply defining the co-domain B to be the range $f(A)$.

For example if we redefine the above function as

$$f\colon \mathbb{R} \to [0, \infty)$$
$$x \mapsto x^2$$

then this **is** surjective.

Bijectivity and inverses

A function f is said to be **bijective** if it is **both** injective **and** surjective.

A function $f\colon A \to B$ is said to be invertible with inverse $g\colon B \to A$ if for all $x \in A$

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

We normally write the **inverse function** g as f^{-1}.

A function f has an inverse if and only if it is bijective.

A useful technique for finding the inverse of a bijective function is to introduce a new variable, y say and change the subject of the equation $y = f(x)$. For example, to find the inverse of $f(x) = x^3 + 1$,

$$\begin{aligned} y &= x^3 + 1 \\ y - 1 &= x^3 \\ (y-1)^{1/3} &= x \end{aligned}$$

So that $f^{-1}(y) = (y-1)^{1/3}$.

Increasing and decreasing functions

A function $f\colon A \to \mathbb{R}$, where $A \subset \mathbb{R}$ defined to be:

- **inceasing** if whenever $a > b$, we have that $f(a) \geq f(b)$;
- **strictly increasing** if whenever $a > b$, we have that $f(a) > f(b)$;
- **decreasing** if whenever $a > b$, we have that $f(a) \leq f(b)$;
- **strictly decreasing** if whenever $a > b$, we have that $f(a) < f(b)$.

A function that satifies any of the conditions above is said to be **monotone**.

A function that is either stricly increasing or strictly decreasing is said to be **strictly monotone**.

Continuous functions - the basic idea

A function such as

$$\begin{aligned} f\colon \mathbb{R} &\to \mathbb{R} \\ x &\mapsto \begin{cases} x^2, & \text{if } x < 2; \\ 4, & \text{if } x \geq 2 \end{cases} \end{aligned}$$

is said to be **continuous** since its graph consists of a single connected curve (see Figure 6).

If the graph of a function is **not** a single connected curve then the function is said to be **discon-**

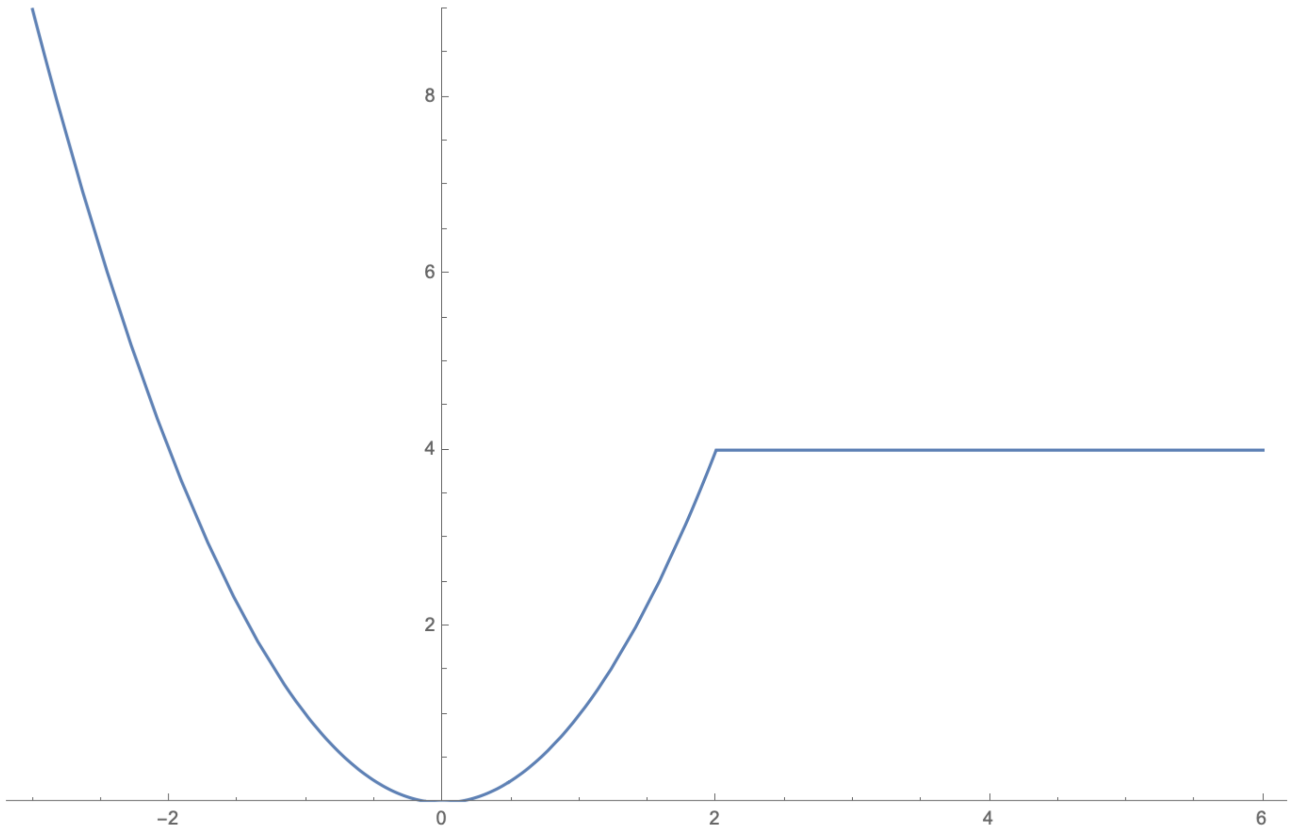


Figure 6: The graph of a continuous functio

tinuous. For example,

$$\begin{aligned} f\colon \mathbb{R} &\to \mathbb{R} \\ x &\mapsto \begin{cases} x^2, & \text{if } x < 2; \\ 2, & \text{if } x \geq 2 \end{cases} \end{aligned}$$

is a discontinuous function (Figure 7).

Suppose the domain of f is an interval.

Note: if f is both strictly monotone and continuous then it is injective.

This informal definition of continuity if fine to start with but what happens when we have more complicated functions. For example is the following function continuous? Its graph is given in Figure 8.

$$\begin{aligned} f\colon \mathbb{R} &\to [-1, 1] \\ x &\mapsto \begin{cases} \cos(1/x) & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases} \end{aligned}$$

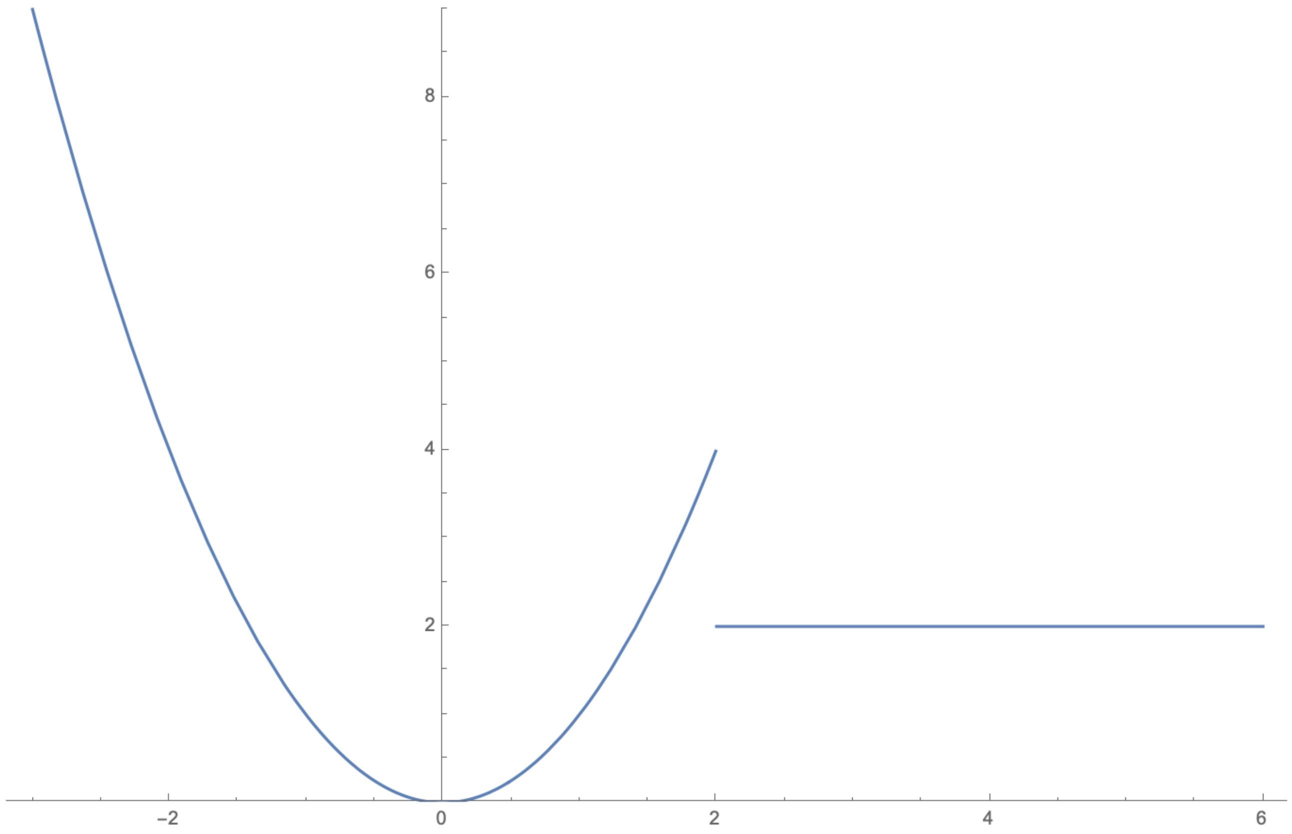


Figure 7: The graph of a discontinuous function

Limits

To understand continuity better we need to understand limits.

The function $f\colon [0,2] \to \mathbb{R}$ is defined to be

$$f(x) = \begin{cases} x^3, & x \neq 1 \\ 2, & x = 1 \end{cases}$$

The graph of this is given in Figure 9.

If x approaches 1 from either side, $f(x)$ approaches 1 also (even through $f(1) = 2$). We say that **the limit of $f(x)$ as x tends to 1**. This is written as

$$\lim_{x \to 1} f(x) = 1.$$

In general we can find the limit of a function at any point and this is written

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

and read **the limit of $f(x)$ as x tends to a**.

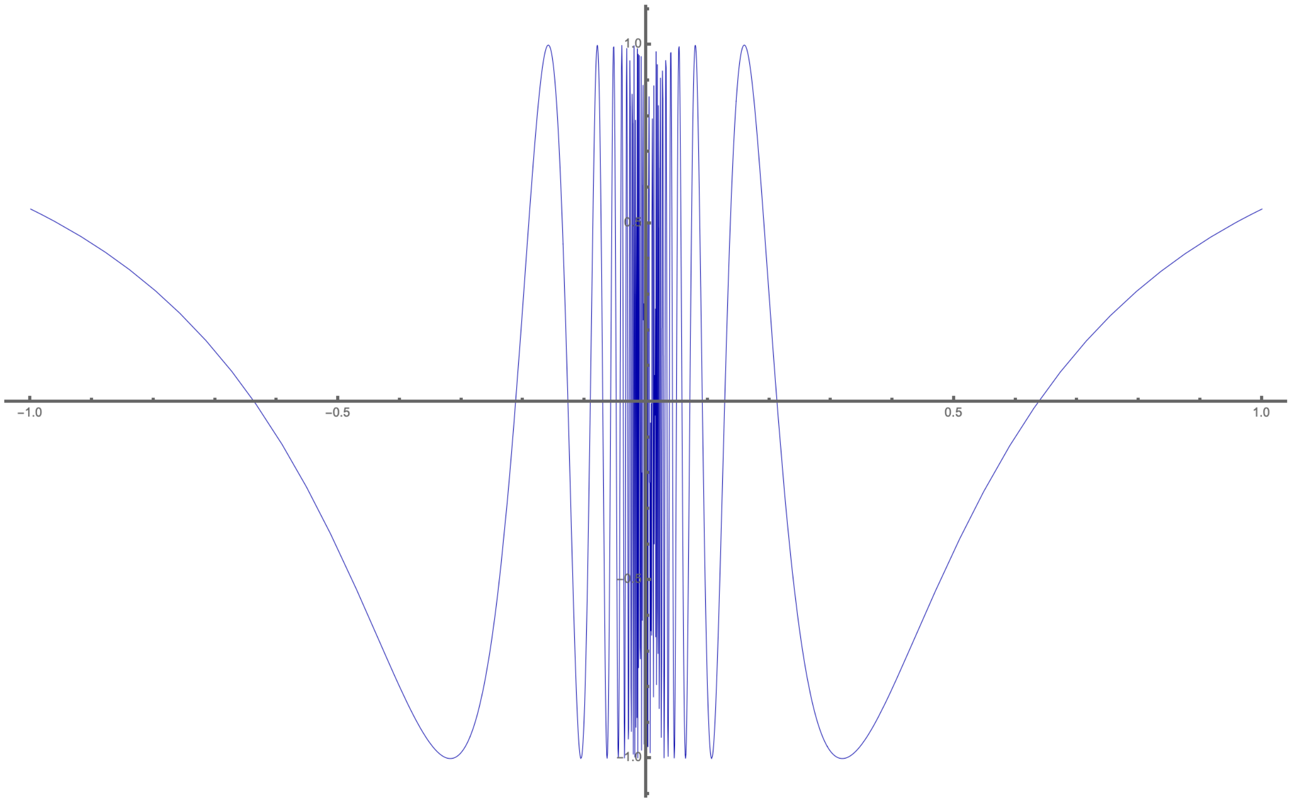


Figure 8: The graph of cos$(1/x)$

The idea of the limit is that the values $f(x)$ should be getting closer and closer to the limit as x gets closer and closer to a - although we don't need $f(a)$ to be defined or to be the same as the limit.

Applications of the limit: continuity

A function $f\colon A \to \mathbb{R}$ is said to be **continuous at** a if a is in the domain of f and

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

A function is said to be **continuous** if it is continuous at all points in its domain.

For example, the function we previously defined,

$$f(x) = \begin{cases} x^3, & x \neq 1 \\ 2, & x = 1 \end{cases}$$

$$\lim_{x \to 1} f(x) = 1 \neq f(1).$$

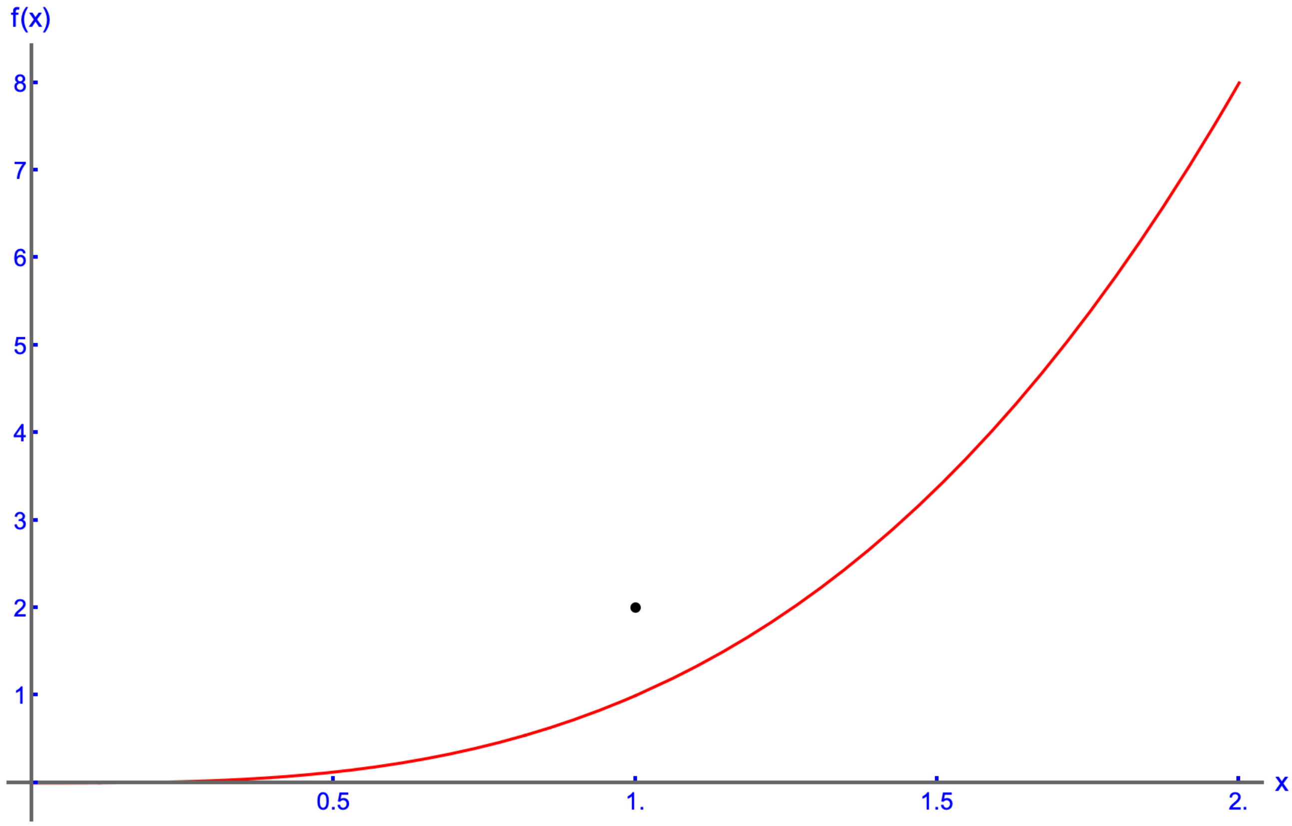


Figure 9: The limit of a function

Working with limits

To work with limits we introduce a number of important facts about them.

Theorem 1

If $L = \lim_{x \to a} f(x)$ then L is unique - i.e. it is not possible for a function to tend to two or more different limits.

Theorem 2

If $L = \lim_{x \to a} f(x)$ and $M = \lim_{x \to a} g(x)$ then:

1. $\lim_{x \to a} f(x) + g(x) = L + M$
2. $\lim_{x \to a} f(x)g(x) = LM$
3. $\lim_{x \to a} f(x)/g(x) = L/M$ as long as $M \neq 0$

For example, since

$$\lim_{x \to 2} x^2 = 4$$

$$\lim_{x \to 2} \cos \pi x = 1$$

we have that

$$\lim_{x \to 2} x^2 \cos \pi x = 4 \times 1 = 4.$$

We can also say that

$$\lim_{x \to 2} \frac{x^2}{\cos \pi x} = \frac{4}{1} = 4.$$

And so on.

Theorem 3

The following families of functions are continuous:

- Polynomials
- Exponential functions and logarithms
- The trigonometric functions sin, cos and tan
- The hyperbolic trigonometric functions sinh, cosh and tanh
- Any arithmetic combination of these, or any composition of functions including these

When limits don't exist

It's possible for a function not to have a limit. For example, define,

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

The graph of this function shown in Figure 10.

As x approaches 0 from the negative side (the left-hand side) $f(x)$ tends to -1.

On the other hand as x approaches 0 from the positive side (the right-hand side) $f(x)$ tends to 1.

This is an example where the limit does not exist at the point $a = 0$.

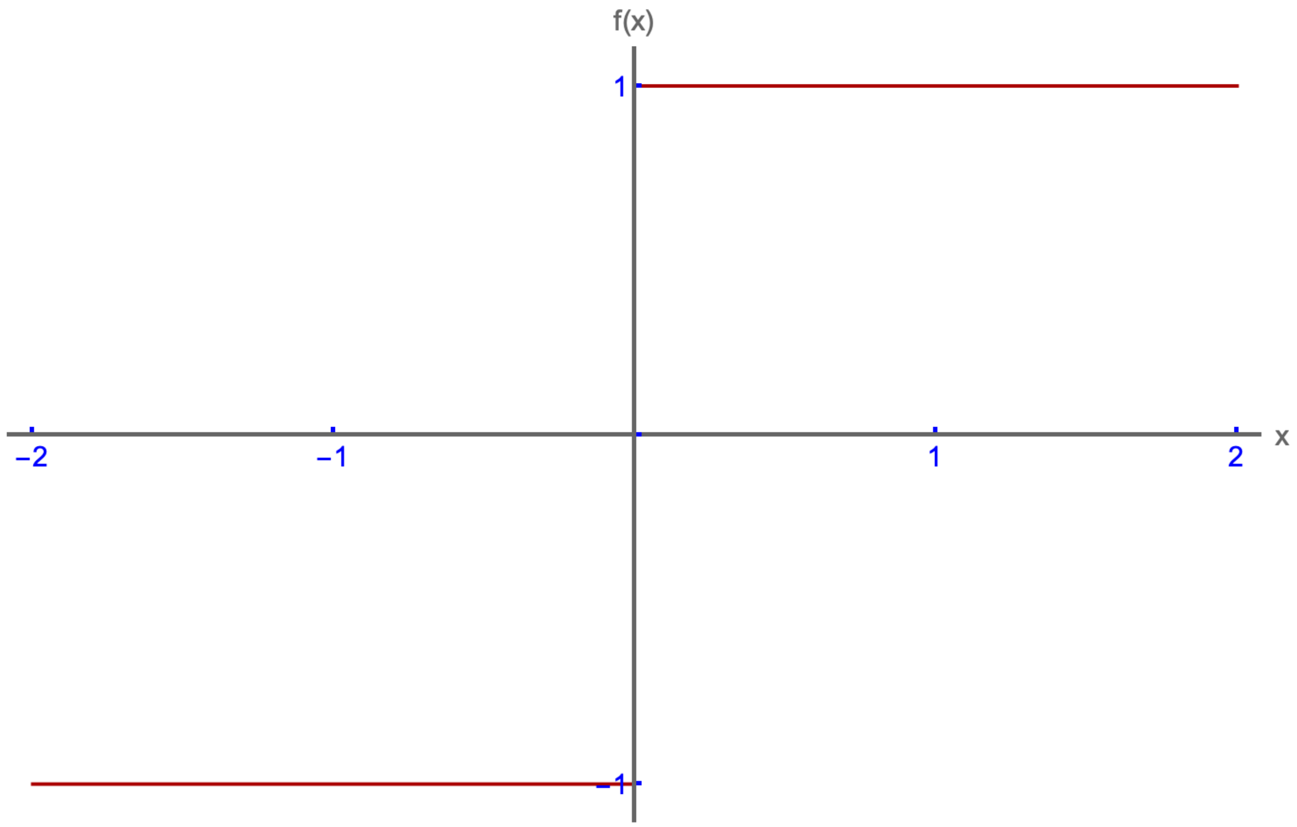


Figure 10: Grpah of a function without a limit

Week 2: Single Variable Calculus

Applications of the limit: differentiation

The derivative of a function $f\colon \mathbb{R} \to \mathbb{R}$ at a point a is the **gradient** of the **tangent** to the function at the point $(a, f(a))$ in its graph (Figure 11).

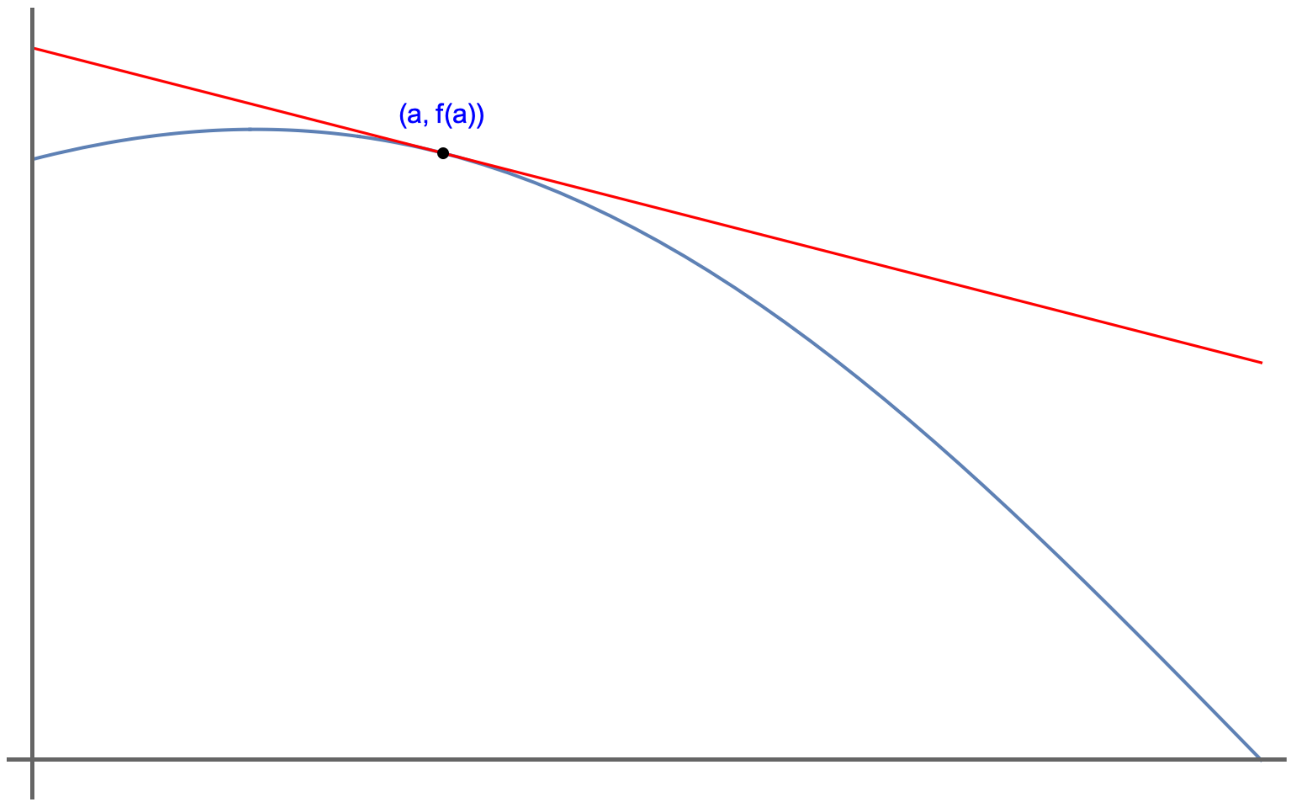


Figure 11: The derivative as the gradient of a tangent

The technique to calculate this is to calculate the gradient of successive approximations of the tangent as in Figure 12.

Take a value $h \neq 0$ and calculate the gradient of the line joining $(a, f(a))$ to $(a+h, f(a+h))$. We can calculate this gradient using the usual formula:

$$\begin{aligned} \text{gradient} &= \frac{\text{difference in } y \text{ values}}{\text{difference in } x \text{ values}} \\ &= \frac{f(a+h)-f(a)}{a+h-a} \\ &= \frac{f(a+h)-f(a)}{h} \end{aligned}$$

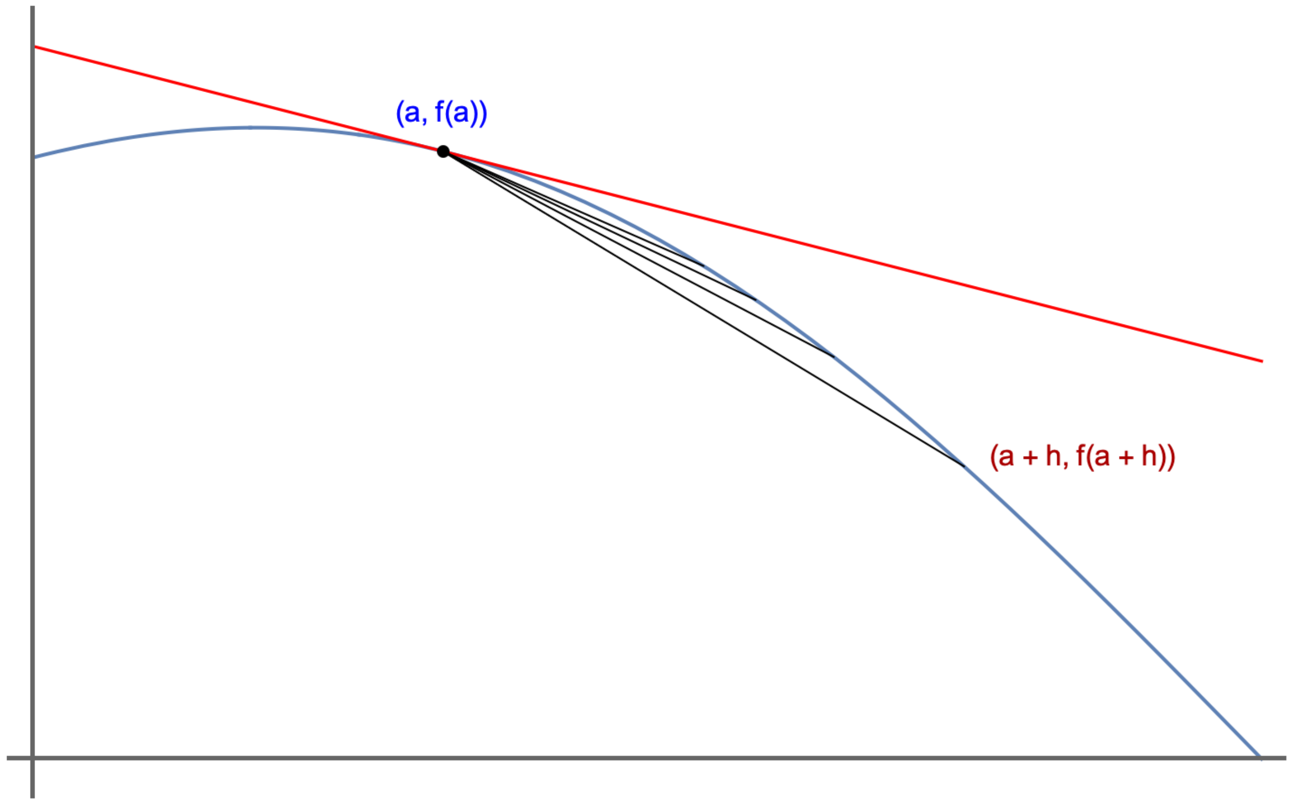


Figure 12: Successive approximations of the gradient

The derivative is then the following limit:

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

If we can find this limit then we call it the **derivative of f at a** and write it as

$$f'(a) \quad \text{or} \quad \frac{\mathrm{d}f}{\mathrm{d}x}(a)$$

For example suppose $f(x) = x^2$. Then to find the the derivative when $a = 2$ we calculate:

$$\begin{aligned}\frac{f(2+h)-f(2)}{h} &= \frac{(2+h)^2-4}{h} \\ &= \frac{4+4h+h^2-4}{h} \\ &= \frac{4h+h^2}{h}\end{aligned}$$

The graph of this function is given in Figure 13.

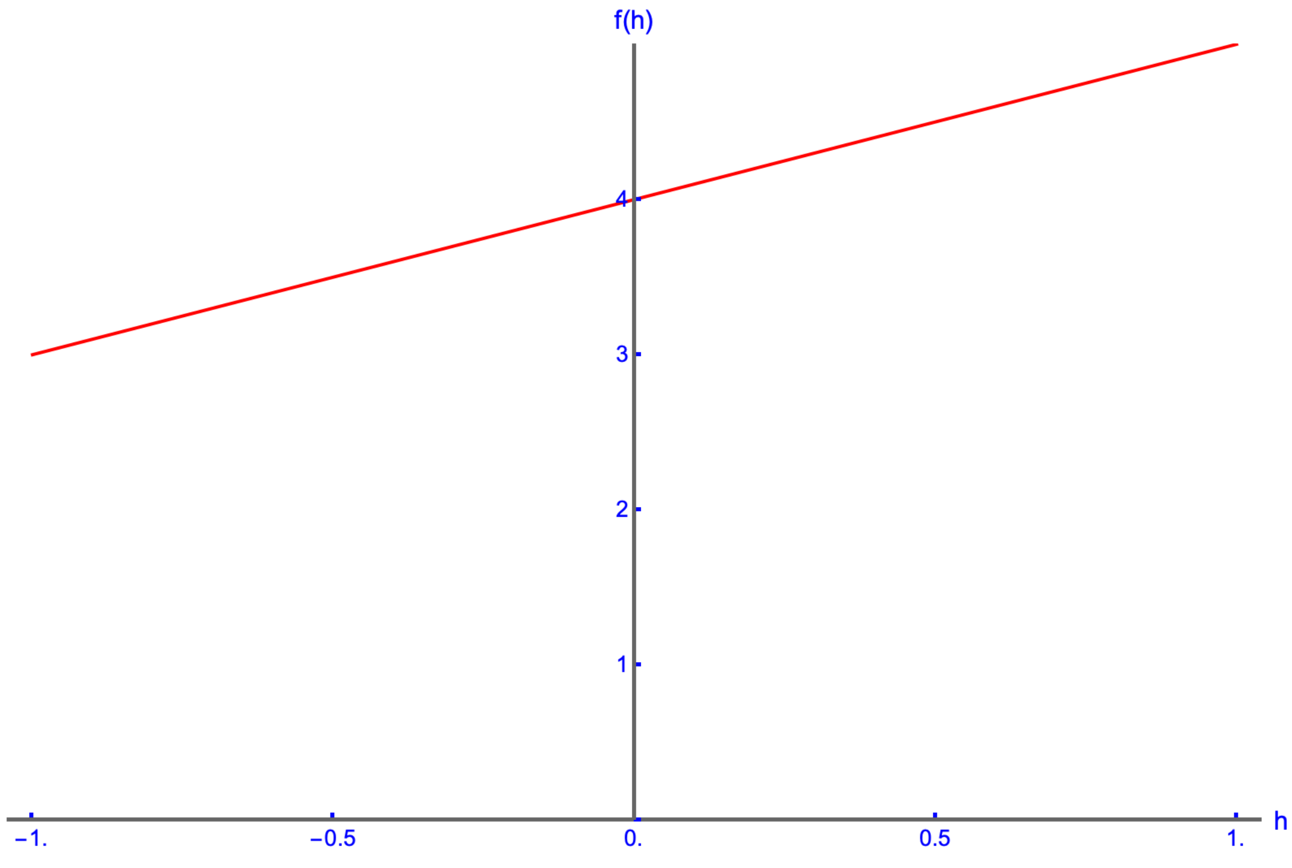


Figure 13: The derivative of the square function

Although this is not defined at $h = 0$, the limit is 4. Hence,

$$f'(2) = 4$$

Note: all the common rules of differentiation follow from this description of the derivative and the definition of the limit.

Suppose we do the same for $g(x) = |x|$. Let $h \neq 0$. Then to find the derivative at $a = 0$ we calculate,

$$\begin{aligned}\frac{g(0+h)-g(0)}{h} &= \frac{|h|}{h} \\ &= \begin{cases} -1, & h < 0; \\ 1, & h > 0. \end{cases}\end{aligned}$$

This is the example we have seen previously. The limit

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

does not exist.

Therefore, $|x|$ does not have a derivative at 0.

In fact this is self-evident if you consider the graph. There is no tangent at the point $(0,0)$ in its graph, Figure 14.

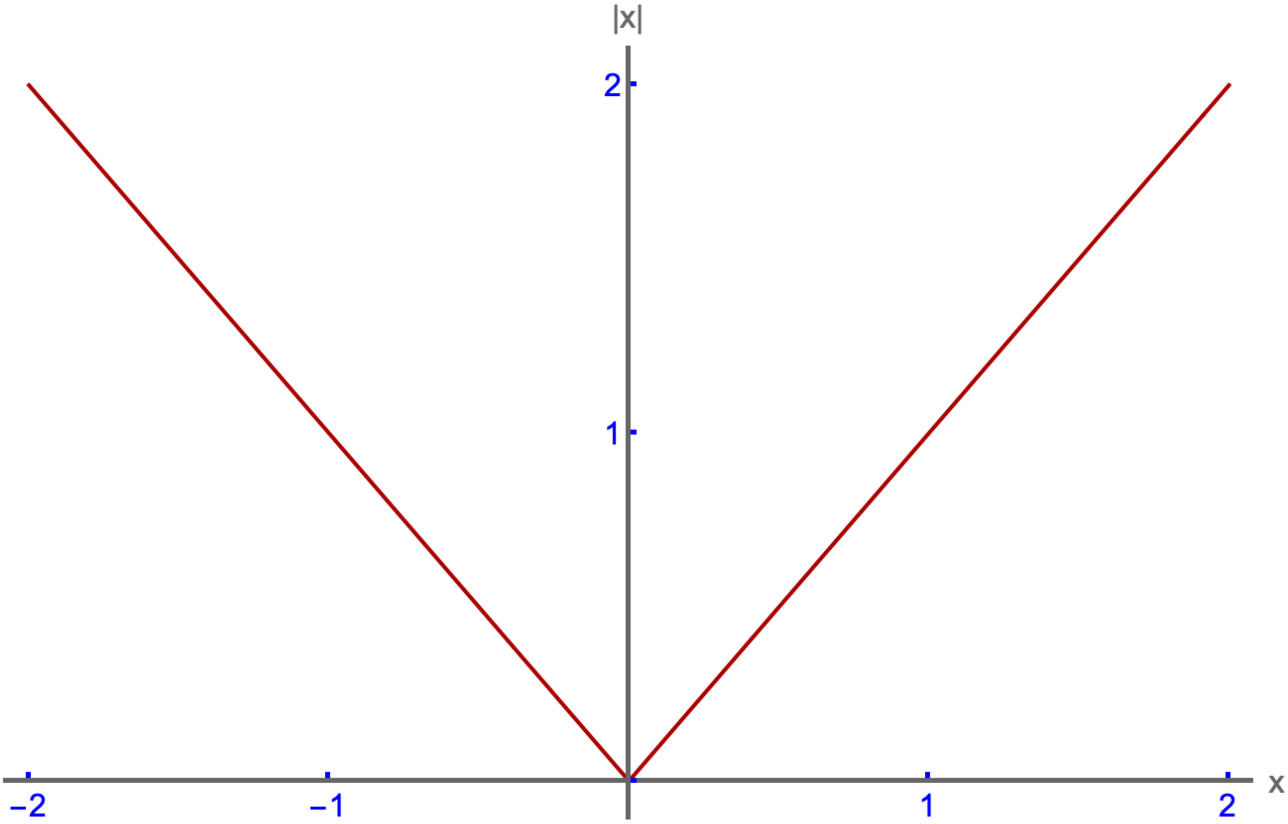


Figure 14: The graph of $|x|$

Differentiability

If a function f has a derivative at a point a then we say it is **differentiable at** a. A function that is differentiable at each point of its domain is called **differentiable**.

L'Hopital's rule

If $\lim_{x \to a} g(x) = 0$ then we can't normally calculate $\lim_{x \to a} f(x)/g(x)$.

However if we also know that $\lim_{x \to a} f(x) = 0$ then we may be able to do something.

For example, if $f(x) = 2x^2$ and $g(x) = x^2$ and $a = 0$. Then

$$\lim_{x \to 0} \frac{f(x)}{g(x)} = 2.$$

Another approach is to use the following result, which works for differentiable functions.

L'Hopital's rule

If

- $\lim_{x \to a} f(x) = 0$,
- $\lim_{x \to a} g(x) = 0$,
- $f'(a)$ and $g'(a)$ exist and $g'(a) \neq 0$,

then

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

To understand why this is true notice that f and g are continuous at a and so $f(a) = g(a) = 0$. And hence

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

For example, if $f(x) = e^x - 1$ and $g(x) = x$ then both these tend to 0 as x tends to 0. Hence,

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

Differentiable implies continuous

If a function has a derivative at a point a, $f'(a)$, then it is continuous there. To see this note that we need to show that $\lim_{x\to a} f(x) = f(a)$, or equivalently that $\lim_{x\to a} f(x) - f(a) = 0$. But

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

Higher order derivatives

The derivative of a differentiable function is itself a function with the same domain as f. If it is differentiable at a point a then we write this as $f''(a)$ and call it the **second derivative**.

In general the nth derivative of a function f at a point a is written as

$$f^{(n)}(a).$$

For the nth derivative to exists, the kth derivative must exist for $k = 1, 2, \dots, n$.

Smoothness of functions

By the previous two sections, if $f^{(n)}(a)$ exists then $f^{(n-1)}(a)$ not only exists, it is continuous there. In general however the derivative of a function need not be continuous. The example

$$\begin{aligned} f\colon \mathbb{R} &\to \mathbb{R}\\ x &\mapsto \begin{cases} x^2\sin(1/x) & \text{if } x \neq 0\\ 0 & \text{if } x = 0,\end{cases}\end{aligned}$$

has derivative $f'(0) = 0$. However, for other values of x,

$$f'(x) = 2x\sin(1/x) - \cos(1/x)$$

which doesn't have a limit at 0.

However for many results to be true we often need the derivative to be continuous. This leads to the following classification of **smoothness**.

Suppose X is an interval (or more generally any subset of $\mathbb{R}$).

- $C(X)$ is the set of all **continuous functions** whose domain is X
- $C^1(X)$ is the set of all **continuously differentiable functions** whose domain is X. Continuously differentiable means it has a derivative **and** the derivative is continuous.
- In general, $C^n(X)$ is the set of all functions that have n continuous derivatives.

- The set $C^{\infty}(X)$ denotes the collection of functions that have derivatives of **all** orders. These functions are often called **smooth** functions.

Approximating differentiable functions

If $f \in C^1(\mathbb{R})$ the equation of a tangent to the curve $y = f(x)$ at the point $(x_0, f(x_0))$ is given by

$$y = f(x_0) + f'(x_0)(x - x_0)$$

Since the tangent approximates the graph of f this gives us the following useful approximation to $f(x)$ for values of x close to x_0

$$f(x) \approx f(x_0) + f'(x_0)(x - x_0).$$

For example, for x near 1,

$$\sqrt{x} \approx 1 + \frac{1}{2}(x - 1)$$

This can be generalised to functions $f \in C^n(X)$. This is called the **Taylor polynomial**.

$$f(x) \approx f(x_0) + f'(x_0)(x - x_0) + \cdots + \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n.$$

For example,

$$\sqrt{x} \approx 1 + \frac{1}{2}(x - 1) - \frac{1}{8}(x - 1)^2 + \frac{1}{16}(x - 1)^3.$$

These can provide useful approximations for functions, but we will see they have much more practical use in obtaining information about functions from their derivatives.

Applications of the limit: integration

The integral of a function between two points a and b is the signed area between the graph of the function and the horizontal axis for $x \in [a, b]$, where "signed" means areas underneath the axis count negatively to the total area, as in Figure 15.

We write it as

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

Note:

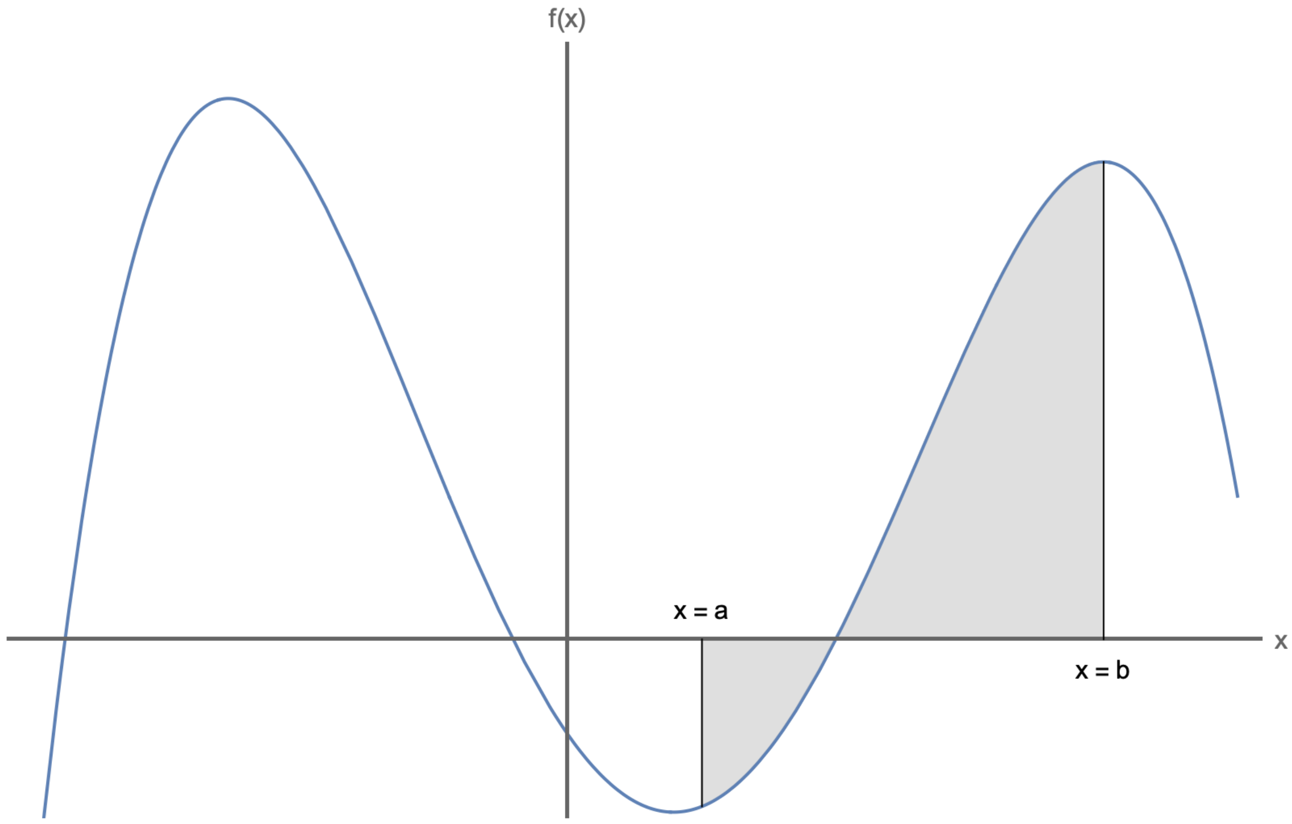


Figure 15: The integral of a function

- We don't define the integral as the opposite of differentiation as you may have done in your previous eduation. There is a link between the two that seems to imply that but it is not true in general.
- The integral is not defined as a limit, but it *almost* is, so we will develop it in this way.
- The area under a curve can be defined for functions that aren't continuous by taking different parts of the interval and calculating the areas separately.

Defining the integral properly is important, but difficult. One way is to partition the interval $[a, b]$ into a sequence of smaller strips as shown in Figure 16.

We then approximate these strips by rectangles. The main difficulty is ensuring that the limit, as the width of the stips tends to 0, is well-defined. This can be shown to be true in many cases and is called **Riemann Integration**, but the machinery you need to develop before you can do this is too far away from what we need to do in this module.

The important thing to understand is that if we think of the integral of a function as the area between its graph and the horizontal azis, then this makes sense even for functions that are not continuous.

Note: all the common rules of integration follow from this description of the integral.

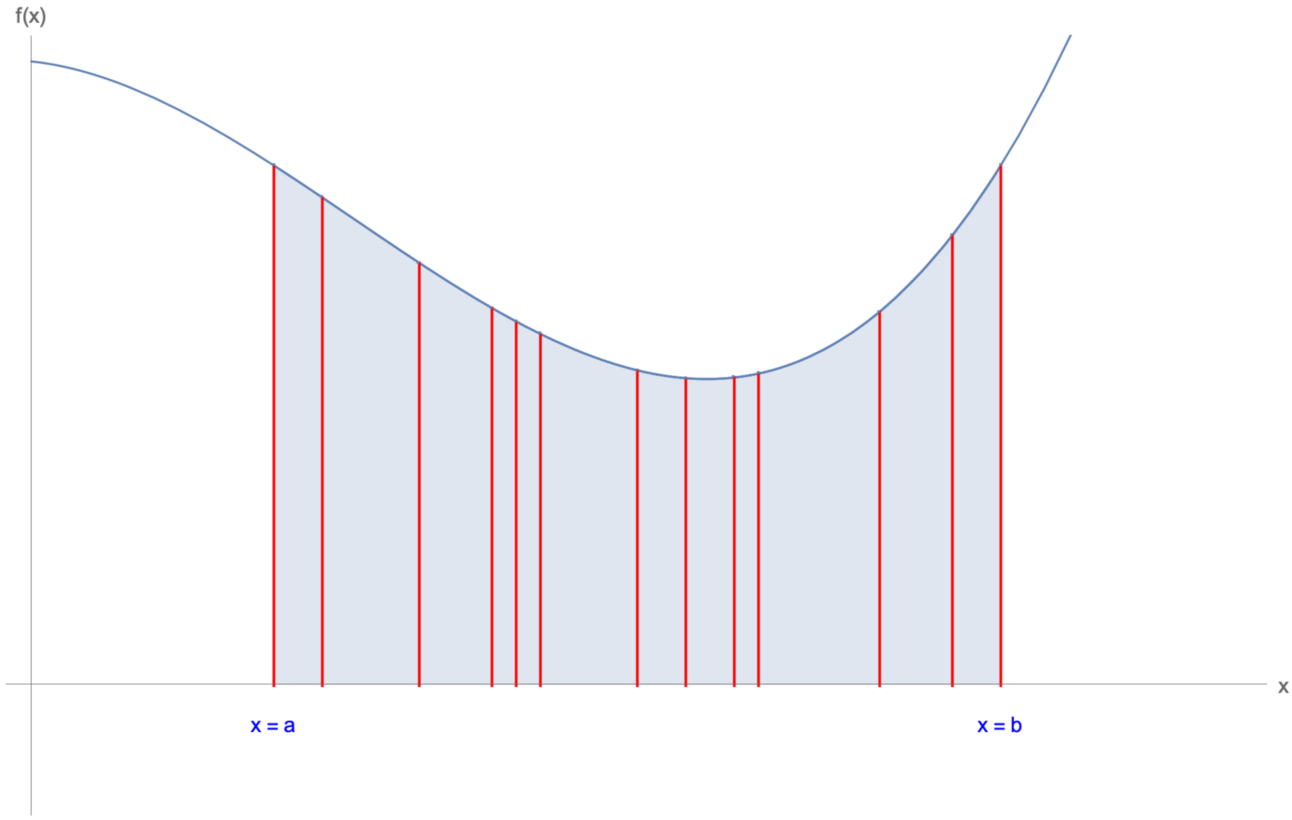


Figure 16: The definition of the integral

The fundamental theorem of calculus

We can bring all of these three ideas together: continuity, differentiation and integration. This highlights the link between integration and differentiation. It is called the Fundamental Theorem of Calculus.

Fundamental Theorem of Calculus

Let $f\colon [a,b] \to \mathbb{R}$, and define

$$G(x) = \int_a^x f(t)\mathrm{d}t$$

for $x \in [a,b]$.

1. G is continuous. Furthermore if f is continuous at $x_0 \in [a,b]$ then $G'(x_0)$ exists and is equal to $f(x_0)$.

2. If there is a function F such that $F'(x) = f(x)$ for all $x \in [a, b]$ then

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

The function F in the second part is called the **anti-derivative** or you may have seen it as the **indefinite integral**.

The Fundamental Theorem of Calculus says quite formally this: if f is **continuous** then it has an antiderivative, F, (i.e. $F' = f$), and

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

However, it gives us the ability to deal with more difficult functions.

For example, suppose,

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

Then f is not continuous, so it has no antiderivative. So, for example, the integral,

$$\int_{-1}^2 f(x)\mathrm{d}x$$

cannot be evaluated using the method you would normaly use, even though calculating the area between the graph and the horizontal axis is easy, and it's 1, see Figure 17

However, if we define $G(x)$ as in the theorem then

$$G(x) = \begin{cases} -x, & x < 0; \\ x, & x \geq 0. \end{cases}$$

You might recognise this as $|x|$. And we have already seen that this can't be differentiated at $x = 0$.

Improper integrals

So far we've seen integrals over intervals $[a, b]$. We also need to understand integrals over an **unbounded** interval - this means integrals where one or both of the limits are $\pm\infty$.

For example, what is

$$\int_1^\infty \frac{1}{x^2}\mathrm{d}x.$$

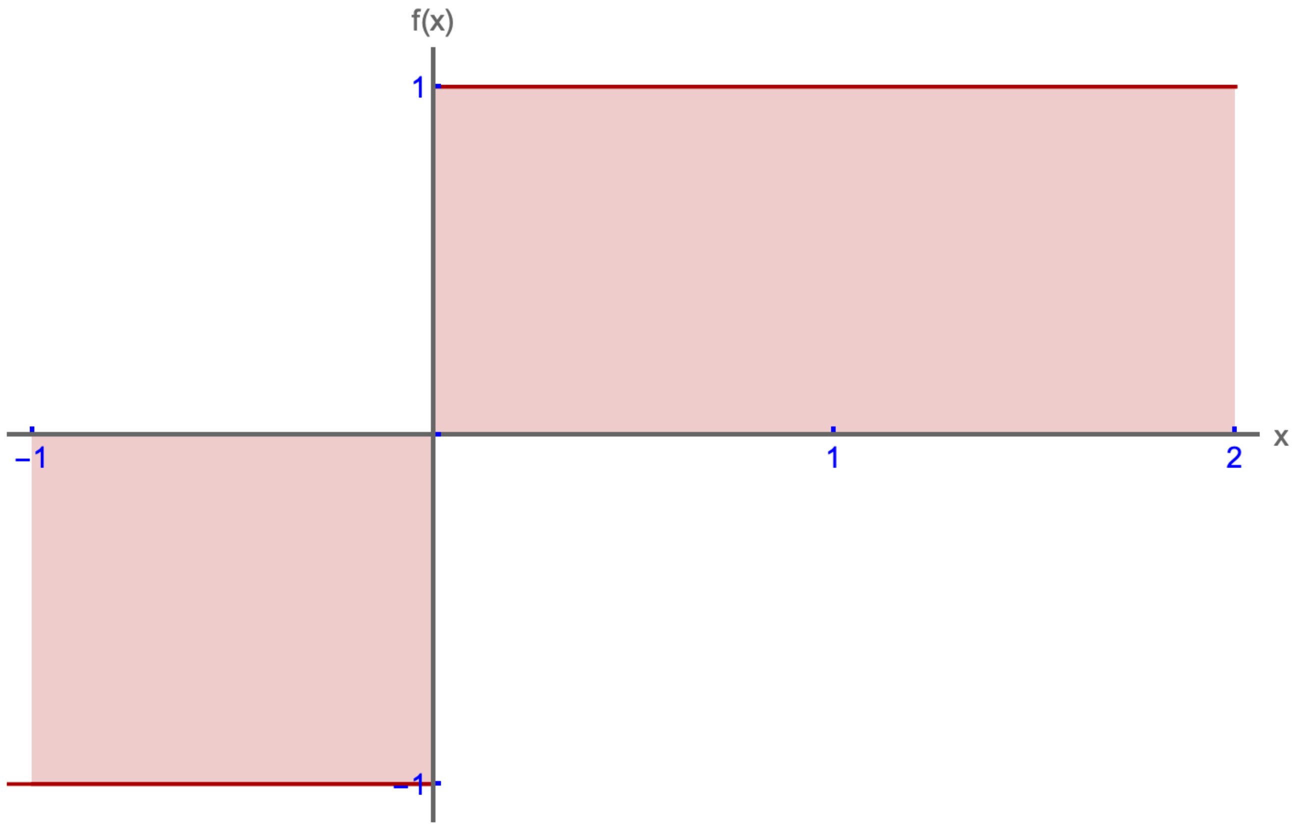


Figure 17: An example of an function without an antiderivative

These kinds of integrals are called **improper integrals**. We will approach this problem by approximating this improper integral by the **proper integral**

$$\int_1^y \frac{1}{x^2}\mathrm{d}x = 1 - \frac{1}{y}.$$

If we think of y as being a very large number then $1/y$ is a very small number. Indeed the bigger y gets, the smaller $1/y$ gets. In the end $1/y$ gets so vanishingly small we can conclude that **in the limit** this integral tends to 1.

We write this as follows:

$$\int_1^\infty \frac{1}{x^2}\mathrm{d}x = \lim_{y\to\infty} \int_1^y \frac{1}{x^2}\mathrm{d}x = 1.$$

Of course we haven't encountered limits at infinity before. So let's pause and gather some information.

Limits at infinity

Suppose $f\colon [0,\infty) \to \mathbb{R}$ is a function. We say that $f(x)$ has a limit at infinity, L say, written

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

if $f(x)$ gets closer and closer to L (and only L) as x gets larger and larger. More formally, if $f(x)$ is as close to L as we like for large values of x.

It is more useful to understand this by studying examples. First we have the following facts about limits.

Theorem 3

If $L = \lim_{x\to\infty} f(x)$ then L is unique - i.e. it is not possible for a function to tend to two or more different limits.

Theorem 4

If $L = \lim_{x\to\infty} f(x)$ and $M = \lim_{x\to\infty} g(x)$ then:

1. $\lim_{x\to\infty} f(x) + g(x) = L + M$
2. $\lim_{x\to\infty} f(x)g(x) = LM$
3. $\lim_{x\to\infty} f(x)/g(x) = L/M$ as long as $M \neq 0$

Example 1: rational functions

A **rational function** is a function of the form $p(x)/q(x)$ where p and q are polynomials.

For example, suppose $f(x) = 1/x^2$. Then as we argued earlier, the value of $1/x^2$ gets smaller and smaller as x gets larger and larger. So for this reason,

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

In general we have for any polynomial, p, of order 1 or above,

$$\lim_{x\to\infty} \frac{1}{p(x)} = 0.$$

We can use this to find the limits of any rational function using Theorem 4 and some algebra. The following example illustrates the technique.

Suppose

$$f(x) = \frac{2x^2 - x + 1}{3x^2 + 2x - 8}.$$

We argue as follows:

$$\begin{aligned}\lim_{x\to\infty} f(x) &= \lim_{x\to\infty} \frac{2x^2 - x + 1}{3x^2 + 2x - 8} \\ &= \lim_{x\to\infty} \frac{(2x^2 - x + 1)/x^2}{(3x^2 + 2x - 8)/x^2} \\ &= \lim_{x\to\infty} \frac{2 - 1/x + 1/x^2}{3 + 2/x - 8/x^2} \\ &= \frac{2 - 0 + 0}{3 + 0 - 0} = \frac{2}{3}\end{aligned}$$

For another example consider

$$g(x) = \frac{2x^2 + 5x}{4x^3 + 2x^2 - 3}$$

Here we have a quadratic divided by a cubic polynomial. We divide numerator and denominator by x^3, thus,

$$\begin{aligned}\lim_{x\to\infty} g(x) &= \lim_{x\to\infty} \frac{(2x^2 + 5x)/x^3}{(4x^3 + 2x^2 - 3)/x^3} \\ &= \lim_{x\to\infty} \frac{2/x + 5/x^2}{4 + 2/x - 3/x^3} \\ &= \frac{0 + 0}{4 + 0 - 0} = 0.\end{aligned}$$

As another example we consider the exponential function $\exp x = e^x$. Quite simply, $\exp -ax$ tends to 0 as x tends to ∞ for any $a > 0$. In fact it tends to 0 so quickly that

$$\lim_{x\to\infty} p(x) \exp(-x) = 0,$$

for **any** polynomial. For example,

$$\lim_{x\to\infty} x^4 \exp(-x) = 0.$$

In the seminar you will develop a library of examples that help you understand how to calculate limits at infinity, and hence improper integrals. We finish with this example.

What is,

$$\int_0^\infty \exp(-2x) \mathrm{d}x?$$

Here we go back to our limit definition of this improper integral.

$$\begin{aligned}\int_0^\infty \operatorname{exp}(-2x)\mathrm{d}x &= \lim_{y\to\infty}\int_0^y \operatorname{exp}(-2x)\mathrm{d}x \\ &= \lim_{y\to\infty}\frac{1}{2}\left(1-\operatorname{exp}(-2y)\right) \\ &= \frac{1}{2}.\end{aligned}$$

Applications of the integral: The mean value

Suppose $f\colon [a,b]\to\mathbb{R}$. The **mean value** of f over the interval $[a,b]$ is given by the formula:

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

For example suppose the interest rate for an inverstment is 2% for the first year and then steadily increases to 3.5% over the next three years, then we can model this as the following function.

$$\begin{aligned} r\colon [0,4] &\to \mathbb{R} \\ t &\mapsto \begin{cases} 2, & t\in[0,1] \\ 2+\frac{1}{2}(t-1) & t\in(1,4] \end{cases}\end{aligned}$$

The average interest rate over the four years is

$$\frac{1}{4}\int_0^4 r(t)\,\mathrm{d}t = \frac{41}{16}\approx 2.56$$

Extension material (optional)

The function

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

approaches -1 as x tends to 0 from the left-hand side. Similarly it approaches 1 as x tends to 0 for the right-hand side.

These two values are called the **left-hand limit** and the **right-hand limit**.

We write these as follows,

$$\begin{aligned}\lim_{x\to 0^-} f(x) &= -1, \\ \lim_{x\to 0^+} f(x) &= 1.\end{aligned}$$

This gives us a bit more flexibility.

Note: the limit exists if and only if **both** the left-hand and right-hand limits exist **and are equal**.

We can also now define the following. A function is said to be:

- **Continuous from the left** or simply **left continuous** at a if, $f(a) = \lim_{x \to a^-} f(x)$;
- **Continuous from the right** or **right continuous** at a if, $f(a) = \lim_{x \to a^+} f(x)$.

For example consider the **floor function**:

$$f(x) = \lfloor x \rfloor$$

which gives the largest integer smaller than x, e.g. $f(2.3) = 2$, $f(-12.3) = -13$, and so on.

This function is **right continuous** but not left continuous. Its graph is shown in Figure 18.

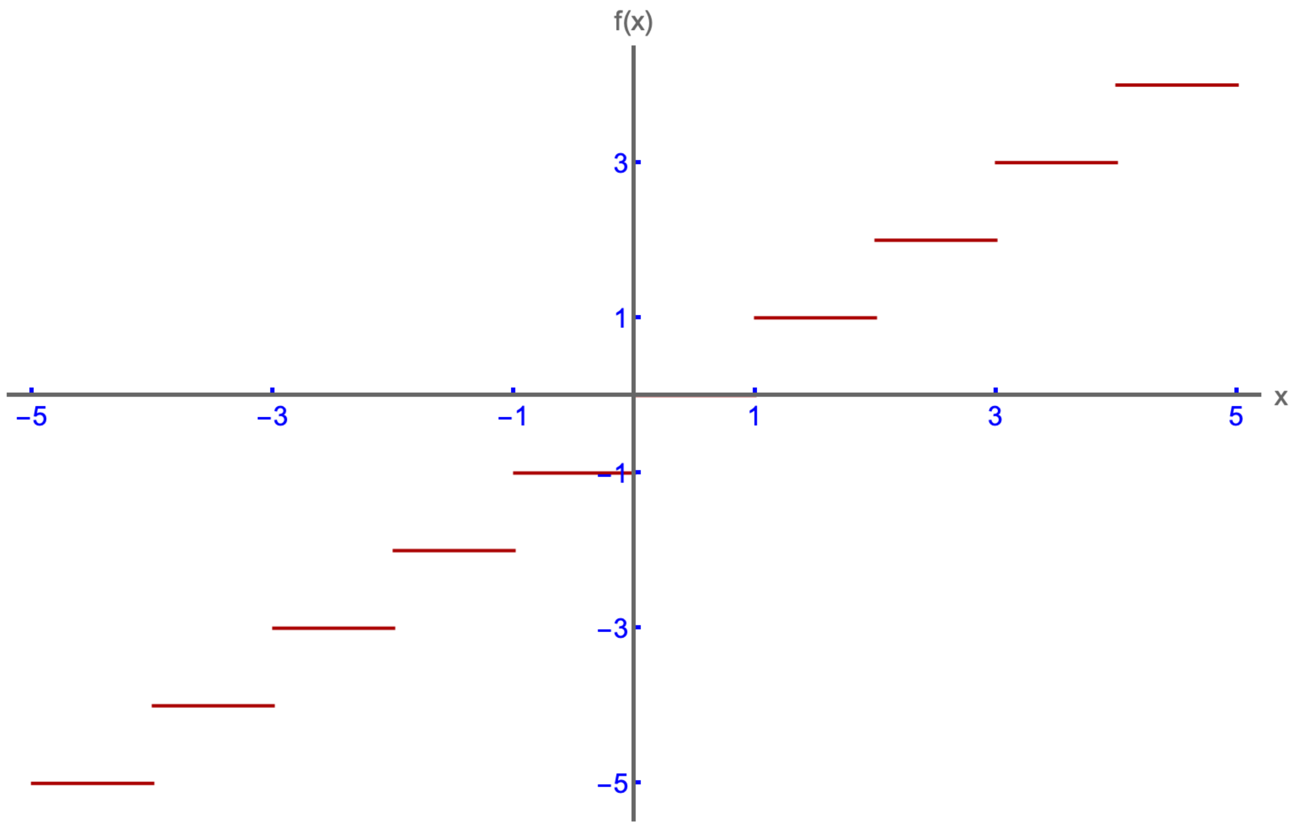


Figure 18: The graph of the floor function

Week 3: Multi-variable calculus

Multivariate functions and graphs

In most real-life applications of calculus to business modelling we need to introduce more than one variable to decribe different quantities. In Data Science we often have hundreds or even thousands of variables that describe some quantity. This means we must generalise the tecnhniques of calculus to these situations.

We have previously introduced $\mathbb{R}^n$ as the set of n-tuples of real numbers. We will also frequently find that it is useful to represent $\mathbb{R}^n$ as the set of $n \times 1$ column vectors.

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

The functions we will study are of the form

$$f \colon \mathbb{R}^n \to \mathbb{R}.$$

For example we might define

$$f(\mathbf{x}) = x^2 + y^2 + z^2$$

where $\mathbf{x} = (x, y, z)^T$ - note we often use the **transpose** to write a column vector on a single line like this.

We will also often abbreviate this to something like this:

$$f(x, y, z) = x^2 + y^2 + z^2$$

although it is **important** to always remember we are talking about vectors here.

Visualising graphs and tangents

Before we develop calculus more generally we will start by taking a visual approach. And to do this we will start by only considering functions $f \colon \mathbb{R}^2 \to \mathbb{R}$.

The graph of a function $f\colon \mathbb{R}^2 \to \mathbb{R}$ is a subset of $\mathbb{R}^2 \times \mathbb{R} = \mathbb{R}^3$.

We can plot the points $z = f(x, y)$ on a 3 dimensional plot - Mathematica does this quite nicely and allows you to move the graph around.

For example the graph of $f(x, y) = \sin(x + y^2)$ is shown in Figure 19.

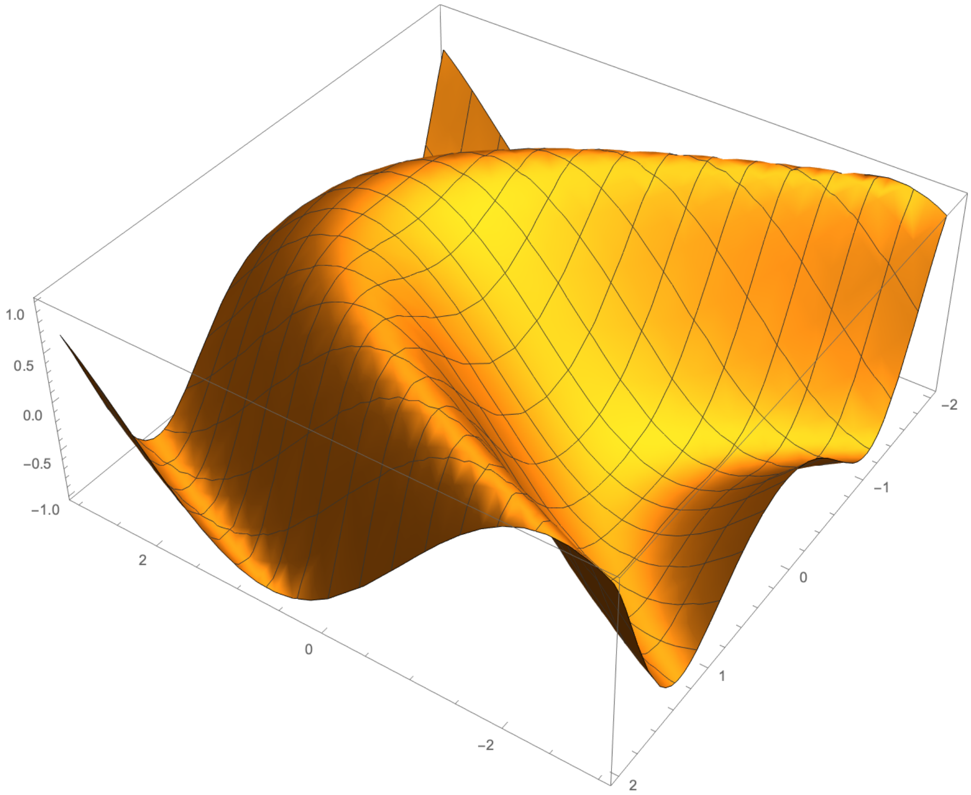


Figure 19: The graph of $z = \sin(x + y^2)$

Take a point on this surface. There is no single line that is tangent to the surface at our point. Instead of a tangent line however we have a **tangent plane**.

In the diagram in Figure 20 the tangent plane to the surface $z = x^2 + y^2$ is drawn at the point $(1, 2, 5)$.

A plane is a **2 dimensional object**. So, as with any plane, we need **two linearly independent** vectors to describe it.

For example, suppose we want to find the tangent plane to $z = x^2 + y^2$ at the point $(1, 2, 5)$ as shown in Figure 20.

The approach we take is to consider the point $(1, 2)$ in $\mathbb{R}^2$ and find the vectors on the tangent plane starting at $(1, 2)$ and moving either in a direction parallel to the x-axis, or x-axis.

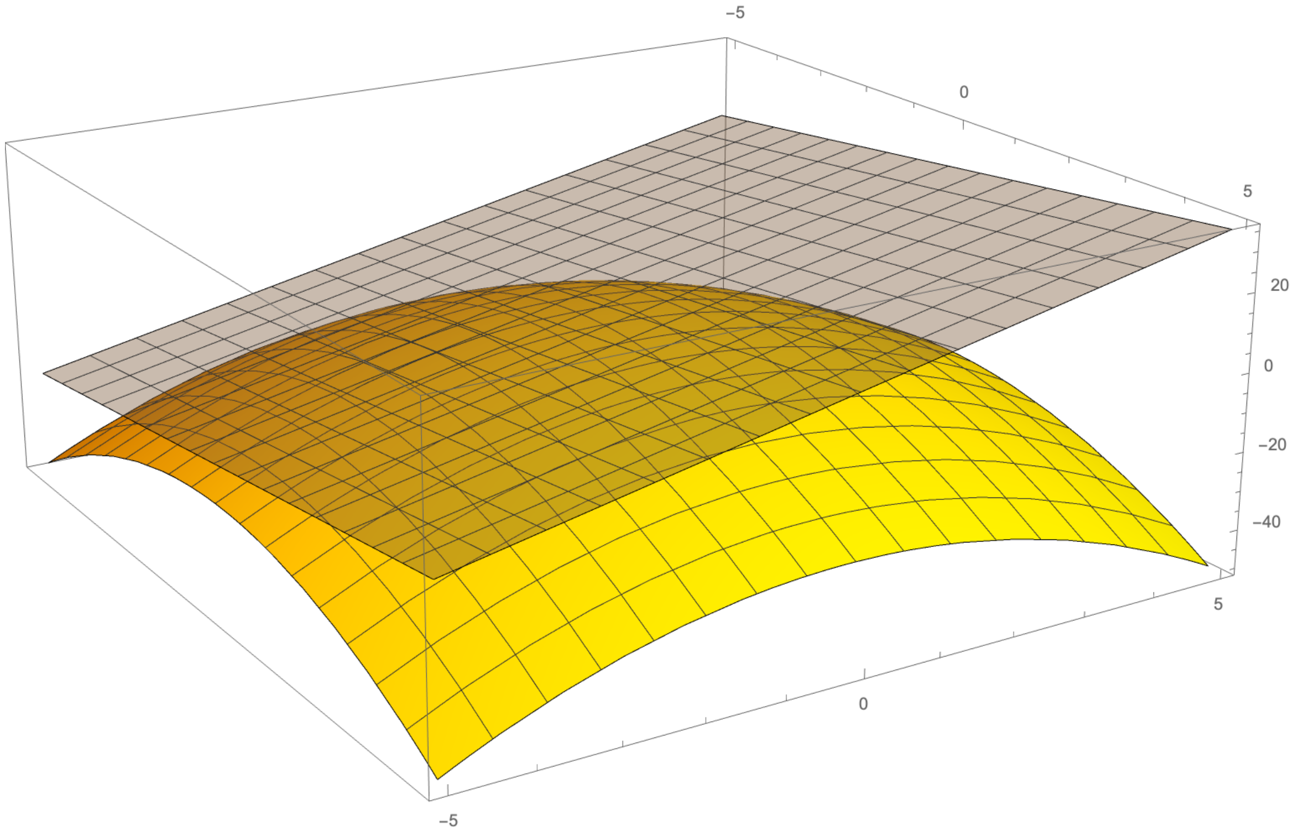


Figure 20: An example of a tangent plane

Partial derivatives

The partial derivatives of a function $f\colon \mathbb{R}^2 \to \mathbb{R}$ are given by the following limits.

$$\frac{\partial f}{\partial x}(a,b) = \lim_{h\to 0} \frac{f(a+h,b) - f(a,b)}{h}$$
$$\frac{\partial f}{\partial y}(a,b) = \lim_{h\to 0} \frac{f(a,b+h) - f(a,b)}{h}$$

These tell us that when calculating, for example, the partial derivative with resepct to x the value of the y-variable is constant. Note the new symbol ∂ that we use when writing partial derivatives.

Consider the example of the function we saw previously.

$$f\colon \mathbb{R}^2 \to \mathbb{R}$$
$$(x,y) \mapsto \sin(x+y^2)$$

The **partial derivative** of f with respect to x can be calculated in practice by treating all expressions involving y as if they are constant:

$$\frac{\partial f}{\partial x} = \cos(x + y^2).$$

Similarly, the partial derivative of f with respect to y is the derivative we obtain by treating expressions involving x as if they are constant.

$$\frac{\partial f}{\partial y} = 2y\cos(x + y^2).$$

The partial derivatives are functions from $\mathbb{R}^2$ to $\mathbb{R}$, the same as the original funciton f.

As a second example suppose $g(x, y) = 3x^2y$. Then

$$\frac{\partial g}{\partial x} = 6xy,$$

$$\frac{\partial g}{\partial y} = 3x^2.$$

Higher order derivatives

As with single-variable calculus we can continue differentiating our partial derivatives leading to higher order derivatives. If $g(x, y) = 3x^2y$ then we saw previously that

$$\frac{\partial g}{\partial x} = 6xy.$$

Differentiating with respect to x again we get

$$\frac{\partial^2 g}{\partial x^2} = 6y.$$

If instead of differentiating with respect to x, we differentiated with respect to y we get

$$\frac{\partial^2 g}{\partial y \partial x} = \frac{\partial}{\partial y}\left(\frac{\partial g}{\partial x}\right) = 6x.$$

Notice that when the variables differ the order of the variables goes from right to left.

Similarly since

$$\frac{\partial g}{\partial y} = 3x^2$$

we can calculate the higher order derivatives

$$\frac{\partial^2 g}{\partial y^2} = 0$$
$$\frac{\partial^2 g}{\partial x \partial y} = 6x$$

Notince that in this example

$$\frac{\partial^2 g}{\partial x \partial y} = \frac{\partial^2 g}{\partial y \partial x}.$$

This isn't a coincidence. It is always true when the function g is **continuously differentiable**.

Smoothness of multivariate functions

If X is a subset of $\mathbb{R}^n$ then we can generalise the smoothness classes you learned about in Week 2 as follows.

- $C(X)$ is the set of continuous function, $f\colon X \to \mathbb{R}$
- $C^1(X)$ is the set of functions for which **all** partial derivatives exist and are continuous
- $C^n(X)$ is the set of functions for which **all** partial derivatives of order n exist and are continuous
- $C^\infty(X)$ is the set of fucntions for which **all** partial derivatives of all orders exist.

The derivative

The derivative of a function $f\colon \mathbb{R}^2 \to \mathbb{R}$ is the function

$$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \end{pmatrix}.$$

For example, if $g(x, y) = 3x^2 y$ then the derivative is

$$\begin{pmatrix} 6xy & 3x^2 \end{pmatrix}.$$

It is written using the usual f' symbol:

$$g'(x, y) = \begin{pmatrix} 6xy & 3x^2 \end{pmatrix}.$$

For example the derivative of g at $(2, 3)$ is

$$g'(2, 3) = \begin{pmatrix} 36 & 12 \end{pmatrix}.$$

Notice that we have switched to a row vector here rather than column vectors.

We have similar rules for differentiation that we do in single-variable calculus. For example if a and b are real numbers then

$$(af + bg)' = af' + bg'$$

We also have the **product rule**

$$(fg)' = fg' + gf'$$

The equation of the tangent plane

In single variable calculus the derivative allows us to find the equation of the tangent line. In a similar way, the derivative of a multivariable function allows us to find the equation of the tangent plane. The tangent plane at the point **p** is given by the formula,

$$z = +f(\mathbf{p}) + f'(\mathbf{p})(\mathbf{x} - \mathbf{p})$$

where $\mathbf{x} = \begin{pmatrix} x & y \end{pmatrix}^T$.

For example for $g(x, y) = 3x^2y$ at $(2, 3)$ the equation can be calculated as follows.

$$\begin{aligned} z &= 36 + \begin{pmatrix} 36 & 12 \end{pmatrix} \begin{pmatrix} x - 2 \\ y - 3 \end{pmatrix} \\ &= 36 + 36(x - 2) + 12(y - 3) \\ &= 36x + 12y - 72 \end{aligned}$$

Note: the matrix multiplication is only possible here because f' is a row vector.

The gradient of a differentiable function

It is often useful to think of the derivative as a column vector rather than a row vector. We write this using the “nabla” symbol,

$$\nabla f = (f')^T$$

For example if $g(x, y) = 3x^2y$ as above, then at $(2, 3)$, $f' = \begin{pmatrix} 36 & 12 \end{pmatrix}$. Hence,

$$\nabla g(2, 3) = \begin{pmatrix} 36 \\ 12 \end{pmatrix}.$$

This now represents a vector.

Directional derivatives

If we take the two basis vectors in $\mathbb{R}^2$, $\mathbf{i} = (1 \quad 0)^T$ and $\mathbf{j} = (0 \quad 1)^T$ then we can recover our partial derivatives using the dot-product:

$$\frac{\partial f}{\partial x} = \nabla f \cdot \mathbf{i}$$

$$\frac{\partial f}{\partial y} = \nabla f \cdot \mathbf{j}$$

As we stated earlier these are the slopes of the lines on the tangent plane parallel to the x and y axes respectively.

If we now want to find the slope of the line on the tangent plane in **any** other particular direction, say $\mathbf{u} = (a, b)^T = a\mathbf{i} + b\mathbf{j}$ then we can use the linearity of the dot-product:

$$\begin{aligned}\nabla f \cdot (a\mathbf{i} + b\mathbf{j}) &= a\nabla f \cdot \mathbf{i} + b\nabla f \cdot \mathbf{j} \\ &= a\frac{\partial f}{\partial x} + b\frac{\partial f}{\partial y}\end{aligned}$$

This can be a useful measure of the slope of the tangent plane. We usually standardise the vector $\mathbf{u}$ by requiring that $|\mathbf{u}| = 1$ (we often call this a **normal vector**) to avoid the size of the vector affecting the size of the slope.

We thus have the following definition.

Given **any** vector $\mathbf{n}$ with $|\mathbf{n}| = 1$ we define the **directional derivative** of f **in the direction of n** as

$$\nabla_{\mathbf{n}} f = (\nabla f) \cdot \mathbf{n}$$

For example suppose $g(x, y) = 3x^2y$. Then we saw that at $(2, 3)$

$$\nabla g(2, 3) = \begin{pmatrix} 36 \\ 12 \end{pmatrix}$$

So if we let $\mathbf{n} = \begin{pmatrix} 1/\sqrt{2} & \sqrt{3}/2 \end{pmatrix}^T$ then the directional derivative in the direction of $\mathbf{n}$ is

$$\begin{aligned}\nabla_{\mathbf{n}} g(2, 3) &= \begin{pmatrix} 36 \\ 12 \end{pmatrix} \cdot \begin{pmatrix} 1/\sqrt{2} \\ \sqrt{3}/2 \end{pmatrix} \\ &= 18\sqrt{2} + 6\sqrt{3}\end{aligned}$$

The direction of greatest descent

Imagine you are standing on our surface and you drop a ball - in which direction will it roll?

After a little throught you should be able to convince yourself that it will be in the direction where the surface is steepest.

This means the directional derivative that has the smallest negative value.

We can find this direction using the definition of the dot-product, as

$$\mathbf{u} \cdot \mathbf{v} = |\mathbf{u}|\,|\mathbf{v}| \cos\theta$$

where θ is the angle between the vectors **u** and **v** if you visualise them as starting at the same position.

Now suppose f is a differentiable function, then we want to find the direction **n** that minimises the directional derivative,

$$\begin{aligned}\nabla_{\mathbf{n}} f &= \nabla f \cdot \mathbf{n} \\ &= |\nabla f|\,|\mathbf{n}| \cos\theta \\ &= |\nabla f| \cos\theta, \qquad \text{since } |\mathbf{n}| = 1.\end{aligned}$$

Now the range of cos is $[-1, 1]$ and so the minimum value of $\cos\theta$ is -1 when $\theta = \pi$.

Geometrically this means that the minimum is obtained when **n** is point at π radians (i.e. the opposite direction) to ∇f.

This is the vector,

$$\mathbf{n} = -\frac{1}{|\nabla f|}\nabla f.$$

So the ball will always roll in the **opposite direction** to the gradient.

For example suppose $g(x, y) = 3x^2y$ and we drop a ball at position $(2, 3, 36)$. Then we previously calculated the gradient here as

$$\nabla g(2, 3) = \begin{pmatrix} 36 \\ 12 \end{pmatrix}$$

Now $|\nabla g(2, 3)| = 12\sqrt{10}$ so that the ball will fall in the direction

$$-\frac{1}{12\sqrt{10}}\nabla g(2, 3) = \begin{pmatrix} -3/\sqrt{10} \\ -1/\sqrt{10} \end{pmatrix}$$

Note: we can also calculate that the direction that maximises the value of $\nabla_{\mathbf{n}} f$. In our previous calculation we used the fact that $\cos\theta$ is minimised when $\theta = \pi$. If we want to find the maximum then this is at $\theta = 0$ when $\cos\theta = 1$. But this direction represents a vector in the same direction as ∇f.

We conclude that the **maximums** value of the directional derivative is in the direction

$$\mathbf{n} = \frac{1}{|\nabla f|}\nabla f,$$

at which

$$\nabla_{\mathbf{n}} f = |\nabla f|.$$

Level curves

Another way of visualising the graph of a function $f\colon \mathbb{R}^2 \to \mathbb{R}$ is to draw its **level curves**. These are the curves in $\mathbb{R}^2$ defined as

$$f(x, y) = c$$

where c takes on different values.

We can plot the level curves of a function for different values of c, normally equally spaced. This produces a **contour plot**.

For example suppose we have the following function,

$$\begin{aligned} f\colon \mathbb{R}^2 &\to \mathbb{R} \\ \begin{pmatrix} x \\ y \end{pmatrix} &\mapsto (x-1)^2 + 4y^2. \end{aligned}$$

The level curves are the curves

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

You should recognise this as the equation of an ellipse, see Figure 21 for an example.

It is often useful to plot a number of these level curves on a single plot which lets us visualise the surface as if we are looking at it from above/below, see Figure 22.

The plot of surface is given in Figure 23.

$$z = 3(1-x)^2 e^{-x^2-(y+1)^2} - (2x - 10x^3 - 10y^5)e^{-x^2-y^2} \\ -\frac{1}{3}e^{-(x+1)^2-y^2}$$

The contour plot of this function is given in Figure 24.

On the contour plot, level curves that are close together represent a steep gradient.

A contour plot can be easier to understand than a 3d plot which can often obscure information.

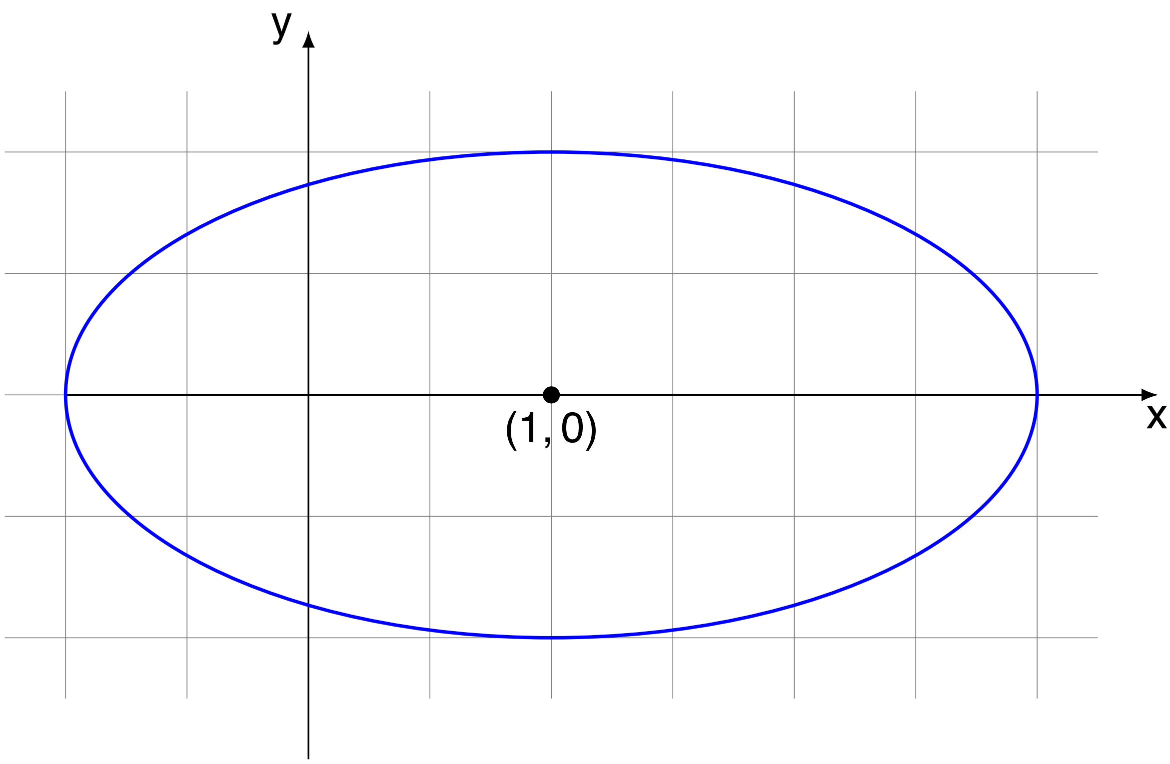


Figure 21: A level curve for $(x-1)^2 + 4y^2$.

An interesting fact about the gradient ∇f and the level curves $f(x, y) = c$ is that the gradient is always **perpendicular** to the level curve. This means if we calculate the graident of a point on the level curve and draw it as a vector starting at the point, then it meets the curve in a right angle.

To illustrate this suppose $f(x, y) = x^2 + 2y^2$. The level curves are then ellipses, and the gradients at various points are drawn in the diagram below.

Here the blue ellipse is the level curve, and the red arrows are the gradient vectors plotted at various points on the curves.

Geometrically this makes sense if we know that the gradient represents the direction of maximum increase.

Higher dimensional

The content of this week's work has so far focussed on functions of 2 variables. We can extend **everything** to functions of more than two variables without much work. In fact I've made sure that the algebra is essentially unchanged.

What does change is the fact that we can't really visualise things very well.

In general a function of n-variables is normally written as

$$f(x_1, x_2, x_3, \dots, x_n)$$

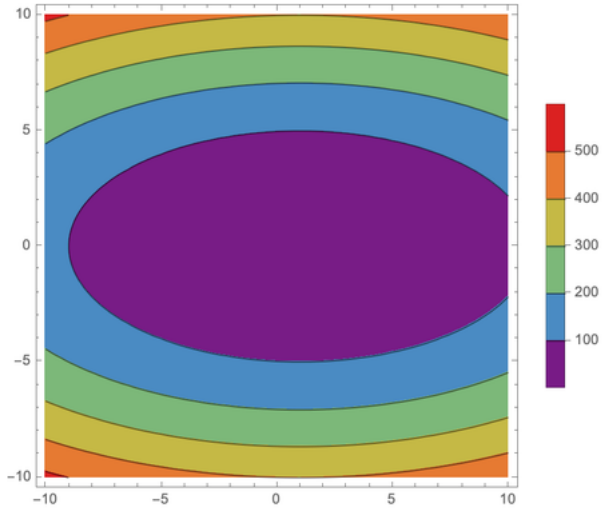


Figure 22: The contour plot for $(x-1)^2 + 4y^2$

were each of the x_i is one of the variables. So we can still define partial derivatives

$$\frac{\partial f}{\partial x_i}$$

by treating all other variables as constants.

The derivative is then defined as

$$\nabla f = \begin{pmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} & \cdots & \frac{\partial f}{\partial x_n} \end{pmatrix}$$

Importantly this can be used to define the **directional derivative**. The difference now is that you have to have a vector **n** $\in \mathbb{R}^n$, but the directional derivative is still defined as

$$\nabla_{\mathbf{n}} f = \nabla f \cdot \mathbf{n}.$$

This represents the gradient in the direction of **n** and, again, the direction of **greatest descent** is

$$-\frac{1}{|\nabla f|}\nabla f.$$

This is the same as before.

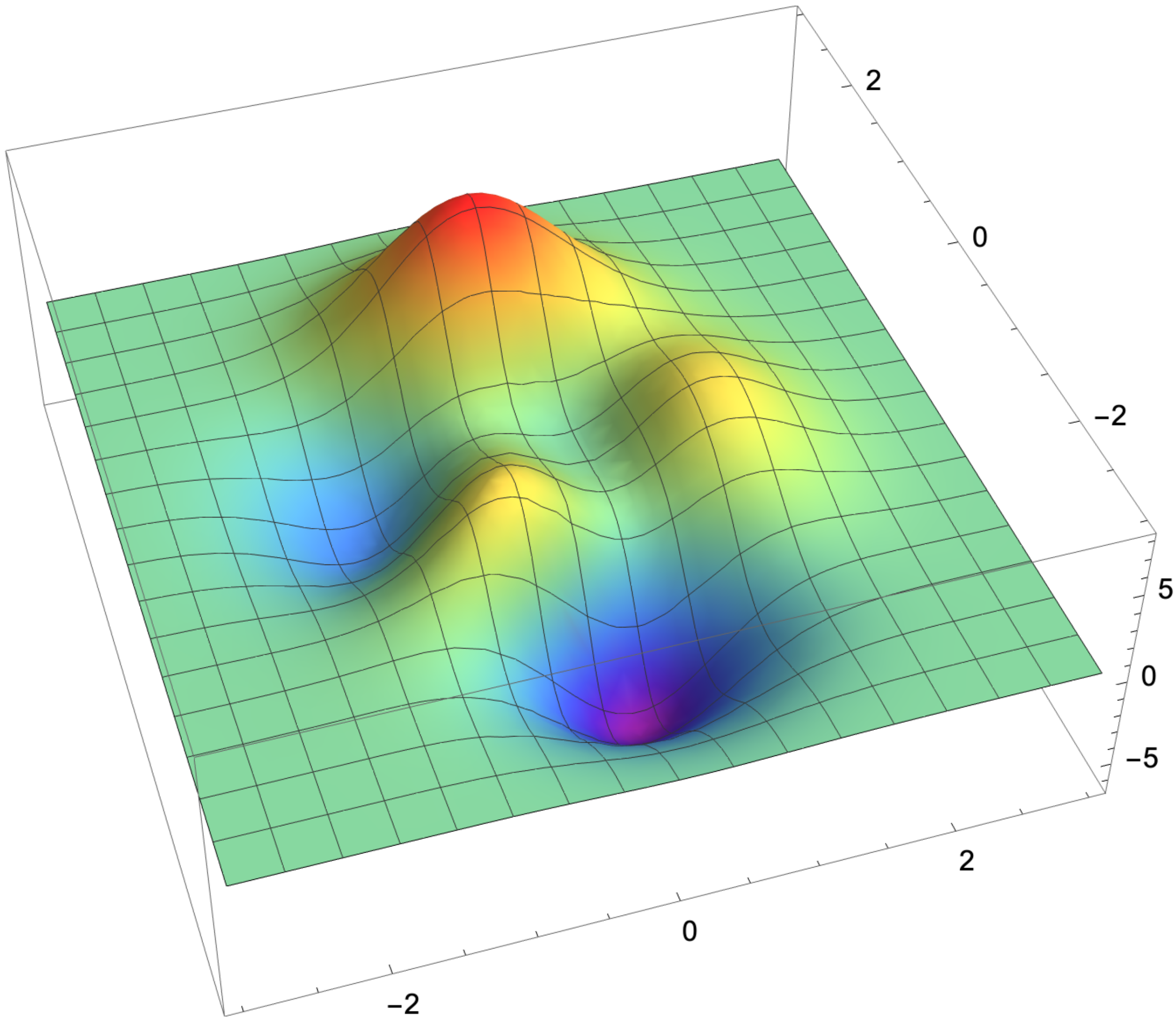


Figure 23: The plot of this complicated function

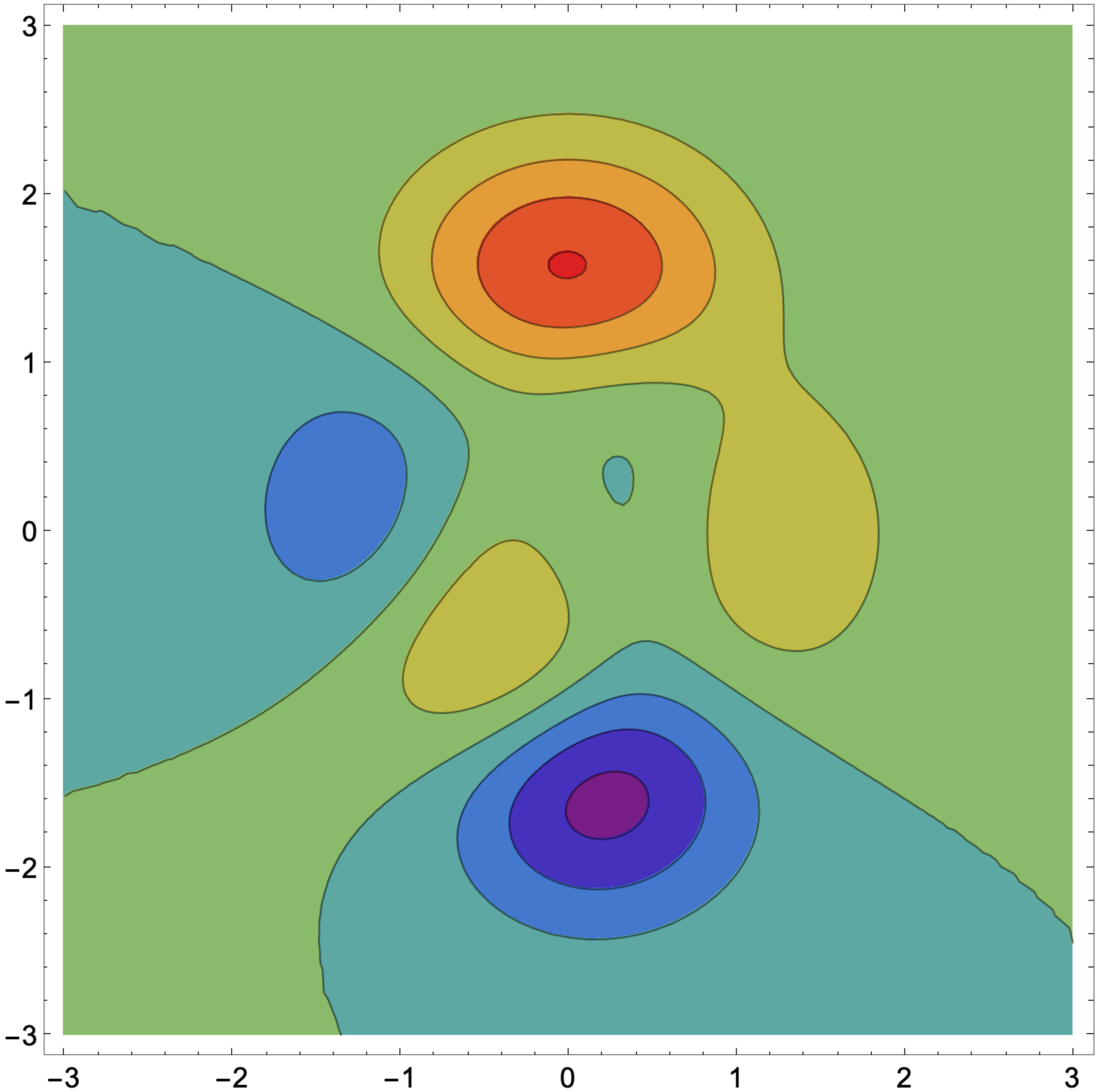


Figure 24: The contour plot of this complicated function

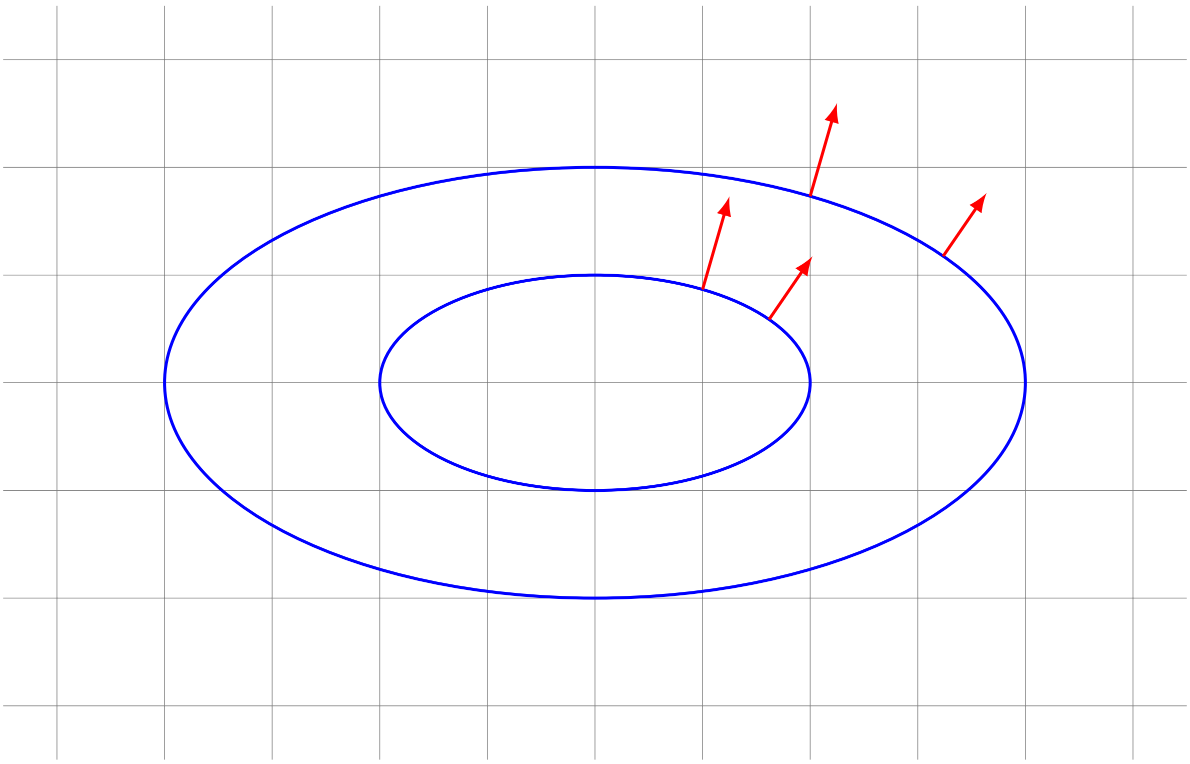


Figure 25: The level curves and perpendicular gradients.

Week 4: Optimisation

Optimisation and business modelling

This course is mainly concerned with using mathematics to model business problems and solve them. For the next two weeks we will learn to model and solve optimisation problems. Optimisation prolems aim to find the maximum or minimum value of a given function.

Unconstrained optimisation for functions of one variable

In an unconstrained optimisation problem we have a function

$$f\colon \mathbb{R} \to \mathbb{R}, \qquad f \text{ continuous}$$

In a maximisation problem, our aim is then to find a value x_0 so that

$$\text{for all } x, \qquad f(x) \leq f(x_0).$$

The point x_0 is called the **optimum point** and we write

$$\max f(x) = f(x_0).$$

For a minimisation problem we reverse the inequality, $f(x) \geq f(x_0)$ and write

$$\min f(x) = f(x_0)$$

Note: There is no discernable difference between the maximisation problem and minimisation problem. In fact since

$$\min f(x) = -\max(-f(x))$$

the two problems are interchangable if we replace $f(x)$ with $-f(x)$.

In general the unconstrained optimisation problem does not have a solution. The following examples demonstrate two ways in which a function may not attain its maximum.

Example 1

The logistic function is given by:

$$f\colon \mathbb{R} \to \mathbb{R}$$
$$x \mapsto \frac{1}{1+e^{-x}}$$

It is strictly increasing for all values of x and its range is $f(\mathbb{R}) = (0, 1)$. There is however no value x_0 so that $f(x) \leq f(x_0)$ for all x.

Example 2

A simple function such as the square function

$$\begin{aligned} f\colon \mathbb{R} &\to \mathbb{R} \\ x &\mapsto x^2 \end{aligned}$$

is **unbounded**: its range is $f(\mathbb{R}) = [0, \infty)$. Therefore, again, the maximisation problem has no solution.

Local maximums and minimums

In general the unconstrained optimisation problem seems hopeless. However there is still some useful information we can get from problems like this.

For a continuous function f we call a point x_0 a **local maximum value** if $f(x) \leq f(x_0)$ for all x 'near' x_0. More precisely we usually require that the inequality is true in an **open** interval (c, d) that contains the point x_0 – although we don't prescribe any particular size of interval.

A similar definition holds for **local minimum value**.

One way of finding local optimal points is to find solve the equation

$$f'(x) = 0$$

for x, so-called **stationary points**. This is unsatisfactory however since it pre-supposes the derivative exists. Since the optimisation problem only assumes our function f is continuous this technique is quite limited. In the next section we will solve this entirely.

Constrained optimisation for functions of one variable

In contrast to unconstrained optimisation problems the constrained problem attempts to find an optimum point in a given interval, $[a, b]$.

Given a continuous function $f\colon \mathbb{R} \to \mathbb{R}$ the constrained maximisation problem is to find a value $x_0 \in [a, b]$ so that

$$\text{for all } x \in [a, b], \qquad f(x) \leq f(x_0).$$

This is written

$$\max_{x \in [a,b]} f(x) = f(x_0)$$

Important: constrained optimisation problems **can always be solved**. This is a dramatic difference to the unconstrained problem, and so we will learn a technique for finding the optimal point.

First we will examine an example.

Example

Suppose we want to find the maximum and minimum of the function

$$f(x) = x + x|\cos x|, \qquad \pi \le x \le 2\pi.$$

The graph of this function is in figure 26.

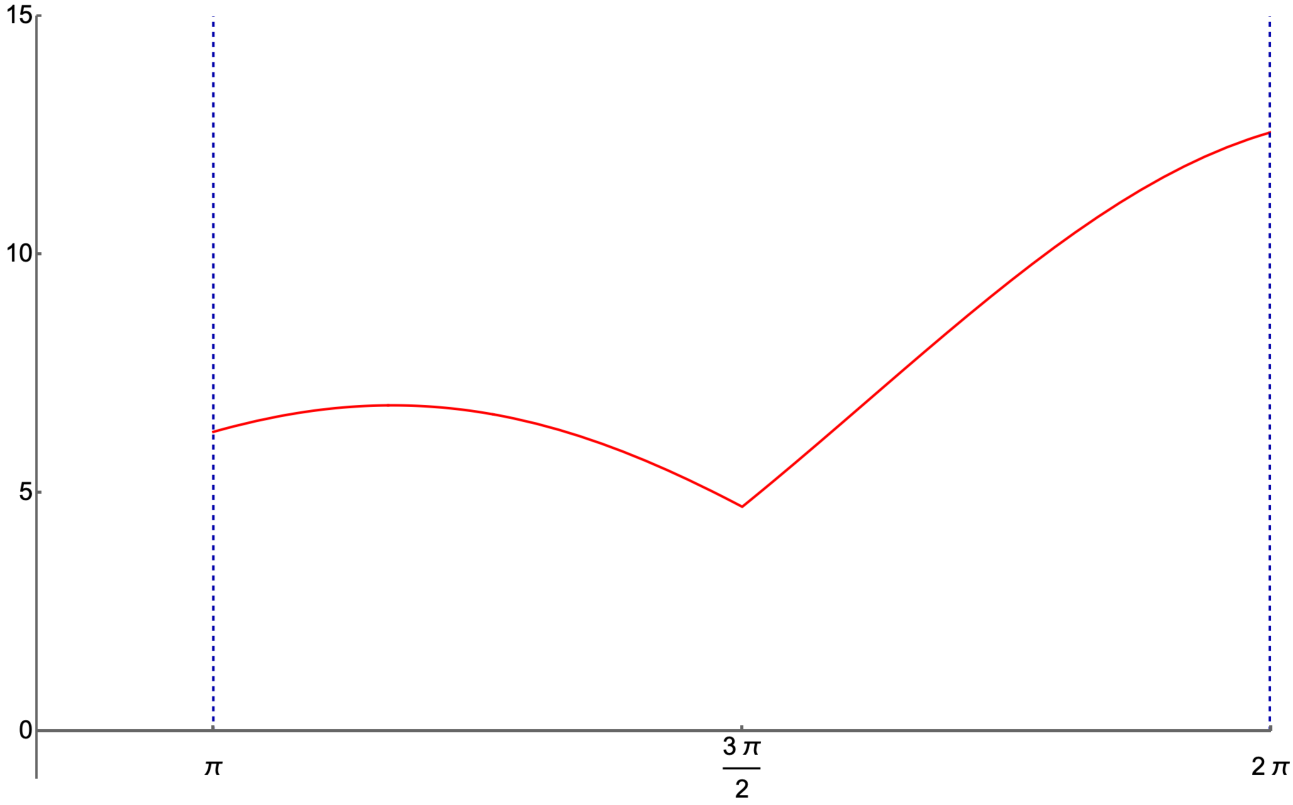


Figure 26: An example of a constrained optimistion problem

From the graph we can see that

$$\min_{x\in[\pi,2\pi]} f(x) = f\left(\frac{3\pi}{2}\right) = \frac{3\pi}{2},$$

and

$$\max_{x\in[\pi,2\pi]} f(x) = f(2\pi) = 4\pi$$

If you examine the graph of this function you should see the following interesting properties:

- The maximum point, 2π is not a point where the derivative is 0
- The minimum point, $3\pi/2$ is a point where f' does not exist, since there's no tangent to the curve at this point
- There is a local maximum between π and $3\pi/2$ but this is not the maximum value

All these suggest a broader approach to finding optimal points is needed.

Critical points

Suppose $f\colon \mathbb{R} \to \mathbb{R}$ is continuous, then we define a **critical point** on the graph of f to be any point $(x, f(x))$ such that

1. f is **not differentiable** at x; or
2. f is **differentiable** at x and $f'(x) = 0$.

The second of these are commonly called **stationary points**.

In the example above there were two critical points, one stationary point between π and $3\pi/2$ and one non-differentiable point when $x = 3\pi/2$.

Critical points allow us to solve constrained optimisation problems.

Fermat's Theorem

Suppose $f\colon \mathbb{R} \to \mathbb{R}$ is continuous. Then for $a \leq x \leq b$ there is an optimal point x_0. Furthermore x_0 is eithe a critical point of f or is one of the endpoints a or b.

This means we can solve constrained optimisation problems with the following technique:

1. Find all the critical points of f in $[a, b]$, write these as $x_0, x_1, \dots, x_n$
2. Evaluate $f(x_0), f(x_1), \dots, f(x_n)$ and $f(a)$, $f(b)$
3. Find
$$\max\{f(x_0), f(x_1), \dots, f(x_n), f(a), f(b)\}$$
or the minimum.

Example

Find the maximum point for the function

$$\begin{aligned} f\colon \mathbb{R} &\to \mathbb{R} \\ x &\mapsto 1 - x + |2x^2 - 13x + 15| \end{aligned}$$

for $0 \leq x \leq 4$.

1. First we find the points where this function is not differentiable. These are the points where $|2x^2 - 13x + 15| = 0$. Factorising we have $2x^2 - 13x + 15 = (2x - 3)(x - 5)$. The critical point that lies in $[0, 4]$ is at $x_0 = 1.5$.
2. Next we find the derivative of this function. Note that The quadratic expression $2x^2 - 13x + 15$ is positive for $0 < x < 3/2$ and negative for $3/2 < x < 4$. We take these two cases separately,

$0 < x < 3/2$

Here $2x^2 - 13x + 15 > 0$ so $|2x^2 - 13x + 15| = 2x^2 - 13x + 15$ and we can write,

$$f(x) = 1 - x + 2x^2 - 13x + 15 = 2x^2 - 14x + 16.$$

Solving $f'(x) = 0$ we find

$$4x - 14 = 0, \qquad \text{so that } x = 7/2$$

This value is outside the interval $(0, 3/2)$ so is not a critical point.

$3/2 < x < 4$

Here $2x^2 - 13x + 15 < 0$ so $|2x^2 - 13x + 15| = -(2x^2 - 13x + 15)$ and we may write,

$$f(x) = 1 - x - (2x^2 - 13x + 15) = -2x^2 + 12x - 14.$$

Solving $f'(x) = 0$ we find

$$-4x + 12 = 0, \qquad \text{so that } x = 3.$$

This is our second critical point $x_1 = 3$.

3. Finally we evaluate f at each of the critical points and at the endpoints 0 and 4.

x	$f(x)$
1.5	-0.5
3	4
0	16
4	2

From the table we see that

$$\max_{x \in [0,4]} f(x) = f(0) = 16$$

and

$$\min_{x \in [0,4]} f(x) = f(1.5) = -0.5$$

Unconstrained optimisation for multivariate functions

Suppose now we want to solve the same problem but for a multivariate function:

$$f \colon \mathbb{R}^n \to \mathbb{R}$$

Note: in order to avoid unnecessary complications when we learn about optimisation for multivariate functions we will assume they are **differentiable**. This is stronger than our assumption for functions of one variable, but it does mean we don't have to deal with non-differentiable points like we did there.

Local maximums and minimums and stationary points

In the same way as with functions of one variable we can make progress towards finding optimal points by finding local maximums and local minimums. Since we are assuming our functions are differentiable these are now **stationary points**, i.e. points where

$$f'(\mathbf{x}) = \mathbf{0}.$$

Suppose for simplicity $f\colon \mathbb{R}^2 \to \mathbb{R}$ and write it as $f(x, y)$. Then the equation above can be written,

$$\begin{pmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \end{pmatrix} = \begin{pmatrix} 0 & 0 \end{pmatrix}.$$

This, in turn, can be written as the simultaneous equations

$$\frac{\partial f}{\partial x} = 0, \qquad \frac{\partial f}{\partial y} = 0.$$

Example

Suppose $f(x, y) = 2x^2 + 2xy - y^2 - 8x + 2y$.

We calculate the partial derivatives first and set the equal to 0,

$$\frac{\partial f}{\partial x} = 4x + 2y - 8 = 0$$
$$\frac{\partial f}{\partial y} = 2x - 2y + 2 = 0$$

If we add both equations we find that

$$6x - 6 = 0, \qquad \text{so that} \quad x = 1.$$

Substituting back into the first equation we find that

$$4 + 2y - 8 = 0, \qquad \text{and so} \quad y = 2.$$

The stationary point is therefore at $(1, 2)$.

For surfaces $z = f(x, y)$ we can visualise the three possible types of stationary points. They are the **local maximum**, Figure 27, the **local minimum** Figure 28, and the **local saddle**, Figure 29.

It is possible to classify a stationary point using the second derivatives. Unfortunately this requires us to further restrict the generality of the problem since we have to assume that our functions have second order derivatives.

Nevertheless we define the Hessian of the function f as the matrix,

$$\text{Hess } f = \begin{pmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial x^2} \end{pmatrix}$$

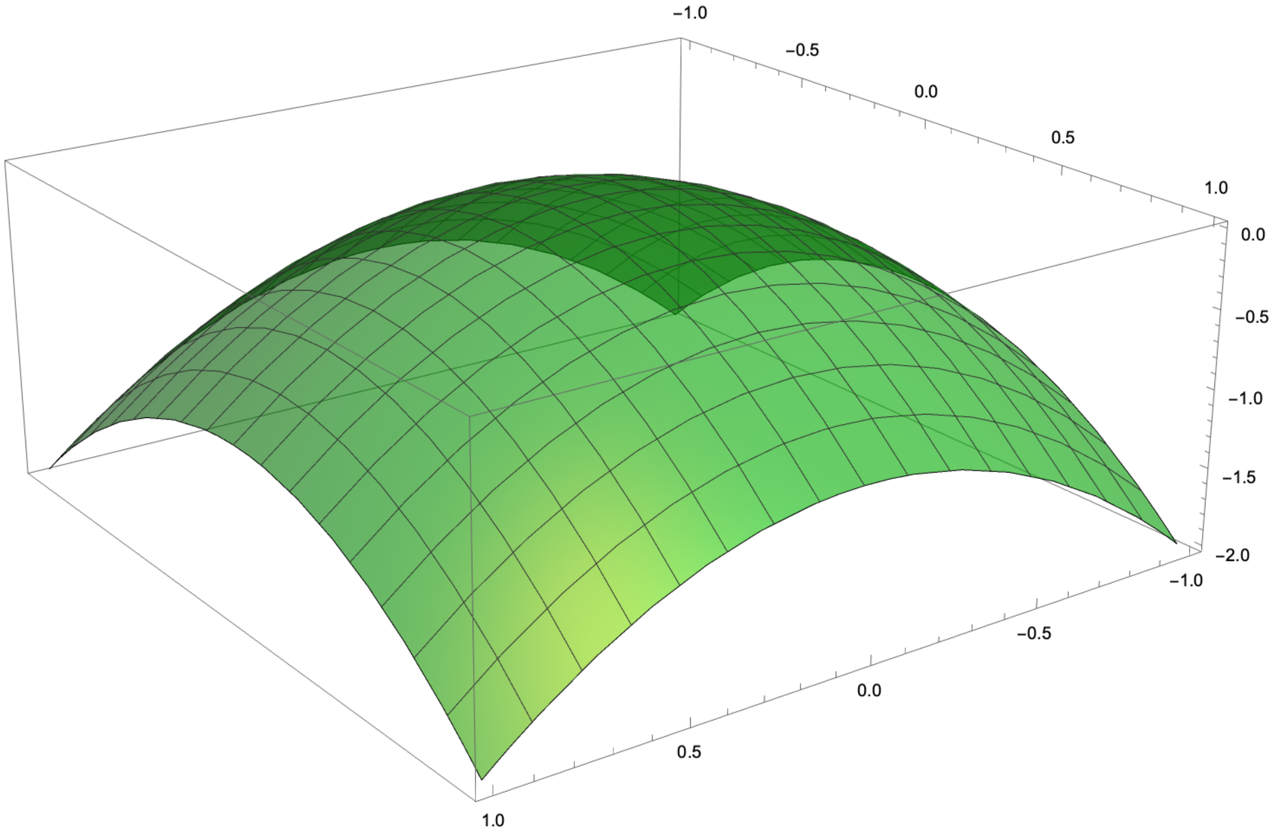


Figure 27: Local maximum

If we evaluate the Hessian at each of the stationary point (a, b) and determine the eigenvalues we have the following:

- If all eigenvalues of Hess $f(a, b)$ are negative then (a, b) represents a local maximum
- If all eigenvalues of Hess $f(a, b)$ are positive then (a, b) represents a local minimum
- If the eigenvalues of Hess $f(a, b)$ are a mixture of positive and negative values then (a, b) represents a local saddle

Example (continued)

With $f(x, y) = 2x^2 + 2xy - y^2 - 8x + 2y$ we found that

$$\frac{\partial f}{\partial x} = 4x + 2y - 8$$
$$\frac{\partial f}{\partial y} = 2x - 2y + 2$$

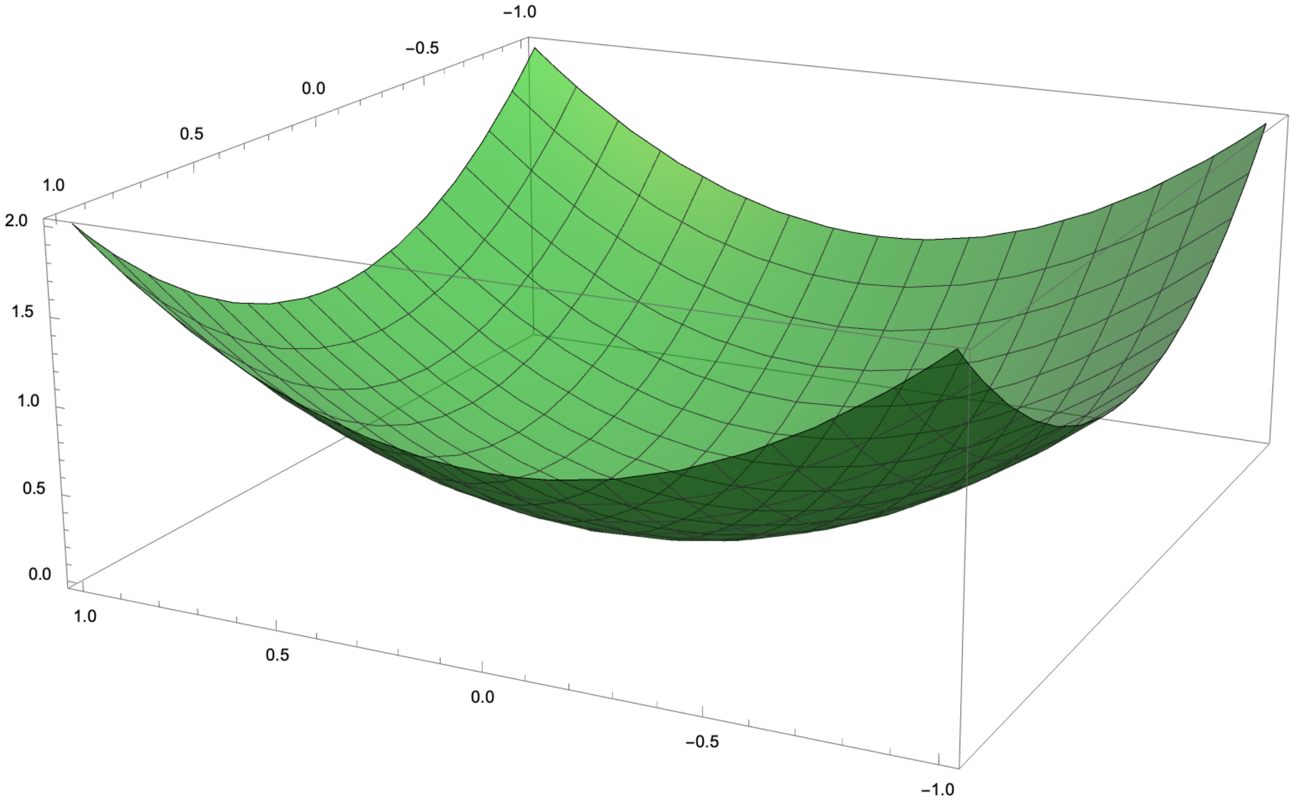


Figure 28: Local minimum

Calculating the second derivatives,

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

Hence,

$$\text{Hess } f(1,2) = \begin{pmatrix} 4 & 2 \\ 2 & -2 \end{pmatrix}$$

To calculate the eigenvalues, we find the following determinant,

$$\begin{aligned} 0 = \det \begin{pmatrix} 4-\lambda & 2 \\ 2 & -2-\lambda \end{pmatrix} &= (4-\lambda)(-2-\lambda) - 4 \\ &= \lambda^2 - 2\lambda - 12 \end{aligned}$$

Solving we find $\lambda = 1 \pm \sqrt{13}$. One of these is positive and one is negative, meaning that the point $(1,2)$ represents a saddle point.

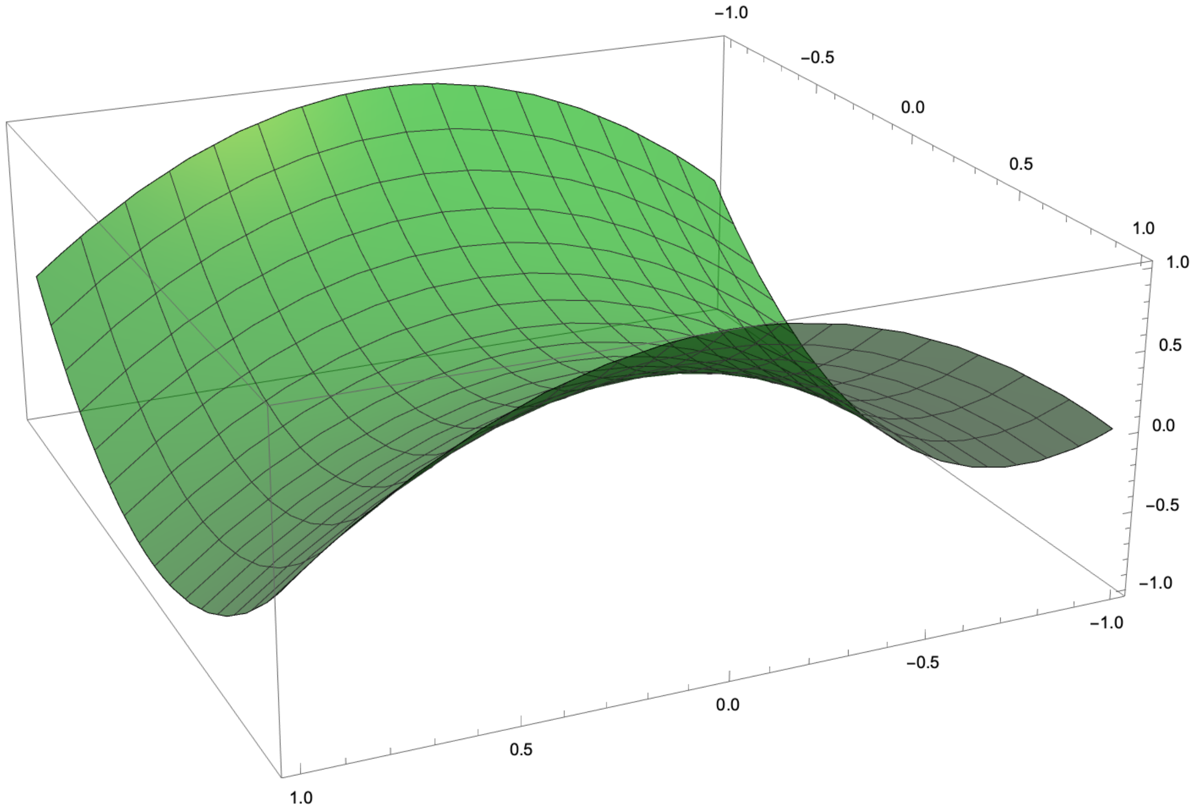


Figure 29: Local saddle

In general, for $f\colon \mathbb{R}^n \to \mathbb{R}$, the Hessian is given by,

$$\text{Hess } f = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix}$$

The classification above generalises to this case.

Special classification method for functions of two variables

The technique for classifying can be simplified when f is a function of two variables. However this doesn't generalise so it is very much a special case. Nevertheless many books use this technique and it is easier than calculating eigenvalues.

First we find the **discriminant**

$$D = \frac{\partial^2 f}{\partial x^2}\frac{\partial^2 f}{\partial y^2} - \left(\frac{\partial^2 f}{\partial x \partial y}\right)^2$$

Then we follow the following decision tree.

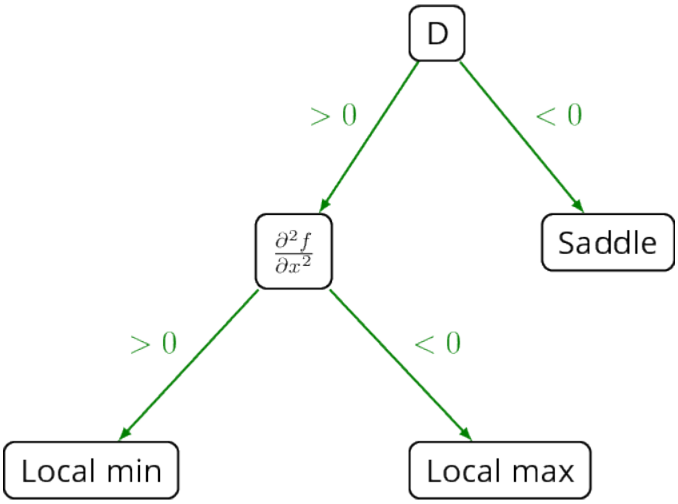


Figure 30: Decision tree for classifying stationary points

Convex functions and convex optimisation

For unconstrained problems with differentiable functions finding and classifying stationary points can be useful since an optimal point if it exists is certainly a stationary point. However since it is not always clear whether or not an optimal point exists at all it can be a bit misleading to simply rely on this.

In this section we will introduce a special class of unconstrained problems that have the very special property that local optimum points are also global optimum points. This means finding stationary points and classifying them will solve the optimisation problem. These are such an important class of problems that much of data science, machine learning, statistics set up optimisation problems with this specific property in mind.

Convex sets

A set S is convex if the line joining any two points A and B in S is entirely contained in S.

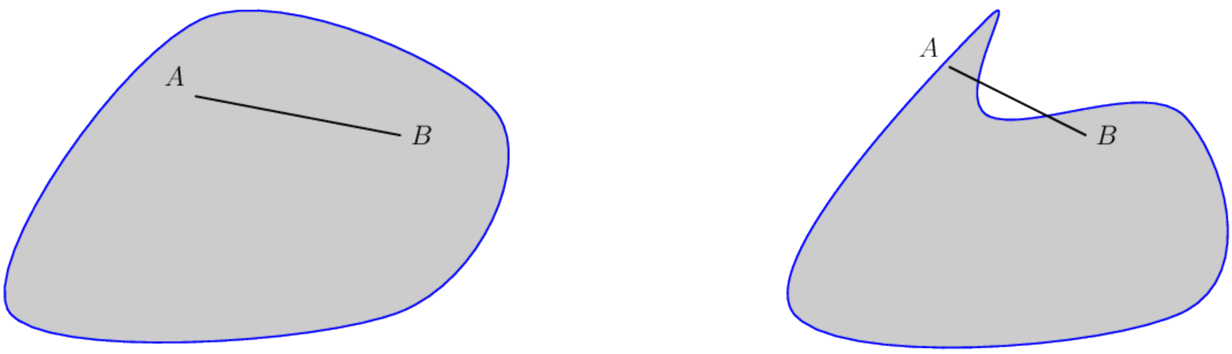
Figure 31 shows an example of a convex set and a non-convex set.

Convex functions

A function is said to be a **convex function** if the region **above** its graph is a convex set. The archetype example is the function $x \mapsto x^2$.

However the following functions are also convex:

- $\exp x$



Convex set **Non-convex set**

Figure 31: Convex sets

- Any line $a + bx$ for $b \neq 0$
- Any quadratic $ax^2 + bx + x$ for $a > 0$

For multivariate functions it's more difficult to determine whether or not a function is convex. If the second order derivatives exist then the following is true:

- If all the eigenvalues of the Hessian are non-negative at all points then f is convex

Of course, this can be difficult to verify in general. However the following can be verified to be convex functions

- Linear functions $a + b_1x_1 + \cdots + b_nx_n$ as long as at least one of the b's is non-zero
- Quadratic functions of the form $(a_1x_1 + b_1)^2 + (a_2x_2 + b_2)^2 + \cdots + (a_nx_n + b_n)^2$

Concave functions

In contrast to convex functions, a **concave function** is one where the region **below** its graph is a convex set. Again the archetype example is a quadraric, $f(x) = -x^2$, whose graph is given in Figure 33.

In general the following functions are concave:

- $\log x$
- Any line $a + bx$ for $b \neq 0$
- Any quadratic $ax^2 + bx + x$ for $a < 0$

Again in general for multivariate functions with second order derivatives we can use the Hessian to determine if the function is concave:

- If all the eigenvalues of the Hessian are non-positive at all points then f is concave

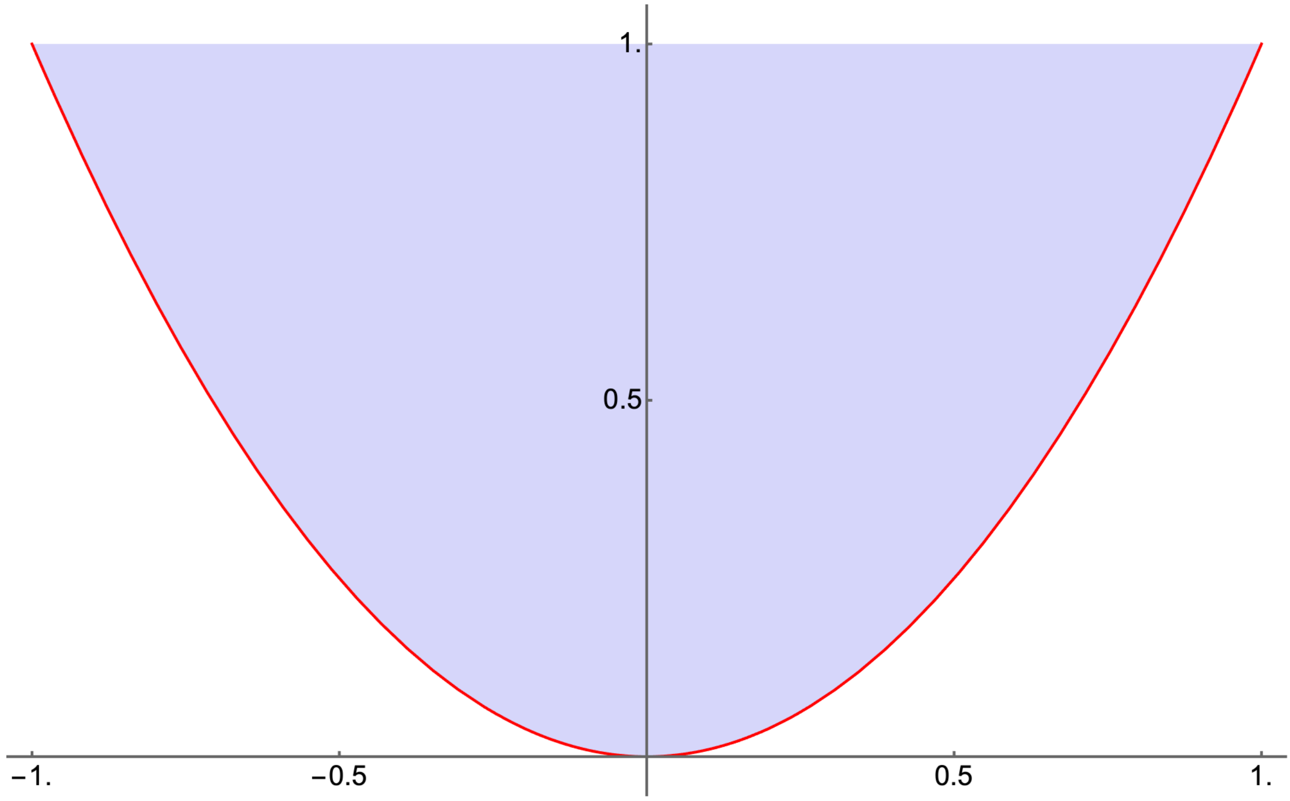


Figure 32: The graph of the convex function x^2

However as with convex functions we can verify that the following functions are concave,

- Linear functions $a + b_1x_1 + \cdots + b_nx_n$ as long as at least one of the b's is non-zero
- Quadratic functions of the form $-(a_1x_1 + b_1)^2 - (a_2x_2 + b_2)^2 - \cdots - (a_nx_n + b_n)^2$

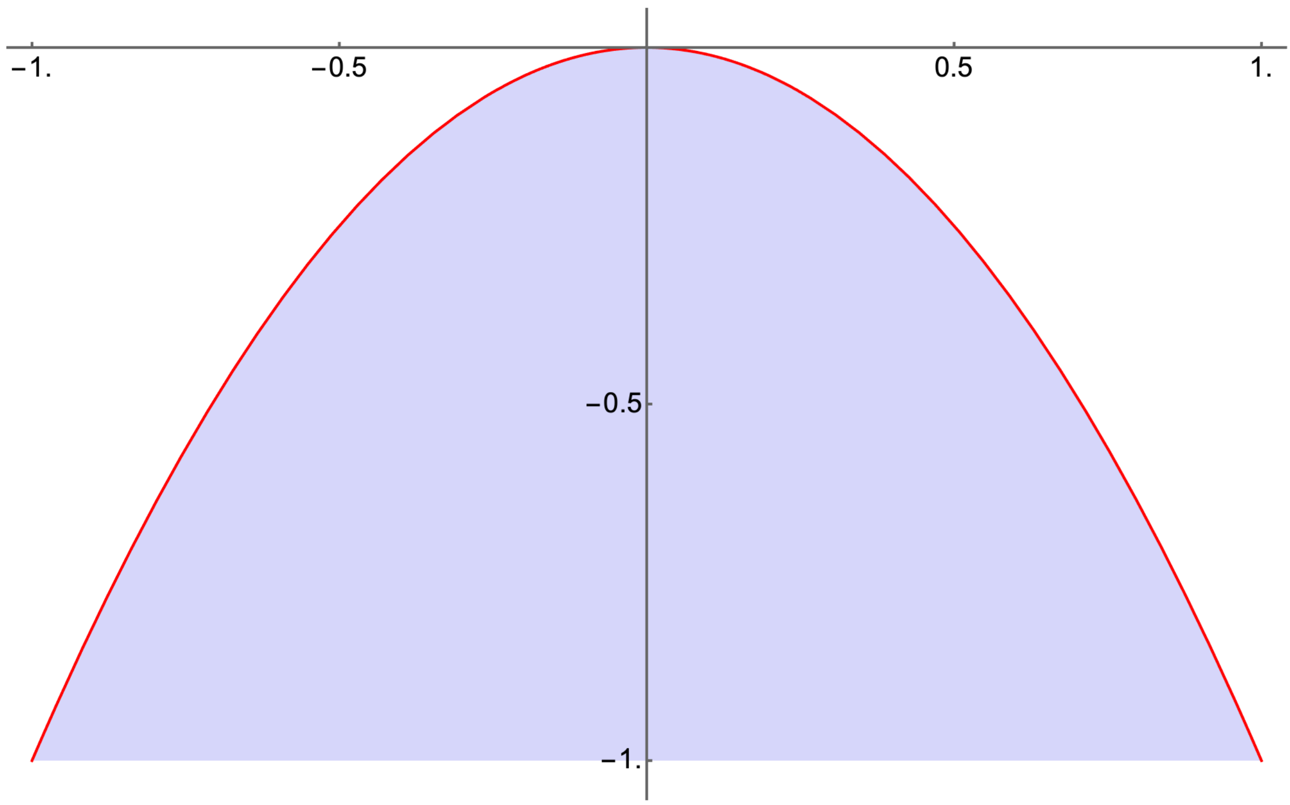


Figure 33: The graph of the concave function $-x^2$

Week 5: Convex and Constrained Optimisation

Convex optimisation

Last time we learned how to find and classify stationary points for a differentiable multivariate function. This was of limited use in general since it only allowed us to find **local maximums** or **local minimums**.

However if the function we want to find an optimal point for is **convex** or **concave** then local maximums and minimums are also **global** maximums and minimums. This simplifies the problem significantly.

The study of optimisation problems where the function is convex or concave is called **convex optimisation**. An example of this is Least Squares is statistics and Linear Programming. The difficulty of solving such problems is significantly reduced. For example the following is a very popular algorithm for solving very large scale convex optimisation problems.

The Gradient Descent Algorithm

For problems with very large numbers of variables the common technique for finding stationary points becomes intractible. This is often the case in practical applications of optimisation in, say, data science and statistics.

The **Gradient Descent Algorithm** provides a way of finding, approximately, the position of a stationary point for a function $f\colon \mathbb{R}^n \to \mathbb{R}$. Recall the gradient for such a function is

$$\nabla f(x_1, x_2, \dots, x_n) = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix}.$$

We introduced the gradient in Week 3. We learned that if we stand at a point on a surface and drop a ball then it falls in the direction directly opposite the gradient. The Gradient Descent

Algorithm uses the idea of dropping a ball and follows its path to find where it comes to rest. The thinking is that if it comes to rest then it must do so at a local minimum.

In Figure 34 we have drawn the contour plot of a function. Suppose we drop the ball at the point $\mathbf{x}_0$. Then this falls in the direction opposite the gradient, so following a vector of the form $-\eta\nabla f(\mathbf{x}_0)$ for some fixed value $\eta > 0$.

The algorithm thus updates $\mathbf{x}_0$ to give the next guess,

$$\mathbf{x}_1 = \mathbf{x}_0 - \eta\nabla f(\mathbf{x}_0).$$

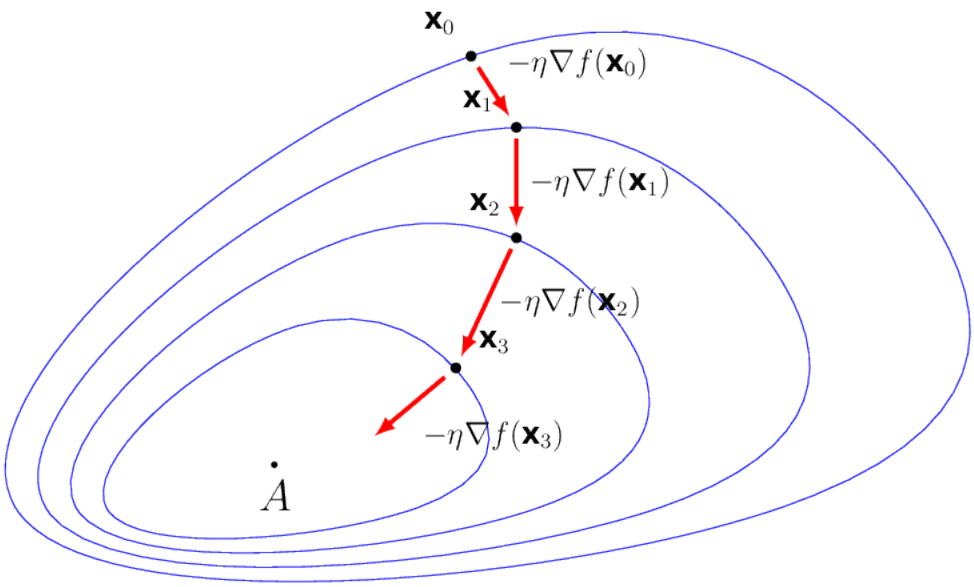


Figure 34: The Gradient Descent Algorithm

We continue in this way, producing better and better guesses. After a number of iterations we eventually reach the ‘bottom’ of the surface, i.e. approximately at the local minimum.

Since we can’t always expect to exactly reach the stationary point, we normally just specify a level of precision $\delta > 0$ that is acceptable. The algorithm is written like this.

Input: $f\colon \mathbb{R}^n \to \mathbb{R}$; $\eta > 0$; $\delta > 0$, $\mathbf{x}_0$

Algorithm:

1. $t \leftarrow 0$
2. Define $\mathbf{x}_{t+1} = \mathbf{x}_t - \eta\nabla f(\mathbf{x}_t)$
3. If $|\mathbf{x}_t - \mathbf{x}_{t+1}| > \delta$ let $t \leftarrow t+1$ and go to step 2.

Output: $\mathbf{x}_t$

The value of $\eta > 0$ is called the **learning rate**. In this version of the Gradient Descent Algorithm it is fixed.

This is the basic variation of the algorithm and there are many improvements that can be made. We will discuss these in the seminars.

Example

Suppose we have four points in $\mathbb{R}^2$:

$$(1,0) \quad (1,1) \quad (0,1) \quad (0,0)$$

What is the closest point to all four points?

Here the euclidead distance from a point (x, y) to (a, b) is

$$\sqrt{(x-a)^2 + (y-b)^2}.$$

Since the square root is an increasing function, finding the minimum value of this is equivalent to finding the minimum value of

$$(x-a)^2 + (y-b)^2.$$

If we want to find the minimum distance to **all four** points then we can minimise the function:

$$\begin{aligned} f(x,y) &= ((x-1)^2 + y^2) + ((x-1)^2 + (y-1)^2) \\ &\quad + (x^2 + (y-1)^2) + (x^2 + y^2) \\ &= 2x^2 + 2(x-1)^2 + 2y^2 + 2(y-1)^2 \\ &= 4\left(x^2 - x + y^2 - y + 1\right) \end{aligned}$$

Here

$$\nabla f(x,y) = 4\begin{pmatrix} 2x-1 \\ 2y-1 \end{pmatrix}$$

So if we start at $(1, 1)$ and take $\eta = 0.1$ then we may proceed as follows to calculate the next estimate:

$$\mathbf{x}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} - 0.1 \times 4 \begin{pmatrix} 2\times 1 - 1 \\ 2 \times 1 - 1 \end{pmatrix} = \begin{pmatrix} 0.6 \\ 0.6 \end{pmatrix}$$

Continuing, the next iteration is

$$\mathbf{x}_2 = \begin{pmatrix} 0.6 \\ 0.6 \end{pmatrix} - 0.1 \times 4 \begin{pmatrix} 2\times 0.6 - 1 \\ 2 \times 0.6 - 1 \end{pmatrix} = \begin{pmatrix} 0.52 \\ 0.52 \end{pmatrix}.$$

The next two values are:

$$\mathbf{x}_2 = \begin{pmatrix} 0.504 \\ 0.504 \end{pmatrix}, \quad \mathbf{x}_3 = \begin{pmatrix} 0.5008 \\ 0.5008 \end{pmatrix}.$$

The values **converge** to the local minimum $\begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix}$.

Convergence of Gradient Descent

If $\mathbf{x}_t \to \mathbf{x}$ as $t \to \infty$ we say that the algorithm converges. This means that

$$|\mathbf{x}_t - \mathbf{x}| \to 0, \qquad t \to \infty.$$

It doesn't always converge because there isn't always a minimum.

The Gradient Descent Algorithm as it is given above is in its most basic form. It is possible to improve the algorithm in different ways.

As for any numerical algorithm there are a number of questions that need to be understood:

- Under what circumstances does the algorithm converge to the right answer?
- What is the effect of the parameters on the algorithm, specifically η here?
- Can the algorithm be improved by modifying parts of it?

These questions will be studied in the seminars.

Constrained optimisation

We now turn our attention to **constrained optimisation** for multivariate functions.

Profit seeking businesses are often dealing with multiple constraints that affect levels of production, such as availability of resources, levels of demand and many other factors. So it is often unrealistic to model problems as unconstrained optimisation problems.

Suppose for example that a manufacturing company makes two products A and B. The profit made from selling the items, after taking costs into account, is given by the function

$$f(x, y) = 2x + 3y,$$

where x is the quantity of product A and y product B.

There may be a number of constraints to production. With functions of one variable constraints manifested themselves simply as restrictions to the domain of f. For multivariate functions we have more scope to be precise about the constraints. For example suppose the amount of available resources used when making products is

$$(2x^2 - xy + y^2)\text{kg}$$

and we have maxmimum of $10,000$kg available. This then becomes the **constraint**

$$2x^2 - xy + y^2 = 10000.$$

So we might model this problem as follows:

$$\begin{aligned} \text{maximise} \quad & f(x, y) = 2x + 3y \\ \text{subject to} \quad & 2x^2 - xy + y^2 = 10000 \end{aligned}$$

Analysing contrained optimisation problems

We will develop a method for solving constrained optimisation problems called **Lagrange Multipliers** after the Italian mathematician Joseph-Louis Lagrange.

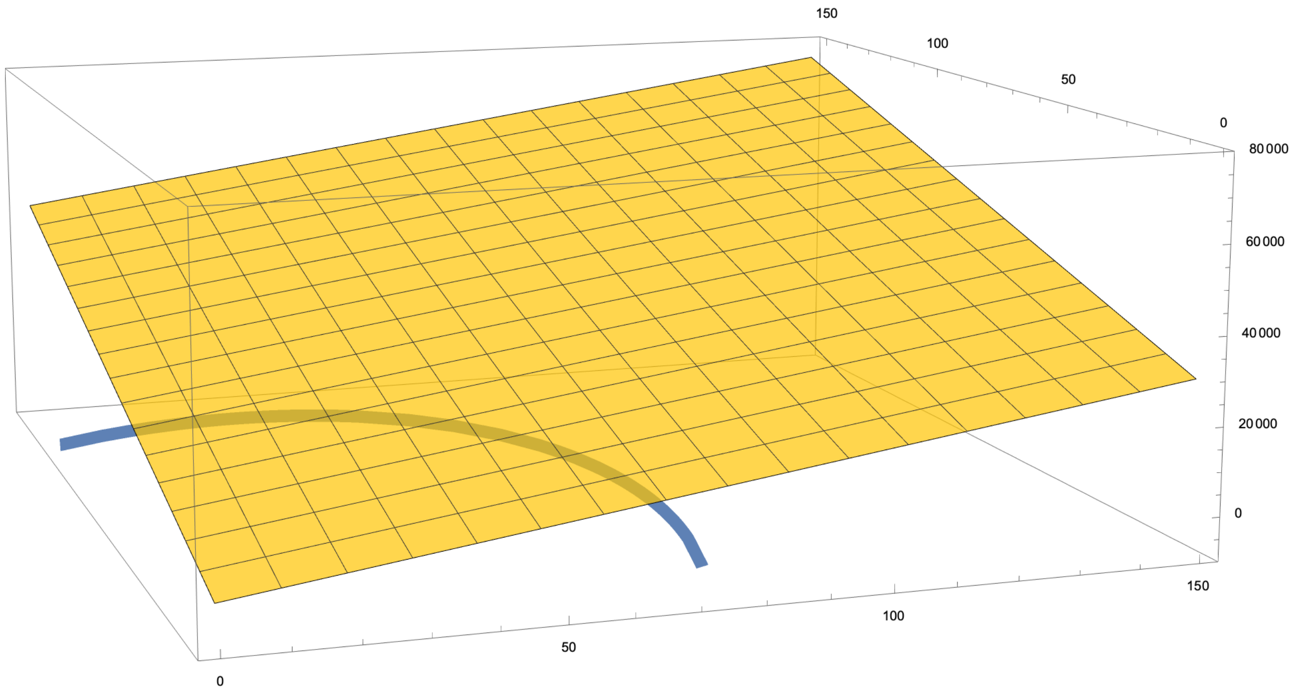


Figure 35: Constrained optimisation

In Figure 35 the surface $z = f(x, y)$ is drawn. The constraint, which we've written as

$$2x^2 - xy + y^2 = 10000$$

is the curve on the $x - y$ plane underneath the surface.

So the problem is to find the maximum z value on the surface 'above' the curve. Figure 36 projects the curve directly upwards. We're thus trying to find the largest z-coordinate on the intersection of this projected curve and the surface.

Lagrange multipliers - an explanation

Draw the contour plot of the surface, as in Figure 37, where the red curve is the curve implicitly defined by the constraint

$$2x^2 - xy + y^2 = 10000.$$

The green lines are the level curves

$$f(x, y) = c$$

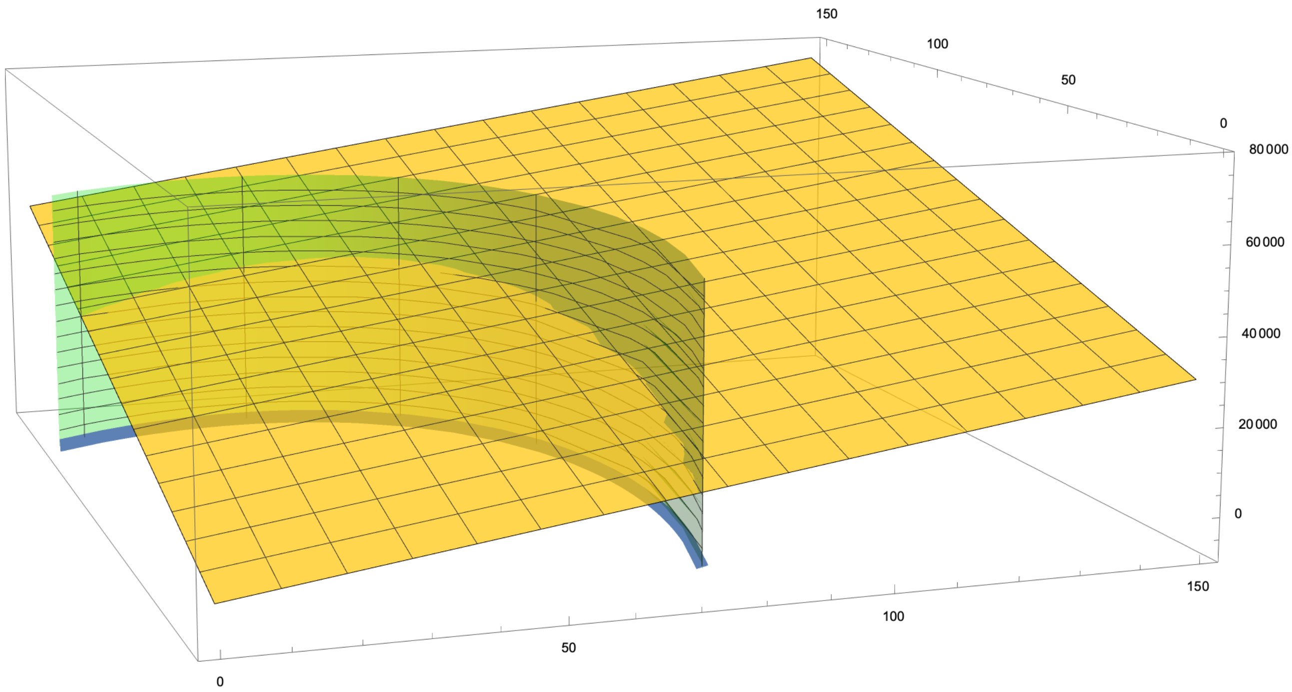


Figure 36: Constrained optimisation - the maximum value

for $c = 20000, 30000, 40000$, and 50000.

The key concept to notice is that at the level curve that represents the maximum is the one that is tangent to the constraint curve.

Suppose the optimal point is (x_m, y_m). Then we learned in Week 3 that the gradient $\nabla f(x_m, y_m)$ is a vector perpendicular to the level curve.

Now let $g(x, y) = 2x^2 - xy + y^2$ then the constraint can **also** be thought of as a level curve,

$$g(x, y) = 10000.$$

From what we have said, since this level curve is tangent to the level curve for f at (x_m, y_m) you would expcect the gradient $\nabla g(x_m, y_m)$ to be a vector pointing in the same direction as $\nabla f(x_m, y_m)$.

We write this as

$$\nabla f(x_m, y_m) = \lambda \nabla g(x_m, y_m)$$

where λ is a scalar called the **Lagrange multiplier**.

Now we can calculate these gradients:

$$\nabla f = \begin{pmatrix} 2 \\ 3 \end{pmatrix}, \qquad \nabla g = \begin{pmatrix} 4x - y \\ -x + 2y \end{pmatrix}$$

So if we want to find (x_m, y_m) we need to solve the equations

$$\begin{pmatrix} 2 \\ 3 \end{pmatrix} = \lambda \begin{pmatrix} 4x - y \\ -x + 2y \end{pmatrix}.$$

This can be written as two equations

$$\begin{aligned} 2 &= (4x - y)\lambda \\ 3 &= (-x + 2y)\lambda \end{aligned}$$

We have here three unknown variables and two equations, so this might seem like a problem, until we remember that we also have a third equation, the original constraint

$$2x^2 - xy + y^2 = 10000.$$

If we solve these equations we can find the optimal point.

First rewrite the two original equations,

$$\begin{aligned} \lambda^{-1} &= (4x - y)/2 \\ \lambda^{-1} &= (-x + 2y)/3 \end{aligned}$$

So $(4x - y)/2 = (-x + 2y)/3$ from which we can simplify to get

$$y = 2x.$$

In the constraint equation this becomes:

$$\begin{aligned} 10000 &= 2x^2 - x(2x) + (2x)^2 \\ &= 4x^2 \end{aligned}$$

Hence $x = \pm 50$, and $y = \pm 100$. We can also calculate

$$\lambda = \pm \frac{1}{50}.$$

To decide which of these points is the maximum (or minimum) we can evaluate the objective function f,

$$f(-50, -100) = -400, \qquad f(50, 100) = 400.$$

Therefore the optimal point is $(50, 100)$.

The method of Lagrange multipliers

The method of Lagrange multipliers simplifies the above method by changing out **constrained optimisation** problem into an **unconstrained optimisation** problem where we simply need to find stationary points.

Suppose we want to find the optimal point for a function $f(x, y)$ subject to the constraint

$$g(x, y) = 0.$$

Here $f, g \in C^1(\mathbb{R}^2)$.

The method of Lagrange multipliers is as follows:

1. Define the function $\mathcal{L}(x, y, \lambda) = f(x, y) - \lambda g(x, y)$
2. Find the stationary points for $\mathcal{L}$
3. For each stationary point (x, y, λ), evaluate $f(x, y)$ and determine the optimal value

Example

What is the largest rectangle that we can fit into an ellipse of minor radius 1 and major radius 2?

The equation for the ellipse is

$$x^2 + \frac{y^2}{4} = 1.$$

If we let (x, y) be an arbitrary point on this ellipse then we may form the rectangle by joining the points $(\pm x, \pm y)$. This has area $4xy$.

So let

$$\mathcal{L}(x, y, \lambda) = 4xy - \lambda \left(x^2 + \frac{y^2}{4} - 1 \right)$$

Finding the partial derivatives and setting them to 0,

$$\frac{\partial \mathcal{L}}{\partial x} = 4y - 2x\lambda = 0$$
$$\frac{\partial \mathcal{L}}{\partial x} = 4x - \frac{y\lambda}{2} = 0$$
$$\frac{\partial \mathcal{L}}{\partial \lambda} = x^2 + \frac{y^2}{4} - 1 = 0$$

From the first two equations we get

$$\lambda = \frac{2y}{x} = \frac{8x}{y}.$$

So that

$$y^2 = 4x^2.$$

In the constraint equation, the third equation,

$$2x^2 - 1 = 0$$

or $x = \pm \frac{1}{\sqrt{2}}$.

Substituting back in we find

$$y^2 = 4x^2 = 2$$

so that $y = \pm\sqrt{2}$.

We can also find $\lambda = \pm 4$ here although it is immaterial.

The largest area is when

$$(x, y) = \left(\frac{1}{\sqrt{2}}, \sqrt{2}\right)$$

and the largest rectangle has area 4.

Extensions of Lagrange multipliers

The method can be extended to any number of variables $f\colon \mathbb{R}^n \to \mathbb{R}$ by writing

$$\mathcal{L}(x_1, x_2, \dots, x_n, \lambda) = f(x_1, x_2, \dots, x_n) - \lambda(g(x_1, x_2, \dots, x_n) - c).$$

It can also be extended to any number of constraints.

Suppose we want to

$$\begin{aligned} \text{maximise} \quad & f\colon \mathbb{R}^n \to \mathbb{R} \\ \text{subject to} \quad & g_1(\mathbf{x}) = 0 \\ & g_2(\mathbf{x}) = 0 \\ & \vdots \\ & g_m(\mathbf{x}) = 0 \end{aligned}$$

Then the generalised method of Lagrange multipliers finds the stationary points of the function

$$\mathcal{L}(x_1, x_2, \dots, x_n, \lambda_1, \lambda_2, \dots, \lambda_m) = f(x_1, x_2, \dots, x_n) - \sum_{k=1}^{m} \lambda_k g_k(x_1, x_2, \dots, x_n).$$

Lagrange multipliers and Gradient Descent

It might be tempting to think that for large and complex constrained optimisation problems can be solved approximately using the Gradient Descent Algorithm. After all the function $\mathcal{L}$ has $m + n$ variables.

Unfortunately it turns out that the stationary points for the function $\mathcal{L}$ are saddles so that the algorithm won't work.

There are other methods that can be used however. One such method is **Newton's method**.

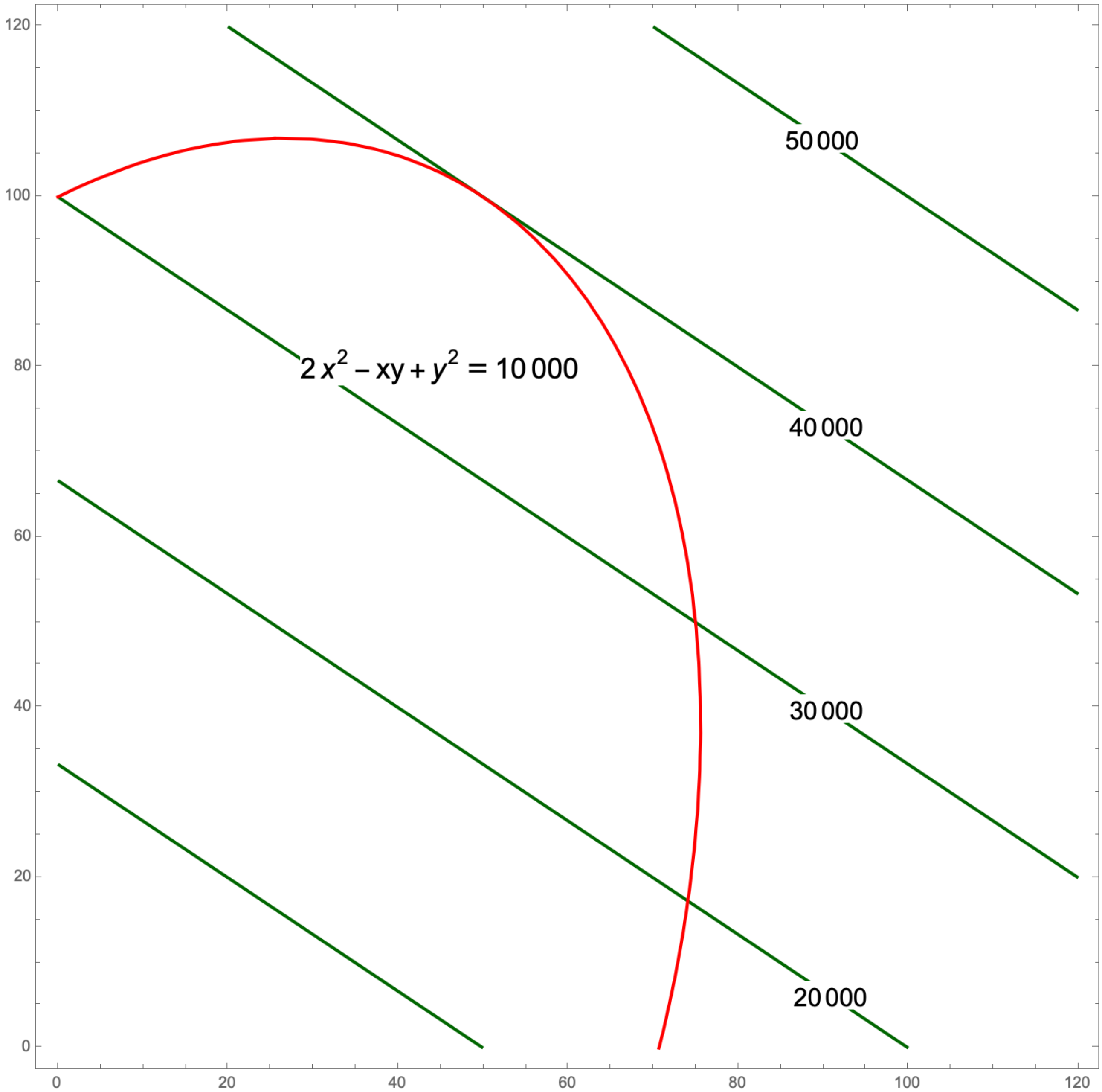


Figure 37: The contour plot and constraint curve

Week 6 Differential Equations

Modelling

Mathematical Modelling is the process of finding a mathematical description of a business problem. The purpose is not always to *solve* the problem, it is often a useful method for developing a clearer understanding of the main constituents of the scenario you are modelling.

Example - free market economics

The economist Adam Smith (not your MSIN0010 lecturer) said that an economy is a **free market** if the following three laws are the only ones that rule the change in price and excess demand (the difference between supply and demand):

If at a given time,

1. demand is less than supply, then prices decrease so the demand increased;
2. demand is greater than supply, then prices increase, so the supply increases
3. demand equals supply, then prices are constant, so the excess deman remains constant

In order to model this we first need to introduce some notation. We'll call $s(t)$ the supply, $d(t)$ the demand, and $r(t) = d(t) - s(t)$ the excess demand. Assume these functions are differentiable, the laws then become:

1. If $r(t) < 0$ then $p'(t) < 0$ and so $r'(t) > 0$;
2. If $r(t) > 0$ then $p'(t) > 0$ and so $r'(t) < 0$;
3. If $r(t) = 0$ then $p'(t) = 0$ and so $r'(t) = 0$.

So we can deal only with r and r' and p'. Looking at these rules we can see that r' and r always have the opposite sign.

We might model this by saying that there are constants $a > 0$ and $b > 0$ so that

$$r'(t) = -ar(t)$$
$$p'(t) = br(t)$$

This example gives us an interesting (but probably not hugely convincing) mathematical model of a free market.

Notice a few typical properties of the model:

- We have quantities that change, often with time but not always. We write these as fuctions.
- We introduce some relationships that we believe might link the quantities. We write these are equations between the functions, variables, derivatives etc.

Differential Equations and their solutions

The equations given in the example are called **differential equations** since they describe the derivative of the function in terms of the variable and the function.

A **solution** to a differential equation is a function that satisfies the equation.

Example - free market continued

If we consider the first equation we had in the free market model, $r'(t) = -ar(t)$ we can check that the function $r(t) = e^{-at}$ satisfies this differential equation since

$$r'(t) = -ae^{-at} = -ar(t).$$

The second equation was $p'(t) = br(t)$ which we can now write as $p'(t) = be^{-at}$.

We can integrate this to get

$$p(t) = \int be^{-at}\,\mathrm{d}t = -\frac{b}{a}e^{-at}.$$

These aren't the only solutions however, the function, $r(t) = 2e^{-at}$ is also a solution to the first equation. So is $r(t) = -10e^{-at}$. In fact for any real number A the function

$$r(t) = Ae^{-at}$$

is a solution.

Similarly you may have noticed we didn't include the constant of integration in the calculation of $p'(t)$. If we do, and call it B, then we get

$$p(t) = \int bAe^{-at}\,\mathrm{d}t = -A\frac{b}{a}e^{-at} + B.$$

These two solutions are called the **general solution** to the differential equation.

Example - probability

Suppose at a particular coffee shop the length of time it takes for you to get served is random. You notice that if you know that it is going to take at least xs then the probability that it takes an extra ys is independent. In other words,

$$P(X > x + y \mid X > x) = P(X > y)$$

This is the so-called **memoryless** property since knowing that $X > x$ means any additional wait in the queue is independent.

If we let $g(x) = P(X > x)$ then since $X \geq 0$, the domain of g is $[0, \infty)$. We also know,

$$g(0) = 1, \qquad \lim_{x\to\infty} g(x) = 0$$

Using Bayes's Theorem

$$\begin{aligned}
P(X > x + y \mid X > x) &= \frac{P(X > x + y \wedge X > x)}{P(X > x)} \\
&= \frac{P(X > x + y)}{P(X > x)} = \frac{g(x+y)}{g(x)}
\end{aligned}$$

Hence the memoryless property becomes $\dfrac{g(x+y)}{g(x)} = g(y)$, or simply

$$g(x + y) = g(x)g(y)$$

Now we can calculate the derivative of this function at x

$$\begin{aligned}
g'(x) &= \lim_{y\to 0} \frac{g(x+y) - g(x)}{y} = \lim_{y\to 0} \frac{g(x)g(y) - g(x)}{y} \\
&= g(x) \lim_{y\to 0} \frac{g(y) - 1}{y} = g(x) \lim_{y\to 0} \frac{g(y) - g(0)}{y} \\
&= g'(0)g(x)
\end{aligned}$$

It can be shown that $g'(0) < 0$, we will write $-\lambda = g'(0)$. So g is a solution to the differential equation

$$g' = -\lambda g.$$

This is the same equation we had before, its solution is

$$g(x) = Ae^{-\lambda x}.$$

This is the **general solution** to this differential equation.

In this example we can do better. We know that $g(0) = 1$. So

$$1 = Ae^{-\lambda \times 0} = A.$$

Therefore,

$$g(x) = e^{-\lambda x}.$$

The condition $g(0) = 1$ is called an **initial condition** since it gives the value of the solution at the initial point, $x = 0$. It allows us to find the value(s) of constants in the general solution.

Continuing this example, the cummulative distribution function is $F(x) = 1 - P(X > x) = 1 - g(x)$. So $F(x) = 1 - e^{-\lambda x}$ and therefore X is **exponentially distributed** $X \sim Exp(\lambda)$.

These two examples show that differential equations are an important tool for examining problems that occur in all sorts of areas.

First order differential equtions

A **first order ordinary differential equation** is an equation of the form

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

where F is a function - it is called first order since it only includes the first derivative y'. Here x is called the **independent variable** and y is the **dependent variable**.

A **solution** is a function $f(x)$ so that

$$F(x, f(x), f'(x)) = 0$$

for all x.

Some examples are given below (a and r are constants):

$$\begin{aligned} y' &= k(M - y) && \text{Newton's law of cooling} \\ y' &= ay && \text{exponential decay/growth} \\ y' &= ry(1 - y) && \text{logistic growth} \end{aligned}$$

Note: since both dependent variable y and independent variable x often appear together in a differential equation do make sure you know which is which. Often for example if the independent variable is time, it is denoted by the letter t.

A differential equation together with an initial condition is called an **initial value problem**.

Solving differential equations - integration

If a differential equation is of the form

$$y' = f(x)$$

where f is continuous, then we can integrate directly:

Example

Solve the initial value problem

$$y' = x(x - 1), \qquad y(0) = 1$$

Since the right hand side does not include the dependent variable y, we can integrate this to find the **general solution**,

$$\begin{aligned} y &= \int x(x-1)\,\mathrm{d}x = \int x^2 - x\,\mathrm{d}x \\ &= \frac{x^3}{3} - \frac{x^2}{2} + c \end{aligned}$$

We can now use the initial condition $y(0) = 1$,

$$1 = \frac{0^3}{3} - \frac{0^2}{2} + c = c$$

So the **solution** is

$$y = \frac{x^3}{3} - \frac{x^2}{2} + 1$$

Solving differential equations - separation of variables

If a differential equation is of the form

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

where both f and g are continuous, then we can solve this using the chain rule for integration.

$$\begin{aligned} \frac{\mathrm{d}y}{\mathrm{d}x} &= f(x)g(y) \\ \frac{1}{g(y)}\frac{\mathrm{d}y}{\mathrm{d}x} &= f(x) \\ \int \frac{1}{g(y)}\frac{\mathrm{d}y}{\mathrm{d}x}\mathrm{d}x &= \int f(x)\mathrm{d}x \\ \int \frac{1}{g(y)}\mathrm{d}y &= \int f(x)\mathrm{d}x \end{aligned}$$

This is often written as follows:

$$\begin{aligned} \frac{\mathrm{d}y}{\mathrm{d}x} &= f(x)g(y) \\ \frac{1}{g(y)}\mathrm{d}y &= f(x)\mathrm{d}x \\ \int \frac{1}{g(y)}\mathrm{d}y &= \int f(x)\mathrm{d}x \end{aligned}$$

It is important however to realise the notation $\frac{\mathrm{d}y}{\mathrm{d}x}$ is **not** a fraction and can't be manipulated like this. You should however appreciate that the notation for the derivative is clever enough to make it look like you can.

Example

Solve the initial value problem,

$$y' = 3x^2y, \qquad y(0) = 4.$$

Here,

$$\begin{aligned}
\frac{\mathrm{d}y}{\mathrm{d}x} &= 3x^2y \\
y^{-1}\mathrm{d}y &= 3x^2\mathrm{d}x \\
\int y^{-1}\mathrm{d}y &= \int 3x^2\mathrm{d}x \\
\log y &= x^3 + c \\
y &= e^c e^{x^3} = Ae^{x^3}
\end{aligned}$$

where I've written $A = e^c$ to simplify the calculation. This is the **general solution**.

Using the initial condition, $y(0) = 4$

$$4 = Ae^{0^3} = A$$

so the solution is

$$y = 4e^{x^3}.$$

Example - logistic population growth

A standard model for the growth of the population is the **logistic growth model**,

$$p' = kp(1-p), \qquad p(0) = p_0.$$

Here $k > 0$ is a constant, $p(t)$ is the size of the population at time t, and p_0 is the initial number of people in the population at time $t = 0$.

To find the general solution,

$$\begin{aligned}
\frac{\mathrm{d}p}{\mathrm{d}t} &= kp(1-p) \\
\frac{1}{p(1-p)}\mathrm{d}p &= k\,\mathrm{d}t \\
\log\frac{p}{1-p} &= kt + c \\
\frac{p}{1-p} &= Ae^{kt}, \qquad A = e^c, \\
p &= \frac{1}{1+Ae^{-kt}}.
\end{aligned}$$

Now, using the initial condition $p(0) = p_0$ we have

$$p_0 = \frac{1}{1+Ae^0} = \frac{1}{1+A}$$

so that,

$$A = \frac{1-p_0}{p_0}.$$

And finally by substituting this in and simplifying we get the solution,

$$p(t) = \frac{p_0}{p_0 + (1-p_0)e^{-kt}}.$$

Solving differential equations - integrating factors

A differential equation is called **linear** if it is of the form

$$y' + py = q$$

where p and q are functions only of the independent variable x. A linear differential equation can be solved by finding a function $F(x)$ so that

$$p(x) = \frac{F'(x)}{F(x)} = \frac{\mathrm{d}}{\mathrm{d}x}(\log F(x))$$

This is called an **integrating factor** for the differential equation. The reason it is useful is that we have the following,

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}x}(F(x)y) &= F(x)y' + F'(x)y \\ &= F(x)\left(y' + \frac{F'(x)}{F(x)}y\right) \\ &= F(x)\left(y' + p(x)y\right) \\ &= F(x)q(x).\end{aligned}$$

So we now have a differential equation of the type we saw initially (i.e. the right hand side does not include the dependent variable y). We can solve this by integration.

To find the integrating factor $F(x)$ we integrate

$$p(x) = \frac{\mathrm{d}}{\mathrm{d}x}(\log F(x))$$

to get

$$F(x) = \exp\left(\int p(x)\,\mathrm{d}x\right)$$

Example

Solve the initial value problem

$$y' + 2xy = x, \qquad y(0) = 1$$

Here $p(x) = 2x$ so an integrating factor is

$$F(x) = \exp \int 2x \, \mathrm{d}x = e^{x^2}$$

Note: we don't need the constant of integration, since we only need one integrating factor.

Now we solve the equation

$$\frac{\mathrm{d}}{\mathrm{d}x}(F(x)y) = F(x)q(x)$$

or,

$$\frac{\mathrm{d}}{\mathrm{d}x}(e^{x^2}y) = xe^{x^2}$$

Integrating we get

$$e^{x^2}y = \int xe^{x^2} \, \mathrm{d}x = \frac{1}{2}e^{x^2} + c$$

Dividing by e^{x^2} we find that the general solution is

$$y = \frac{1}{2} + ce^{-x^2}$$

Using the initial condition, $y(0) = 1$,

$$1 = \frac{1}{2} + ce^0 = \frac{1}{2} + c$$

so that $c = 1/2$ and the solution is

$$y = \frac{1}{2} + \frac{1}{2}e^{-x^2}.$$

Example - exponential growth/decay

A very common model of growth or decay is given by the equation

$$y' = ay, \qquad y(0) = y_0.$$

We first write this in the form

$$y' - ay = 0$$

so that $p(x) = -a$ and $q(x) = 0$.

Then an integrating factor is

$$F(x) = e^{-ax}.$$

And the general solution is found by solving

$$\frac{\mathrm{d}}{\mathrm{d}x}(F(x)y) = F(x)q(x)$$

For this equation the right hand side is 0 and so integrating we only get a constant, i.e. $e^{-ax}y = c$ and so the general solution is

$$y = ce^{ax}.$$

Substituting the initial condition in we find the solution is

$$y = y_0 e^{ax}.$$

Week 7 - Second order differential equations

A second order ordinary differential equation is an equation of the form

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

where F is a function. It is called second order because there is a second order derivative.

The scope of these types of differential equations is very broad and there are only a few methods for solving specific types of equation, unlike the first order equations. So will only deal with a very specific form of equation.

Linear equations with constant coefficients

A linear second order ordinary differential equation is an equation of the form

$$y'' + a(x)y' + b(x)y = r(x)$$

where a, b, r are continuous functions of x.

A **linear second order ordinary differential equation with constant coefficients** is an equation of the form

$$y'' + ay' + by = r(x)$$

where a, b are constants, and r is a continuous fuction of x.

For example,

$$y'' + y = 0 \qquad \text{simple harmonic motion}$$

The key to linear equations is that if $f(x)$ and $g(x)$ are two solutions for a linear equation, then so is

$$Af(x) + Bg(x)$$

for any constants A and B. This means that to find the **general solutions** of a linear equation we only need to find two separate solutions f and g.

Homogeneous equations

If $r(x) = 0$ then we get the **homogeneous** equation

$$y'' + ay' + by = 0.$$

To solve, we need to find two different functions that form solutions. The standard approach is to try a function of the form

$$f(x) = e^{\lambda x}$$

Here

$$f'(x) = \lambda e^{\lambda x}, \qquad f''(x) = \lambda^2 e^{\lambda x}.$$

So $f(x)$ is a solution if

$$\lambda^2 e^{\lambda x} + a\lambda e^{\lambda x} + be^{\lambda x} = 0$$

or, dividing by $e^{\lambda x}$,

$$\lambda^2 + a\lambda + b = 0.$$

It follows that $f(x) = e^{\lambda x}$ is a solution if and only if $\lambda^2 + a\lambda + b = 0$. This last equation is called the **auxiliary equation** and $\lambda^2 + a\lambda + b$ the **auxiliary function**.

You know that a quadratic equation like this has two solutions counting multiplicity. There are three possibilities:

1. Two distinct real roots
2. One real root of multiplicity 2
3. Two complex conjugate roots

We will treat these separately, the objective being to find two distinct solutions that we can combine to make the general solution of the equation.

Two distinct roots

The most straightforward possibility is that the auxiliary function has distinct real roots, α and β. In this case the general solution is

$$Ae^{\alpha x} + Be^{\beta x}$$

Example

Solve the homogeneous equation

$$y'' + 5y' + 6y = 0.$$

Here the auxiliary function is

$$\lambda^2 + 5\lambda + 6 = (\lambda + 2)(\lambda + 3).$$

This has roots $\lambda = -2$ and $\lambda = -3$. So the general solution is

$$y = Ae^{-2x} + Be^{-3x}.$$

On root of multiplicity 2

If the auxiliary function is of the form $(x-\alpha)^2$ then the general solution is

$$(A+Bx)e^{\alpha x}.$$

Example

Solve the homogeneous equation

$$y'' + 2y' + y = 0.$$

Here the auxiliary function is

$$\lambda^2 + 2\lambda + 1 = (\lambda+1)^2$$

This has a root at $\lambda = -1$ of multiplicity 2 so the general solution is

$$y = (A+Bx)e^{-x}.$$

Complex conjugate roots

If the auxiliary function has two complex conjugate roots $\alpha \pm \beta i$ then the general solution is

$$Ae^{\alpha x}\cos\beta x + Be^{\alpha x}\sin\beta x.$$

You may wonder where this comes from. It is in fact the same as the general solution $Ce^{\alpha+i\beta} + De^{\alpha-i\beta}$ but we have simplified it using Euler's formula $e^{i\beta} = \cos\beta x + i\sin\beta x$.

Example

Solve the homogeneous equation

$$y'' + 4y' + 5y = 0.$$

The auxiliary function here is

$$\lambda^2 + 4\lambda + 5.$$

The roots are

$$\lambda = 2 \pm i$$

so here the general solution is

$$y = Ae^{2x}\cos x + Be^{2x}\sin x$$

Nonhomogeneous equations

A **nonhomogeneous** linear differential equation with constant coefficients is an equation of the form

$$y'' + ay' + by = r(x)$$

where r is a continuous function of the independent variable x.

The general solution to a nonhomogeneous equation is given by

$$c(x) + p(x)$$

where $c(x)$ is the general solution to the homogeneous equation (called the **complementary function**)

$$y'' + ay' + by = 0$$

and $p(x)$ is a **particular solution** of the nonhomogeneous equation

$$y'' + ay' + by = r(x)$$

A particular solution is simply any solution to the equation, i.e. it must satisfy

$$p''(x) + ap'(x) + bp(x) = r(x).$$

Since we're not interesting in finding all possible solutions (that's what the complementary function is) we only need one particular example.

We will consider a number of examples that will demonstrate an approach to finding a particular solution. The general idea is that we use $r(x)$ to inform the type of function that $p(x)$ must be in order for the left hand side to be $r(x)$.

Example

Find the general solution of

$$y'' + 9y = e^x.$$

First the complementary function is

$$c(x) = A\cos 3x + B\sin 3x.$$

To find the particular solution we see we must choose $p(x)$ so that we end up with e^x on the right hand side.

A sensible guess will be a function of the form

$$p(x) = de^x$$

for some constant d. In order to find d, we substitute p into the left hand side, thus.

$$\begin{aligned} p''(x) + 9p(x) &= e^x \\ de^x + 9de^x &= e^x \\ 10de^x &= e^x \end{aligned}$$

So for $p(x)$ to be a particular solution we must have $d = 1/10$,

$$p(x) = \frac{1}{10}e^x.$$

Hence, the general solution to this equation is $c(x) + p(x)$, or

$$y = A\cos 3x + B\sin 3x + \frac{1}{10}e^x$$

Example

Find the general solution of

$$y'' - y = e^x.$$

The complementary function here is

$$c(x) = Ae^x + Be^{-x}.$$

Like the previous example we might now want to try $p(x) = de^x$. However note that e^x is already one of the two solutions we used to make up the complementary function. This means we can't use this again, so instead we try $p(x) = dxe^x$.

Now,

$$p'(x) = d(x+1)e^x, \qquad p''(x) = d(x+2)e^x.$$

So for $p(x)$ to be a particular solution it must satisfy

$$\begin{aligned} p''(x) - p(x) &= e^x \\ d(x+2)e^x - dxe^x &= e^x \\ 2de^x &= e^x. \end{aligned}$$

So for $p(x)$ to be a particular solution, $d = 1/2$, i.e.

$$p(x) = \frac{1}{2}xe^x.$$

The general solution is therefore

$$y = Ae^x + Be^{-x} + \frac{1}{2}xe^x.$$

Example

Find the general solution to the equation

$$y'' - 6y' + 8y = \sin 2x.$$

Here the auxhiliary equation is $\lambda^2 - 6\lambda + 8 = 0$ which has solutions $\lambda = 2$ and $\lambda = 4$. The complementary function is therefore

$$c(x) = Ae^{2x} + Be^{4x}.$$

This time $r(x) = \sin 2x$ so we will try

$$p(x) = j \sin 2x + k \cos 2x.$$

Now

$$\begin{aligned} p'(x) &= 2j \cos 2x - 2k \sin 2x \\ p''(x) &= -4j \sin 2x - 4k \cos 2x. \end{aligned}$$

Hence for $p(x)$ to be a particular solution we must have

$$\begin{aligned} \begin{matrix} (-4j \sin 2x - 4k \cos 2x) - 6(2j \cos 2x - 2k \sin 2x) \\ +8(j \sin 2x + k \cos 2x) \end{matrix} &= \sin 2x \\ (-4j + 12k + 8j) \sin 2x + (-4k - 12j + 8k) \cos 2x &= \sin 2x \\ (4j + 12k) \sin 2x + (4k - 12j) \cos 2x &= \sin 2x \end{aligned}$$

So for $p(x)$ to be a particular solution we must have

$$\begin{aligned} 4j + 12k &= 1 \\ -12j + 4k &= 0. \end{aligned}$$

The solution to these simultaneous equations is

$$j = \frac{1}{40}, \qquad k = \frac{3}{40}.$$

The general solution is therefore

$$y = Ae^{2x} + Be^{4x} + \frac{1}{40}(\sin 2x + 3 \cos 2x).$$

In summary we pick $p(x)$ so that 1. when differentiated once or twice, $p(x)$ includes terms in $r(x)$, 2. it is different to the terms in the complimentary function, if not then try multiplying by powers of x

Determining values of constants

As we did with first order differential equations we can find the values of any constants in the general solution if we are given more information about the solution.

With second order differential equations we normally get two constants so we need two pieces of information. These come in two forms:

1. **Initial value problems** - these are problems that include values of the solution at their initial value, normally when the independent variable is 0. They have the form
$$y(0) = y_0, \qquad y'(0) = y_1.$$
2. **Boundary value problems** - here the solution $y(x)$ has domain $[0, x_E]$, and the conditions are boundary conditions,
$$y(0) = y_0, \qquad y(x_E) = y_E.$$

Example

Find the solution to the initial value problem,
$$y'' - 6y' + 8y = \sin 2x, \qquad y(0) = 0, y'(0) = 1.$$
We found the general solution to this was
$$y = Ae^{2x} + Be^{4x} + \frac{1}{40}(\sin 2x + 3\cos 2x).$$
Hence if $y(0) = 0$,
$$0 = A + B + \frac{3}{40}.$$
Differentiating the general solution gives
$$y' = 2Ae^{2x} + 4Be^{4x} + \frac{1}{40}(2\cos 2x - 3\sin 2x).$$
So if $y'(0) = 1$,
$$1 = 2A + 4B + \frac{1}{20}.$$
These can be solved to give
$$A = -\frac{5}{8}, \qquad B = \frac{11}{20}.$$
Hence the solution is
$$y = -\frac{5}{8}e^{2x} + \frac{11}{20}e^{4x} + \frac{1}{40}(\sin 2x + 3\cos 2x).$$

Example

Find the solution to the boundary value problem,
$$y'' + 9y = e^x, \qquad y(0) = 0, y(\pi/2) = 0.$$
Previously we found the general solution was
$$y = A\cos 3x + B\sin 3x + \frac{1}{10}e^x.$$

First, $y(0) = 0$,

$$0 = A + \frac{1}{10}$$

so that $A = -1/10$. Second, $y(\pi/2) = 0$,

$$\begin{aligned} 0 &= A \cos 3\pi/2 + B \sin 3\pi/2 + \frac{1}{10} e^{\pi/2} \\ &= -B + \frac{1}{10} e^{\pi/2} \end{aligned}$$

so that $B = e^{\pi/2}/10$.

The solution is therefore

$$y = -\frac{1}{10} \cos 3x + \frac{e^{\pi/2}}{10} \sin 3x + \frac{1}{10} e^x.$$

Week 8 - Multiple Integration

In this final week's work you will learn how to integrate functions

$$f\colon \mathbb{R}^2 \to \mathbb{R}.$$

The methods can be generalised to any number of variables, but the calculations are quite awkward. So we'll stick to two variables.

Introduction to integration

First let's review some important concepts from integration. First we define the integral to be the area beneath the graph of a function f between two points a and b.

The notation for this is normally

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

however we might also write it as

$$\int_{[a,b]} f(x)\,\mathrm{d}x$$

to make clear that it is the area under the curve for all values in the closed interval $[a, b]$. In general we can, if we wanted, replace $[a, b]$ with any nice enough set E and write this as

$$\int_E f(x)\,\mathrm{d}x.$$

For functions of two variables we do the same thing. We start with a set $E \subset \mathbb{R}^2$ and define the integral to be the **volume under the surface**. In Figure 39 we have drawn a typical example, with the set E being a square on the x, y-plane.

We call this a **double integral** and write it as

$$\iint_E f(x,y)\,\mathrm{d}x\mathrm{d}y.$$

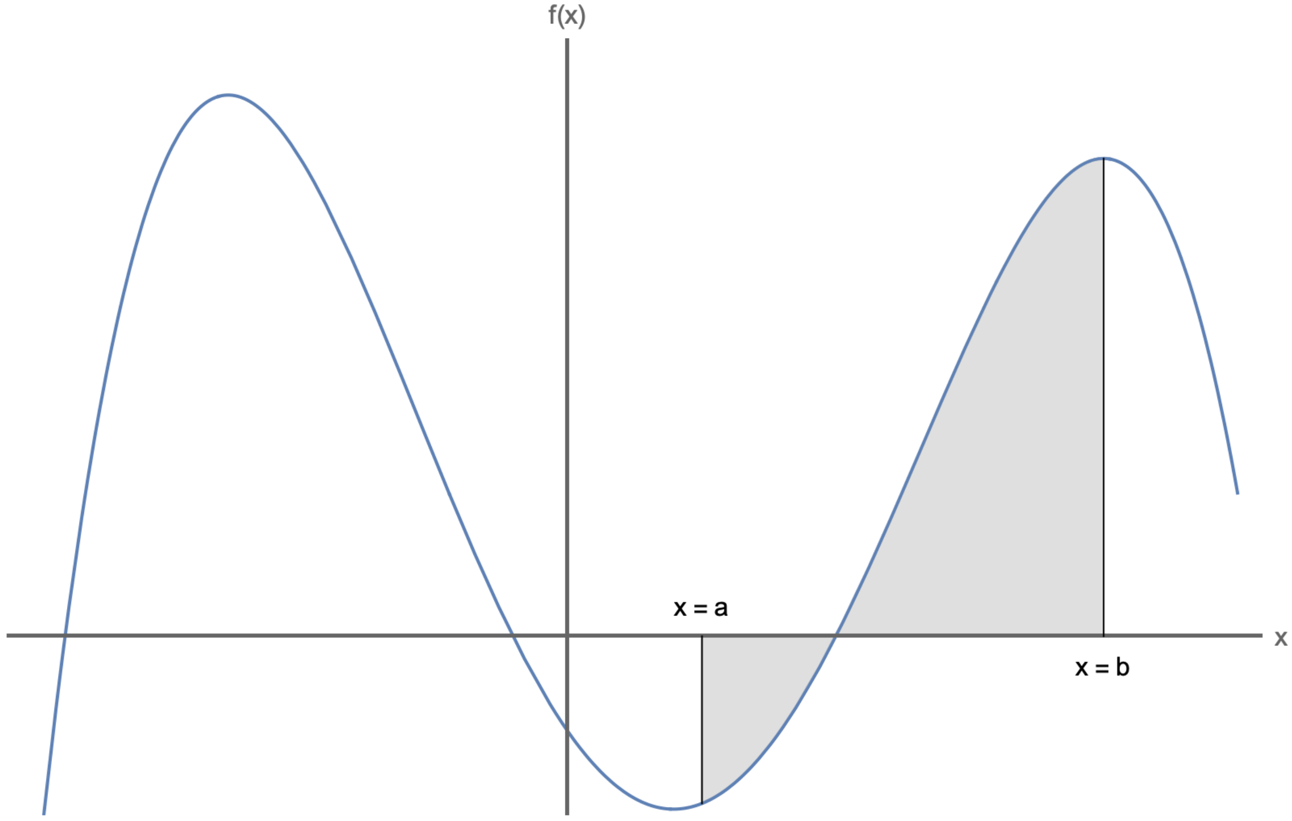


Figure 38: The intgral

The integral over a rectangle

To illustrate the method we start with a rectangle,

$$E = \{(x, y) \colon 1 \leq x \leq 2,\ 2 \leq y \leq 4\}.$$

This is drawn in Figure 40.

To integrate the function $f(x, y) = x^2 y$ over this rectangle we treat each variable separately, reminiscent of when we were calculating partial derivatives. So with $E = \{(x, y) \colon 1 \leq x \leq 2,\ 2 \leq y \leq 4\}$.

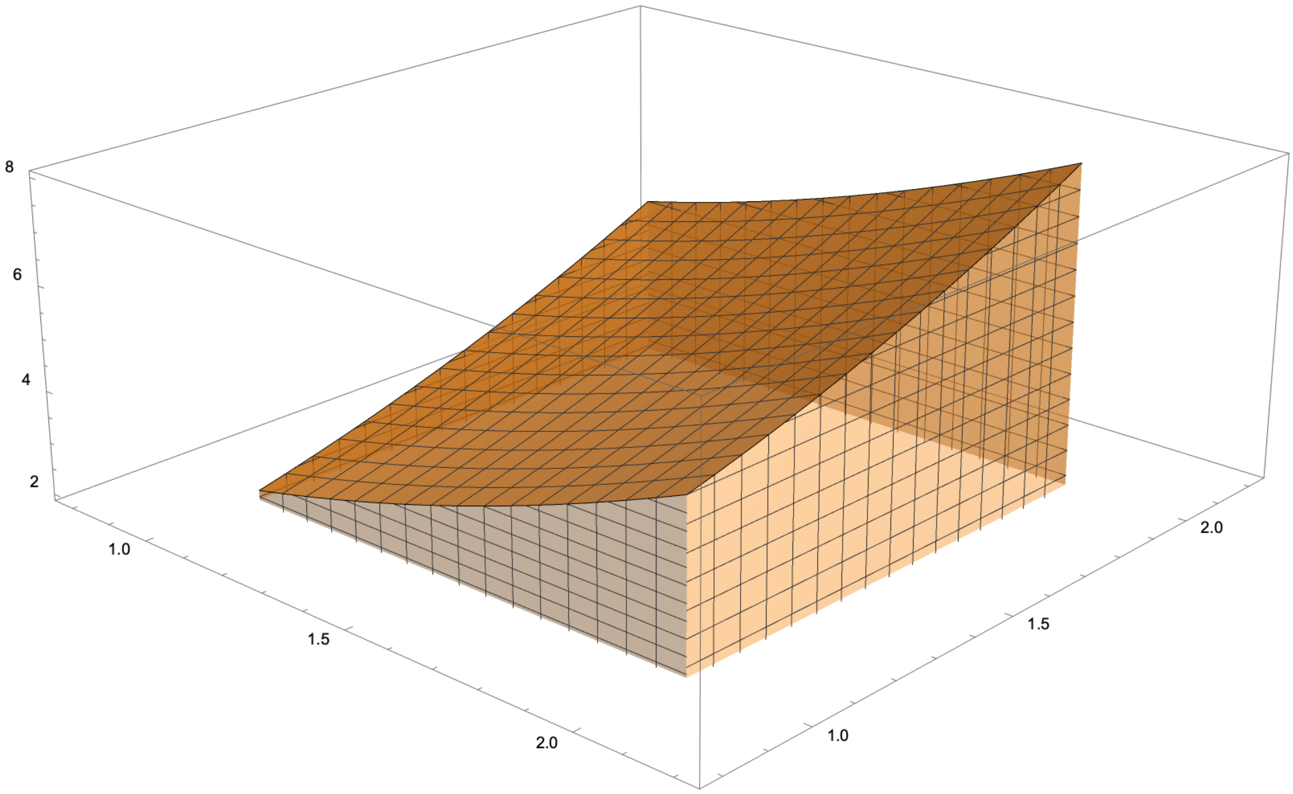


Figure 39: The volume under a surface

$$\begin{aligned}
\iint_E x^2 y \, \mathrm{d}x \mathrm{d}y &= \int_2^4 \int_1^2 x^2 y \, \mathrm{d}x \mathrm{d}y \\
&= \int_2^4 \left(\int_1^2 x^2 y \, \mathrm{d}x \right) \mathrm{d}y \\
&= \int_2^4 \left[\frac{x^3 y}{3} \right]_1^2 \mathrm{d}y \\
&= \int_2^4 \frac{7y}{3} \, \mathrm{d}y \\
&= \left[\frac{7y^2}{6} \right]_2^4 \\
&= \frac{84}{6} = 14
\end{aligned}$$

Note: when integrating over a rectangle be careful to match the limits on the single integrals with the differential $\mathrm{d}x$ or $\mathrm{d}y$.

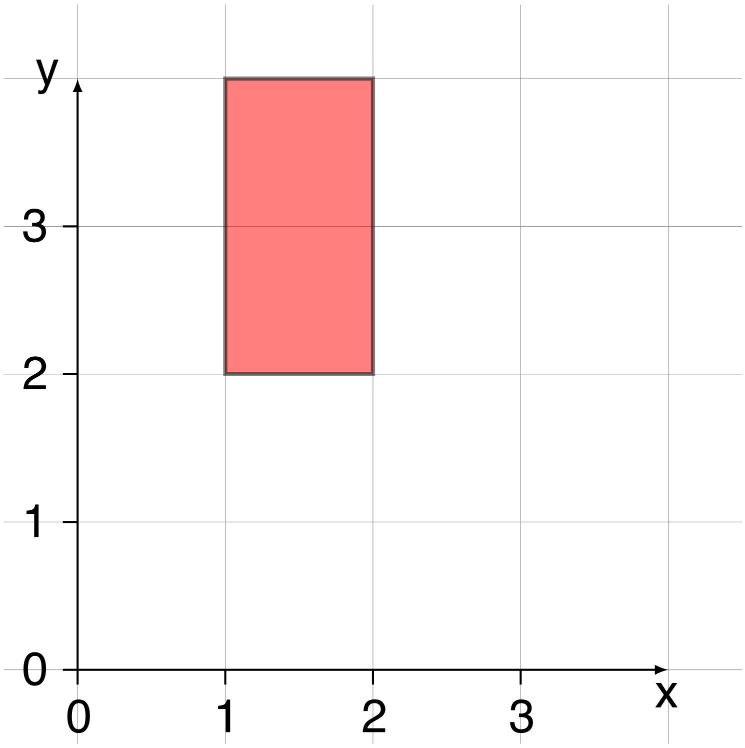


Figure 40: The rectangle E

Note: For a rectangle it does not matter which variable you integrate with respect to first, the answer is the same. Again, reminiscent of the mixed second partial derivatives. This is called **Fubini's Theorem** in its full glory and is so far beyond the scope of this course I'd have to spend years teaching all the incredible theory of integration before we even got to the point where you could revel in it.

Sets between a curve and the horizontal axis

The next level of complexity for our set E is when it is between two curves. In this case we have to be careful how we write the limits.

Example

Suppose we want to find the volume beneath the surface $z = x^2 y$ over the set E, drawn in Figure 41. The set E is the area between the x-axis and the curve $y = -x(x-4)$ between $x = 0$ and $x = 4$.

To do this we need limits for the integral with respect to x and limits for the integral with respect to y.

The set E can be described as

$$E = \{(x, y) \colon 0 \leq x \leq 4,\ 0 \leq y \leq -x(x-4)\}.$$

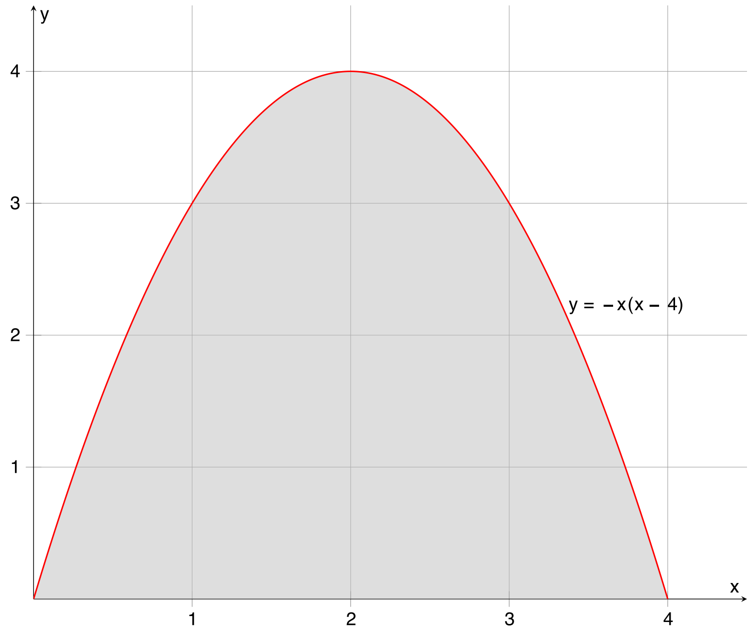


Figure 41: The area between the x-axis and the curve $y = -x(x-4)$

Notice how the y limits depend on x. This means we will need to integrate with respect to y first. Therefore,

$$\begin{aligned}
\iint_E x^2 y \,\mathrm{d}x\mathrm{d}y &= \int_0^4 \int_0^{-x(x-4)} x^2 y \,\mathrm{d}y\mathrm{d}x \\
&= \int_0^4 \left[\frac{x^2y^2}{2}\right]_0^{-x(x-4)} \mathrm{d}x \\
&= \int_0^4 \frac{1}{2}x^4(x-4)^2 \,\mathrm{d}x = \int_0^4 \frac{1}{2}x^6 - 4x^5 + 8x^4 \,\mathrm{d}x \\
&= \left[\frac{1}{14}x^7 - \frac{2}{3}x^6 + \frac{8}{5}x^5\right]_0^4 \\
&= \frac{8192}{105} \approx 78.02 \qquad \text{to 2 d.p.}
\end{aligned}$$

Integrating over areas between two curves

Suppose we want to find the volume of the region beneath the surface $z = x^2y$ above the set E shown in Figure 42.

This set can be described as

$$E = \{(x, y) \colon 0 \leq x \leq 4,\ -x(x-4)/2 \leq y \leq -x(x-4)\}.$$

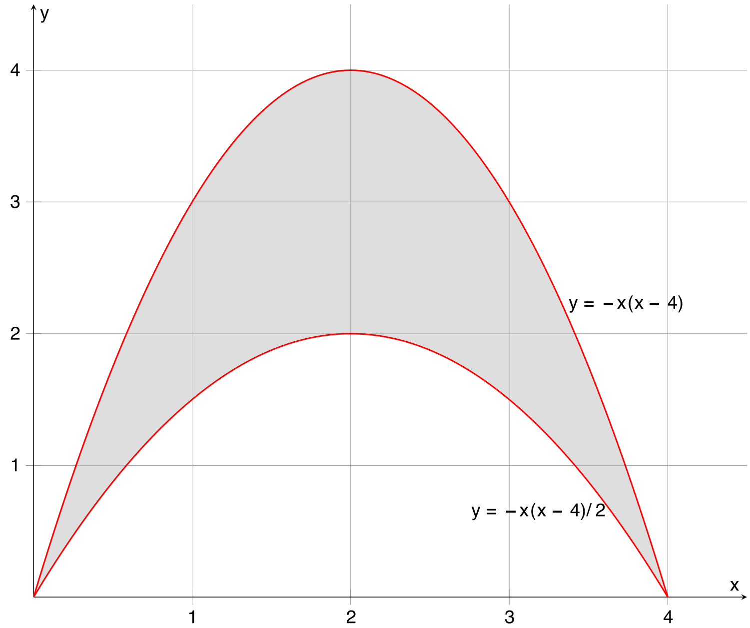


Figure 42: The area between the curves $y = -x(x-4)$ and $y = -x(x-4)/2$

As before we must integrate with respect to y first since its limits involve the variable x.

$$\begin{aligned}
\iint_E x^2 y \, \mathrm{d}x\mathrm{d}y &= \int_0^4 \int_{-x(x-4)/2}^{-x(x-4)} x^2 y \, \mathrm{d}y\mathrm{d}x \\
&= \int_0^4 \left[\frac{x^2y^2}{2}\right]_{-x(x-4)/2}^{-x(x-4)} \mathrm{d}x \\
&= \int_0^4 \frac{3}{8}x^4(x-4)^2 \, \mathrm{d}x = \int_0^4 \frac{3x^6}{8} - 3x^5 + 6x^4 \, \mathrm{d}x \\
&= \left[\frac{3}{56}x^7 - \frac{1}{2}x^6 + \frac{6}{5}x^5\right]_0^4 \\
&= \frac{2048}{35} \approx 58.51 \qquad \text{to 2 d.p.}
\end{aligned}$$

Integrals over unbounded sets

It is possible for the set E over which we are integrating to be **unbounded**. For example we might have

$$E = \{(x, y) : 0 \leq x < \infty, \ -1 \leq y \leq 1\}$$

The work we did dealing with **improper integrals** in Week 2 can be used here.

Example

With E above calculate

$$\iint_E e^{-(x+y)}\,\mathrm{d}x\mathrm{d}y$$

Here we follow the same approach as above,

$$\begin{aligned}
\iint_E e^{-(x+y)}\,\mathrm{d}x\mathrm{d}y &= \int_{-1}^{1}\int_{0}^{\infty} e^{-(x+y)}\,\mathrm{d}x\mathrm{d}y \\
&= \int_{-1}^{1} \lim_{z\to\infty} \left[-e^{-(x+y)}\right]_0^z \,\mathrm{d}y \\
&= \int_{-1}^{1} \lim_{z\to\infty} \left(e^{-y} - e^{-(z+y)}\right) \mathrm{d}y \\
&= \int_{-1}^{1} e^{-y} \lim_{z\to\infty} \left(1 - e^{-z}\right) \mathrm{d}y \\
&= \int_{-1}^{1} e^{-y} \times 1\,\mathrm{d}y = \int_{-1}^{1} e^{-y}\,\mathrm{d}y \\
&= \left[-e^{-y}\right]_{-1}^{1} \\
&= e - \frac{1}{e}.
\end{aligned}$$

Changing variables from cartesian to polar coordinates

Suppose now that E is given in Figure 43. This is a region between two quarter circles of radius 1 and 3.

We could try to decribe this in our usual way, however note that if we change to **polar coordinates** then this can be more simply described as

$$E = \{(r,\theta)\colon 1 \le r \le 3,\ 0 \le \theta \le \pi/2\}.$$

If we now want to find the integral

$$\iint_E x^2 y\,\mathrm{d}x\mathrm{d}y$$

we must change from using (x, y) coordinates to (r, θ). This is called a **change of variables** and we will learn how to do it in this section.

Recall for single variable integrals if we change from x to another variable u using a change of

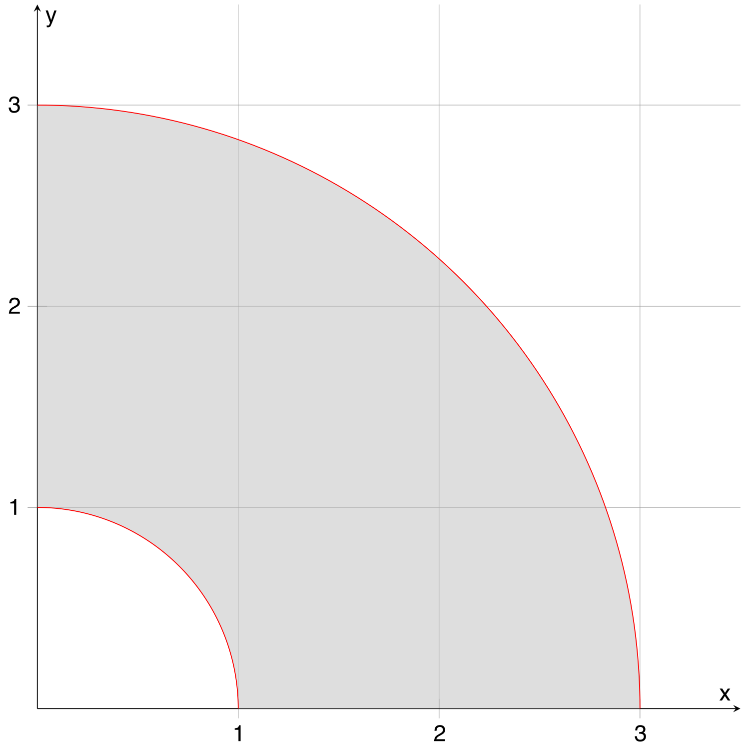


Figure 43: A quarter circle region

variable $x = g(u)$, then

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

Note: The function g must have an inverse for us to be able to use this formula.

The effect of changing from x to u in this integral introduces a *distortion* to the value of the integral. Multiplying by the derivative $\dfrac{\mathrm{d}x}{\mathrm{d}u}$ compensates for this distortion.

When we change the variables in multiple integrals we also have to do the same thing. This time we must multiply by the **Jacobian determinant**.

Suppose we change from (x, y)-coordinates to (r, θ)-coordinates (polar coordinates). This is done using the equations

$$x = r\cos\theta$$
$$y = r\sin\theta$$

The Jacobian matrix is the 2×2 matrix

$$\frac{\partial(x,y)}{\partial(r,\theta)} = \begin{pmatrix} \dfrac{\partial x}{\partial r} & \dfrac{\partial x}{\partial \theta} \\ \dfrac{\partial y}{\partial r} & \dfrac{\partial y}{\partial \theta} \end{pmatrix}$$

Here we may calculate this as

$$\frac{\partial(x,y)}{\partial(r,\theta)} = \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix}$$

In general a matrix, thought of as a transformation from $\mathbb{R}^2$ to $\mathbb{R}^2$ maps a square of area 1 onto a parallelogram whose area is the determinant of the matrix.

This leads to the formula for changing the variables from (x,y)-coordinates to (r,θ)-coordinates is

$$\iint_E f(x,y)\,\mathrm{d}x\mathrm{d}y = \iint_E f(r\cos\theta, r\sin\theta)\left|\det\frac{\partial(x,y)}{\partial(r,\theta)}\right|\,\mathrm{d}r\mathrm{d}\theta$$

Now

$$\det\frac{\partial(x,y)}{\partial(r,\theta)} = \det\begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix} = r(\cos^2\theta + \sin^2\theta) = r.$$

Hence the following formula holds,

$$\iint_E f(x,y)\,\mathrm{d}x\mathrm{d}y = \iint_{E'} f(r\cos\theta, r\sin\theta)\,r\mathrm{d}r\mathrm{d}\theta$$

where we write E' to be the set E but in polar coordinates.

Let us calculate the integral of x^2y over the region E, where

$$E = \{(r,\theta)\colon 1 \le r \le 3,\ 0 \le \theta \le \pi/2\}.$$

Therefore,

$$\begin{aligned}\iint_E x^2y\,\mathrm{d}x\mathrm{d}y &= \int_0^{\pi/2}\int_1^3 (r\cos\theta)^2(r\sin\theta)\,r\,\mathrm{d}r\mathrm{d}\theta \\ &= \int_0^{\pi/2}\int_1^3 r^4\cos^2\theta\sin\theta\,\mathrm{d}r\mathrm{d}\theta \\ &= \int_0^{\pi/2}\left[\frac{r^5}{5}\cos^2\theta\sin\theta\right]_1^3\mathrm{d}\theta \\ &= \frac{242}{5}\int_0^{\pi/2}\cos^2\theta\sin\theta\,\mathrm{d}\theta \\ &= \frac{242}{5}\left[-\frac{1}{3}\cos^3\theta\right]_0^{\pi/2} \\ &= \frac{242}{15}\end{aligned}$$

Changing variables in general

Suppose we can write the coordinates (x, y) as functions of a second coordinate system, (u, v), as follows.

$$(x, y) = (f_1(u, v), f_2(u, v)).$$

Suppose also that E can be described in the new coordinate system as the set F. Then we have the change of variable formula:

$$\iint_E f(x, y)\, \mathrm{d}x\mathrm{d}y = \iint_F f(u, v) \left|\det \frac{\partial(x, y)}{\partial(u, v)}\right| \mathrm{d}u\mathrm{d}v$$

where

$$\frac{\partial(x, y)}{\partial(u, v)} = \begin{pmatrix} \frac{\partial f_1}{\partial u} & \frac{\partial f_1}{\partial v} \\ \frac{\partial f_2}{\partial u} & \frac{\partial f_2}{\partial v} \end{pmatrix}$$

Points to note about this:

- You must write x and y as functions of (u, v)
- The function that maps (u, v) to (x, y) must be invertible

Example

Calculate the integral

$$\iint_E xy^2\, \mathrm{d}x\mathrm{d}y$$

where E is the region shown in Figure 45 bounded by the straight lines

$$y = -2x + 2, \quad y = -2x + 3, \quad y = x, \quad y = x + 2.$$

If we look at the equations of the lines bounding the region then we can write them as

$$y + 2x = 2, \quad y + 2x = 3, \quad y - x = 0, \quad y - x = 2.$$

This suggests a useful change of variable would be

$$u = y + 2x, \qquad v = y - x.$$

Then in the (u, v) coordinate system, E becomes the rectangle,

$$F = \{(u, v) \colon 2 \le u \le 3,\ 0 \le v \le 2\}.$$

Now, we need to write x and y as functions of (u, v). Write

$$u = y + 2x$$
$$v = y - x$$

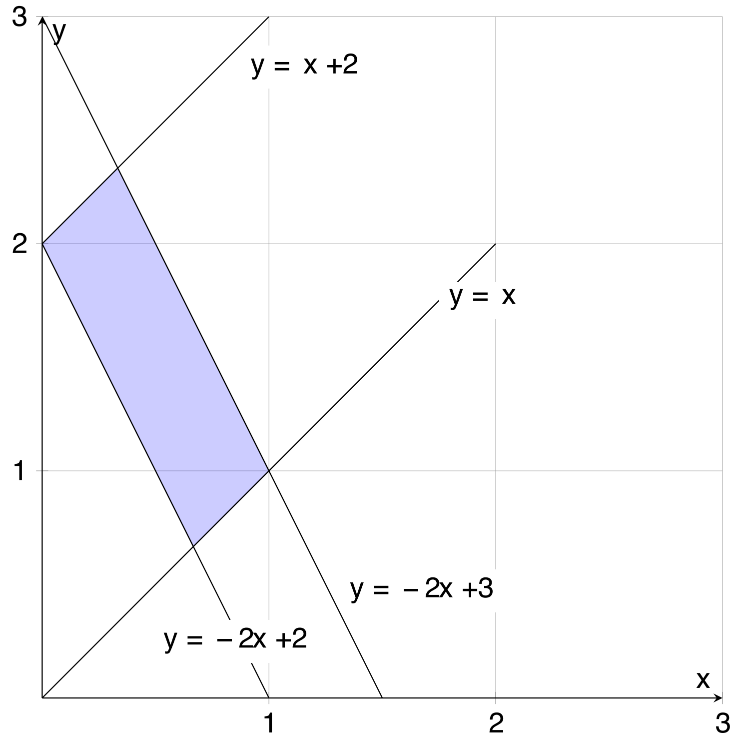


Figure 44: The set E

Subtracting the second from the first equation we get

$$u - v = 3x$$

so that $x = (u - v)/3$. By the second equation,

$$y = v + x = v + \frac{u-v}{3} = \frac{u+2v}{3}.$$

So that our change of variables is

$$x = \frac{u-v}{3}, \qquad y = \frac{u+2v}{3}.$$

Next the Jacobian,

$$\begin{aligned} \frac{\partial(x,y)}{\partial(u,v)} &= \begin{pmatrix} \frac{\partial f_1}{\partial u} & \frac{\partial f_1}{\partial v} \\ \frac{\partial f_2}{\partial u} & \frac{\partial f_2}{\partial v} \end{pmatrix} \\ &= \begin{pmatrix} 1/3 & -1/3 \\ 1/3 & 2/3 \end{pmatrix} \end{aligned}$$

And so its determinant is

$$\det \begin{pmatrix} 1/3 & -1/3 \\ 1/3 & 2/3 \end{pmatrix} = \frac{1}{3}.$$

Finally,

$$xy^2 = \frac{(u-v)(u+2v)^2}{27} = \frac{u^3+3u^2v-4v^3}{27}$$

Hence, brining this all together,

$$\begin{aligned}
\iint_E xy^2 \,\mathrm{d}x\mathrm{d}y &= \iint_F \frac{1}{27}(u^3+3u^2v-4v^3)\,\frac{1}{3}\mathrm{d}u\mathrm{d}v \\
&= \frac{1}{81}\int_0^2\int_2^3 u^3+3u^2v-4v^3 \,\mathrm{d}u\mathrm{d}v \\
&= \frac{1}{81}\int_0^2 \left[\frac{1}{4}u^4+u^3v-4v^3u\right]_2^3 \mathrm{d}v \\
&= \frac{1}{81}\int_0^2 \frac{65}{4}+19v-4v^3 \,\mathrm{d}v \\
&= \frac{1}{81}\left[\frac{65}{4}v+\frac{19}{2}v^2-v^4\right]_0^2 \\
&= \frac{1}{81}\left(\frac{65}{2}+38-16\right) \\
&= \frac{109}{162} \approx 0.673 \qquad \text{to 3 s.f.}
\end{aligned}$$

Example

Define E to be the region bounded by the curves

$$y=x^2, \quad y=2x^2, \quad y=4-x^2, \quad y=6-x^2.$$

Calculate

$$\iint_E \frac{y}{x}\,\mathrm{d}x\mathrm{d}y$$

We may write the bounding curves as

$$\frac{y}{x^2}=1, \quad \frac{y}{x^2}=2, \quad y+x^2=4, \quad y+x^2=6.$$

This suggests the change of variables,

$$u=\frac{y}{x^2}, \qquad v=y+x^2.$$

Then in the (u,v) coordinate system E becomes,

$$F=\{(u,v)\colon 1\le u\le 2,\ 4\le v\le 6\}.$$

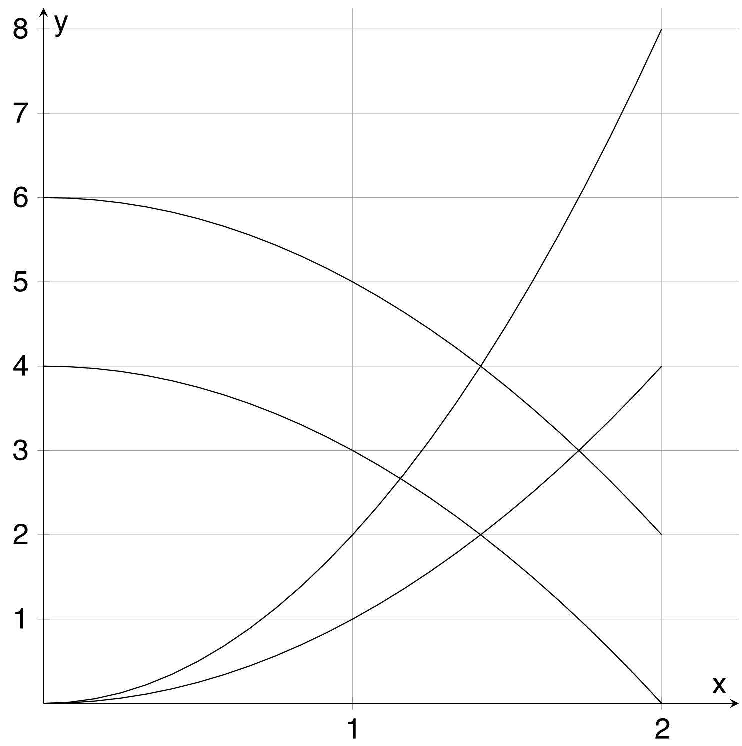


Figure 45: The set E

Now write x and y as functions of (u, v).

First we have that $x^2 = y/u$ so that

$$v = y + \frac{y}{u} = y\left(1 + \frac{1}{u}\right)$$

Hence

$$y = \frac{uv}{1+u}$$

Therefore

$$x^2 = \frac{y}{u} = \frac{v}{1+u}$$

and we have

$$x = \left(\frac{v}{1+u}\right)^{1/2}.$$

Now

$$\frac{\partial x}{\partial u} = -\frac{v^{1/2}}{2(u+1)^{3/2}}, \qquad \frac{\partial x}{\partial v} = \frac{1}{2\sqrt{v(1+u)}}$$

and

$$\frac{\partial y}{\partial u} = \frac{v}{(u+1)^2}, \qquad \frac{\partial y}{\partial v} = \frac{u}{u+1}$$

So the determinant of the Jacobian matrix is

$$-\frac{v^{1/2}}{2(u+1)^{3/2}}\frac{u}{u+1}-\frac{v}{(u+1)^2}\frac{1}{2\sqrt{v(1+u)}}=-\frac{\sqrt{v}}{2(u+1)^{3/2}}$$

Now

$$\begin{aligned}\frac{y}{x}&=\frac{uv}{1+u}\Big/\left(\frac{v}{1+u}\right)^{1/2}\\&=u\sqrt{v(1+u)}\end{aligned}$$

Putting all this together

$$\begin{aligned}\iint_E \frac{y}{x}\,\mathrm{d}x\mathrm{d}y&=\iint_F u\sqrt{v(1+u)}\left|-\frac{\sqrt{v}}{2(u+1)^{3/2}}\right|\mathrm{d}u\mathrm{d}v\\&=\int_4^6\int_1^2\frac{uv}{1+u}\,\mathrm{d}u\mathrm{d}v\\&=\int_4^6\int_1^2\frac{v(1+u)-v}{1+u}\,\mathrm{d}u\mathrm{d}v\\&=\int_4^6\int_1^2 v-\frac{v}{1+u}\,\mathrm{d}u\mathrm{d}v\\&=\int_4^6\left[uv-v\log(1+u)\right]_1^2\,\mathrm{d}v\\&=\int_4^6 v\left(1-\log\frac{3}{2}\right)\mathrm{d}v\\&=\left(1-\log\frac{3}{2}\right)\left[\frac{v^2}{2}\right]_4^6\\&=10\left(1-\log\frac{3}{2}\right)\approx 5.95\qquad\text{to 3 s.f.}\end{aligned}$$

Extra: Changing the order of integration

(This section is not assessed) Consider the example where E was the set given in Figure 41. This is the set we described as

$$E=\{(x,y)\colon 0\le x\le 4,\ 0\le y\le -x(x-4)\}$$

We previously integrated with respect to y first.

Suppose we instead want to integrate with respect to x first. Then we need to change the way we have described the set. We need to now have a fixed range for y and a changing range for x that may depend on y.

To do this first we look at Figure 41 and note that the range of y values is between 0 and 4.

To find the range for x we consider a fixed y and think of the horizontal line that intersects the set E. This intersects the curve $y = -x(x-4)$ in two places. These will be the lower and upper limits for x.

To calculate this we need to solve

$$y = -x(x-4)$$

for x. This is a quadratic equation which, when expanded and written in standard form, is

$$x^2 - 4x + y = 0$$

The two solutions are

$$x = \frac{4 \pm \sqrt{16-4y}}{2} = 2 \pm \sqrt{4-y}.$$

Therefore we have shown that we can describe E as

$$E = \{(x,y)\colon 0 \le y \le 4,\ 2 - \sqrt{4-y} \le x \le 2 + \sqrt{4-x}\}.$$

Therefore we can calculate the integral as follows.

$$\begin{aligned}
\iint_E x^2 y \, \mathrm{d}x\mathrm{d}y &= \int_0^4 \int_{2-\sqrt{4-y}}^{2+\sqrt{4-y}} x^2 y \, \mathrm{d}x\mathrm{d}y \\
&= \int_0^4 \left[\frac{x^3 y}{3}\right]_{2-\sqrt{4-y}}^{2+\sqrt{4-y}} \mathrm{d}y \\
&= \int_0^4 \frac{y}{3}\left((2+\sqrt{4-y})^3 - (2-\sqrt{4-y})^3\right) \mathrm{d}y \\
&= \int_0^4 \frac{2}{3} y (4-y)^{1/2}(16-y) \, \mathrm{d}y
\end{aligned}$$

If we use the substitution $u = 4 - y$ then $\mathrm{d}y = -\mathrm{d}u$. Also when $y = 0$, $u = 4$, and when $y = 4$, $u = 0$ (i.e. it swaps the limits around). Then, continuing,

$$\begin{aligned}
\iint_E x^2 y \, \mathrm{d}x\mathrm{d}y &= \int_0^4 \frac{2}{3}(4-u)u^{1/2}(12+u) \, \mathrm{d}u \\
&= \int_0^4 32u^{1/2} - \frac{16}{3}u^{3/2} - \frac{2}{3}u^{5/2} \, \mathrm{d}u \\
&= \left[\frac{64u^{3/2}}{3} - \frac{32u^{5/2}}{15} - \frac{4u^{7/2}}{21}\right]_0^4 \\
&= \frac{8192}{105}
\end{aligned}$$

which is the same answer we calculated before.