MULTIVARIABLE CALCULUS

Sam C. Collingbourne *
Department of Mathematics, Columbia University, New York

January 17, 2023

Overview and Motivation

Beware: These notes are incomplete and subject to change! If you're reading ahead of lectures, it is very likely that there are mistakes. Use with care and please inform me of typos/mistakes. Comments are also welcome.

A first course in calculus studies *univariate* calculus. Univariate means that one studies functions that give a rule from the real line $\mathbb{R}$ to itself, i.e. one has a function f that takes a point/variable $x \in \mathbb{R}$ to $f(x) \in \mathbb{R}$, whilst the calculus part means that one studies limits, derivatives and integrals of such functions. This course studies the generalisation of these concepts to vector-valued *multivariable* functions. In plain terms, this means that the functions one studies depend on more than one variable and can produce values which cannot be expressed as a single number but only as a ordered list of numbers. More precisely, a vector-valued multivariable function is a map/rule f which takes n inputs, $(x_1, ..., x_n)$, with $x_i \in \mathbb{R}$ for $i = 1, ..., n$ and outputs m real numbers $(y_1, ..., y_m)$. In this course, the multivariable functions studied will typically have $m, n \leq 3$. For example, one could define f by

$$(x, y) \mapsto f(x, y) = x^2 + y^2. \tag{1}$$

A natural question to ask is why one would want to study such things. Essentially this is motivated by the physical world. Multivariable calculus has applications everywhere in physics, mathematics, economics and engineering to name only a few. Below are some examples:

- **Physical positions require more than one number:** Putting string theories aside, we perceive 3 dimensions (plus time, so in reality 4). You can move forwards/backwards, left/right and up/down. Suppose you wanted to specify a place to meet someone in Manhattan (or on Earth in general), you need 2 bits of information: a street and an avenue or you need to tell them how far North and how far West (a longitude and a latitude). For example, $(40°48'2''N, 73°57'43''W)$ is Columbia University. Suppose further you wanted to meet them on the fifth floor of a building then you need a third number. You might even want to add a time giving a fourth number.
- **Many quantities change in space and time:** If you wanted to arrive at the above place by a certain time you need to know at what velocity you need to travel. Velocity is a directed rate of change of distance over time, i.e. you need a direction (which is more than one number) and a speed. Therefore, studying geometry in 2 and 3 dimensions and rates of change (i.e. derivatives) is important to understand physical problems.
- **Physical objects have 2 or 3-dimensional extent:** understanding areas/volumes of objects or how quantities leave a areas/volume (say energy of a water wave leaving a circular disk) requires integration in 2 and 3 dimensions.
- **Partial differential equations:** You may/may have come across 'partial differential equations' which allow one to study many physical systems such as electrodynamics, fluid mechanics, differential game theory and gravity (both Newtonian and Einstein's theory of general relativity). In aiming to study such equations one requires a good working knowledge of multivariable calculus.

*sc5197@columbia.edu

Contents

Notation

The following is a list of notation that is used in lectures and these notes:

- $\{...\}$: denotes a set or a collection of elements, usually numbers, i.e. $\{1, 3, 7, \sqrt{2}\}$, $\{1, 2, 3, 4, ...\}$. One can have a set with a condition, this is written with a colon as follows:
$$\{\text{elements} : \text{condition}\}. \tag{2}$$
For example
$$\{x \in \mathbb{R} : x > 0\} \tag{3}$$
is the set of positive reals.
- $\{$ can denote a set of equations you usually want solve simultaneuously, i.e.
$$\begin{cases} ax + by + cz + d = 0 \\ ex + fy + gz + h = 0. \end{cases} \tag{4}$$
- $\mathbb{R}$: the set/collection of real numbers, i.e. (inprecisely) the set or collection of all numbers which have an infinite decimal expansion. Examples would be 1, 2, -7, $\pi = 3.1415...$, $\sqrt{2}$ and (uncountably) many more.
- $\in$: belongs to, i.e. $a \in \mathbb{R}$, a belongs to the reals.
- $A \subseteq B$: denotes that A is a 'subset' of B meaning that all elements of A are contained in B. For example, if $A = \{1, 2, 3\}$ and $B = \{1, 2, 3, 4\}$ then $A \subseteq B$. Note that A and B can be equal, i.e. in the above example $A = \{1, 2, 3, 4\}$ is a subset of B.
- (a, b) is the open interval from a to b, i.e. it is the set of all real numbers between a and b, excluding a and b.
- $(a, b]$ is a half-open interval from a to b, i.e. it is the set of all real numbers between a and b, excluding a but containing b.
- $[a, b)$ is a half-open interval from a to b, i.e. it is the set of all real numbers between a and b, including a but excluding b.
- $[a, b]$ is the closed interval from a to b, i.e. it is the set of all real numbers between a and b, including a and b.
- $A \times B$ is called the Cartesian product of A and B, it is the set of pairs (a, b) such that $a \in A$ and $b \in B$. In set notation:
$$\{(a, b) : a \in A, b \in B\} \tag{5}$$
- $\mathbb{R}^2$: denotes the Cartesian product $\mathbb{R} \times \mathbb{R}$, i.e. it is the set of ordered pairs of real numbers, (a, b), with $a, b \in \mathbb{R}$. Notationally, this is written,
$$\mathbb{R}^2 = \{(a, b) : a \in \mathbb{R}, b \in \mathbb{R}\}, \tag{6}$$
where the colon : denotes 'such that' or equivalently that a condition follows.
- $\mathbb{R}^3$: denotes the Cartesian product $\mathbb{R} \times \mathbb{R} \times \mathbb{R}$, i.e. the set of ordered triples of real numbers, (a, b, c), with $a, b, c \in \mathbb{R}$. Notationally, this is written as follows:
$$\mathbb{R}^3 = \{(a, b, c) : a, b, c \in \mathbb{R}\}, \tag{7}$$

- $\mathbb{R}^n$ for $n \geq 1$: denotes the n-fold Cartesian product $\mathbb{R} \times ... \times \mathbb{R}$, i.e. the set of ordered n-tuples of real numbers, $(x_1, ..., x_n)$, with $x_i \in \mathbb{R}$ for all $i = 1, ..., n$. Notationally, this is written as follows:
$$\mathbb{R}^n = \{(x_1, ..., x_n) : x_i \in \mathbb{R} \text{ for all } i = 1, ..., n\}. \tag{8}$$

- $\mathbb{S}^1_R$: denotes a circle of radius R. This has equation $x^2 + y^2 = R^2$. Notationally, this is written as follows:
$$\mathbb{S}^1_R = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = R^2\}. \tag{9}$$

- $\mathbb{S}^2_R$: denotes a sphere of radius R. This has equation $x^2 + y^2 + z^2 = R^2$. Notationally, this is written as follows:
$$\mathbb{S}^2_R = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 = R^2\}. \tag{10}$$

- $f : A \to B$: denotes a function taking inputs from a set A (called its domain) and outputing elements of B (called its codomain).

- $\sum_i$: denotes a sum, i.e. $\sum_{i=1}^n a_i = a_1 + a_2 + ... + a_n$.

- $\pm, \mp$: denotes $+$ or $-$ compactly, i.e. ± 1 means plus or minus 1, $f_\pm(x) = \pm\sqrt{1 - x^2}$ means two functions defined simultaneously $f_+(x) = +\sqrt{1 - x^2}$, $f_-(x) = -\sqrt{1 - x^2}$. This notation can be used in more elaborate ways:
$$f_\pm(x) = a \mp b \pm cx^7 \tag{11}$$
which translated to
$$f_+(x) = a - b + cx^7 \qquad f_-(x) = a + b - cx^7. \tag{12}$$

- $\implies$: implies, i.e. $A \implies B$ for example $x^2 = a \implies x = \pm a$.

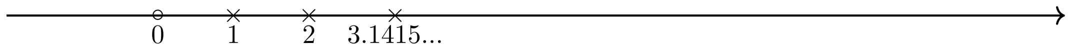
- $\doteq$: definition via an equality, i.e. define a function $f(x) \doteq ....$ For example $f_\pm(x) \doteq \pm\sqrt{1 - x^2}$ defines two functions f_+ f_- by the right-hand side with their respective signs.

- $\mapsto$: Arrow notation defines the rule of a function inline, without requiring a name to be given to the function. For example $(x, y) \mapsto x^2 - y^2$ should be read take (x, y) in the domain and map them to $x^2 - y^2$ in the codomain $\mathbb{R}$.

- γ: usually a line or a curve.

1 Coordinate Systems on the Plane and $3D$ Space

When you think of the real numbers $\mathbb{R}$ you probably, quite involintarily, think of an infinite straight line with each point representing an infinite decimal expansion:



This lecture is about how draw pairs or triples of numbers, which requires a coordinate system.

A **coordinate system** is a one-to-one assignment of a **coordinate**, which is an *ordered* list of n numbers (often called an n-tuple) to points on a surface or space. In plain terms a coordinate system is a way of labelling points on a surface or space with numbers. The notion of one-to-one just means that one assigns distinct elements to distinct elements: here one assigns a distinct coordinate to a distinct point on the surface or space. The reason for this is you want to uniquely label your points.

In this course, the number n will be 2 or 3 and the surface or space will be the 2-dimensional Euclidean plane, or 3-dimensional Euclidean space. These are often called just 'the plane' or '$3D$ space' respectively.

Remark 1.1. *The reader should be aware that there are notions of $2D$ or $3D$ spaces that are not Euclidean. For example, the surface of a ball, also called a sphere, is a two dimensional space (a given point on the surface of a sphere can be specified by two numbers) that is not Euclidean.*

1.1 Coordinate Systems on the Plane

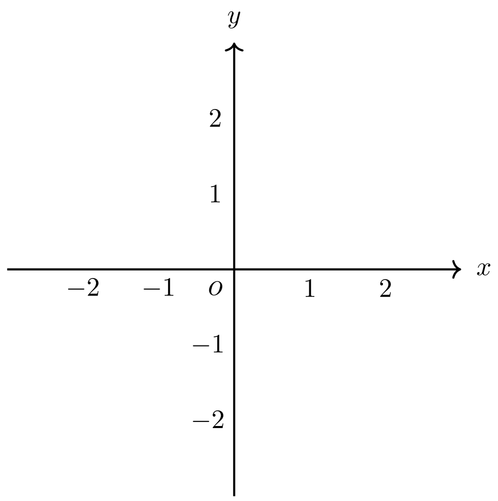
1.1.1 Cartesian Coordinates

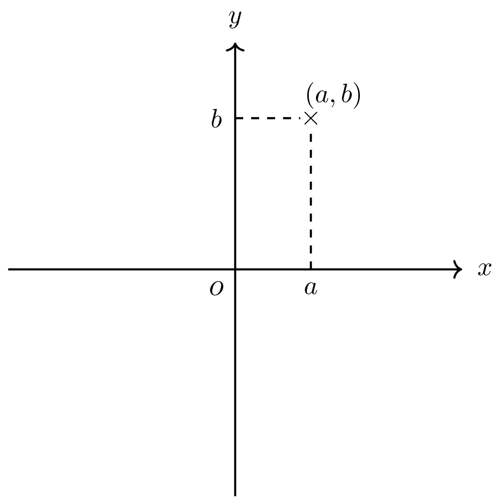
On the plane

$$\mathbb{R}^2 = \mathbb{R} \times \mathbb{R} = \{(x, y) : x \in \mathbb{R}, y \in \mathbb{R}\}, \tag{13}$$

one can set up **Cartesian coordinates** (due to French mathematician René Descartes). There is not much to do here: if $p \in \mathbb{R}^2$ then Cartesian coordinates simply its ordered pair (a, b).

Just as the real numbers $\mathbb{R}$ can be visualised as points on a straight infinite line, the plane can be visualised with points on a flat two-dimensional surface with infinite extent. To draw the Cartesian coordinate system, one picks an a point for $o = (0, 0)$ the origin and sets up two perpendicular axes or lines which intersect at $(0, 0)$: conventionally one horizontally, called the x-axis and one vertically, called the y-axis. One identifies each of these lines with the real numbers $\mathbb{R}$ such that the real numbers labelling the x-axis increase to the right and the real numbers labelling the y-axis increase upwards. This is easiest to visualise with a picture:

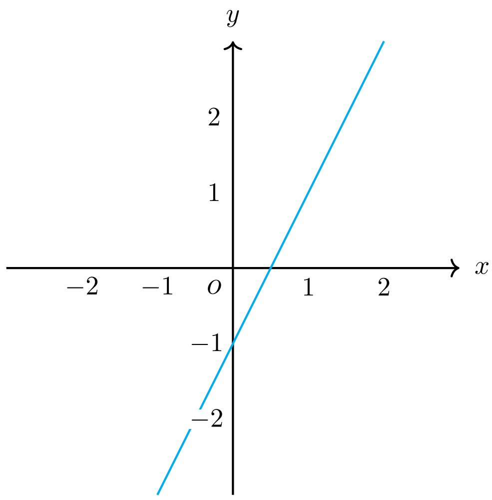




Given a point in the above diagram its Cartesian coordinates be determined by the following procedure (shown in the right hand figure of the above diagram):

1. Draw a line perpendicular to the x-axis through p and another line perpendicular to the y-axis through p.
2. a is given by the point at which line perpendicular to the x-axis meets the x-axis.
3. b is given by the point at which line perpendicular to the y-axis meets the y-axis.

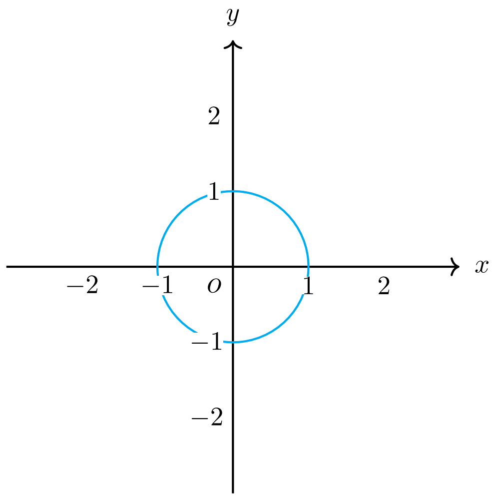
It is very easy to describe straight lines in Cartesian coordinates, for example the straight line determined by the equation $y = 2x - 1$:



Round shapes are slightly tricker. The unit circle is the subset of $\mathbb{R}^2$:

$$\mathbb{S}^1 \doteq \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}. \tag{14}$$

This is drawn below:



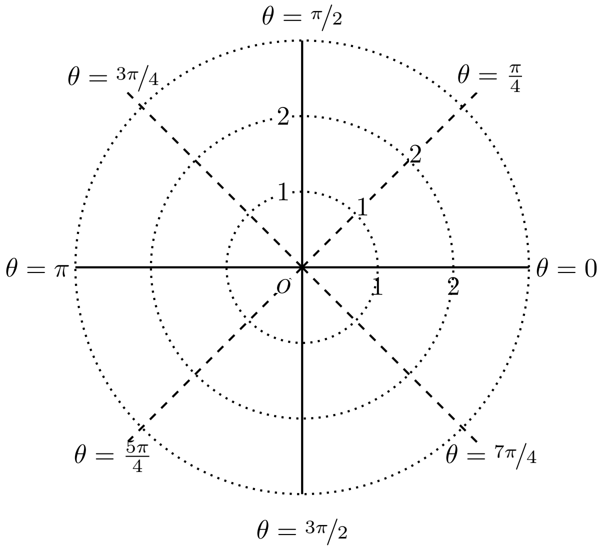
Polar coordinates on $\mathbb{R}^2$ give a cleaner description of round shapes.

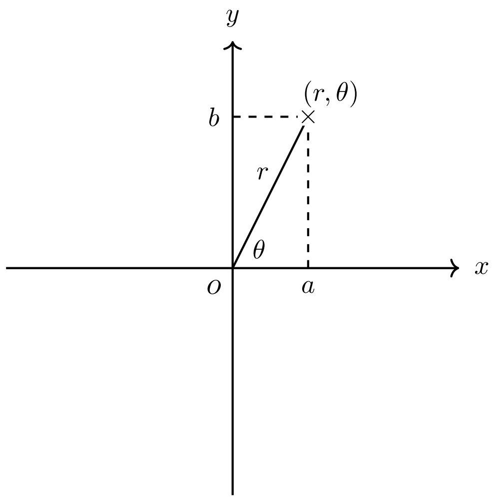
1.1.2 Polar Coordinates

Polar coordinates provide a one-to-one map from $(0, \infty) \times [0, 2\pi)$ tow $\mathbb{R}^2 \setminus \{p\}$ where p is usually chosen to be the origin o in Cartesian coordinates, i.e. $(0, 0)$.[1] Concretely, one takes the origin in $\mathbb{R}^2$ and the positive part of the x-axis, i.e. the set $\{(x, y) : x \in (0, \infty), y = 0\}$. This is often called the polar axis. Now to

[1] For those unfamiliar with set notation $\mathbb{R}^2 \setminus \{p\}$ this means $\mathbb{R}^2$ without the point p.

construct the polar coordinates of a point $q \in \mathbb{R}^2$ one take the distance r from the origin o to q and the anticlockwise angle between the polar axis and the line/ray from o to q. So the coordinate $r \in (0, \infty)$ and $\theta \in [0, 2\pi)$. Again, this is best visualised with a figure as follows:





Remark 1.2. *Whilst these comments may seem pedantic there are technical situations where they can be important. In general you will not have to worry about them but it is good to be aware.*

1. *Notice that polar coordinates here have been defined on $\mathbb{R}^2 \setminus (0, 0)$. This is because the point $(0, 0)$ can be represented in many different ways as $(0, \theta)$ for all $\theta \in [0, 2\pi)$. This is non-unique and therefore we don't have a one-to-one correspondence between a polar coordinate and the origin. One can either ignore this and 'represent o as $(0, \theta)$' or one could state that $(r, \theta) = (0, 0)$ is the origin. Both of these solutions are mostly fine (especially when integrating) but have problems when you want to take limits at the origin such as the ones that arise in taking derivatives.*

2. *Related to point 1 is values of θ outside of $[0, 2\pi)$. Note that any other length 2π interval is fine; you can pick your favourite. The reason for the restriction is to have a unique θ coordinate for each point in $\mathbb{R}^2 \setminus \{(0, 0)\}$.*

   Intuitively. one can image revolving around the origin any number of times and assigning a that θ value to represent the θ coordinate of a point. Say $r = 1$ then $\theta = \frac{\pi}{2}$ can be represented by $\theta = 5\pi/2$ or $\theta = 9\pi/2$ or $\theta = -3\pi/2$ etc. However, one runs into the same non-uniqueness issue as 1 if you allow this. So either one restricts θ to an interval of 2π length or periodically identifies θ, i.e. one declares θ is equivalent to $\theta + 2n\pi$ for $n \in \mathbb{Z}$ (the integers). One can represent this in notation as $\theta \sim \theta + 2n\pi$ for $n \in \mathbb{Z}$. In practise what this means is the following: You compute the angle to be $\tilde{\theta}$. Then one finds the n such that $\theta \doteq \tilde{\theta} + 2n\pi \in [0, 2\pi)$. This is the θ coordinate one assigns to that point.

Relations between Cartesian and Polar Coordinates
To find the Cartesian coordinates give polar coordinates one uses,

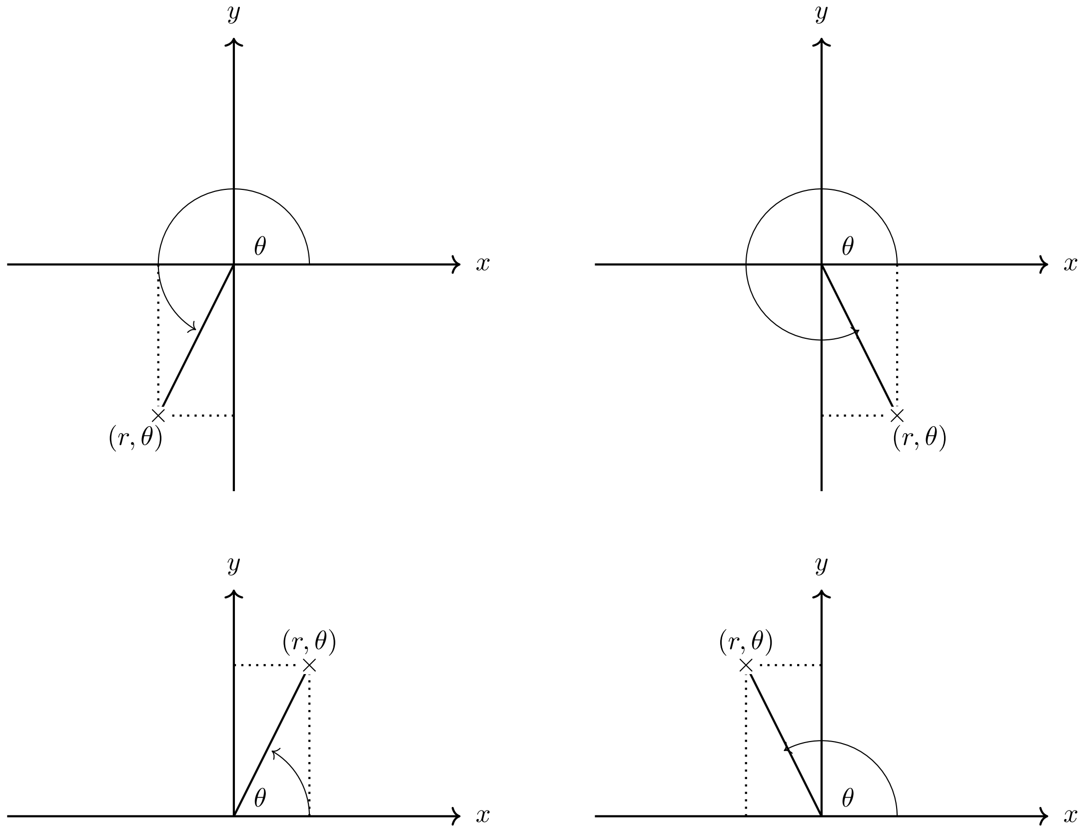
$$x = r\cos\theta, \qquad y = r\sin\theta. \tag{15}$$

These can be inverted to find the polar coordinates from Cartesian coordinates. The resulting formulas are the following,

$$r = \sqrt{x^2 + y^2}, \qquad \theta = \begin{cases} \arctan\left(\frac{y}{x}\right) & x > 0, y \geq 0 \\ \arctan\left(\frac{y}{x}\right) + 2\pi & x > 0, y < 0 \\ \arctan\left(\frac{y}{x}\right) + \pi & x < 0 \\ \frac{\pi}{2} & x = 0, y > 0 \\ \frac{3\pi}{2} & x = 0, y < 0 \\ \text{undefined} & x = 0, y = 0. \end{cases} \tag{16}$$

The reason for the cases for θ is due to $\arctan : \mathbb{R} \to (-\frac{\pi}{2}, \frac{\pi}{2})$.

One can see why these formulas in (15) are true from using SOHCAHTOA plus trig.-identities on the triangles in the following pictures



Polar Curves

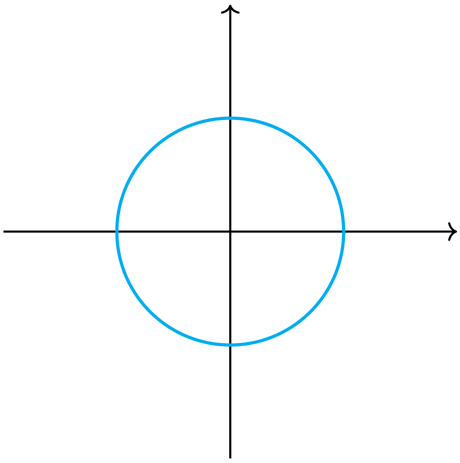
In Cartesian coordinates one can consider functions that depend on both x and y, i.e. $F = F(x, y)$. For example,

$$(x, y) \mapsto x^2 + y^2 - 1. \tag{17}$$

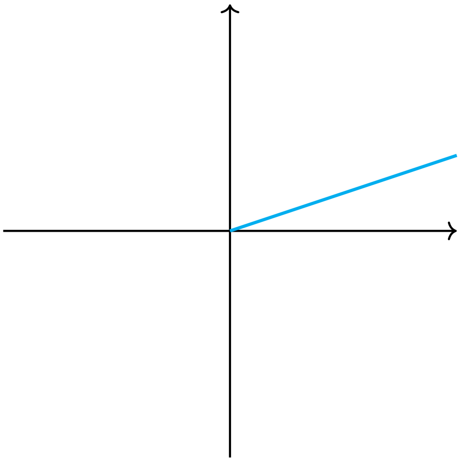
One can do the same in terms of polar coordinates. Then one writes $F(r, \theta)$. One can consider the **level set** of a such a function which is the set of points in $\mathbb{R}^2$ such that $F(r, \theta) = 0$. One can consider plots of such a set or the plot of a polar curve.

Example 1.1. *Here are some examples of plot of polar curves:*

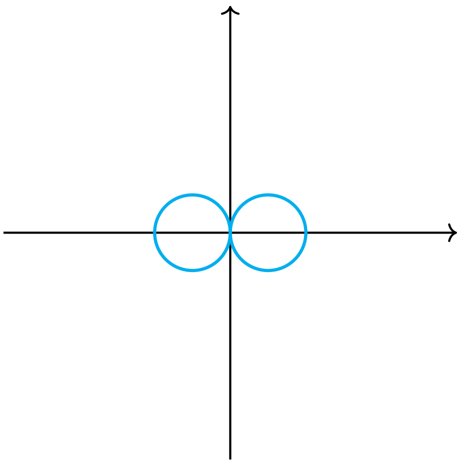
1. $r = c$ *where* $c \in (0, \infty)$*:*



2. $\theta = \theta_0$, for $\theta_0 \leq \pi/4$:



3. $r = a|\cos\theta|$,



1.2 Coordinate Systems on $3D$ Euclidean space

The Euclidean 3-space $\mathbb{R}^3$ is the set of all ordered *triples* of real numbers:

$$\mathbb{R}^3 = \mathbb{R} \times \mathbb{R} \times \mathbb{R} = \{(x, y, z) : x \in \mathbb{R}, y \in \mathbb{R}, z \in \mathbb{R}\}. \tag{18}$$

One visualises this as flat 3-dimensional space with infinite extent in all directions. It is often used as the model for the physical world, i.e. where all physical processes take place.

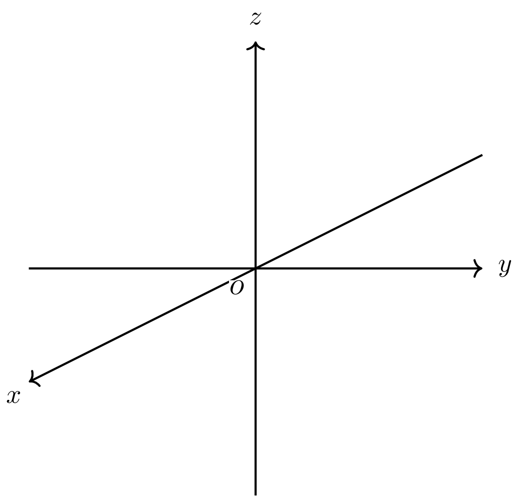
1.2.1 Cartesian Coordinates

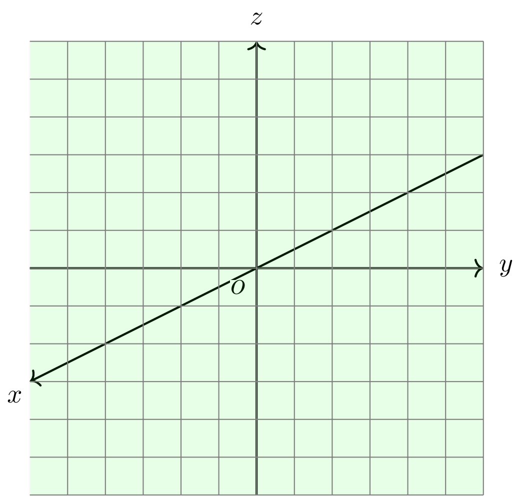
If $p \in \mathbb{R}^3$ then **Cartesian coordinates** simply its ordered triple (a, b, c).

One can draw Cartesian coordinates for $\mathbb{R}^3$ in the following way:

1. Pick a point for origin $o = (0, 0, 0)$.
2. Draw three perpendicular axes intersecting at o according to the right-hand rule: label the axis associated to your thumb 'z', the axis associated to your index finger 'x' and the axis associated to your middle finger 'y'.
3. Identify each of these lines with the real numbers $\mathbb{R}$.

This construction is shown one the left hand side of the following diagram:

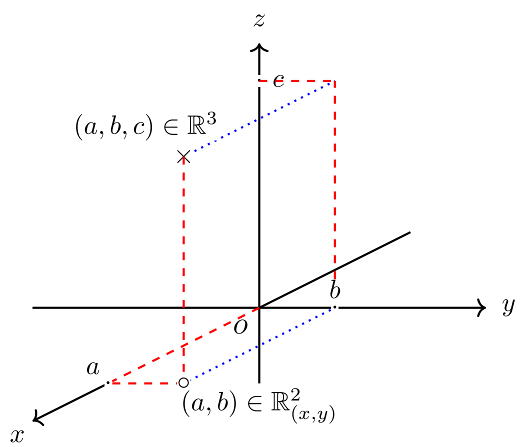




To each pair of axes one can associate a plane or $\mathbb{R}^2$. For example the yz-plane is depicted above on the right. One also has a xy-plane and a xz-plane. One then draws the point (a, b, c) in $\mathbb{R}^3$ as follows:

1. one goes a along the x axis.
2. one draws a line parallel to the y-axis in the xy-plane emanating from a and goes directed distance b along this line. This gives the point (a, b) in the xy-plane.
3. From (a, b) in the xy-plane one goes c along a line parallel to the z-axis through (a, b) in the xy-plane.

As usual this is best visualised in a diagram as follows:

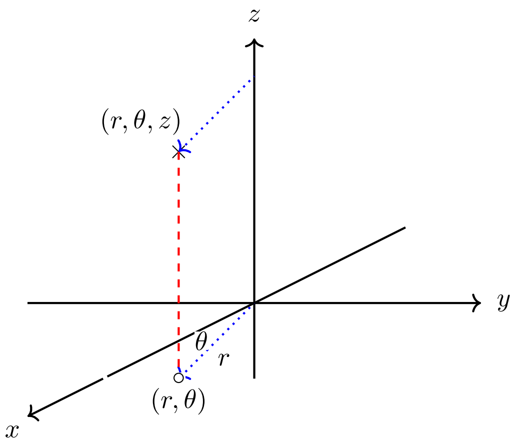


1.2.2 Cylindrical Coordinates

Another set of coordinates on $\mathbb{R}^3$ are cylindrical coordinates, which effectively extend polar coordinates on the xy-plane all of $\mathbb{R}^3$ by using the standard Cartesian z coordinate to complete the triple. More precisely, $p \in \mathbb{R}^3$ is assigned cylindrical coordinates as follows:

1. project p to the xy-plane and assign the usual polar coordinates (r, θ) to the projection of p.
2. z is then the directed distance of p from the xy-plane.
3. p is then given the cylindrical coordinates (r, θ, z).

This can be visualised as follows:



To convert between cylindrical and Cartesian coordinates one uses the relations (15) and (16) with $z = z$.

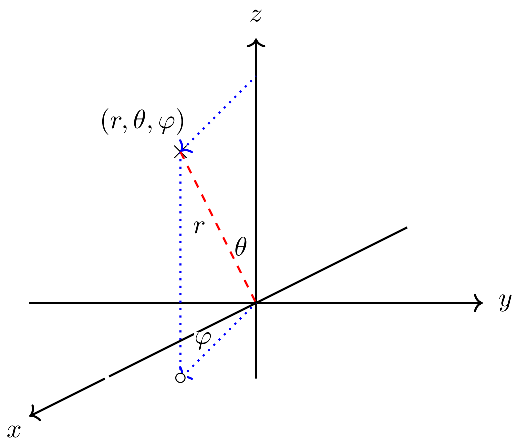
Remark 1.3. *Since polar coordinates on the plane do not cover the origin, cylindrical coordinates for $\mathbb{R}^3$ do not cover the z-axis. So one can think of cylindrical coordinates mapping $\mathbb{R}^3 \setminus \{(0, 0, z) : z \in \mathbb{R}\}$ to $(0, \infty) \times [0, \pi) \times \mathbb{R}$.*

1.2.3 Spherical Coordinates

Spherical coordinates on $\mathbb{R}^3$ are often very convenient for problems with symmetry about a point. They are a natural generalisation of polar coordinates on the plane. In particular, if one has a $p \in \mathbb{R}^3 \setminus \{(0, 0, z) : z \in \mathbb{R}\}$ (the usual remarks about the set $\{(0, 0, z) : z \in \mathbb{R}\}$ apply with spherical coordinates) one assigns spherical coordinates as follows:

1. Let r be the distance from the origin to p. Then one needs two angles (θ, φ) to give a unique representation of the point.
2. Let θ be the angle between the line segment from the origin o to p and the positive z-axis. Thus $\theta \in (0, \pi)$
3. Let φ be the angle in the xy-plane of the projection of p to the xy-plane as measured from the positive x-axis, i.e. the usual polar coordinate. Therefore, $\varphi \in [0, 2\pi)$.

The point $p \in \mathbb{R}^3$ then has polar coordinates (r, θ, φ). This can be drawn as follows:



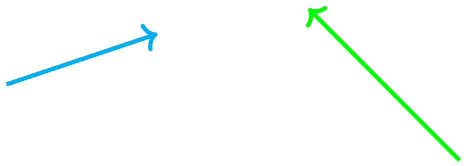
Remark 1.4. *It is very common for authors to make the swap the labelling of θ and φ. Stewart for example uses this convention.*

The relationship between Cartesian coordinates an spherical coordinates is

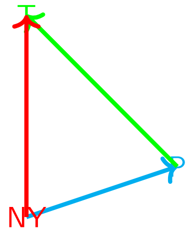
$$x = r \sin\theta \cos\varphi, \qquad y = r \sin\theta \sin\varphi, \qquad z = r\cos\theta. \tag{19}$$

2 Vectors in $\mathbb{R}^2$ and $\mathbb{R}^3$

One often associates the term vector to a quantity (often physical, i.e. velocity) that has both a magnitude and direction. This is often depicted with an arrow which has a length that depicts the magnitude of the vector and which has a orientation depicting the direction. For example:

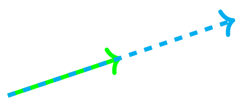


For example, this could represent 'displacement' (a distance one has to travel in a certain direction), i.e. New York to Providence, approximately $x$km ENE. Further, one could travel on to Toronto which you could represent with an arrow of length $y$km from Providence NW. You could then draw a third arrow which is your final displacement from NY to Toronto which would be the sum of the displacement from NY-Providence and the displacement from Providence-Toronto:



This requires a notion of **vector addition**.

Suppose you're driving along a straight section of motorway on your way to Providence. For simplicity, say it runs ENE. So you have some velocity (the directed rate of change of distance over time) that points ENE and has say a length of 100km/h. Suppose someone breaks in front of you and you have to slow down to 50km/h. The velocity that results has the same direction but half the magnitude, i.e. the arrow that one draws would be **scaled** by 50%:



These notions of **scalar multiplication** and **addition** are the key 'vector operations' that must satisfy certain properties. This lecture is about how to make these notions mathematically precise.

One can think about the content of this lecture in the following alternative way: at this stage you're happy to add single numbers $a + b$. How would you add ordered pairs $(a, b) + (c, d)$ of numbers? You're also probably happy to multiply numbers. What about multiplying a pair (a, b) by a number c? Very simply put this lecture is about how to define such operations and what properties these operations have.

2.1 The Algebraic Approach

In lecture 1, $\mathbb{R}^2$ and $\mathbb{R}^3$ were discussed. Recall that $\mathbb{R}^2$, which is the set of all ordered pairs (x, y), and $\mathbb{R}^3$, which is the set of all ordered triples (x, y, z), can be thought of as a plane or as 'ordinary space' respectively. To ease notation one often denotes an element of $\mathbb{R}^2$ (or $\mathbb{R}^3$) with a single boldface letter, i.e. $\mathbf{x} = (x, y)$. Other common notations are an single underlined letter, i.e. $\underline{x} = (x, y)$.[2] One can define notions of addition and 'scalar' multiplication on $\mathbb{R}^2$ (or $\mathbb{R}^3$) as follows. These give $\mathbb{R}^2$ and $\mathbb{R}^3$ a 'vector space' structure.

[2]Often authors do not distinguish at all and simply write $x \in \mathbb{R}^2$.

Definition 2.1. *The* **addition** *of ordered pairs* $\mathbf{x}_1 = (x_1, y_1)$, $\mathbf{x}_2 = (x_2, y_2)$ *in* $\mathbb{R}^2$ *is defined by adding corresponding (Cartesian) coordinates, i.e.*

$$(x_1, y_1) + (x_2, y_2) = (x_1 + x_2, y_1 + y_2). \tag{20}$$

The resultant ordered pair $(x_1 + x_2, y_1 + y_2)$ *is denoted* $\mathbf{x}_1 + \mathbf{x}_2$.

The **scalar multiplication** *of a ordered pair* $\mathbf{x} = (x, y) \in \mathbb{R}^2$ *by a number (here called a scalar)* $\lambda \in \mathbb{R}$ *is defined by multiplying each (Cartesian) coordinate by* λ*, i.e.*

$$\lambda(x, y) = (\lambda x, \lambda y). \tag{21}$$

The resultant ordered pair $(\lambda x, \lambda y)$ *is denoted* $\lambda\mathbf{x}$.

Analogous statements hold for ordered triples in $\mathbb{R}^3$ *(in fact this extends to all* $\mathbb{R}^n$*).*

One can prove that the sets $\mathbb{R}^2$ and $\mathbb{R}^3$ with addition and scalar multiplication defined above satisfy the following **axioms of a vector space**[3]:

(i) **Commutativity** of addition: $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$ for all $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$.

(ii) **Associativity** of addition and scalar multiplication: $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$ and $(ab)\mathbf{v} = a(b\mathbf{v})$ for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ and $a, b \in \mathbb{R}$.

(iii) The existence of an **additive identity**: there exists $\mathbf{0} \in \mathbb{R}^n$ such that $\mathbf{v} + \mathbf{0} = \mathbf{v}$ for all $\mathbf{v} \in \mathbb{R}^n$.

(iv) The existence of an **additive inverse**: for all $\mathbf{x} \in \mathbb{R}^n$ there exists $-\mathbf{x} \in \mathbb{R}^n$ such that $\mathbf{x} + (-\mathbf{x}) = \mathbf{0}$. For example for $\mathbb{R}^2$, $-\mathbf{x} = (-x, -y)$.

(v) The existence of a **multiplicative identity**: there exists $1 \in \mathbb{R}$ such that $1\mathbf{v} = \mathbf{v}$ for all $\mathbf{v} \in \mathbb{R}^n$.

(vi) **Distributive** properties: $a(\mathbf{v} + \mathbf{u}) = a\mathbf{v} + a\mathbf{u}$ and $(a + b)\mathbf{v} = a\mathbf{v} + b\mathbf{v}$.

So when we think of $\mathbb{R}^n$ in this manner (i.e. with the notions of addition and scalar multiplication defined as above), $\mathbb{R}^n$ is a vector space and any element of $\mathbf{x} \in \mathbb{R}^n$ is a **vector**. The ordered list $(x_1, x_2, ..., x_n)$ is known as the components of $\mathbf{x}$, i.e. x_1 is the first component of the vector $\mathbf{x}$. Additionally, one can define the **difference** of two vectors $\mathbf{u}, \mathbf{v}$ by

$$\mathbf{u} - \mathbf{v} = \mathbf{u} + (-\mathbf{v}). \tag{22}$$

Remark 2.1. *Stewart uses the notation* $\langle x_1, x_2, ..., x_n\rangle$ *to write a vector in its components. This notation could lead to confusion in other classes (especially if you take a class where no boldface is used to distinguish vectors from numbers, i.e.* x *represents vectors rather than* $\mathbf{x}$*): if one is dealing with* $\mathbb{R}^2$ *then* $\langle x_1, x_2\rangle$ *could mean the 'inner product' of two* <u>*vectors*</u>*. Moreover,* $\langle x_1, x_2, ..., x_n\rangle$ *is common notation for something known as the 'span' of the vectors* $x_1, x_2, ..., x_n$*. In this class you can use the notation* $\langle\rangle$ *notation if you like it or* $()$ *or even* $[]$ *but just be aware that the first is quite uncommon and used for other things (do not use* $\{\}$ *as this is typically set notation).*

One can define lots of additional structure on $\mathbb{R}^n$ (as will be covered in lectures 3 and 4). One useful notation is the length/magnitude/norm of a vector $\mathbf{v}$ in $\mathbb{R}^2$ or $\mathbb{R}^3$:

$$||\mathbf{v}|| \doteq \begin{cases} \sqrt{v_1^2 + v_2^2} & \text{for } \mathbb{R}^2, \\ \sqrt{v_1^2 + v_2^2 + v_3^2} & \text{for } \mathbb{R}^3, \\ \sqrt{\sum_{i=1}^n v_i^2} & \text{for } \mathbb{R}^n. \end{cases} \tag{23}$$

[3] To define a vector space in a more abstract sense one needs a notion of vector addition and scalar multiplication such that these axioms are satisfied (see any course on Linear Algebra for more).

The vector space $\mathbb{R}^n$ with the norm defined in equation (23) gives $\mathbb{R}^n$ a **normed space structure**, which can be defined more abstractly. We call a vector $\mathbf{v}$ with norm 1 a unit vector. Any non-zero vector can be **normalised** to give a vector $\hat{\mathbf{v}} = (v_1, v_2, v_3)$ of length 1 by dividing by its norm:

$$\hat{\mathbf{v}} = \frac{\mathbf{v}}{||\mathbf{v}||}. \tag{24}$$

There are two vectors that are often distinguished in $\mathbb{R}^2$:

$$\mathbf{e_1} \doteq (1,0) \qquad \mathbf{e_2} \doteq (0,1). \tag{25}$$

These are the standard basis (unit) vectors for $\mathbb{R}^2$. Other common notation is $\mathbf{i}$ and $\mathbf{j}$. In $\mathbb{R}^3$ this generalises to

$$\mathbf{e_1} \doteq (1,0,0), \qquad \mathbf{e_2} \doteq (0,1,0), \qquad \mathbf{e_3} \doteq (0,0,1), \tag{26}$$

or $\mathbf{i}$, $\mathbf{j}$ and $\mathbf{k}$. For $\mathbb{R}^n$ with $n \geq 3$ this generalises to

$$\mathbf{e_1} \doteq (1,0,...,0), \qquad \mathbf{e_2} \doteq (0,1,...,0), \qquad \ldots \qquad \mathbf{e_n} \doteq (0,0,...,1). \tag{27}$$

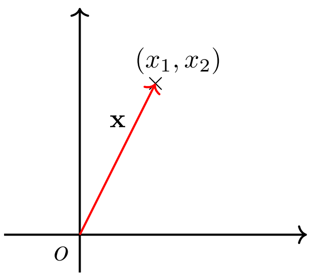
With the rules of vector addition laid out above one can express the vector $\mathbf{v} = (v_1, v_2, v_3)$ as

$$\mathbf{v} = v_1\mathbf{e_1} + v_2\mathbf{e_2} + v_3\mathbf{e_3} \tag{28}$$

in the basis $\mathbf{e_1}, \mathbf{e_2}, \mathbf{e_3}$.

2.2 The Geometric Approach/Drawing Vectors in $\mathbb{R}^2$ and $\mathbb{R}^3$

As before one can sketch $\mathbb{R}^2$ on a piece of paper or blackboard. When we discussed Cartesian coordinates on $\mathbb{R}^2$ we represented an ordered pair $\mathbf{x} = (x_1, x_2)$ with a point on the plane. Instead one can think of $\mathbf{x}$ as an arrow from the origin of $\mathbb{R}^2$ to the point (x_1, x_2) as drawn below:

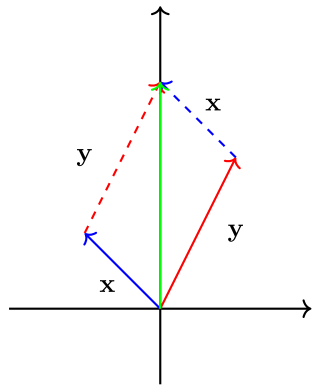


When one thinks of $\mathbf{x}$ as an arrow one calls it a **vector**.

Vector addition in this pictorial sense works as follows. Suppose you have two vectors $\mathbf{x} = (x_1, x_2)$ and $\mathbf{y} = (y_1, y_2)$ in $\mathbb{R}^2$. Vector addition gives you the vector $\mathbf{x}+\mathbf{y} = (x_1+y_1, x_2+y_2)$. So you draw an arrow from $\mathbf{0}$ to $\mathbf{x}+\mathbf{y}$. In practise what you can do is:

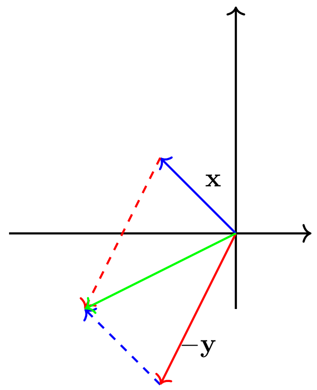
1. Draw the vector $\mathbf{x}$ (or $\mathbf{y}$) and move $\mathbf{y}$ parallel to itself to place the tail of $\mathbf{y}$ (or $\mathbf{x}$) at the tip of $\mathbf{x}$ (or $\mathbf{y}$).
2. $\mathbf{x}+\mathbf{y}$ is then the vector that goes from the tail of $\mathbf{x}$ (or $\mathbf{y}$) to the tip of $\mathbf{y}$ (or $\mathbf{x}$).

Here $\mathbf{x}+\mathbf{y}$ (or $\mathbf{y}+\mathbf{x}$) is drawn in green:



Note that this figure illustrates the commutativity of vector addition.

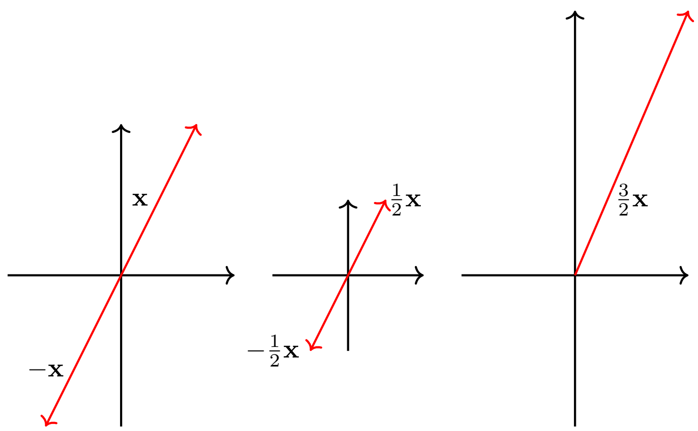
Using the definition of the difference of two vectors $\mathbf{x}$ and $\mathbf{y}$ one can draw $\mathbf{x}-\mathbf{y}$ using the addition prescription above:



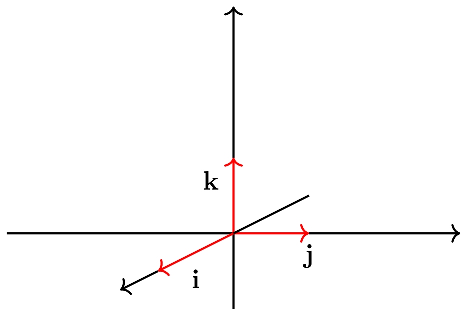
Scalar multiplication of a vector $\mathbf{v} \in \mathbb{R}^2$ by a scalar $a \in \mathbb{R}$ pictorially is carried out as follows:

1. Take the length of $\mathbf{v}$ and multiply it by $|a|$ call this b.
2. If $a > 0$ then $a\mathbf{v}$ is the vector of length b in the direction of $\mathbf{v}$.
3. If $a < 0$ then $a\mathbf{v}$ is the vector of length b in the opposite direction to $\mathbf{v}$.

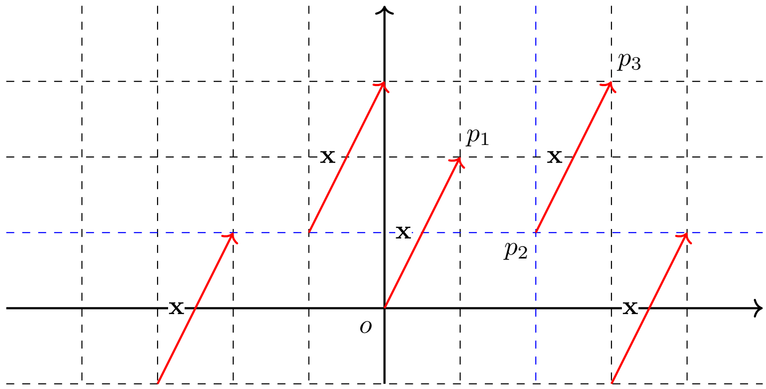
For example, pictorially one has:



All of the above is drawn in $\mathbb{R}^2$ but extends readily to $\mathbb{R}^3$. So that there is an example of drawing in $\mathbb{R}^3$, lets draw the unit basis vectors $\mathbf{i}$, $\mathbf{j}$ and $\mathbf{k}$:



As defined in section 2.1, all vectors emanate from $(0,0)$. However, you may be wondering what about drawing arrows as depicted in the following diagram:



In this diagram the arrow representing the vector $\mathbf{x}$ has been translated without changing its length or direction. In this course we will say that these arrows (which we will again call vectors) are equivalent to (or representations of) the original vector (the one emanating from o).

Remark 2.2. *(Do not worry about this comment too much.) Strictly speaking, the arrows that are drawn from any point other than o belong to different copies of $\mathbb{R}^n$ (when viewed as a vector space). As defined in section 2.1, all vectors emanate from $(0,0)$. However, as represented in blue above one could realign the origin with $p_2 = (0,0)$ and draw $\mathbf{x}$. In technical terms, which you don't need to worry about, you are using something known as the 'affine structure' of $\mathbb{R}^n$. You may see this in further courses on linear algebra.*

One can often come across various terms associated to such vectors. For example, if a particle moves from p_2 to p_3 then it's **displacement vector** is $\mathbf{x}$. Sometimes one will denote this with $\overrightarrow{p_2p_3}$ instead of $\mathbf{x}$. The displacement vector $\overrightarrow{op_1}$ is often distinguished further and called the **position vector** of p_1. If one is given the coordinates of the points $p_3 = (3,3)$ and $p_2 = (2,1)$ then the displacement vector is computed by taking the difference of the position vectors of p_2 and p_3:

$$\mathbf{x} = \overrightarrow{p_2p_3} = (3,3) - (2,1) = (1,2). \tag{29}$$

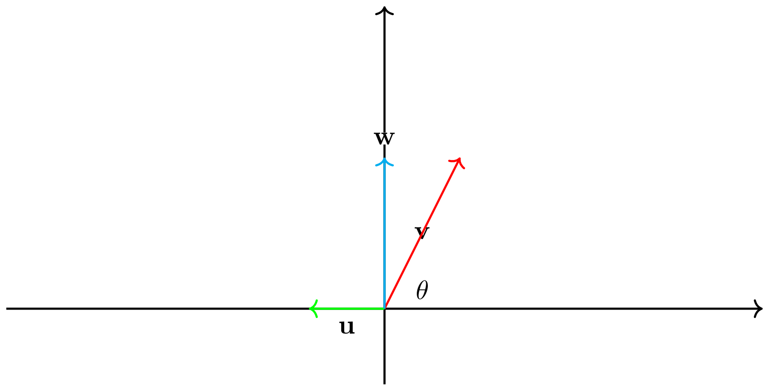
This tells us that to reach p_3 from p_2 one must travel across 1 and up 2. Note that our origin is at p_2 in this case.

2.3 Application: Crossing a River

Considering the following example of a computing velocity vector:

Example 2.1. *Suppose a woman launches a boat from the south shore of a straight river that flows directly west at 4km/h. She wants to land at the point directly across on the opposite shore. If the speed of the boat (relative to the water) is 8km/h, in what direction should she steer the boat in order to arrive at the desired landing point?*

Lets align the x-axis with the south shore of the river and the y-axis pointing across the river meeting the x-axis at the launching point. The boats velocity $\mathbf{v} = 8(\cos\theta, \sin\theta)$. The water velocity is $\mathbf{u} = (-4,0) = -4\mathbf{i}$. We want the resultant velocity $\mathbf{w}$ to be $\mathbf{w} = \omega\mathbf{j}$ for $\omega > 0$.



So,

$$\mathbf{w} = 4(2\cos\theta - 1, 2\sin\theta) = (0, \omega), \tag{30}$$

which has solution $\theta = \frac{2\pi}{3}$.

3 The Scalar Product

On the vector space $\mathbb{R}^n$, one can define even more structure. In particular, one can define a notion of multiplication of two vectors which called the **dot product** or **scalar product**. This is an example of an **inner product** on a vector space. Vector spaces with an inner product (known as inner product spaces) arise everywhere in physics. For example, the mathematical foundations of quantum mechanics are based on the theory of a particular type of inner product spaces called Hilbert spaces.

3.1 Definition and Properties

Definition 3.1 (Scalar Product)**.** *The* **scalar product** *or* **dot product** *of two vectors* $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ *is defined by*

$$\mathbf{u} \cdot \mathbf{v} \doteq \sum_{i=1}^{n} u_i v_i. \tag{31}$$

Remark 3.1. *Common notation for an scalar product is* $\langle \mathbf{u}, \mathbf{v} \rangle$*. You can use either in this course.*

Example 3.1. *Let's do an example: Suppose one has* $\mathbf{u} = (2, 5, -1)$ *and* $\mathbf{v} = (-3, 1, 0)$ *then one can compute using the formula* (31) *that*

$$\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u} \cdot \mathbf{v} = 2 \times (-3) + 5 \times 1 + (-1) \times 0 \tag{32}$$

$$= -6 + 5 + 0 = -1. \tag{33}$$

One can see that the scalar product satisfies the following axioms of an inner product:

1. **Symmetry**: $\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle$ for all $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$.
2. **Linearity** in the first argument: $\langle a\mathbf{v} + b\mathbf{w}, \mathbf{u} \rangle = a\langle \mathbf{v}, \mathbf{u} \rangle + b\langle \mathbf{w}, \mathbf{u} \rangle$ for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ and $a, b \in \mathbb{R}$.
3. **Positive semi-definiteness**: $\langle \mathbf{u}, \mathbf{u} \rangle \geq 0$ for all $\mathbf{u} \in \mathbb{R}^n$ with equality if and only if $\mathbf{u} = 0$.

The scalar product (and indeed any inner product) satisfies the following additional properties:

1. **Linearity** in the second argument: $\langle \mathbf{u}, a\mathbf{v} + b\mathbf{w} \rangle = a\langle \mathbf{u}, \mathbf{v} \rangle + b\langle \mathbf{u}, \mathbf{w} \rangle$ for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ and $a, b \in \mathbb{R}$.
2. $\langle \mathbf{0}, \mathbf{u} \rangle = \langle \mathbf{u}, \mathbf{0} \rangle = 0$.

Let's check the positive semi-definiteness property (you should check the rest of the above properties):

Proof. One can compute with formula (31) that

$$\langle \mathbf{u}, \mathbf{u} \rangle = \sum_{i=1}^{n} u_i u_i = \sum_{i=1}^{n} u_i^2. \tag{34}$$

So since $u_i^2 \geq 0$, $\langle \mathbf{u}, \mathbf{u} \rangle \geq 0$. If $u_i = 0$ for all $i = 1, 2, ..., n$, then $\langle \mathbf{u}, \mathbf{u} \rangle = 0$ from the right-hand side of (34). On the other hand, if $\langle \mathbf{u}, \mathbf{u} \rangle = 0$ then, by the right-hand side of equation (34), $\sum_i u_i^2 = 0$ which implies $u_i^2 = 0$ for all $i = 1, 2, ..., n$. This implies $u_i = 0$ for all $i = 1, ..., n$. If a vectors components vanish then it is the zero vector. □

Note that the right-hand side of equation (34) is $||\mathbf{u}||^2$ where $|| \cdot ||$ is the norm defined in equation (23). So,

$$||\mathbf{u}|| = \sqrt{\langle \mathbf{u}, \mathbf{u} \rangle}, \tag{35}$$

which is well-defined because $\langle \mathbf{u}, \mathbf{u} \rangle \geq 0$.

Remark 3.2. *This is a general property of an inner product: an inner product induces a norm on a vector space.*

This allows one to prove the following proposition:

Proposition 3.1 (Polarisation Identity). *For* $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$,

$$\langle \mathbf{u}, \mathbf{v} \rangle = \frac{1}{2}\Big(||\mathbf{u}+\mathbf{v}||^2 - ||\mathbf{u}||^2 - |||\mathbf{v}||^2\Big). \tag{36}$$

Proof. Expand $||\mathbf{u}+\mathbf{v}||^2$ using equation (35) as

$$||\mathbf{u}+\mathbf{v}||^2 = \langle \mathbf{u}+\mathbf{v}, \mathbf{u}+\mathbf{v} \rangle. \tag{37}$$

Using linearity in both arguments

$$||\mathbf{u}+\mathbf{v}||^2 = \langle \mathbf{u}, \mathbf{u} \rangle + \langle \mathbf{v}, \mathbf{v} \rangle + \langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{v}, \mathbf{u} \rangle. \tag{38}$$

Using equation (35) and symmetry of the scalar product one has

$$||\mathbf{u}+\mathbf{v}||^2 = ||\mathbf{u}||^2 + ||\mathbf{v}||^2 + 2\langle \mathbf{u}, \mathbf{v} \rangle. \tag{39}$$

Rearranging gives the result. □

Another useful notion associated to the scalar product is orthogonality of vectors:

Definition 3.2 (Orthogonal Vectors). *Two vectors* $\mathbf{u}, \mathbf{v}$ *in* $\mathbb{R}^n$ *are said to be orthogonal if*

$$\langle \mathbf{u}, \mathbf{v} \rangle = 0. \tag{40}$$

3.2 Angles

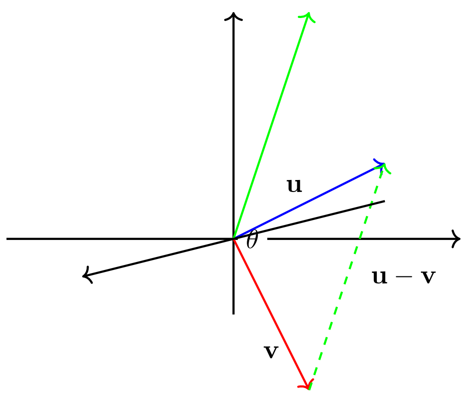
The angle between vectors can be related to their scalar product as follows:

Proposition 3.2. *Let* $\mathbf{u}, \mathbf{v}$ *be vectors in* $\mathbb{R}^n$ *then*

$$\langle \mathbf{u}, \mathbf{v} \rangle = ||\mathbf{u}||||\mathbf{v}|| \cos\theta, \tag{41}$$

where $\theta \in [0, \pi]$ *is the angle between* $\mathbf{u}, \mathbf{v}$.

Proof. Lets prove this in $\mathbb{R}^3$. As usual its helpful to draw a picture:



One now has a triangle with sides of length $||\mathbf{u}||$, $||\mathbf{v}||$ and $||\mathbf{u}-\mathbf{v}||$. The law of cosines gives

$$||\mathbf{u}-\mathbf{v}||^2 = ||\mathbf{u}||^2 + ||\mathbf{v}||^2 - 2||\mathbf{u}||||\mathbf{v}|| \cos\theta. \tag{42}$$

Replacing $\mathbf{v}$ with $-\mathbf{v}$ in the polarisation identity gives

$$\langle \mathbf{u}, \mathbf{v} \rangle = \frac{1}{2}\Big(||\mathbf{u}||^2 + |||\mathbf{v}||^2 - ||\mathbf{u}-\mathbf{v}||^2\Big). \tag{43}$$

Substituting $||\mathbf{u}-\mathbf{v}||^2$ from equation (42) gives the result. □

Remark 3.3. *The same proof works in $\mathbb{R}^2$. In fact this is the essential argument since any two vectors can be thought to lie in a plane in $\mathbb{R}^n$.*

Corollary 3.1. *If $\mathbf{u}, \mathbf{v} \neq 0$ then the angle between $\mathbf{u}, \mathbf{v} \neq 0$ is given by*

$$\cos\theta = \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{||\mathbf{u}|| ||\mathbf{v}||} = \langle \hat{\mathbf{u}}, \hat{\mathbf{v}} \rangle. \tag{44}$$

If $\langle \hat{\mathbf{u}}, \hat{\mathbf{v}} \rangle > 0$ then $\theta \in [0, \frac{\pi}{2})$ with $\theta = 0$ when $\mathbf{v} = c\mathbf{u}$ with $c > 0$. If $\langle \hat{\mathbf{u}}, \hat{\mathbf{v}} \rangle < 0$ then $\theta \in (\frac{\pi}{2}, \pi]$ with $\theta = \pi$ when $\mathbf{v} = c\mathbf{u}$ with $c < 0$. If $\langle \hat{\mathbf{u}}, \hat{\mathbf{v}} \rangle = 0$ then $\theta = \frac{\pi}{2}$ which means our terminology of orthogonality makes sense!

The following is are two **very** useful inequalities:

Proposition 3.3 (Cauchy-Schwarz Inequality)**.** *Let $\mathbf{u}, \mathbf{v}$ be vectors in $\mathbb{R}^n$ then*

$$|\langle \mathbf{u}, \mathbf{v} \rangle| \leq ||\mathbf{u}|| ||\mathbf{v}||. \tag{45}$$

Proof. By proposition 3.2 one has

$$|\langle \mathbf{u}, \mathbf{v} \rangle| = ||\mathbf{u}|| ||\mathbf{v}|| |\cos\theta|. \tag{46}$$

Now $|\cos\theta| \leq 1$, so

$$|\langle \mathbf{u}, \mathbf{v} \rangle| \leq ||\mathbf{u}|| ||\mathbf{v}||. \tag{47}$$

□

Proposition 3.4 (Triangle Inequality)**.** *Let $\mathbf{u}, \mathbf{v}$ be vectors in $\mathbb{R}^n$ then*

$$||\mathbf{u} + \mathbf{v}|| \leq ||\mathbf{u}|| + ||\mathbf{v}||. \tag{48}$$

Proof. See the problem sheet. □

3.3 Projections

In some cases you may wish to know how much some vector points along another vector. This leads to the definition of projection.

Definition 3.3 (Projection)**.** *Let $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$. Then the* **projection** *(or* **vector projection***) of $\mathbf{u}$ onto $\mathbf{v}$ is given by*

$$\mathrm{proj}_{\mathbf{v}}(\mathbf{u}) = \Pi_{\mathbf{v}}(\mathbf{u}) \doteq \langle \hat{\mathbf{v}}, \mathbf{u} \rangle \hat{\mathbf{v}}. \tag{49}$$

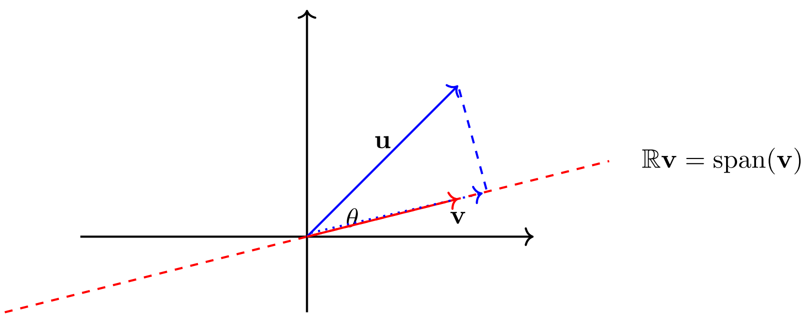
The **component** *(or* **scalar projection***) of $\mathbf{u}$ along $\mathbf{v}$ is given by*

$$\mathrm{comp}_{\mathbf{v}}(\mathbf{u}) \doteq \langle \hat{\mathbf{v}}, \mathbf{u} \rangle. \tag{50}$$

Notice that

$$\mathrm{comp}_{\mathbf{v}}(\mathbf{u}) = ||\mathbf{u}|| \cos\theta \tag{51}$$

where θ is the angle between $\mathbf{u}$ and $\mathbf{v}$ as drawn in the following diagram (the dotted blue arrow is the projection of $\mathbf{u}$ onto $\mathbf{v}$):



Therefore, projection is doing what we want, i.e. it tells you how much some vector points along another vector.

4 The Cross Product

The cross product is a very special operation defined **only** for vectors in $\mathbb{R}^3$. Its usefulness is in the fact that if one is given two vectors $\mathbf{u}$, $\mathbf{v}$, the cross product of these vectors is orthogonal to both $\mathbf{u}$ and $\mathbf{v}$.

Definition 4.1 (Cross Product)**.** *Let* $\mathbf{u}, \mathbf{v} \in \mathbb{R}^3$ *then the* **cross product** *or* **vector product** *of* $\mathbf{u}$ *and* $\mathbf{v}$ *is*

$$\mathbf{u} \times \mathbf{v} = (u_2v_3 - u_3v_2, u_3v_1 - u_1v_3, u_1v_2 - u_2v_1). \tag{52}$$

You may be wondering how does one remember this formula. There are a few options. The first is with **determinants**.[4]

Definition 4.2 (Determinant of 2×2 Matrix)**.** *The determinant of a* 2×2 *matrix,*

$$\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \tag{53}$$

is

$$\det(\mathbf{A}) = \begin{vmatrix} a & b \\ c & d \end{vmatrix} \doteq ad - bc. \tag{54}$$

Remark 4.1. *The determinant operation is effectively to multiply across diagonals and then subtract.*

One can extend this definition to 3×3 matrices with the following:

Definition 4.3 (Determinant of 3×3 Matrix)**.** *The determinant of a* 3×3 *matrix,*

$$\mathbf{A} = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix}, \tag{55}$$

is

$$\det(\mathbf{A}) = \begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} \doteq a\begin{vmatrix} e & f \\ h & i \end{vmatrix} - b\begin{vmatrix} d & f \\ g & i \end{vmatrix} + c\begin{vmatrix} d & e \\ g & h \end{vmatrix}. \tag{56}$$

The matrices

$$\mathbf{M}_{1,1} \doteq \begin{pmatrix} e & f \\ h & i \end{pmatrix}, \qquad \mathbf{M}_{1,2} \doteq \begin{pmatrix} d & f \\ g & i \end{pmatrix}, \qquad \mathbf{M}_{1,3} \doteq \begin{pmatrix} d & e \\ g & h \end{pmatrix}, \tag{57}$$

are called the minors and result from removing the first row and the j^{th} *(*$j = 1, 2, 3$ *respectively as the notation suggests) column of the matrix* $\mathbf{A}$*.*

Remark 4.2. *One can generalise further to* $n \times n$ *matrices but there is no need in this course.*

The following is a method to remember/derive the formula for the cross product:

1. Write $\mathbf{u}$ and $\mathbf{v}$ in terms of the standard basis vectors $\mathbf{i}$, $\mathbf{j}$ and $\mathbf{k}$, i.e.

$$\mathbf{u} = u_1\mathbf{i} + u_2\mathbf{j} + u_3\mathbf{k}, \tag{58}$$
$$\mathbf{v} = v_1\mathbf{i} + v_2\mathbf{j} + v_3\mathbf{k}. \tag{59}$$

2. Construct the following matrix:

$$\mathbf{C} = \begin{pmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{pmatrix} \tag{60}$$

   and treat $\mathbf{i}, \mathbf{j}, \mathbf{k}$ as elements of a matrix, not vectors.

[4]Determinants crop up in linear algebra. For example a famous result is that if a square matrix has non-zero determinant then it is invertible.

3. Then $\mathbf{u} \times \mathbf{v} = \det(\mathbf{C})$. In otherwords, we can now rewrite the formula for the cross product in terms of the determinant of the 3×3 matrix $\mathbf{C}$:

$$\mathbf{u} \times \mathbf{v} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{vmatrix}. \tag{61}$$

Explicitly,

$$\mathbf{u} \times \mathbf{v} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{vmatrix} = \mathbf{i}\begin{vmatrix} u_2 & u_3 \\ v_2 & v_3 \end{vmatrix} - \mathbf{j}\begin{vmatrix} u_1 & u_3 \\ v_1 & v_3 \end{vmatrix} + \mathbf{k}\begin{vmatrix} u_1 & u_2 \\ v_1 & v_2 \end{vmatrix} \tag{62}$$

$$= (u_2v_3 - u_3v_2)\mathbf{i} - (u_1v_3 - v_1u_3)\mathbf{j} + (u_1v_2 - u_2v_1)\mathbf{k}, \tag{63}$$

which is precisely formula (52) when written in the standard basis.

The following alternative is **non-examinable** but may be of interest to some. One can write the components of the cross product with a succinct formula:

$$(\mathbf{u} \times \mathbf{v})_i = \sum_{j=1}^{3}\sum_{k=1}^{3} \varepsilon_{ijk} u_j v_k \tag{64}$$

where ε_{ijk} is known as the (3-dimensional) Levi-Civita symbol which is defined as

$$\varepsilon_{ijk} = \begin{cases} (-1)^p & \text{if} \quad i \neq j \neq k \\ 0 & \text{otherwise} \end{cases} \tag{65}$$

where p is the number of pairwise interchanges of indices necessary to unscramble i, j, k to $1, 2, 3$. So,

$$\varepsilon_{123} = 1, \qquad \varepsilon_{231} = 1, \qquad \varepsilon_{312} = 1 \tag{66}$$

$$\varepsilon_{213} = -1, \qquad \varepsilon_{132} = -1, \qquad \varepsilon_{321} = -1, \tag{67}$$

whilst all other selection of indices vanish ($\varepsilon_{11j} = 0$ etc). Lets compute $(\mathbf{u} \times \mathbf{v})_1$. So,

$$(\mathbf{u} \times \mathbf{v})_1 = \sum_{j=1}^{3}\sum_{k=1}^{3} \varepsilon_{1jk} u_j v_k. \tag{68}$$

Now the only non-zero options for ε_{1jk} are $\varepsilon_{123} = 1$ and $\varepsilon_{132} = -1$. Therefore, the double sum in (68) only has two terms:

$$(\mathbf{u} \times \mathbf{v})_1 = \varepsilon_{123}u_2v_3 + \varepsilon_{132}u_3v_2 \tag{69}$$

$$= u_2v_3 - u_3v_2, \tag{70}$$

which if you compare to formula (52) is the correct result for $(\mathbf{u} \times \mathbf{v})_1$.

4.1 Properties

Proposition 4.1. *Let $\mathbf{u}, \mathbf{v} \in \mathbb{R}^3$ then the cross product $\mathbf{u} \times \mathbf{v}$ of $\mathbf{u}$ and $\mathbf{v}$ is orthogonal to both $\mathbf{u}$ and $\mathbf{v}$.*

Proof. Lets compute directly the scalar product

$$\langle \mathbf{u} \times \mathbf{v}, \mathbf{u} \rangle = (u_2v_3 - u_3v_2)u_1 + (u_3v_1 - u_1v_3)u_2 + (u_1v_2 - u_2v_1)u_3. \tag{71}$$

Upon expanding the first and the fourth term cancel. Likewise, the second and the fifth term cancel and the third and the last term cancel. Computing $\langle \mathbf{u} \times \mathbf{v}, \mathbf{v} \rangle$ is completely analogous. □

*Proof (***Non-examinable***).* The formula (64) gives a slick proof that $\mathbf{u}$ and $\mathbf{v}$ are orthogonal to $\mathbf{u} \times \mathbf{v}$. The scalar product of $\mathbf{v}$ and $\mathbf{u} \times \mathbf{v}$ in components is

$$\langle \mathbf{v}, \mathbf{u} \times \mathbf{v} \rangle = \sum_i v_i (\mathbf{u} \times \mathbf{v})_i = \sum_{i,j,k=1}^{3} \varepsilon_{ijk} u_j v_i v_k. \tag{72}$$

Observe that ε_{ijk} is totally antisymmetric. In particular, $\varepsilon_{ijk} = -\varepsilon_{kji}$. Therefore,

$$\langle \mathbf{v}, \mathbf{u} \times \mathbf{v} \rangle = \frac{1}{2} \sum_{i,j,k=1}^{3} \varepsilon_{ijk} u_j v_i v_k - \frac{1}{2} \sum_{i,j,k=1}^{3} \varepsilon_{kji} u_j v_i v_k. \tag{73}$$

You can now relabel the indices $i \leftrightarrow k$ since these both occur on $\mathbf{v}$ and they are summed over to find:

$$\langle \mathbf{v}, \mathbf{u} \times \mathbf{v} \rangle = \frac{1}{2} \sum_{i,j,k=1}^{3} \varepsilon_{ijk} u_j v_i v_k - \frac{1}{2} \sum_{i,j,k=1}^{3} \varepsilon_{ijk} u_j v_k v_i = 0. \tag{74}$$

□

Example 4.1. *Lets do some examples for some practise:*

(a) Compute the cross product of the standard basis vectors:

$$\mathbf{i} \times \mathbf{j} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{vmatrix} = \mathbf{k}, \quad \mathbf{j} \times \mathbf{k} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix} = \mathbf{i}, \quad \mathbf{k} \times \mathbf{i} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{vmatrix} = \mathbf{j}. \tag{75}$$

(b) The cross product of any vector with itself vanishes:

$$\mathbf{u} \times \mathbf{u} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ u_1 & u_2 & u_3 \\ u_1 & u_2 & u_3 \end{vmatrix} = \mathbf{i}(u_2 u_3 - u_3 u_2) - \mathbf{j}(u_1 u_3 - u_3 u_1) + \mathbf{k}(u_1 u_2 - u_1 u_2) = \mathbf{0}. \tag{76}$$

(c) Let $\mathbf{u} = (1, -1, 5)$ and $\mathbf{v} = (2, 1, -2)$. The cross product of $\mathbf{u}$ and $\mathbf{v}$ is

$$\mathbf{u} \times \mathbf{v} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 1 & -1 & 5 \\ 2 & 1 & -2 \end{vmatrix} = (2-5)\mathbf{i} - (-2-10)\mathbf{j} + (1+2)\mathbf{k} = -3\mathbf{i} + 12\mathbf{j} + 3\mathbf{k}. \tag{77}$$

Proposition 4.2 (Properties of $\times$)**.** *Let $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^3$ and $\mathbf{a} \in \mathbb{R}$. The cross product satisfies the following properties:*

(a) $\mathbf{u} \times \mathbf{u} = \mathbf{0}$.

(b) Anticommutative: $\mathbf{u} \times \mathbf{v} = -\mathbf{v} \times \mathbf{u}$.

(c) Distributive over vector addition: $\mathbf{u} \times (\mathbf{v} + \mathbf{w}) = \mathbf{u} \times \mathbf{v} + \mathbf{u} \times \mathbf{w}$.

(d) Compatible with scalar multiplication: $(a\mathbf{u}) \times \mathbf{v} = a(\mathbf{u} \times \mathbf{v}) = \mathbf{u} \times (a\mathbf{v})$.

(e) The **scalar triple product** $\langle \mathbf{u}, (\mathbf{v} \times \mathbf{w}) \rangle$ *satisfies* $\langle \mathbf{u}, (\mathbf{v} \times \mathbf{w}) \rangle = \langle (\mathbf{u} \times \mathbf{v}), \mathbf{w} \rangle$.

(f) The **vector triple product** $\mathbf{u} \times (\mathbf{v} \times \mathbf{w})$ *satisfies* $\mathbf{u} \times (\mathbf{v} \times \mathbf{w}) = \langle \mathbf{u}, \mathbf{w} \rangle \mathbf{v} - \langle \mathbf{u}, \mathbf{v} \rangle \mathbf{w}$.

Proof. Property a) is proved above in example 4.1. Most are left to you to check yourself. Here is a proof of property f). After computing $\mathbf{v} \times \mathbf{w}$ one has

$$\mathbf{u} \times (\mathbf{v} \times \mathbf{w}) = (u_1\mathbf{i} + u_2\mathbf{j} + u_3\mathbf{k}) \times (\mathbf{i}(v_2w_3 - v_3w_2) - \mathbf{j}(v_1w_3 - v_3w_1) + \mathbf{k}(v_1w_2 - w_1v_2)) \tag{78}$$

By using the distributive property, the anticommutivity property in conjunction with the distributive property and property a) one has

$$\begin{aligned}\mathbf{u} \times (\mathbf{v} \times \mathbf{w}) = &-\mathbf{i} \times \mathbf{j}u_1(v_1w_3 - v_3w_1) + \mathbf{i} \times \mathbf{k}u_1(v_1w_2 - w_1v_2)\\ &+ \mathbf{j} \times \mathbf{i}u_2(v_2w_3 - v_3w_2) + \mathbf{j} \times \mathbf{k}u_2(v_1w_2 - w_1v_2))\\ &+ \mathbf{k} \times \mathbf{i}u_3(v_2w_3 - v_3w_2) - \mathbf{k} \times \mathbf{j}u_3(v_1w_3 - v_3w_1).\end{aligned} \tag{79}$$

Using the formulas from example 4.1 part a) and the anticommutative property one has

$$\begin{aligned}\mathbf{u} \times (\mathbf{v} \times \mathbf{w}) = &-\mathbf{k}u_1(v_1w_3 - v_3w_1) - \mathbf{j}u_1(v_1w_2 - w_1v_2)\\ &- \mathbf{k}u_2(v_2w_3 - v_3w_2) + \mathbf{i}u_2(v_1w_2 - w_1v_2)\\ &+ \mathbf{j}u_3(v_2w_3 - v_3w_2) + \mathbf{i}u_3(v_1w_3 - v_3w_1).\end{aligned} \tag{80}$$

Suggestively collecting terms gives:

$$\begin{aligned}\mathbf{u} \times (\mathbf{v} \times \mathbf{w}) = &\, u_1v_1(-\mathbf{k}w_3 - w_2\mathbf{j}) + u_1w_1(v_3\mathbf{k} + v_2\mathbf{j})\\ &+ u_2v_2(-\mathbf{k}w_3 - \mathbf{i}w_1) + u_2w_2(v_3\mathbf{k} + v_1\mathbf{i})\\ &+ u_3w_3(v_2\mathbf{j} + v_1\mathbf{i}) + u_3v_3(-w_2\mathbf{j} - w_1\mathbf{i}).\end{aligned} \tag{81}$$

This can be rewritten as

$$\begin{aligned}\mathbf{u} \times (\mathbf{v} \times \mathbf{w}) = &\, u_1v_1(-\mathbf{w} + \mathbf{i}w_1) + u_1w_1(\mathbf{v} - v_1\mathbf{i})\\ &+ u_2v_2(-\mathbf{w} + \mathbf{j}w_2) + u_2w_2(\mathbf{v} - \mathbf{j}w_2)\\ &+ u_3w_3(\mathbf{v} - \mathbf{k}v_3) + u_3v_3(-\mathbf{w} + \mathbf{k}w_3)\end{aligned} \tag{82}$$

$$= (u_1w_1 + u_2w_2 + u_3w_3)\mathbf{v} - (u_1v_1 + u_2v_2 + u_3v_3)\mathbf{w}. \tag{83}$$

□

Remark 4.3. *Note that the associative property from ordinary multiplication does* **not** *hold for the cross product, i.e.*

$$(\mathbf{u} \times \mathbf{v}) \times \mathbf{w} \neq \mathbf{u} \times (\mathbf{v} \times \mathbf{w}). \tag{84}$$

For example, $\mathbf{i} \times \mathbf{i} = \mathbf{0}$ *so* $(\mathbf{i} \times \mathbf{i}) \times \mathbf{j} = 0$ *but* $\mathbf{i} \times \mathbf{j} = \mathbf{k}$ *so* $\mathbf{i} \times (\mathbf{i} \times \mathbf{j}) = -\mathbf{j}$. *Note that proposition 4.2 and example 4.1 have been used.*

Remark 4.4. *You can check (and you should check) that computing the scalar triple product* $\langle \mathbf{u}, (\mathbf{v} \times \mathbf{w})\rangle$ *is equivalent to computing the determinant of the following matrix:*

$$\mathbf{M} \doteq \begin{pmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{pmatrix}. \tag{85}$$

4.2 Geometric Properties of Cross Product

Proposition 4.3. *Let* θ *denote the angle between* $\mathbf{u}$ *and* $\mathbf{v}$. *The norm/length of the vector* $\mathbf{u} \times \mathbf{v}$ *is then given by*

$$||\mathbf{u} \times \mathbf{v}|| = ||\mathbf{u}||||\mathbf{v}|| \sin\theta. \tag{86}$$

Proof. One can show by a direct computation that

$$||\mathbf{u} \times \mathbf{v}||^2 = ||\mathbf{u}||^2||\mathbf{v}||^2 - |\langle \mathbf{u}, \mathbf{v} \rangle|^2. \tag{87}$$

From proposition 3.2 one has

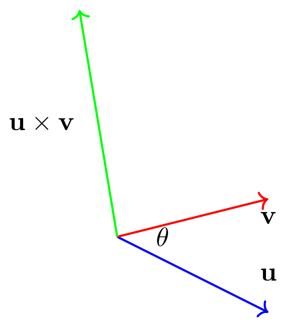
$$||\mathbf{u} \times \mathbf{v}|| = ||\mathbf{u}||||\mathbf{v}||(1 - \cos^2\theta) = ||\mathbf{u}||||\mathbf{v}||\sin^2\theta. \tag{88}$$

The angle θ is the angle between the two vectors, so $\theta \in [0, \pi]$ and therefore, $\sin\theta \geq 0$. Therefore, its square root is well defined on the reals. Additionally, the norm of a vector is a positive quantity. Therefore, we take the positive branch of the square root to complete the proof. □

Corollary 4.1. *Let $\mathbf{u}$, $\mathbf{v}$ be two non-zero vectors. $\mathbf{u}$, $\mathbf{v}$ are parallel if and only if $\mathbf{u} \times \mathbf{v} = \mathbf{0}$.*

Proof. The vectors $\mathbf{u}$, $\mathbf{v}$ are parallel if and only if $\theta = 0$ or $\theta = \pi$. In either case $\sin\theta = 0$. Therefore, by proposition 4.3, $\mathbf{u}$, $\mathbf{v}$ are parallel if and only if $||\mathbf{u} \times \mathbf{v}|| = 0$. From the properties of the norm $||\mathbf{u} \times \mathbf{v}|| = 0$ if and only if $\mathbf{u} \times \mathbf{v} = \mathbf{0}$. □

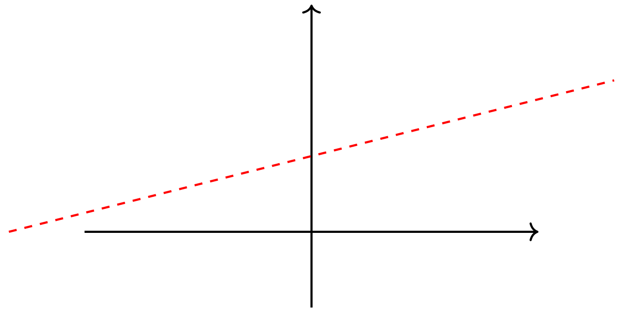
How should you visualise of cross product geometrically? Suppose you have two vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^3$. Then $\mathbf{u} \times \mathbf{v}$ is a vector that points perpendicular to the plane through (spanned by) $\mathbf{u}$ and $\mathbf{v}$. It's direction is determined by the *right-hand rule*. Curl the fingers of your right hand in the direction of the smallest angle from $\mathbf{u}$ to $\mathbf{v}$. The vector $\mathbf{u} \times \mathbf{v}$ is then in the direction of your thumb. This drawn below:



5 Equations of Lines and Planes

5.1 Review: Lines in $\mathbb{R}^2$

Suppose one has a straight line, denoted γ, in the plane $\mathbb{R}^2$ as drawn in red below:



How do you go about describing this mathematically? First of all it's a subset of $\mathbb{R}^2$, i.e. it's a collection of points in the plane. You can determine it uniquely with two points (x_0, y_0) and (x_1, y_1), i.e. the straight line passing through these points is unique. From this you can compute the lines gradient

$$m = \frac{y_1 - y_0}{x_1 - x_0} \tag{89}$$

which is how much you go up for how much you go across. Then equation of a line is

$$y - y_0 = m(x - x_0), \tag{90}$$

which is simply a relation between the x and y coordinates of the points on the line. So γ is the subset:

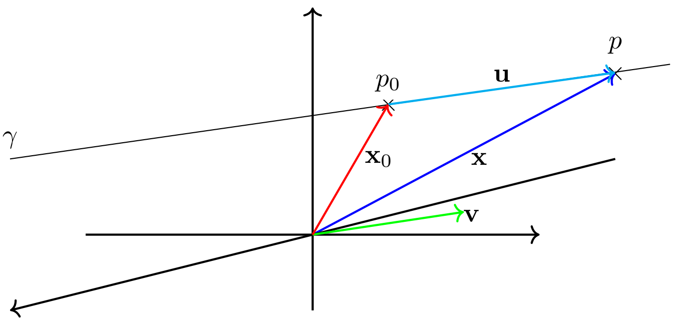
$$\gamma = \Big\{(x, y) \in \mathbb{R}^2 : y - y_0 = m(x - x_0)\Big\}. \tag{91}$$

There is an alternative, one could also describe the line with a point and a direction (which would be a vector in $\mathbb{R}^2$). This is what we are going to do to study lines in $3D$ space $\mathbb{R}^3$ but the same works in $\mathbb{R}^2$ and indeed $\mathbb{R}^n$.

5.2 Lines in $\mathbb{R}^3$

A line γ in $\mathbb{R}^3$ is specified by a point $p_0 \in \mathbb{R}^3$ that γ passes through and a direction for γ. This direction can be described by any vector $\mathbf{v}$ parallel to the line.

The setup is the following: Let $\mathbf{x}_0$ be the position vector of some p_0 on γ and $\mathbf{x}$ be the position vector of another arbitrary point p on γ. Let $\mathbf{u}$ be the vector that points from p_0 to p, i.e. $\mathbf{u} = \overrightarrow{p_0 p}$. Finally, let $\mathbf{v}$ be any vector that is parallel to $\mathbf{u}$. This is drawn below:



So from the discussion in lecture 2 one has that the position vector $\mathbf{x}$ is given by

$$\mathbf{x} = \mathbf{x}_0 + \mathbf{u}. \tag{92}$$

Now, since $\mathbf{u}$ is parallel to $\mathbf{v}$, one can express $\mathbf{u} = \lambda\mathbf{v}$ for $\lambda \in \mathbb{R}$. Hence, be written as equation (92) can be written as

$$\mathbf{x} = \mathbf{x}_0 + \lambda\mathbf{v}, \tag{93}$$

which is the **vector equation** of the line. Notice that $\mathbf{x}$ has dependence on the **parameter** λ; for each λ one has a the position vector of a point on γ. So the tip of the position vector traces out γ as λ varies. One can highlight these ideas by writing

$$\mathbf{x}(\lambda) = \mathbf{x}_0 + \lambda\mathbf{v}. \tag{94}$$

Additionally, one can write the above equation in components

$$(x, y, z) = (x_0 + \lambda v_1, y_0 + \lambda v_2, z_0 + \lambda v_3). \tag{95}$$

This gives the set of **parametric equations** for the line γ:

$$x = x_0 + \lambda v_1, \qquad y = y_0 + \lambda v_2, \qquad z = z_0 + \lambda v_3. \tag{96}$$

Note that $\mathbf{v} \neq 0$ otherwise one has a point in $\mathbb{R}^3$. So at least one of v_1, v_2, v_3 is non-vanishing. Suppose it is v_1 then one can solve for λ,

$$\lambda = \frac{x - x_0}{v_1} \tag{97}$$

and replace λ in (96) to give

$$y - y_0 = \frac{x - x_0}{v_1} v_2, \qquad z - z_0 = \frac{x - x_0}{v_1} v_3. \tag{98}$$

Remark 5.1. *One should compare to equation* (90)*, the equation for a line in* $\mathbb{R}^2$*. Note that if one was dealing with lines in the plane then one would simply have*

$$y - y_0 = \frac{v_2}{v_1}(x - x_0) \tag{99}$$

and no z*-equation. So we have precisely equation* (90) *with the gradient* $m = \frac{v_2}{v_1}$*.*

If $v_2 = 0 = v_3$ then the line one is describing is parallel to the x-axis. Suppose additionally, $v_2 \neq 0$ and $v_3 = 0$ then one can

$$\frac{y - y_0}{v_2} = \frac{x - x_0}{v_1}, \qquad z = z_0, \tag{100}$$

which describes a line in the plane $z = z_0$. Finally if $v_2 \neq 0$ and $v_3 \neq 0$ then one can see

$$\frac{y - y_0}{v_2} = \frac{x - x_0}{v_1} = \frac{z - z_0}{v_3}, \tag{101}$$

which is sometimes called the **symmetric equation** for γ.

In practise if you are asked to determine the equation of a straight line in $\mathbb{R}^3$ from two points $p_0 = (x_0, y_0, z_0)$ and $p_1 = (x_1, y_1, z_1)$ you would compute the following:

1. The position vectors of p_0 and p_1 are $\mathbf{x}_0 = (x_0, y_0, z_0)$ and $\mathbf{x}_1 = (x_1, y_1, z_1)$ respectively.
2. The vector $\mathbf{u} = \mathbf{x}_1 - \mathbf{x}_0$ is parallel to the line.
3. Therefore,

$$\mathbf{x}(\lambda) = \mathbf{x}_0 + \lambda\mathbf{u} \tag{102}$$

is the equation of the line or alternatively, using $\mathbf{u} = \mathbf{x}_1 - \mathbf{x}_0$,

$$\mathbf{x}(\lambda) = (1 - \lambda)\mathbf{x}_0 + \lambda\mathbf{x}_1. \tag{103}$$

Notice that $\mathbf{x}(1) = (x_1, y_1, z_1)$ and $\mathbf{x}(0) = (x_0, y_0, z_0)$. So if one limits the values that λ can take one has a description of a **line segment**, i.e. the line segment from p_0 to p_1 is given by

$$\mathbf{x}(\lambda) = (1-\lambda)\mathbf{x}_0 + \lambda\mathbf{x}_1, \qquad 0 \leq \lambda \leq 1. \tag{104}$$

Lets do an example:

Example 5.1. *Let γ_1 be the line passing through the point $(1, -2, 4)$ with parallel vector $\mathbf{v} = (1, 3, -1)$. Let γ_2 be the line passing through the points $(2, 4, 1)$ and $(0, 3, -3)$.*

1. *Determine their vector equations and parametric equations.*
2. *Determine where they intersect the xy-plane.*
3. *Show that they do not intersect and are not parallel.*

1. Denote the position vector $(1, -2, 4)$ as $\mathbf{x}_{0,1}$. So denoting the position vector of an arbitrary point p along γ_1 as $\mathbf{r}_1$, the vector equation for γ_1 is

$$\mathbf{r}_1(\lambda_1) = \mathbf{x}_{0,1} + \lambda_1\mathbf{v}, \tag{105}$$

which can be written as

$$\mathbf{r}_1(\lambda_1) = (1+\lambda_1, -2+3\lambda_1, 4-\lambda_1). \tag{106}$$

So its parametric equations can be read off as

$$x = 1+\lambda_1, \qquad y = -2+3\lambda_1, \qquad z = 4-\lambda_1. \tag{107}$$

Now one has the position vectors $\mathbf{x}_{1,2} = (2, 4, 1)$ and $\mathbf{x}_{0,2} = (0, 3, -3)$. So denoting the position vector of an arbitrary point p along γ_2 as $\mathbf{r}_2$ and using equation (103) *one has*

$$\mathbf{r}_2(\lambda_2) = (1-\lambda_2)\mathbf{x}_{0,2} + \lambda_2\mathbf{x}_{1,2} = (2\lambda_2, 3+\lambda_2, 4\lambda_2 - 3). \tag{108}$$

So the parametric equations for γ_2 are

$$x = 2\lambda_2, \qquad y = 3+\lambda_2, \qquad z = 4\lambda_2 - 3. \tag{109}$$

2. To determine where γ_1 intersects the xy-plane one checks what value of λ_1 gives $z = 0$. From above, this means $\lambda_1 = 4$. Therefore, γ_1 intersects the xy-plane at $(5, 10, 0)$. For γ_2 one has $\lambda_2 = \frac{3}{4}$ and $(\frac{3}{2}, \frac{15}{4}, 0)$.

3. To be parallel $\mathbf{v}$ and $\mathbf{u} = \mathbf{x}_{1,2} - \mathbf{x}_{0,2} = (2, 1, 4)$ would have to be proportional: one should be able to solve

$$(1, 3, -1) + a(2, 1, 4) = 0 \tag{110}$$

for a, which gives a contradiction. To see this explicitly, note that if they were to intersect one should be able to find λ_1 and λ_2 such that

$$2\lambda_2 = 1+\lambda_1, \qquad 3+\lambda_2 = -2+3\lambda_1, \qquad 4\lambda_2 - 3 = 4-\lambda_1. \tag{111}$$

Solving the first gives

$$\lambda_1 = 2\lambda_2 - 1 \tag{112}$$

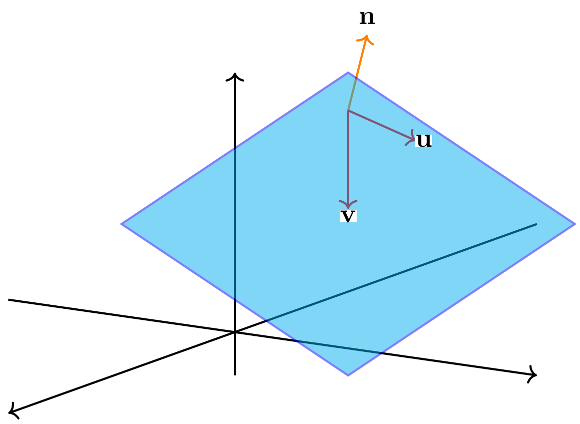
which can be used to eliminate λ_2 in the second and third equations to give:

$$\lambda_2 = \frac{8}{5}, \qquad \lambda_2 = \frac{4}{3} \tag{113}$$

which is inconsistent and therefore, the lines do not intersect.

5.3 Planes

Suppose you have a plane in $\mathbb{R}^3$. How do you describe such and object mathematically? You want to specify a point which is in the plane and how the plane sits in $3D$ space relative to that point. Two non-parallel vectors which lie in the plane would suffice to specify the 'direction' of the plane:



Alternatively, one could work out the planes **normal vector**, $\mathbf{n}$, which is a vector orthogonal to the plane (drawn in orange above). Therefore, any vector $\mathbf{v}$ lying in the plane satisfies

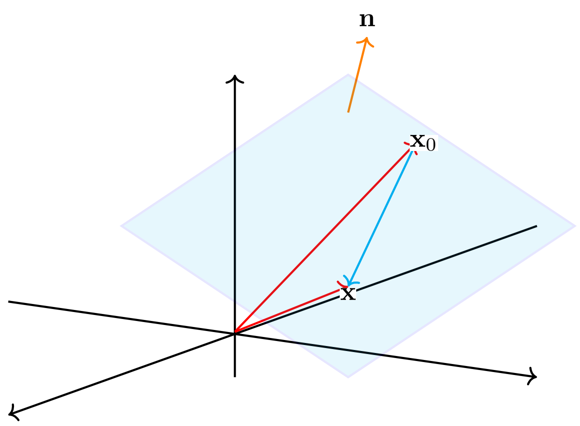
$$\langle \mathbf{n}, \mathbf{v} \rangle = 0. \tag{114}$$

In other words, what this equation is encoding is the $2D$ space of vectors orthogonal to the normal vector, which is precisely the plane you want to describe.

Given any two points p_0 and p in the plane with position vectors $\mathbf{x}_0 = (x_0, y_0, z_0)$ and $\mathbf{x} = (x, y, z)$ respectively. Then $\mathbf{v} = \mathbf{x} - \mathbf{x}_0$ lies in the plane. So,

$$\langle \mathbf{n}, \mathbf{x} - \mathbf{x}_0 \rangle = 0, \tag{115}$$

which is often called the **equation of the plane**. This is drawn below



You can write equation (115) using the definition of the scalar product. Suppose $\mathbf{n} = (n_1, n_2, n_3)$, $\mathbf{x} = (x, y, z)$ and $\mathbf{x}_0 = (x_0, y_0, z_0)$. Then, equation (115) becomes

$$n_1(x - x_0) + n_2(y - y_0) + n_3(z - z_0) = 0. \tag{116}$$

Remark 5.2. *Stewart likes to call* (115) *the* **vector** *equation of a plane and* (116) *the* **scalar** *equation of the plane. This seems a bit nonsensical. First,* (115) *is just compact notation for* (116)*. Secondly, the object one computes in equation* (115) *is the scalar product, which is a number (in particular* 0*)* *not* *a vector!*

Remark 5.3. *The general form of equation* (116) *is*

$$ax + by + cz + d = 0 \tag{117}$$

where we've written $(a, b, c) = (n_1, n_2, n_3)$ *and* $d = -(n_1 x_0 + n_2 y_0 + n_3 z_0)$. *One can view this as the general equation for a plane provided one of* a, b, c *is non-vanishing: Suppose* $a \neq 0$ *(the same argument works with* b *or* c *non-vanishing). Then,*

$$a\Big(x + \frac{d}{a}\Big) + b(y - 0) + c(z - 0) = 0 \tag{118}$$

which can be rewritten as

$$\langle \mathbf{n}, \mathbf{x} - \mathbf{x}_0 \rangle = 0 \tag{119}$$

with $\mathbf{n} = (a, b, c)$ *and* $\mathbf{x}_0 = (-{}^{d}\!/\!_{a}, 0, 0)$.

Remark 5.4. *It is very common for the normal vector to be 'normalised' to be a unit vector,* $\hat{\mathbf{n}}$.

Two planes are **parallel** if their normal vectors are parallel. If the two planes are not parallel then they intersect in a straight line and the **angle between the planes** is defined as the acute angle between their normal. In particular, if one has two planes with normals $\mathbf{n}_1$ and $\mathbf{n}_2$ respectively then the angle between the planes is

$$\theta = \arccos\big(|\langle \hat{\mathbf{n}}_1, \hat{\mathbf{n}}_2 \rangle|\big) \tag{120}$$

Finally, one can consider the distance of a point $p \in \mathbb{R}^3$ to a plane. Let p_0 be some point in the plane and let $\mathbf{x}_0$ be its position vector. Additionally, let $\mathbf{x}$ be the position vector of p. The displacement vector from p_0 to p is then $\mathbf{v} = \mathbf{x} - \mathbf{x}_0$. Then the distance from p to the plane is the absolute value of the component of $\mathbf{v}$ onto $\mathbf{n}$:

$$d = |\mathrm{comp}_{\mathbf{n}} \mathbf{v}| = |\langle \hat{\mathbf{n}}, \mathbf{v} \rangle|. \tag{121}$$

Example 5.2. *Let's do a prototypical practise problem to illustrate all these ideas:*

1. *Suppose* $\mathbf{u} = (-1, 1, 1)$, $\mathbf{v} = (7, -4, -5)$ *lie in the plane which passes through* $(1, 0, 0)$. *Find its normal and therefore it's equation.*
2. *Find the equation of the plane which passes through the points* $(-1, 1, 1)$, $(1, -2, 2)$, $(4, -3, 0)$.
3. *Find the angle between these planes and equation of the line of intersection of the planes.*

To do 1 we need to find a normal. For this we simply need to compute the cross product of $\mathbf{u}$ *and* $\mathbf{v}$ *to find a vector orthogonal to* $\mathbf{u}$ *and* $\mathbf{v}$*:*

$$\mathbf{n}_1 = \mathbf{u} \times \mathbf{v} = \left| \begin{pmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ -1 & 1 & 1 \\ 7 & -4 & -5 \end{pmatrix} \right| = -\mathbf{i} + 2\mathbf{j} - 3\mathbf{k}. \tag{122}$$

Therefore, denoting $\mathbf{x}_0 = (1, 0, 0)$ *one has*

$$\langle \mathbf{n}_1, \mathbf{x} - \mathbf{x}_0 \rangle = 0, \tag{123}$$

which can be expanded as

$$-(x - 1) + 2y - 3z = 0 \Longleftrightarrow x - 2y + 3z = 1. \tag{124}$$

For 2 one can do the following. From the position vectors of the points $(-1, 1, 1)$, $(1, -2, 2)$, $(4, -3, 0)$ *one can construct two vectors in the plane:*

$$\mathbf{u} = (5, -4, -1), \qquad \mathbf{v} = (2, -3, 1). \tag{125}$$

Their cross product gives,

$$\mathbf{n}_2 = (-7, -7, -7). \tag{126}$$

Therefore, if $\mathbf{x}_0 = (-1, 1, 1)$ *the equation of the plane is:*

$$-7(x+1) - 7(y-1) - 7(z-1) = 0 \Longleftrightarrow x + y + z = 1. \tag{127}$$

The angle between the planes is

$$\cos\theta = \frac{|-1+2-3|}{\sqrt{42}} = \frac{2}{\sqrt{42}} \implies \theta = \arccos({}^{2}/\sqrt{42}) \tag{128}$$

To find the line of intersection you need a point on the line an a vector along that line. Set $z = 0$ *in both plane equations gives:*

$$x + y = 1, \qquad x - 2y = 1 \implies y = 0, \quad x = 1. \tag{129}$$

So the point $(1, 0, 0)$ *lies on the line. Now note that the vector along the line must lie in both planes, i.e. it's perpendicular to both normals. Therefore, one can determine it by computing the cross product of the normals:*

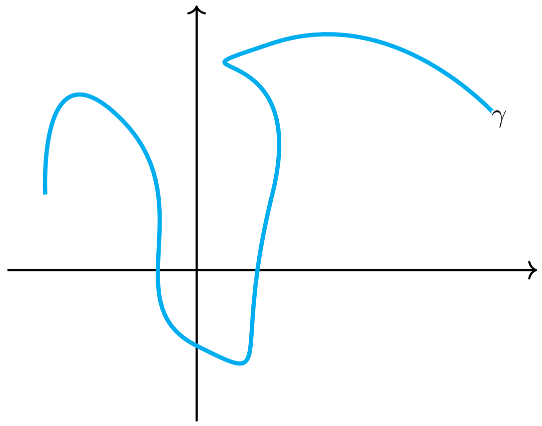
$$\mathbf{w} = \mathbf{n}_1 \times \mathbf{n}_2 = (5, -2, -3). \tag{130}$$

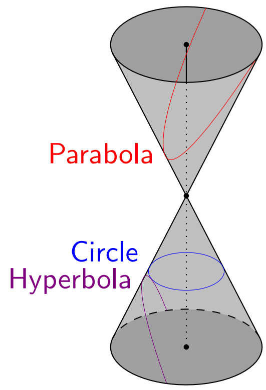
Therefore,

$$\mathbf{x} = (1, 0, 0) + t\mathbf{w}. \tag{131}$$

6 Curves, Conic Sections, Generalised Cylinders and Quardric Surfaces

In last lecture we studied straight lines in $\mathbb{R}^3$. This lecture is about how to write down a mathematical description of lines that are not straight in $\mathbb{R}^2$, i.e. ones that curve in the plane. The first half will cover how to write curves in terms of a parameter and the second half will look at conics. The term conic refers to a particular collection of curves that result from intersecting a plane with a cone: the options are an ellipse (including a circle), a parabola and a hyperbola as drawn below.





6.1 Curves in terms of a Parameter in $\mathbb{R}^2$

In this section we will study how to write down an equation or set of equations for lines that are not straight, i.e. ones that curve, like γ above. Depending on the curve one can sometimes express y as a function of x, i.e. $y = f(x)$. Then what we've drawn above is the **graph** of $f(x)$:

$$\mathrm{Graph}(f) = \{(x, f(x)) : x \in I\}, \tag{132}$$

where I is some interval in $\mathbb{R}$. Therefore, the curve γ above would be precisely this set: the graph of $f(x)$.

In many instances curves cannot be represented as the graph of a single function. For example, even something as simple as a unit circle,

$$\mathbb{S}^1 = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}, \tag{133}$$

cannot be represented as the graph of a single function. In particular, one requires two functions[5]:

$$y = f_\pm(x) = \pm\sqrt{1 - x^2}. \tag{134}$$

One can also abandon trying to relate the x and y coordinates that the curve γ passes through by a function. Instead one can treat x and y independently and allow them to depend on a **parameter**, λ. This involves writing

$$x = f(\lambda), \qquad y = g(\lambda). \tag{135}$$

The equations in (135) are then known as **parametric equations** of the curve γ. Then γ would be the set

$$\gamma = \{(x, y) = (f(\lambda), g(\lambda))\}. \tag{136}$$

The parameter λ may take values in all of the real numbers $\mathbb{R}$ or just some interval I of the real numbers.

Returning to the circle example. A circle of radius R centred at (a, b) can be represented with the parametric equations

$$x = a + R\sin(\lambda), \qquad y = b + R\cos(\lambda), \qquad \lambda \in [0, 2\pi). \tag{137}$$

[5]You may wonder when a curve can be expressed as more than one function *locally*. This is covered by the implicit function theorem.

One can sanity check this,

$$(x-a)^2 + (y-b)^2 = R^2 \sin^2 \lambda + R^2 \cos^2 \lambda = R^2, \tag{138}$$

which is the equation of a circle of radius R centred at (a, b).

Remark 6.1. *Sometimes the parameter has physical meaning, sometimes it does not. For example, if one was modelling the curve traced out by a particle on a plane then the parameter one could use is time, i.e.* $\lambda = t$.

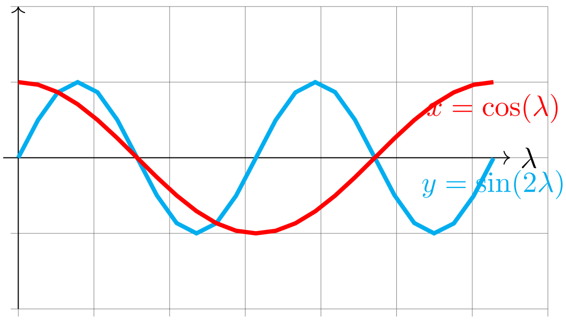
Remark 6.2. *In the circle example the sanity check computation is an example of eliminating the parameter,* λ*. Be warned that it is not always possible to eliminate the parameter.*

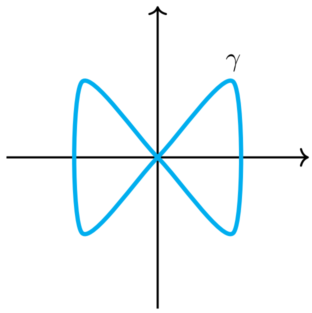
Let's do an example:

Example 6.1. *Consider the curve* γ *defined by*

$$\gamma = \{(x, y) : x = \cos(\lambda), y = \sin(2\lambda))\}. \tag{139}$$

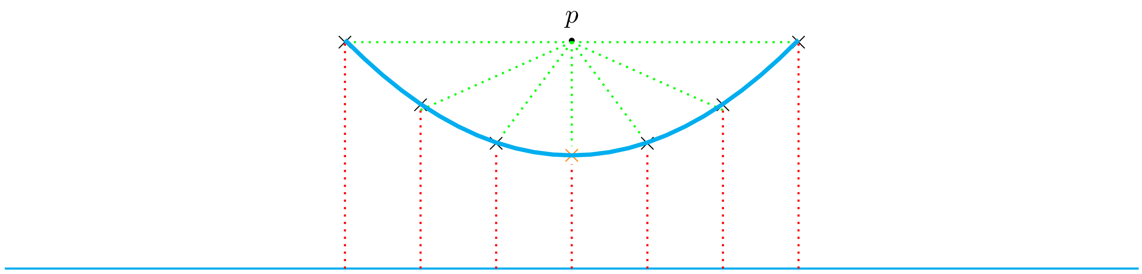
It's useful to plot x *and* y *as functions of* λ *to help you visualise the curve. This done on the left with* γ *on the right:*





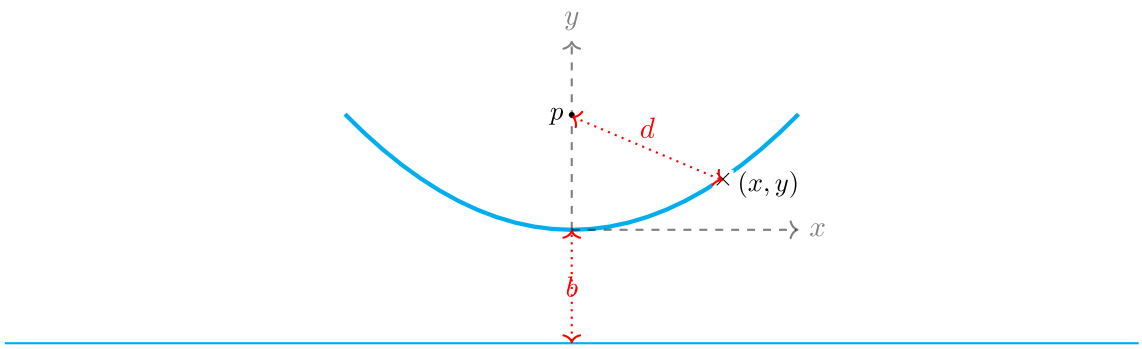
6.2 The Parabola

Suppose you take a line in $\mathbb{R}^2$ and a point p that does not lie on the line:



The curve that lies equidistant from the point and the line is a parabola. The line here is given a special name: **the directrix**. The point p is called the **focus**. The point denoted with a orange cross in the above diagram is called the **vertex**.

Let's work out a simple formula for the parabola by placing the vertex at the origin of our Cartesian coordinates on $\mathbb{R}^2$:



From pythagoras, one has

$$d^2 = (b - y)^2 + x^2. \tag{140}$$

Also, by the definition of the parabola, $|b + y| = d$. Therefore, equating and rearranging gives

$$y = \frac{1}{4b}x^2. \tag{141}$$

So, this is the equation of a parabola with focus at $(0, b)$ and directrix at $y = -b$.

Remark 6.3. *The sign of b controls whether the parabola points up or down, i.e. $b > 0$ then $y \geq 0$, so the parabola points up, $b < 0$ then $y \leq 0$, so the parabola points down.*

As is evident from its drawing the parabola is symmetric with repect to $x \mapsto -x$.

In general, a upright parabola with vertex at (a, c) and focus at $(a, c + b)$ has formula

$$y = \frac{1}{4b}(x - a)^2 + c. \tag{142}$$

6.3 The Ellipse

Take a circle and stretch it, you have an ellipse. The standard equation for an ellipse of height $2b > 0$ and width $2a > 0$ centred at (x_0, y_0) is

$$\frac{(x - x_0)^2}{a^2} + \frac{(y - y_0)^2}{b^2} = 1. \tag{143}$$

Note that for $a = b$ this is the equation of a circle of radius a^2 centred at (x_0, y_0). For simplicity, let's assume that $(x_0, y_0) = (0, 0)$ and that $a \geq b$. Then there are two distinguished points called **foci** $f_\pm$ at $(\pm c, 0)$ with

$$c = \sqrt{a^2 - b^2}. \tag{144}$$

Let p be some point on the ellipse at (x, y), with $x > 0$, $y > 0$. Then the distance from the foci f_+ at $(c, 0)$ to p is

$$d(p, f_+) = \sqrt{(c - x)^2 + y^2}. \tag{145}$$

Similarly, the distance from the foci f_- at $(-c, 0)$ to p is

$$d(p, f_-) = \sqrt{(x + c)^2 + y^2}. \tag{146}$$

Let's compute $d = d(p, f_+) + d(p, f_-)$. Squaring gives

$$\frac{1}{2}d^2 = x^2 + c^2 + y^2 + \sqrt{(c - x)^2 + y^2}\sqrt{(c + x)^2 + y^2}. \tag{147}$$

Solving for the square-root terms and squaring gives:

$$((c - x)^2 + y^2)((c + x)^2 + y^2) - (\frac{1}{2}d^2 - x^2 - c^2 - y^2)^2 = 0, \tag{148}$$

which, using $c^2 = a^2 - b^2$ and the equation (143) reduces to the polynomial:

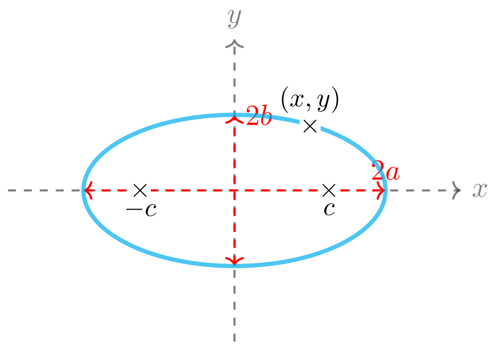
$$\frac{(4a^2 - d^2)(4b^2x^2 + a^2(d^2 - 4x^2))}{4a^2} = 0. \tag{149}$$

This holds for all $x \in (-a, a)$. Hence, the only way this is possible is if

$$d = 2a. \tag{150}$$

This means that the defining feature of an ellipse is that the sum of the distances from each foci to a given point has to be constant (here equal to $2a$).

The line through the foci is called the **(semi)-major axis** and the line perpendicular to this is called the **(semi)-minor axis**. The ellipse and the above discussion is drawn below:



Remark 6.4. *The parametric equations defining the ellipse centred* $(x_0, y_0) = (0, 0)$ *with* $a \geq b$ *are*

$$(x, y) = (a\cos(\lambda), b\sin(\lambda)), \qquad \lambda \in [0, 2\pi). \tag{151}$$

6.4 The Hyperbola

Recall that defining feature of an ellipse is that the sum of the distances from each foci to a given point has to be constant. The hyperbolas defining feature is that the difference of the distances from each foci to a given point has to be constant, i.e if one returns for the above computation

$$d(p, f_+) - d(p, f_-) = \pm a. \tag{152}$$

If one lets $c = \sqrt{a^2 + b^2}$ then the equation for a hyperbola with foci at $(\pm c, 0)$ is

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1. \tag{153}$$

When $y = 0$ then $x = \pm a$. These are the **vertices** of the hyperbola. Also, observe

$$x^2 = a^2 + \frac{a^2}{b^2}y^2 \geq a^2 \implies x \geq a \text{ or } x \leq -a. \tag{154}$$

These are two **branches** of the hyperbola. Finally, note that there is no y-intercept since $y^2 = -b^2 < 0$.

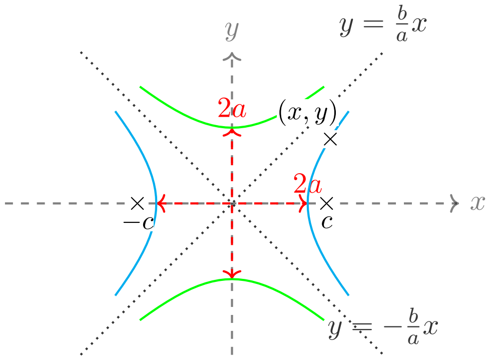
The hyperbola in equation (153) has two asymptotes:

$$y = \pm\frac{b}{a}x. \tag{155}$$

One can reverse the roles of x and y by sending x/a to y/b and visa versa to obtain the equation

$$\frac{y^2}{b^2} - \frac{x^2}{a^2} = 1, \tag{156}$$

which has the same asymptotes and is known as the **conjugate hyperbola**. This is drawn below:



Note that the parametric form of the equation for the hyperbola is

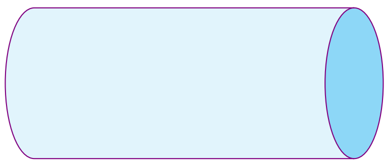
$$x = \pm a\cosh(\lambda), \qquad y = b\sinh(\lambda). \tag{157}$$

6.5 Generalised Cylinders

Up to a rotation and a translation a **cylinder** of radius R is the set of points in $\mathbb{R}^3$:

$$\{(x, y, z) : x^2 + y^2 = R^2\}. \tag{158}$$

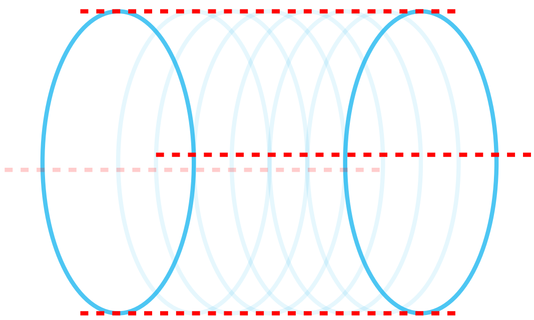
This is drawn below as:



One can think of this is the Cartesian product of the circle times a line:

$$\mathbb{S}^1_R \times \mathbb{R}, \tag{159}$$

where $\mathbb{S}^1_R$ denotes the circle of radius R. In other words, its the surface that results from taking the circle and take all lines that pass through the circle and are parallel to a given line:



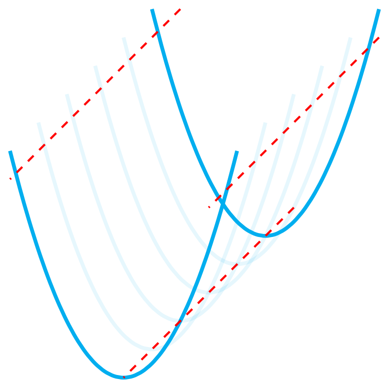
One can use this type of construction to consider more general objects called **generalised cylinders** (Stewart simply calls these cylinders but we will make the distinction). More precisely, a generalised cylinder cylinder is defined as a surface consisting of all the points on all the lines which are parallel to a given line and which pass through a fixed plane curve in a plane not parallel to the given line.

A **cylindric section** is the intersection of the cylinder with a plane. If we orientate the generalised cylinder such that the lines of the cylinder are parallel to one of the axes (say the z axis). Then the **cylindric section** that results from intersecting the generalised cylinder with the coordinate planes (or any parallel plane to the a coordinate plane) is called the **cross-section** or trace.

Example 6.2. *Consider the parabola* $\mathbb{P} = \{(x, y) \in \mathbb{R}^2 : y = x^2\}$*. Now take its Cartesian product with* $\mathbb{R}$ *to give*

$$\mathbb{P} \times \mathbb{R} = \{(x, y, z) \in \mathbb{R}^3 : y = x^2\}. \tag{160}$$

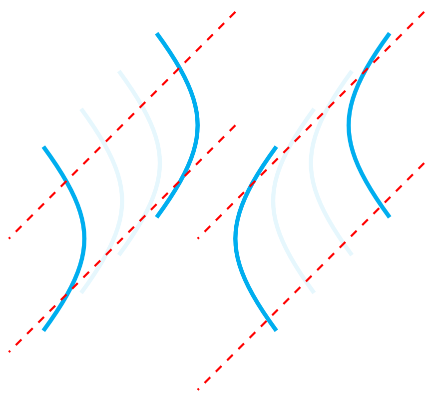
This is a parabolic cylinder, which we can draw as



Example 6.3. *Consider the hyperbola* $\mathbb{H} = \{(x, y) \in \mathbb{R}^2 : x^2 - y^2 = 1\}$*. Now take its Cartesian product with* $\mathbb{R}$ *to give*

$$\mathbb{H} \times \mathbb{R} = \{(x, y, z) \in \mathbb{R}^3 : x^2 - y^2 = 1\}. \tag{161}$$

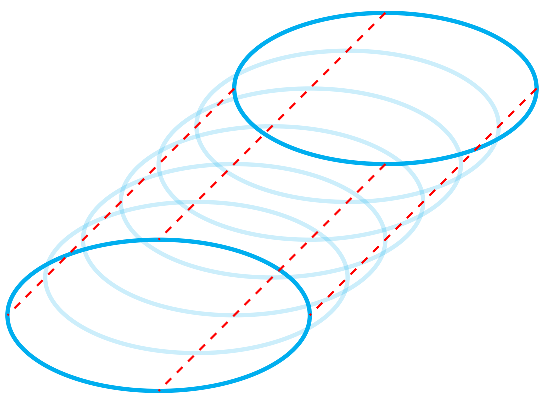
This is a hyperbolic cylinder, which we can draw as



Example 6.4. *Consider the ellipse* $\mathbb{E} = \{(x, y) \in \mathbb{R}^2 : \frac{x^2}{2} + y^2 = 1\}$*. Now take its Cartesian product with* $\mathbb{R}$ *to give*

$$\mathbb{E} \times \mathbb{R} = \{(x, y, z) \in \mathbb{R}^3 : \frac{x^2}{2} + y^2 = 1\}. \tag{162}$$

This is an elliptic cylinder, which we can draw as



6.6 Quadric Surfaces

A **quadric surface**, $\mathcal{Q}$, is a set of points in $\mathbb{R}^3$ which satisfy an equation of the following form:

$$P(x,y,z) \doteq Ax^2 + By^2 + Cz^2 + Dxy + Eyz + Fxz + Gx + Hy + Iz + J = 0, \tag{163}$$

where $A, ..., J$ are constants, i.e. $\mathcal{Q}$ is the zero set of a quadratic equation in 3 variables.[6] By a rotation or translation one can always bring the equation of the quadric into one of the two following 'standard forms':

$$Ax^2 + By^2 + Cz^2 + D = 0, \qquad Ax^2 + By^2 + Cz = 0, \tag{164}$$

where A, B, C, D are not necessarily those above. Compare these equations to the quadratic/ellipse/hyperbola equations:

$$y = ax^2, \qquad \frac{x^2}{a^2} + \frac{y^2}{b^2} = 1, \qquad \frac{x^2}{a^2} - \frac{y^2}{b^2} = 1. \tag{165}$$

One can consider these quadrics a generalisation of the conic sections.

In this course we are going to content ourselves with looking at the standard quadrics. These are listed below:

Ellipsoid	$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1$
Cone	$\frac{x^2}{a^2} + \frac{y^2}{b^2} = \frac{z^2}{c^2}$
Elliptic Paraboloid	$\frac{x^2}{a^2} + \frac{y^2}{b^2} = \frac{z}{c}$
Hyperboloid of One Sheet	$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 1$
Hyperboloid of Two Sheets	$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = -1$
Hyperbolic Paraboloid	$\frac{x^2}{a^2} - \frac{y^2}{b^2} = \frac{z}{c}$

We will come back to ploting these when we consider multivariable functions.

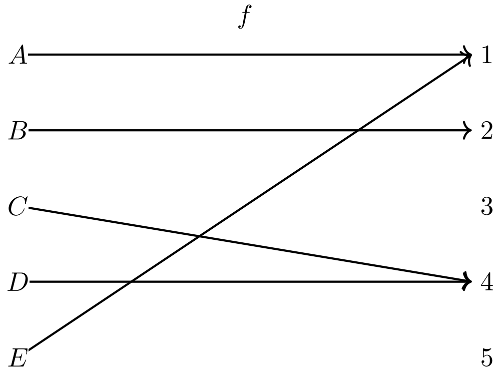
[6] In $\mathbb{R}^n$, $\mathcal{Q}$ is the zero set of a quadratic equation in n variables.

7 Vector-Valued Functions and Curves in Space

A function (or sometimes called a map) is a rule/assignment of each element of one set X to one element of another set Y. This is denoted $f : X \to Y$. The set X is often called the functions domain and Y is called the codomain. Often you will see these denoted $X = \mathrm{dom}(f)$ and $Y = \mathrm{codom}(f)$. The term range can be slightly ambiguous since some people use range to refer to the codomain of a function whilst others use it to mean the image, denoted $\mathrm{im}(f)$, of a function which is the set of all possible values in the codomain reached by f:

$$\mathrm{im}(f) \doteq \{f(x) : x \in \mathrm{dom}(f)\}. \tag{166}$$

Note that the image may not be the whole codomain as demonstrated by this picture:



Here the $\mathrm{dom}(f) = \{A, B, C, D, E\}$, the $\mathrm{im}(f) = \{1, 2, 4\}$ and $\mathrm{codom}(f) = \{1, 2, 3, 4, 5\}$. A functions domain can be specified as part of the definition of a function.

Thus far, you have probably studied functions that map some subset of the real line to another subset of the real line, i.e. $\mathrm{dom}(f) \subseteq \mathbb{R}$ and $\mathrm{im}(f) \subseteq \mathbb{R}$. Such functions typically have a **natural domain** or **domain of definition**,

$$\{x \in \mathbb{R} : f(x) \in \mathbb{R}\}, \tag{167}$$

which is the set of x on which the function is well-defined. In this course, unless otherwise stated, a functions domain will be its natural domain. Here are some examples of functions with various domains and the corresponding images:

$$f(x) = x^2 \qquad \mathrm{dom}(f) = \mathbb{R} \qquad \mathrm{im}(f) = [0, \infty) \tag{168}$$

$$f(x) = x^2 \qquad \mathrm{dom}(f) = [2, 3] \qquad \mathrm{im}(f) = [4, 9] \tag{169}$$

$$f(x) = \frac{1}{x} \qquad \mathrm{dom}(f) = (-\infty, 0) \cup (0, \infty) \qquad \mathrm{im}(f) = (-\infty, 0) \cup (0, \infty) \tag{170}$$

$$f(x) = \begin{cases} \frac{1}{x} & x \neq 0 \\ 0 & x = 0 \end{cases} \qquad \mathrm{dom}(f) = \mathbb{R} \qquad \mathrm{im}(f) = \mathbb{R}. \tag{171}$$

In this lecture we will study **vector-valued functions**. These are functions whose image is a subset of $\mathbb{R}^n$, i.e. the function outputs a vector with real valued components. For now we will consider functions that have a domain which is a subset of $\mathbb{R}$.[7] These functions take as an input some real number $x \in \mathrm{dom}(f)$ and output a vector:

$$\mathbf{f}(x) = (f_1(x), f_2(x), f_3(x)) = f_1(x)\mathbf{i} + f_2(x)\mathbf{j} + f_3(x)\mathbf{k}. \tag{172}$$

Note that f_1, f_2, f_3 are called the component functions of $\mathbf{f}$.

Example 7.1. *Consider the vector-valued function*

$$\mathbf{f}(x) = \Big(\frac{1}{x}, x^2, \sin(x)\Big). \tag{173}$$

Then its domain is $\mathbb{R} \setminus \{0\}$ *and its image is* $\mathbb{R} \setminus \{0\} \times \mathbb{R} \times [-1, 1]$.

[7] Later we will consider functions whose domain is a subset of $\mathbb{R}^m$, i.e multivariable functions.

7.1 Review of Limits for Real-Valued Functions

Recall the definition of limits for functions of single variables

Definition 7.1 (Limits of Scalar Functions)**.** *Let f be a function defined on some open interval $I = (a, b)$, except possibly at some point $c \in I$. Then one says that the limit of f as x tends to c is $L \in \mathbb{R}$ if for all $\epsilon > 0$, there exists a $\delta > 0$ such that if $0 < |x - c| < \delta$ and $x \in I$ implies $|f(x) - L| < \epsilon$. In this case one writes*

$$\lim_{x \to c} f(x) = L. \tag{174}$$

What this definition is saying is that L is the limit of f at c if $f(x)$ gets closer and closer to L as x is gets closer and closer to c.

Recall also the definition of one-sided limits:

Definition 7.2 (One-Sided Limits)**.** *Let f be a function defined on some open interval $I = (a, b)$, except possibly at some point $c \in I$. Then one says that the limit from below of f as x tends to c is $L^- \in \mathbb{R}$ if for all $\epsilon > 0$, there exists a $\delta > 0$ such that if $c - \delta < x < c$ then $|f(x) - L^-| < \epsilon$. In this case one writes*

$$\lim_{x \to c^-} f(x) = L^-. \tag{175}$$

Similarly, one says that the limit from above of f as x tends to c is $L^+ \in \mathbb{R}$ if for all $\epsilon > 0$, there exists a $\delta > 0$ such that if $c < x < c + \delta$ then $|f(x) - L^+| < \epsilon$. In this case one writes

$$\lim_{x \to c^+} f(x) = L^+. \tag{176}$$

Recall the useful fact that the limit of definition 7.3 exists if and only if both one sided limits exist and are equal, i.e.

$$\lim_{x \to c} f(x) = L \iff \lim_{x \to c^-} f(x) = L^- = L^+ = \lim_{x \to c^+} f(x). \tag{177}$$

This is often a good way to show the **limit does not exist**.

Example 7.2. *Define the* $\text{sign} : \mathbb{R} \to \mathbb{R}$ *function*

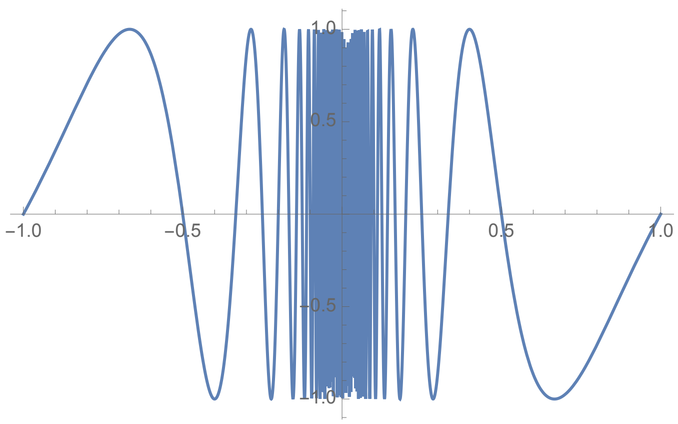
$$\text{sign}(x) = \begin{cases} 1 & x > 0 \\ 0 & x = 0 \\ -1 & x < 0 \end{cases}. \tag{178}$$

Then $\lim_{x \to 0^+} \text{sign}(x) = 1$, $\lim_{x \to 0^-} \text{sign}(x) = -1$. *So,* $\lim_{x \to 0} \text{sign}(x)$ *does not exist.*

What are other ways limits fail to exist? One can have a function that increasingly rapidly oscillates the closer you get to c and therefore it does not approach a fixed number. Such a function would be

$$f(x) = \sin\left(\frac{\pi}{x}\right), \tag{179}$$

which oscillates between 1 and -1 increasingly rapidly as $x \to 0$. This is plotted below:



To see directly that the limit does not exist consider the sequences $x_n = \frac{1}{2n}$ and $x'_n = \frac{1}{2n+\frac{1}{2}}$. Then

$$f(x_n) = 0 \quad \forall n \qquad f(x'_n) = 1 \quad \forall n. \tag{180}$$

Therefore, $\lim_{n\to\infty} f(x_n) = 0$ and $\lim_{n\to\infty} f(x'_n) = 1$ but $x_n \to 0$ and $x'_n \to 0$, i.e. one has a contradiction to $\epsilon = 1/2$ since $|f(x'_n)| = 1 > 1/2$ but n can always be chosen to satisfy $|x'_n| < \delta$ for any $\delta > 0$.

Additionally, a limits fail can to exist if the function diverges, i.e. the function values grow arbitrarily large. For example the function

$$f(x) = \frac{1}{x}. \tag{181}$$

One can introduce the notion of an infinite limit in this case (note we do not think of the limit as existing in this case since $\pm\infty$ is not a number, it is a symbol to denote that a function becomes unbounded):

Definition 7.3 (Infinite Limit of Single Variable Function)**.** *Let f be a function defined on some open interval $I = (a, b)$, except possibly at some point $c \in I$. Then one says that the limit of f as x tends to c is ∞ ($-\infty$) if for all $M > 0$, there exists a $\delta > 0$ such that if $0 < |x - c| < \delta$ then $f(x) > M$ ($f(x) < -M$ resp.). In this case one writes*

$$\lim_{x\to c} f(x) = \infty \qquad \left(\lim_{x\to c} f(x) = -\infty \quad \textit{resp.}\right) \tag{182}$$

Remark 7.1. *One can generalise the definition of infinite limit to one-sided limits.*

Recall some properties of limits:

Proposition 7.1 (Limit Properties)**.** *Suppose f and g are defined on $I = (a, b)$, except possibly at some point $c \in I$. Further suppose,*

$$\lim_{x\to c} f(x), \qquad \lim_{x\to c} g(x) \tag{183}$$

exist and $k \in \mathbb{R}$. Then

1. $\lim_{x\to c}(f(x) + g(x)) = \lim_{x\to c} f(x) + \lim_{x\to c} g(x)$.
2. $\lim_{x\to c}(f(x) - g(x)) = \lim_{x\to c} f(x) - \lim_{x\to c} g(x)$.
3. $\lim_{x\to c}(kf(x)) = k \lim_{x\to c} f(x)$.
4. $\lim_{x\to c}(f(x)g(x)) = \lim_{x\to c} f(x) \cdot \lim_{x\to c} g(x)$.
5. $\lim_{x\to c}(f(x)/g(x)) = \lim_{x\to c} f(x)/\lim_{x\to c} g(x)$ *if* $\lim_{x\to c} g(x) \neq 0$.

Proof. You can find a proof of these statements can be found in Stewart. □

Finally let's note some helpful theorems and tips for finding limits. The first is l'Hôpital's rule which is applicable in the '$0/0$' or the '∞/∞' situation:

Theorem 7.1 (L'Hôpital's Rule)**.** *Let I an open interval and $a \in I$. Suppose f and g are differentiable and $g'(x) \neq 0$ on I except possibly at a. Further suppose,*

$$\lim_{x\to a} f(x) = 0 \quad \textit{and} \quad \lim_{x\to a} g(x) = 0 \tag{184}$$

or

$$\lim_{x\to a} f(x) = \pm\infty \quad \textit{and} \quad \lim_{x\to a} g(x) = \pm\infty. \tag{185}$$

Then,

$$\lim_{x\to a} \frac{f(x)}{g(x)} = \lim_{x\to a} \frac{f'(x)}{g'(x)} \tag{186}$$

if the limit on the right-hand side exists.

Example 7.3. *Evaluate the limit*

$$\lim_{x \to 0} \frac{\sin(x)}{x}. \tag{187}$$

Using l'Hôpital's rule one has

$$\lim_{x \to 0} \frac{\sin(x)}{x} = \lim_{x \to 0} \frac{\cos(x)}{1} = 1. \tag{188}$$

Another useful theorem is the 'Squeeze Theorem':

Theorem 7.2 (Squeeze Theorem)**.** *Let I an open interval and $a \in I$. Suppose f, g and h are functions defined on I except possibly at a. Suppose, $f(x) \leq g(x) \leq h(x)$ when $x \neq a$ and*

$$\lim_{x \to a} f(x) = L = \lim_{x \to a} h(x). \tag{189}$$

Then,

$$\lim_{x \to a} g(x) = L. \tag{190}$$

Example 7.4. *Show that*

$$\lim_{t \to 0} t^2 \sin\left(\frac{1}{t}\right) = 0. \tag{191}$$

Above we noted that $\lim \sin\left(\frac{1}{t}\right)$ does not exist. So one can't just use the product of limits rule. However, for all $t \neq 0$, $\sin(\frac{1}{t})$ is bounded above by 1 and below by -1. So, for all $t \neq 0$,

$$-t^2 \leq t^2 \sin\left(\frac{1}{t}\right) \leq t^2. \tag{192}$$

Applying the squeeze theorem gives the result.

7.2 Review of Continuity for Real-Valued Functions

Lastly here is a brief review of continuity:

Definition 7.4 (Continuity)**.** *Let f be a function defined on some an open interval $I = (a, b)$ of the real line $\mathbb{R}$. Then one says that f* **is continuous at** *$c \in I$ if*

$$\lim_{x \to c} f(x) = f(c). \tag{193}$$

The function $f : I \to \mathbb{R}$ is continuous if it is continuous at every $c \in I$.

Note that this definition assumes that $c \in \mathrm{dom}(f)$, i.e. $f(c)$ exists. Additionally it assumes that the limit on the left-hand side exists and is equal to $f(c)$.[8]

Example 7.5. *Take the function $f(x) = x$, this has that $\lim_{x \to a} x = a$ and $f(a) = a$. Therefore, it is continuous at $a \in \mathbb{R}$. In fact $f : \mathbb{R} \to \mathbb{R}$ is continuous.*

On the other hand take:

$$f(x) = \begin{cases} 0 & x \in (-\infty, 0] \\ 1 & x = (0, \infty). \end{cases} \tag{194}$$

Then $f(0) = 0$ by definition but $\lim_{x \to 0^+} f(x) = 1$ and $\lim_{x \to 0^-} f(x) = 0$. So f is not continuous at 0. However, it is continuous everywhere else.

[8] One can unpack this in $\epsilon - \delta$ form, as follows: Let $I = (a, b) \subseteq \mathbb{R}$ and $c \in I$. Let $f : I \to \mathbb{R}$. Then we say that f is continuous at c if, for all $\epsilon > 0$, there exists a $\delta > 0$ such that for all $y \in I$ with $|y - c| < \delta$, we have $|f(y) - f(c)| < \epsilon$.

Finally let's note that the following proposition:

Proposition 7.2. *The following functions are continuous on all of* $\mathbb{R}$*:* e^t, $\sin$, $\cos$ *and polynomial functions. The function* $\ln(x)$ *is continuous on for* $x > 0$ *and a rational function*

$$R(x) = \frac{P(x)}{Q(x)} \tag{195}$$

is continuous on the set of values for which $Q(x) \neq 0$*.*

Example 7.6. *Where is*

$$f(x) \doteq \frac{t^2 \ln(t)}{t-1} \tag{196}$$

continuous?

It's continuous where its defined: $\ln$ *restricts this to* $t > 0$ *and* $\frac{1}{t-1}$ *restricts this to* $t \neq 1$*. So, its continuous on* $(0,1) \cup (1,\infty)$*.*

7.3 Vector-Valued Functions

One can give two definitions of limits for vector valued functions. Let's start with the $\epsilon - \delta$ definition:

Definition 7.5 (Limit of a Vector-Valued Function)**.** *Let* $\mathbf{f}$ *be a function defined on some open interval* $I = (a,b)$*, except possibly at some point* $c \in I$*. Then one says that the limit of* $\mathbf{f}$ *as* x *tends to* c *is* $\mathbf{L} \in \mathbb{R}^n$ *if for all* $\epsilon > 0$*, there exists a* $\delta > 0$ *such that if* $0 < |x-c| < \delta$ *and* $x \in I$ *implies* $||\mathbf{f}(x) - \mathbf{L}|| < \epsilon$*. In this case one writes*

$$\lim_{x \to c} \mathbf{f}(x) = \mathbf{L}. \tag{197}$$

Alternatively,

Definition 7.6 (Limit of a Vector-Valued Function)**.** *Let* $\mathbf{f}$ *be a function defined on some open interval* $I = (a,b)$*, except possibly at some point* $c \in I$*. Then the limit of* $\mathbf{f}$ *as* x *tends to* c *is defined as*

$$\lim_{x \to c} \mathbf{f}(x) = \Big(\lim_{x \to c} f_1(x), \lim_{x \to c} f_2(x), \lim_{x \to c} f_3(x) \Big), \tag{198}$$

provided the limits of the component functions exist.

Both of these are natural extensions of the definition of limits for scalar valued functions and are equivalent, which you should try to prove.

Continuity generalises naturally:

Definition 7.7 (Continuity)**.** *Let* $\mathbf{f}$ *be a vector-valued function defined on an interval* $I = (a,b)$ *of the real line* $\mathbb{R}$*. Then one says that* $\mathbf{f}$ **is continuous at** $c \in I$ *if*

$$\lim_{x \to c} \mathbf{f}(x) = \mathbf{f}(c). \tag{199}$$

The function $\mathbf{f} : I \to \mathbb{R}$ *is continuous if it is continuous at every* $c \in I$*.*

7.4 Application: Curves in Space

Suppose $\mathbf{r}$ is a continuous vector-valued function with domain $I = (a,b)$ and let $t \in I$. The set of points (x,y,z) where

$$x = r_1(t), \qquad y = r_2(t), \qquad z = r_3(t) \tag{200}$$

is a curve in $\mathbb{R}^3$. The equations in (200) are known as the parametric equations and t is called the parameter. You can think of of this curve being traced out by the tip of a vector from the origin to $(r_1(t), r_2(t), r_3(t))$.

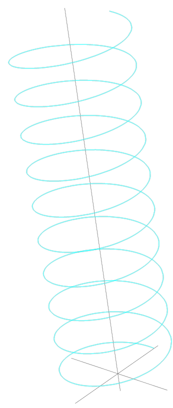
Example 7.7. *Let's sketch the curve*

$$\mathbf{r}(t) = \cos(t)\mathbf{i} + \sin(t)\mathbf{j} + t\mathbf{k}. \tag{201}$$

In the xy-plane the motion is circular since

$$x^2 + y^2 = \cos^2(t) + \sin^2(t) = 1. \tag{202}$$

However, the z component means that the one moves upwards with circular motion, i.e. one has a helix:



8 Derivatives and Integration of Vector-Valued Functions

8.1 Differentiation

Recall that the idea behind differentiability is the 'rate of change' of a function with respect to its variable(s). For functions of a single variable one has the following definition:

Definition 8.1 (Derivative)**.** *A function f of a single variable x is differentiable at $a \in \mathbb{R}$ if*

$$\lim_{h\to 0} \frac{f(a+h)-f(a)}{h} \tag{203}$$

exists. In this case, we define the derivative of f at a, denoted $f'(a)$ or ${}^{df}/{}_{dx}(a)$, as

$$f'(a) \doteq \lim_{h\to 0} \frac{f(a+h)-f(a)}{h} \quad \text{or} \quad \frac{df}{dx}(a) \doteq \lim_{h\to 0} \frac{f(a+h)-f(a)}{h} \tag{204}$$

One can extend this naturally to vector-valued functions:

Definition 8.2 (Derivative of a Vector-Valued Function)**.** *A vector-valued function $\mathbf{f}$ of a single variable x is differentiable at $a \in \mathbb{R}$ if*

$$\lim_{h\to 0} \frac{\mathbf{f}(a+h)-\mathbf{f}(a)}{h}, \tag{205}$$

exists. In this case, we define the derivative of $\mathbf{f}$ at a, denoted $\mathbf{f}'(a)$ or $\frac{d\mathbf{f}}{dx}(a)$, as

$$\mathbf{f}'(a) \doteq \lim_{h\to 0} \frac{\mathbf{f}(a+h)-\mathbf{f}(a)}{h} \quad \text{or} \quad \frac{d\mathbf{f}}{dx}(a) \doteq \lim_{h\to 0} \frac{\mathbf{f}(a+h)-\mathbf{f}(a)}{h}. \tag{206}$$

Recall that the limit of a vector-valued function can be defined through the limits of its components. Therefore, a vector-valued function is differentiable if and only if its components are differentiable, i.e. if $\mathbf{f} = (f_1, f_2, f_3)$ and f_1, f_2, f_3 are differentiable then

$$\frac{d\mathbf{f}}{dx}(a) = \Big(\lim_{h\to 0} \frac{f_1(a+h)-f_1(a)}{h}, \lim_{h\to 0} \frac{f_2(a+h)-f_2(a)}{h}, \lim_{h\to 0} \frac{f_3(a+h)-f_3(a)}{h}\Big), \tag{207}$$

which is the same as

$$\frac{d\mathbf{f}}{dx}(a) = \Big(\frac{df_1}{dx}(a), \frac{df_2}{dx}(a), \frac{df_3}{dx}(a)\Big) = \frac{df_1}{dx}(a)\mathbf{i} + \frac{df_2}{dx}(a)\mathbf{j} + \frac{df_3}{dx}(a)\mathbf{k}. \tag{208}$$

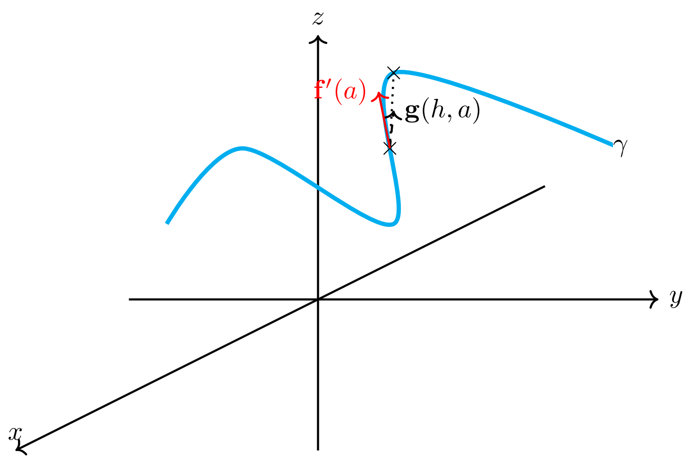
How do we visualise this this definition? Suppose we consider the curve γ defined by a vector-valued function $\mathbf{f}(t)$:

$$\gamma = \{\mathbf{x} = (x, y, z) \in \mathbb{R}^3 : \mathbf{x} = \mathbf{f}(t), t \in (a, b)\}. \tag{209}$$

Let p be the point with position vector $\mathbf{f}(a)$ and q be the point with position vector $\mathbf{f}(a+h)$. Then,

$$\mathbf{g}(h, a) \doteq \frac{1}{h}\big(\mathbf{f}(a+h) - \mathbf{f}(a)\big), \tag{210}$$

is a vector that points in the direction of the displacement vector $\overrightarrow{pq}$. As $h \to 0$, $\mathbf{g}(h, a)$ tends to a vector that is tangent to the curve defined by $\mathbf{f}$ at p. Therefore, $\mathbf{f}'(t)$ is often called the **tangent vector** to the curve defined by $\mathbf{f}$. As usual this is best draw as below:



Remark 8.1. *Just as for scalar functions the n^{th}-derivative of a vector-valued function (if it exists) is the derivative of the $(n-1)^{\text{th}}$ derivative (if it exists), i.e. the second derivative $\mathbf{f}'' = (\mathbf{f}')'$.*

Example 8.1. *Let $\mathbf{r}(t) = (7t^3, (1-\frac{1}{2}t)e^t, \cosh(t))$.*

1. *Find the derivative of $\mathbf{r}(t)$.*
2. *Find the unit tangent vector at the point with at $t = 0$.*

One can differentiate each component in turn

$$\mathbf{r}'(t) = \Big(21t^2, \frac{1}{2}(1-t)e^t, \sinh(t)\Big). \tag{211}$$

The tangent vector at the point with at $t = 0$ is $\mathbf{r}'(0) = (0, \frac{1}{2}, 0)$. The unit is $\widehat{\mathbf{r}'(0)} = (0, 1, 0)$.

Proposition 8.1. *Let c be a scalar and let h be a real valued function. Suppose the vector-valued functions $\mathbf{f}$ and $\mathbf{g}$ are differentiable in some interval (a, b). Then one has*

1. $\frac{d}{dx}(\mathbf{f}(x) + \mathbf{g}(x)) = \frac{d}{dx}\mathbf{f}(x) + \frac{d}{dx}\mathbf{g}(x)$.
2. $\frac{d}{dx}(c\mathbf{f}(x)) = c\frac{d}{dx}\mathbf{f}(x)$.
3. $\frac{d}{dx}(h(x)\mathbf{f}(x)) = \mathbf{f}(x)\frac{dh}{dx}(x) + h(x)\frac{d}{dx}\mathbf{f}(x)$.
4. $\frac{d}{dx}\langle \mathbf{f}(x), \mathbf{g}(x)\rangle = \langle \frac{d}{dx}\mathbf{f}(x), \mathbf{g}(x)\rangle + \langle \mathbf{f}(x), \frac{d}{dx}\mathbf{g}(x)\rangle$.
5. $\frac{d}{dx}(\mathbf{f}(x) \times \mathbf{g}(x)) = \frac{d}{dx}\mathbf{f}(x) \times \mathbf{g}(x) + \mathbf{f}(x) \times \frac{d}{dx}\mathbf{g}(x)$ *for $\mathbf{f}$ and $\mathbf{g}$ with image in $\mathbb{R}^3$.*
6. *the chain rule:* $\frac{d}{dx}\mathbf{f}(h(x)) = \frac{d\mathbf{f}}{dh}(h(x))\frac{dh}{dx}(x)$.

Proof. The proof of all these statements follows from writing the vector-valued function in terms of components and using the above properties for real-valued functions. □

Proposition 8.2. *Suppose $\mathbf{r}(t)$ has constant norm for all t, then $\frac{d}{dt}\mathbf{r}(t)$ is orthogonal to $\mathbf{r}(t)$ for all t.*

Proof. If $||\mathbf{r}(t)|| = k$ for $k \in \mathbb{R}$ then

$$\langle \mathbf{r}(t), \mathbf{r}(t)\rangle = ||\mathbf{r}(t)||^2 = k^2. \tag{212}$$

Therefore, using proposition 8.1 one has

$$\langle \mathbf{r}'(t), \mathbf{r}(t)\rangle + \langle \mathbf{r}(t), \mathbf{r}'(t)\rangle = 0. \tag{213}$$

The symmetry of the scalar product then gives

$$2\langle \mathbf{r}'(t), \mathbf{r}(t)\rangle = 0, \tag{214}$$

or in other words $\mathbf{r}$ and $\mathbf{r}'$ are orthogonal. □

8.2 Integration

Definition 8.3. *Let $\mathbf{f}(t) = (f_1(t), f_2(t), f_3(t))$ where $f_1(t)$, $f_2(t)$ and $f_3(t)$ are continuous real valued functions on the interval $[a, b]$. Then we define the integral of $\mathbf{f}$ as the vector*

$$\int_a^b \mathbf{f}(x)dx = \Big(\int_a^b f_1(x)dx\Big)\mathbf{i} + \Big(\int_a^b f_2(x)dx\Big)\mathbf{j} + \Big(\int_a^b f_3(x)dx\Big)\mathbf{k}. \tag{215}$$

We now extend the fundamental theorem of calculus to vector-valued functions:

Theorem 8.1. *Let $\mathbf{f} : [a,b] \to \mathbb{R}^n$ be continuous. For $x \in [a,b]$ define the* **antiderivative**

$$\mathbf{F}(x) \doteq \int_a^x \mathbf{f}(x')dx'. \tag{216}$$

Then $\mathbf{F}$ is continuous on $[a,b]$ and differentiable with $\mathbf{F}'(x) = \mathbf{f}(x)$ for every $x \in (a,b)$.

Remark 8.2. *Note that*

$$\mathbf{F}(b) - \mathbf{F}(a) = \int_a^b \mathbf{f}(x')dx'. \tag{217}$$

Remark 8.3. *Often we will use the notation $\int \mathbf{f}(x)dx$ for an indefinite integral.*

Example 8.2. *Let $\mathbf{r}(t) = (7t^3, (1 - \frac{1}{2}t)e^t, \cosh(t))$.*

1. *Find the indefinite integral of $\mathbf{r}(t)$.*
2. *Find the definite integral from 0 to 1.*

The indefinite integral is computed component-wise but one now must add a vector of constants $\mathbf{c} = (c_1, c_2, c_3)$:

$$\mathbf{F}(x) = \frac{7}{4}x^4\mathbf{i} + \Big(\frac{3}{2} - \frac{1}{2}t\Big)e^t\mathbf{j} + \sinh(t)\mathbf{k} + \mathbf{c}, \tag{218}$$

where we've used integration by part on $(1 - {}^1\!/_2t)e^t$. The definite integral is

$$\mathbf{F}(1) - \mathbf{F}(0) = \frac{7}{4}\mathbf{i} + \Big(e - \frac{3}{2}\Big)\mathbf{j} + \sinh(1)\mathbf{k}. \tag{219}$$

8.3 Application: Motion in $\mathbb{R}^3$

Suppose $\mathbf{r}$ is a continuous vector-valued function with domain $I = (a,b)$ and let $t \in I$. Suppose $\mathbf{r}$ models the motion of a particle in space, i.e. its position vector. If $\mathbf{r}(t)$ is differentiable, the particles velocity at t, $\mathbf{v}(t)$, is the first derivative of $\mathbf{r}$, i.e.

$$\mathbf{v}(t) = \mathbf{r}'(t). \tag{220}$$

The particles **speed** is the norm of its velocity,

$$s(t) = ||\mathbf{v}(t)||. \tag{221}$$

If $\mathbf{r}(t)$ is twice differentiable, the particles acceleration at t, $\mathbf{a}(t)$, is the second derivative of $\mathbf{r}$, i.e.

$$\mathbf{a}(t) = \mathbf{v}'(t). \tag{222}$$

Let's do some examples:

Example 8.3. *The helix:*

$$\mathbf{r}(t) = \cos(t)\mathbf{i} + \sin(t)\mathbf{j} + t\mathbf{k}. \tag{223}$$

Let's compute $\mathbf{v}(t)$ and $\mathbf{a}(t)$:

$$\mathbf{v}(t) = \mathbf{r}'(t) = -\sin(t)\mathbf{i} + \cos(t)\mathbf{j} + \mathbf{k}, \tag{224}$$
$$\mathbf{a}(t) = \mathbf{r}''(t) = -\cos(t)\mathbf{i} - \sin(t)\mathbf{j}. \tag{225}$$

Example 8.4. *Suppose a particle has acceleration vector*

$$\mathbf{a}(t) = (3t, -2t, 1), \tag{226}$$

with initial velocity $\mathbf{v}(0) = \mathbf{i} + \mathbf{j} - \mathbf{k}$ and position $\mathbf{r}(0) = (0,1,0)$. Find its velocity as a function of t.

9 Multivariable Functions I: Introduction and Limits

9.1 Introduction to Multivariable Functions

In lecture 7, we discussed domains, codomains and images of functions. We want to consider 'multivariable functions', i.e. functions that take more than one input.

Definition 9.1. *Let D be a subset of $\mathbb{R}^n$ then a (real-valued) multivariable function f is a rule that assigns to each ordered n-tuple $(x_1, ..., x_n)$ a real number denoted $f(x_1, ..., x_n)$. In other words, the domain of f, $\mathrm{dom}(f) = D$, is a subset of $\mathbb{R}^n$ and the image of $\mathrm{im}(f)$ is a subset of $\mathbb{R}$. In notation,*

$$f : D \to \mathbb{R}. \tag{227}$$

For us we are going to restrict $n \leq 3$, i.e. functions of two variables $f(x, y)$ and functions of three variables $f(x, y, z)$. Note that we are not going to allow for *vector-valued* multivariable functions, i.e. functions with domain in $\mathbb{R}^n$ *and* codomain $\mathbb{R}^m$ for $m > 1$. All functions will map to a subset of the real numbers ($m = 1$). Let's do some examples:

Example 9.1. *Define $f : D \subseteq \mathbb{R}^2 \to \mathbb{R}$ by*

$$f(x, y) = x^2 + y^2. \tag{228}$$

One could pick $D = [1, 2] \times [1, 10]$. However, its domain of definition is $\mathbb{R}^2$.

Example 9.2. *Define $f : D \subseteq \mathbb{R}^2 \to \mathbb{R}$ by*

$$f(x, y) = \frac{1}{x^2 + y^2}. \tag{229}$$

This function is well-defined as long as its denominator does not vanish, i.e. as long as $x \neq y \neq 0$. So its domain of definition is $\mathbb{R}^2 \setminus \{(0, 0)\}$.

Example 9.3. *Define $f : D \subseteq \mathbb{R}^2 \to \mathbb{R}$ by*

$$f(x, y) = \frac{1}{x^2 - y^2}. \tag{230}$$

This function is well-defined as long as its denominator does not vanish, i.e. as long as $x^2 - y^2 \neq 0$. One can factorise to find the set of ill-definition:

$$(x - y)(x + y) = 0 \iff x = y, \quad x = -y. \tag{231}$$

So its domain of definition is

$$\{(x, y, z) \in \mathbb{R}^3 : x \neq y, x \neq -y\}. \tag{232}$$

Example 9.4. *Define $f : D \subseteq \mathbb{R}^2 \to \mathbb{R}$ by*

$$f(x, y) = \log(x - y). \tag{233}$$

The natural logarithm is well-defined for it's argument in $(0, \infty)$. Therefore, for $f(x, y)$ to be well-defined $x - y > 0$. So, its domain of definition is

$$\{(x, y, z) \in \mathbb{R}^3 : x - y > 0\}. \tag{234}$$

Example 9.5. *Define $f : D \subseteq \mathbb{R}^3 \to \mathbb{R}$ by*

$$f(x, y, z) = \frac{\sqrt{y}}{z} e^{-x^2}. \tag{235}$$

The exponential function and $-x^2$ are well-defined on all of $\mathbb{R}$. The only issues appearing is when $z = 0$ and when $y < 0$ (due to the square-root). So its domain of definition is the set

$$\{(x, y, z) \in \mathbb{R}^3 : y \geq 0\} \setminus \{(x, y, 0)\} \tag{236}$$

(i.e. $\mathbb{R}^3$ without a the $z = 0$ plane and restricted to $y \geq 0$).

There are two groups of functions that are very common:

Definition 9.2 (Polynomial/Rational Functions)**.** *A polynomial function of n-variables $(x_1, ..., x_n)$ is a (finite) sum of terms of the form*

$$cx_1^{m_1}...x_n^{m_n}, \tag{237}$$

for $m_1, ..., m_n \in \mathbb{N}_0$ (the natural numbers including zero, $\mathbb{N}_0 = \{0, 1, 2, 3, 4, ...\}$).

A rational function is a ratio of two polynomials.

Example 9.6. *Two examples of polynomials of two variables are*

$$P(x, y) = x^5 - y^2 + 7, \qquad Q(x, y) = x^7y^3 + y^5 + xy - 3y - 1. \tag{238}$$

An example of a rational function would be the ratio of P and Q,

$$R(x, y) = \frac{P(x, y)}{Q(x, y)} = \frac{x^5 - y^2 + 7}{x^7y^3 + y^5 + xy - 3y - 1}. \tag{239}$$

9.2 Drawing Multivariable Functions: Graphs

To visualise and draw functions we need to extend the idea of graphs:

Definition 9.3. *Let $f : \mathrm{dom}(f) \subseteq \mathbb{R}^n \to \mathbb{R}$ and denote $\mathbf{x} = (x_1, ..., x_n)$ then the graph of f, $\mathrm{Graph}(f)$, is the set*

$$\{(x_1, ..., x_{n+1}) \in \mathbb{R}^{n+1} : x_{n+1} = f(\mathbf{x}), \mathbf{x} \in \mathrm{dom}(f)\} = \{(\mathbf{x}, f(\mathbf{x})) : \mathbf{x} \in \mathrm{dom}(f)\}. \tag{240}$$

Note that this is a subset of $\mathbb{R}^{n+1}$, which in general is not within our capability to visualise. However, if $n \leq 2$ this is possible, i.e. functions of two variables are still within our capability to visualise. These sets can be drawn as surfaces in $\mathbb{R}^3$. Again, lets do some examples

Example 9.7. *Sketch the graph of the function $f(x, y) = 1 - 7x - y$.*

The graph of f is

$$\mathrm{Graph}(f) = \{(x, y, f(x, y)) : (x, y) \in \mathrm{dom}(f)\}. \tag{241}$$

The domain of definition for f is $(x, y) \in \mathbb{R}^2$. The graph is determined by $z = 1 - 7x - y$, which is the equation of a plane. Let's find three points on this plane:

$$z = 0 = y \implies x = \frac{1}{7} \tag{242}$$

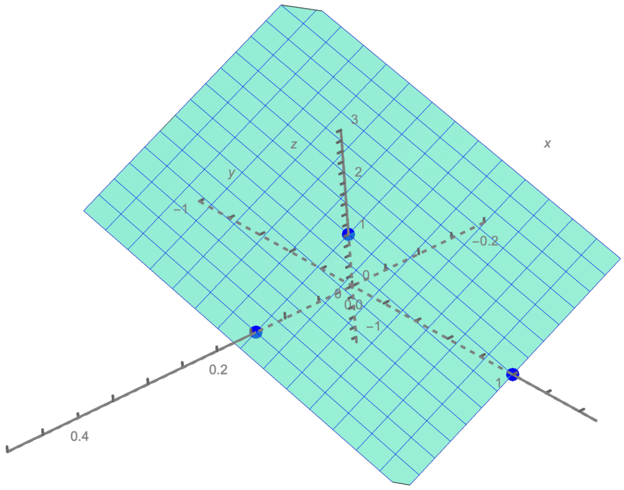
$$z = 0 = x \implies y = 1 \tag{243}$$

$$x = 0 = y \implies z = 1 \tag{244}$$

Therefore, we have three points in $\mathbb{R}^3$ in the plane:

$$\mathbf{x}_1 = (1/7, 0, 0), \qquad \mathbf{x}_2 = (0, 1, 0), \qquad \mathbf{x}_3 = (0, 0, 1). \tag{245}$$

The plot of the graph is then:



Let's look at some of the standard quadrics from lecture 6. Note that sometimes one cannot write them as the graph of a single function but two functions will suffice if necessary.

Example 9.8. *Sketch the graph of the functions $f_{\pm}(x,y) = \pm\sqrt{1 - x^2 - \frac{y^2}{10}}$.*

The graph of $f_{\pm}$ is

$$\mathrm{Graph}(f_{\pm}) = \{(x, y, f_{\pm}(x,y)) : (x,y) \in \mathrm{dom}(f_{\pm})\}. \tag{246}$$

The domain of definition for $f_{\pm}$ is $\{(x,y) \in \mathbb{R}^2 : 1 - x^2 - \frac{y^2}{10} \geq 0\}$, which is the interior (including the boundary) of the ellipse:

$$x^2 + \frac{y^2}{10} = 1. \tag{247}$$

The graph of $f_{\pm}$ is determined by $z = \pm\sqrt{1 - x^2 - \frac{y^2}{10}}$. Let's find some points on this surface:

$$x = 1 \implies y = 0, z = 0 \tag{248}$$

$$y = \sqrt{10} \implies x = 0, z = 0 \tag{249}$$

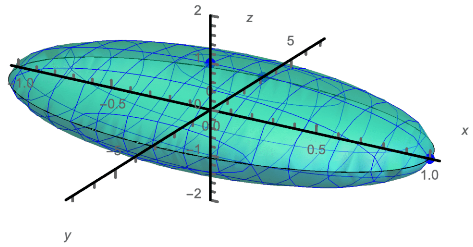
$$z = 0 \implies x^2 + \frac{y^2}{10} = 1 \tag{250}$$

$$x = 0 = y \implies z = \pm 1. \tag{251}$$

Take $z = f_{\pm} = k_{\pm} = const.$ with $0 < k_+ < 1$, $-1 < k_- < 0$ then,

$$\frac{x^2}{1 - k_{\pm}^2} + \frac{y^2}{10(1 - k_{\pm}^2)} = 1, \tag{252}$$

which is the equation of an ellipse. So, for k_+ increasing from 0 to 1 we have a smaller and smaller ellipse at each $z = k_+$. Similarly, for k_- decreasing from 0 to -1 we have a smaller and smaller ellipse at each $z = k_-$. The plot of the combined graphs of $f_{\pm}$ is then:



This is an ellipsoid.

Example 9.9. *Sketch the graph of the functions* $f(x,y) = x^2 + \frac{y^2}{10}$.

The graph of f *is*

$$\mathrm{Graph}(f) = \{(x, y, f(x,y)) : (x,y) \in \mathrm{dom}(f)\}. \tag{253}$$

The domain of definition for f *is* $\mathbb{R}^2$. *The graph of* f *is determined by* $z = x^2 + \frac{y^2}{10}$. *Note that* $z \geq 0$ *and at* $z = 0$ $x = 0 = y$. *Take* $z = f = k = const.$ *with* $0 < k$, *then,*

$$\frac{x^2}{k} + \frac{y^2}{10k} = 1, \tag{254}$$

which is the equation of an ellipse. So, for k *increasing from* 0 *we have a larger and larger ellipse at each* $z = k$.

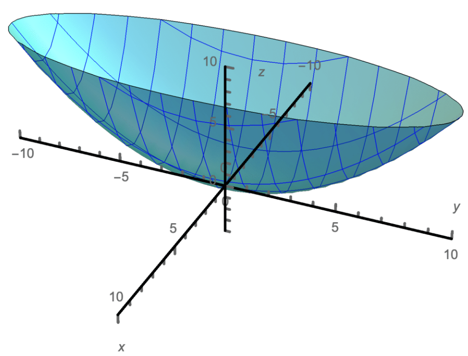
If we take $y = k$ *then*

$$z = \frac{k^2}{10} + x^2 \tag{255}$$

which is the equation of a parabola with variable x. *Similarly, if* $x = k$ *then*

$$z = k^2 + \frac{y^2}{10} \tag{256}$$

which is also a parabola in y. *The plot of the graphs of* f *is then:*



This is an elliptic paraboloid.

Remark 9.1. *The process employed above of considering $z = k$, $y = k$, $x = k$ means we are looking at intersections of the surface with planes parallel to the coordinate planes xy, zx and yz. The curves that result from such intersections are called* **cross-sections**.

Example 9.10. *Sketch the graph of the function $f(x, y) = x^2 - \frac{y^2}{10}$.*

The graph of f is

$$\mathrm{Graph}(f) = \{(x, y, f(x, y)) : (x, y) \in \mathrm{dom}(f)\}. \tag{257}$$

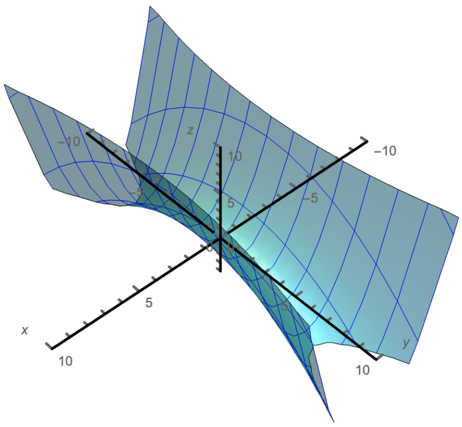
The domain of definition for f is $\mathbb{R}^2$. The graph of f is determined by $z = x^2 - \frac{y^2}{10}$. Take $z = f = k = const.$ with $0 < k$, then,

$$\frac{x^2}{k} - \frac{y^2}{10k} = 1, \tag{258}$$

which is the equation of a hyperbola. Take $z = f = k = const.$ with $0 > k$, then,

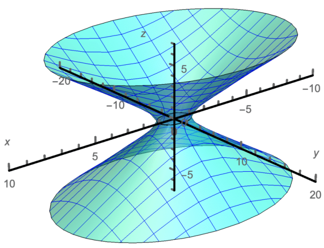
$$\frac{y^2}{10|k|} - \frac{x^2}{|k|} = 1, \tag{259}$$

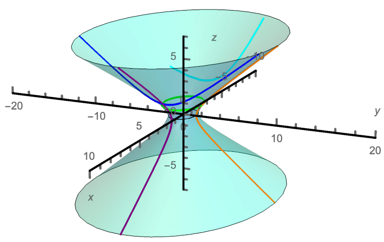
which is also the equation of the conjugate hyperbola. So, for k increasing from 0 we have hyperbola at each $z = k$ and for k decreasing from 0 we have the conjugate hyperbola at each $z = k$. If one looks at $x = k$ and $y = k$ one will find parabola. The plot of the graphs of f is then:



This is an hyperbolic paraboloid.

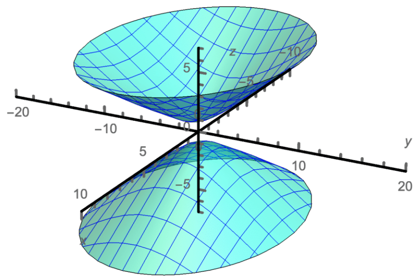
Example 9.11. *Sketch the graph of the functions $f_{\pm}(x, y) = \pm\sqrt{-1 + x^2 + \frac{y^2}{2}}$..*





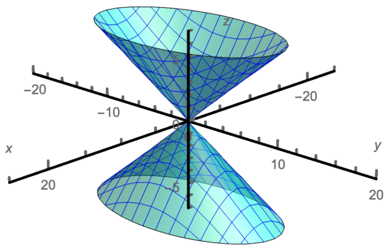
This is an hyperbola of one sheet.

Example 9.12. *Sketch the graph of the functions* $f_{\pm}(x, y) = \pm\sqrt{1 + x^2 + \frac{y^2}{2}}$.



This is an hyperbola of two sheets.

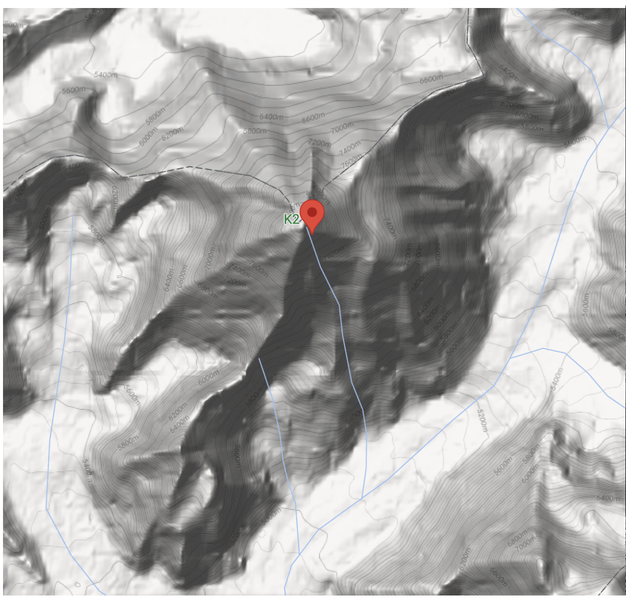
Example 9.13. *Sketch the graph of the functions* $f_{\pm}(x, y) = \pm\sqrt{x^2 + \frac{y^2}{2}}$.



This is a (elliptic) cone.

9.3 Drawing Multivariable Functions: Level Sets/Surfaces/Curves

So far we have been ploting graphs of functions. There is an alternative: level sets or curves. You've probably encountered such drawings in maps: contour/topographic maps are drawings of level curves of some region on Earth. Here is the topographic map of K2 (left) vs the plot/picture of its graph (right):



The topographic map marks different contant height levels on the map with the grey lines: these are the level sets or level curves.

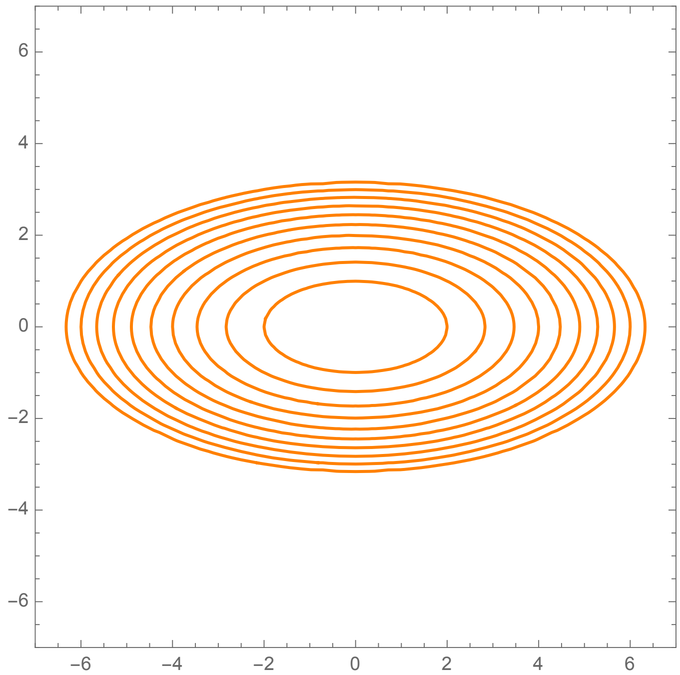
Definition 9.4. *The level sets of a function f of n-variables $(x_1, ..., x_n)$ are the sets*

$$\{(x_1, ..., x_n) : f(x_1, ..., x_n) = k\} \tag{260}$$

for k constant. If f is a function of two variables we call these level curves. If f is a function of three variables we call these level surfaces.

Example 9.14. *Sketch the level curves of the function $f(x, y) = \frac{x^2}{4} + y^2$.*

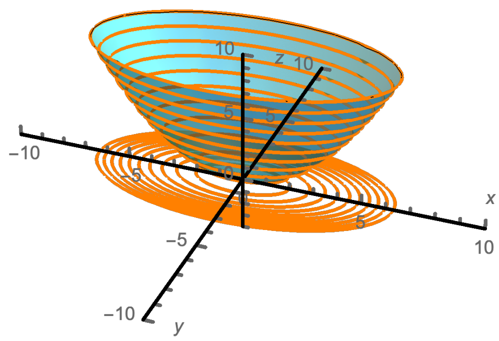
We look at the curves $\frac{x^2}{4} + y^2 = k$ for k constant. For solution $k \geq 0$ so, these are ellipses, which are plotted as follows:



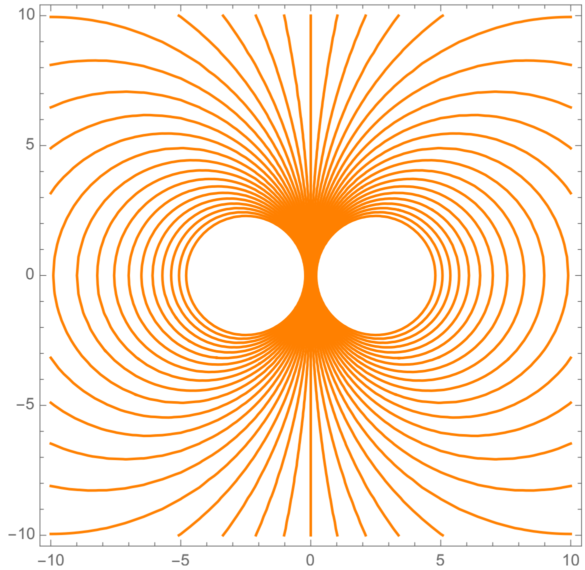
Lets view k as a z coordinate then $f(x, y) = k$ can be rewritten as

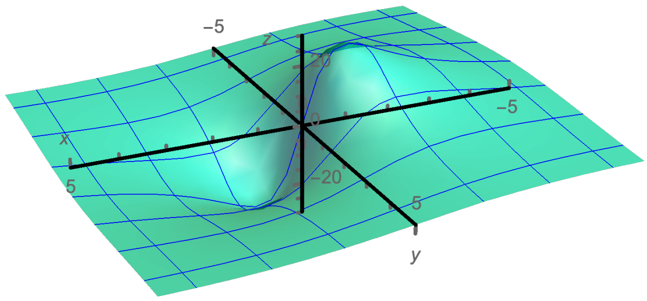
$$\frac{x^2}{4} + y^2 - z = 0 \tag{261}$$

which is the equation of a elliptic paraboloid. One can view the contour plot in the plane as a projection of the curves resulting from intersection the elliptic paraboloid with planes $\{z = k\}$, i.e. the z cross-sections. This is drawn below:

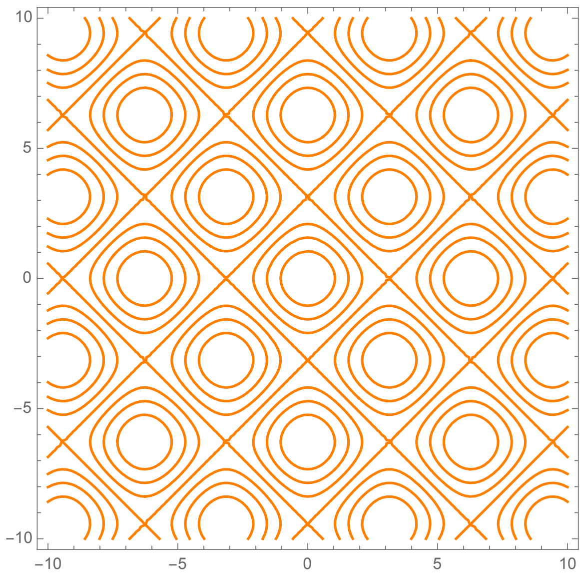


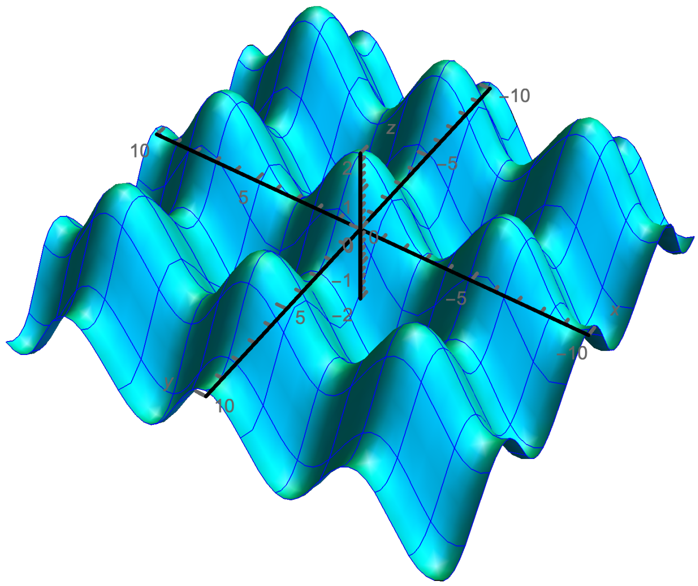
Example 9.15. *The following is a computer generated sketch of the level curves of the function* $f(x,y) = -\frac{50x}{x^2+y^2+1}$ *with the corresponding surface plot of* $(x, y, f(x,y))$*:*





Example 9.16. *The following is a computer generated sketch of the level curves of the function* $f(x,y) = \cos(x) + \cos(y)$ *with the corresponding surface plot of* $(x, y, f(x,y))$*:*





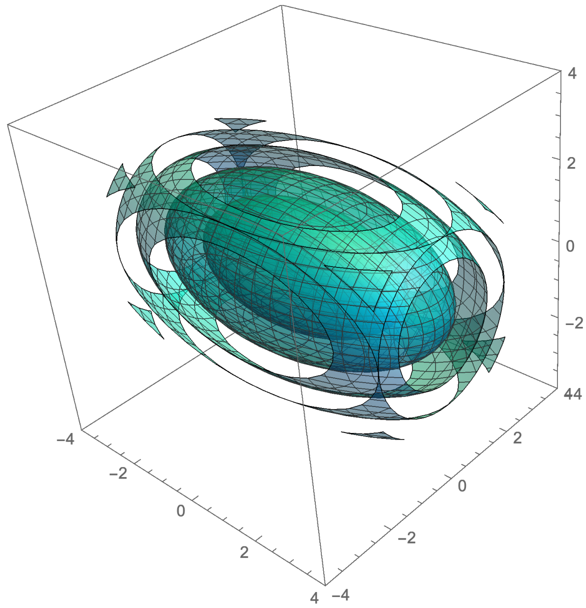
Example 9.17. *A simple example in* $3D$ *would be the level surfaces of*

$$f(x,y,z) = \frac{x^2}{4} + y^2 + z^2. \tag{262}$$

We cannot visualise the graph $\{(x,y,z,f(x,y,z))\}$ *since this requires* 4*-dimensions. However we can look at* $f(x,y,z) = k$ *for* k *constant. This gives,*

$$\frac{x^2}{4k} + \frac{y^2}{k} + \frac{z^2}{k} = 1 \tag{263}$$

for $k > 0$ *(which is nessecary otherwise* $x = y = z = 0$*). This gives concentric ellipsoids:*



9.4 Limits of Multivariable Functions

We want to talk about how functions behave as their variables approach certain values. We need to generalise our notion of limit:

Definition 9.5. *Let* $f : \mathrm{dom}(f) \to \mathbb{R}$ *be a function and* $\mathrm{dom}(f)$ *include all points arbitrarily close to* $\mathbf{a} = (a_1, ..., a_n)$*, i.e. for any* $\tilde{\epsilon} > 0$*, there exists a* $\mathbf{x} = (x_1, ..., x_n) \in \mathrm{dom}(f)$ *such that* $||\mathbf{x} - \mathbf{a}|| < \tilde{\epsilon}$*. We say that the limit of* f *as* $\mathbf{x}$ *goes to* $\mathbf{a}$ *is* L *if for all* $\epsilon > 0$*, there exists a* $\delta > 0$ *such that if*

$$0 < ||\mathbf{x} - \mathbf{a}|| < \delta \tag{264}$$

and $\mathbf{x} \in \mathrm{dom}(f)$ *then*

$$|f(\mathbf{x}) - L| < \epsilon \tag{265}$$

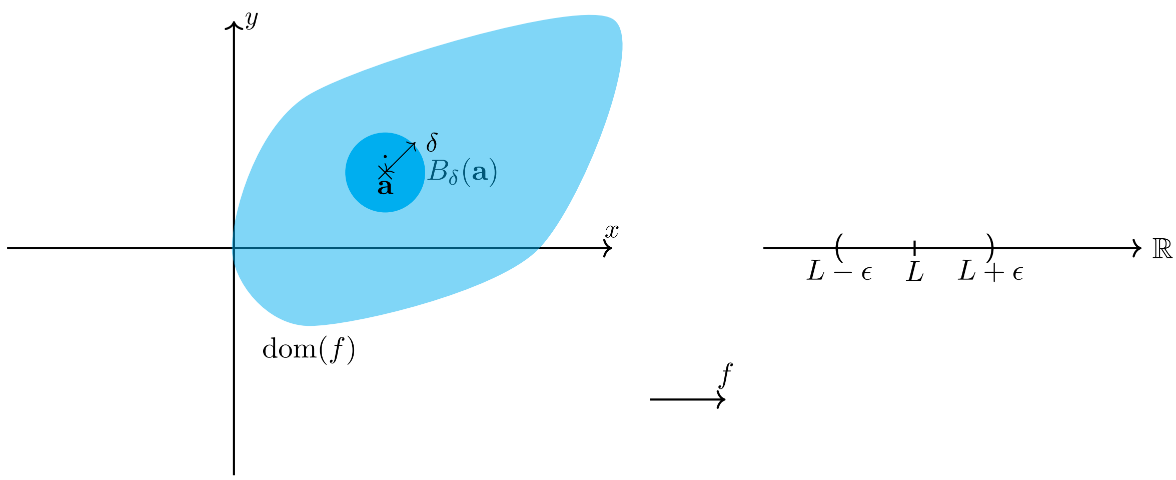
In notation, one writes

$$\lim_{\mathbf{x} \to \mathbf{a}} f(\mathbf{x}) = L \tag{266}$$

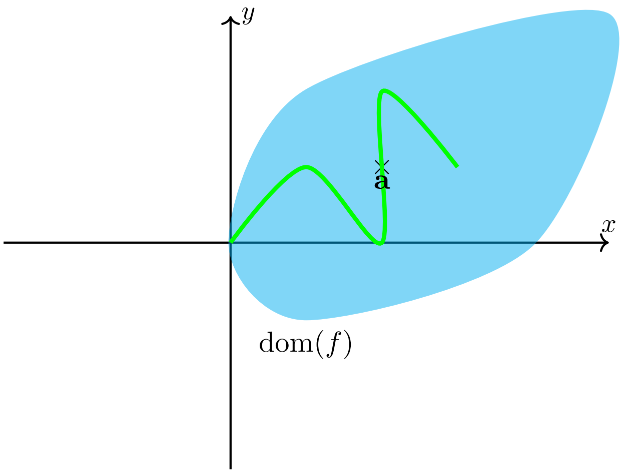
What this says is that given a small interval around L, $I = (L - \epsilon, L + \epsilon)$, then one can find a small (open) ball/disk of radius δ centered at $\mathbf{a}$,

$$B_\delta(\mathbf{a}) \doteq \{\mathbf{x} \in \mathbb{R}^n : ||\mathbf{x} - \mathbf{a}|| < \delta\}, \tag{267}$$

which f maps into I (except possibly $\mathbf{a}$). This is pictured for $\mathbb{R}^2$ below:



When we discussed limits for a function f of single variables we recalled that for the limit $\lim_{x\to a} f(x)$ to exist the above and below limits must exist. There we were constrained to a line to approach the point for which we wanted to investigate the limit of f. In this situation the analogous statement is that approaching $\mathbf{a}$ on **any curve** should give the same limit. **Be careful**, this does not mean that if the limits agree on every line through a point then that is the limit, it must also agree on a parabolic curve, a hyperbolic curve or any weird path you may choose (as long as it lies in the domain of definition). This is illustrated below:



Picking various paths can be very useful for showing that limits do not exist.

9.4.1 Limits Not Existing

Example 9.18. *Show that*

$$\lim_{\mathbf{x}\to\mathbf{0}} \frac{x^2-y^2}{x^2+y^2} \tag{268}$$

does not exist.

Take a path along the x-axis, i.e. set $y=0$, $x\neq 0$, then

$$\lim_{\mathbf{x}\to\mathbf{0}} \frac{x^2-y^2}{x^2+y^2} = 1, \tag{269}$$

along the x-axis. Take a path along the y-axis, i.e. set $x=0$, $y\neq 0$, then

$$\lim_{\mathbf{x}\to\mathbf{0}} \frac{x^2-y^2}{x^2+y^2} = -1, \tag{270}$$

along the y-axis. This is a contradiction.

Example 9.19. *Show that*

$$\lim_{\mathbf{x}\to\mathbf{0}} \frac{xy}{x^2+y^2} \tag{271}$$

does not exist.

Take a path along the x-axis, i.e. set $y=0$, $x\neq 0$, then

$$\lim_{\mathbf{x}\to\mathbf{0}} \frac{xy}{x^2+y^2} = 0, \tag{272}$$

along the x-axis. Take a path along the y-axis, i.e. set $x=0$, $y\neq 0$, then

$$\lim_{\mathbf{x}\to\mathbf{0}} \frac{xy}{x^2+y^2} = 0, \tag{273}$$

along the y-axis. So far, there is no contradiction. We can now look at other lines through $(0,0)$.

Take a path along the the line $y=x$ then

$$\lim_{\mathbf{x}\to\mathbf{0}} \frac{xy}{x^2+y^2} = \frac{1}{2}, \tag{274}$$

along the $y=x$ line. This gives us a contradiction.

Example 9.20. *Show that*

$$\lim_{\mathbf{x}\to(1,1)} \frac{y-x}{1-y+\ln(x)} \tag{275}$$

does not exist.

First take the path along $y=x$ then

$$\lim_{\mathbf{x}\to(1,1)} \frac{y-x}{1-y+\ln(x)} = \lim_{x\to 1} \frac{0}{1-x+\ln(x)} = 0. \tag{276}$$

Consider now the path $y=1$, then

$$\lim_{\mathbf{x}\to(1,1)} \frac{y-x}{1-y+\ln(x)} = \lim_{x\to 1} \frac{1-x}{\ln(x)} = \lim_{x\to 1} \frac{\frac{d}{dx}(1-x)}{\frac{d}{dx}\ln(x)} = \lim_{x\to 1}(-x) = -1, \tag{277}$$

where the second equality is by L'Hôpital's rule. This gives the desired contradiction.

9.5 Properties of Limits

Proposition 9.1 (Limit Properties)**.** *Suppose f and g are defined on $D \subseteq \mathbb{R}^n$, except possibly at some point $\mathbf{a} \in D$. Further suppose,*

$$\lim_{\mathbf{x}\to\mathbf{a}} f(\mathbf{x}), \qquad \lim_{\mathbf{x}\to\mathbf{a}} g(\mathbf{x}) \tag{278}$$

exist and $k \in \mathbb{R}$. Then

1. $\lim_{\mathbf{x}\to\mathbf{a}}(f(\mathbf{x})+g(\mathbf{x})) = \lim_{\mathbf{x}\to\mathbf{a}} f(x) + \lim_{\mathbf{x}\to\mathbf{a}} g(\mathbf{x})$.
2. $\lim_{\mathbf{x}\to\mathbf{a}}(f(\mathbf{x})-g(\mathbf{x})) = \lim_{\mathbf{x}\to\mathbf{a}} f(\mathbf{x}) - \lim_{\mathbf{x}\to\mathbf{a}} g(\mathbf{x})$.
3. $\lim_{\mathbf{x}\to\mathbf{a}}(kf(\mathbf{x})) = k\lim_{\mathbf{x}\to\mathbf{a}} f(\mathbf{x})$.
4. $\lim_{\mathbf{x}\to\mathbf{a}}(f(\mathbf{x})g(\mathbf{x})) = \lim_{\mathbf{x}\to\mathbf{a}} f(\mathbf{x}) \cdot \lim_{\mathbf{x}\to\mathbf{a}} g(\mathbf{x})$.
5. $\lim_{\mathbf{x}\to\mathbf{a}}({}^{f(\mathbf{x})}/_{g(\mathbf{x})}) = {}^{\lim_{\mathbf{x}\to\mathbf{a}} f(\mathbf{x})}/_{\lim_{\mathbf{x}\to\mathbf{a}} g(\mathbf{x})}$ *if* $\lim_{\mathbf{x}\to\mathbf{a}} g(\mathbf{x}) \neq 0$.

Under the (somewhat obvious) conclusions that, for $\mathbf{x}, \mathbf{a} \in \mathbb{R}^n$,

$$\lim_{\mathbf{x}\to\mathbf{a}} x_i = a_i, \qquad \lim_{\mathbf{x}\to\mathbf{a}} c = c, \tag{279}$$

one can see that limit laws imply that limit of any polynomial P can be evaluated by direct substitution:

$$\lim_{\mathbf{x}\to\mathbf{a}} P(\mathbf{x}) = P(\mathbf{a}). \tag{280}$$

Similarly, limit laws imply that limit of any rational function $R = {}^{P}/_{Q}$ can be evaluated by direct substitution with the caveat that the point $\mathbf{x}=\mathbf{a}$ must be in the domain of definition of R, i.e. Q cannot have a root there. Let's do some examples:

Example 9.21. *Evaluate*

$$\lim_{(x,y)\to(-2,3)} \frac{3x^2y+1}{x^3y^2-2x}, \tag{281}$$

if it exists.

So, $x^3y^2 - 2x$ evalutated at $(-2,3)$ gives $-68 \neq 0$. Therefore, $\mathbf{x} = (-2,3)$ is in the domain of the definition of the rational function. Hence, the limit exists and is

$$\lim_{(x,y)\to(-2,3)} \frac{3x^2y+1}{x^3y^2-2x} = -\frac{37}{68} \tag{282}$$

Example 9.22. *Evaluate*

$$\lim_{(x,y)\to(0,0)} \frac{3x^2y}{x^2+y^2}, \tag{283}$$

if it exists.

Consider $y = mx$ for $m \in \mathbb{R}$. Then

$$\frac{3x^2y}{x^2+y^2} = \frac{3mx}{(1+m^2)} \to 0 \tag{284}$$

as $x \to 0$. So on any line the limit through the origin is $(0,0)$. Along the parabolas $y = x^2$ and $x = y^2$ one has

$$\frac{3x^2y}{x^2+y^2} = \frac{3x^2}{2} \to 0, \qquad \frac{3x^2y}{x^2+y^2} = \frac{3y^3}{(1+y^2)} \to 0 \tag{285}$$

as $x \to 0$ and $y \to 0$ respectively.

Let's attempt to prove the limit exists and is 0. Let's give ourselves an $\epsilon > 0$. What we need to show is that there is a $\delta > 0$ such that if $||\mathbf{x} - \mathbf{0}|| < \delta$ then $|{}^{3x^2y}/_{x^2+y^2} - 0| < \epsilon$. First note that

$$||\mathbf{x} - \mathbf{0}|| = \sqrt{x^2+y^2} \tag{286}$$

Now,

$$\left|\frac{3x^2y}{x^2+y^2} - 0\right| = \frac{3x^2|y|}{x^2+y^2} \leq 3|y| = 3\sqrt{y^2} \leq 3\sqrt{x^2+y^2}. \tag{287}$$

Take $\delta < \frac{\epsilon}{3}$ then $3\sqrt{x^2+y^2} < \epsilon$, which then gives,

$$\left|\frac{3x^2y}{x^2+y^2} - 0\right| \leq 3\sqrt{x^2+y^2} < \epsilon. \tag{288}$$

This show that $\lim_{\mathbf{x}\to\mathbf{0}} {}^{3x^2y}/_{x^2+y^2} = 0$.

10 Multivariable Functions II: Continuity, Partial Derivatives and PDE

10.1 Continuity

Now that we have generalised the notion of limits to multivariable functions. We can define what we mean for a multivariable function to be continuous at some point $\mathbf{a} = (a_1, ..., a_n)$.

Definition 10.1 (Continuity)**.** *Let f be a function defined on some subset D of $\mathbb{R}^n$. Then one says that f **is continuous at** $\mathbf{a} \in D$ if*

$$\lim_{\mathbf{x}\to\mathbf{a}} f(\mathbf{x}) = f(\mathbf{a}). \tag{289}$$

The function $f : D \to \mathbb{R}$ is continuous on D if it is continuous at every $\mathbf{a} \in D$.

In plain terms continuity means that if the point $\mathbf{x}$ changes by a little bit, then $f(\mathbf{x})$ changes by a little bit. This means that a surface that is the graph of a continuous function has no hole or break.

Remark 10.1. *When dealing with continuity of multivariable functions one should keep in mind proposition 7.2, with the relevant caveats. For example, that polynomials of many variables are continuous everywhere on $\mathbb{R}^n$ and rational functions are continuous on their domain of definition.*

Let's do some examples:

Example 10.1. *Where is*

$$f_1(x, y) = \frac{x^2 + y^2}{x^2 - y^2} \tag{290}$$

continuous? What about the function

$$f_2(x, y) = \begin{cases} 0 & \text{if } y = \pm x \\ \frac{x^2+y^2}{x^2-y^2} & \text{otherwise} \end{cases} \quad ? \tag{291}$$

The function f_1 is continuous where it is defined since it is a rational function. Therefore, its continuous everywhere on $\mathbb{R}^2$ except where $y = \pm x$.

Now f_2 is f_1 except with a modified definition along $y = \pm x$. Note that taking the limit along $x = 0$ gives

$$\lim_{y\to 0}(-1) = -1, \tag{292}$$

which contradicts the definition of continuity since $f(0) = 0$. We need to check the lines $y = \pm x$ too. This means considering $\lim_{(x,y)\to(c,\pm c)} f_2(x, y)$ for c constant. Consider $y = \pm c$, then

$$\lim_{(x,y)\to(c,\pm c)} f_2(x, y) = \lim_{x\to c} \frac{x^2 + c^2}{x^2 - c^2} = \lim_{x\to c} \frac{x^2 + c^2}{(x - c)(x + c)} = \infty, \tag{293}$$

which again contradicts continuity as $f(0) = 0 \neq \infty$. Therefore, f_2 is continuous everywhere on $\mathbb{R}^2$ except where $y = \pm x$.

Example 10.2. *Let*

$$f(x, y) = \begin{cases} 0 & \text{if } (x, y) = (0, 0) \\ \frac{x^2y^3}{2x^2+y^2} & \text{otherwise.} \end{cases} \tag{294}$$

Where is f continuous?

10.2 Differentiation of Multivariable Functions: Partial Derivatives

Suppose we have a function of two variables $f(x,y)$. We could fix $y=b$, consider $f(x,a)$ as a function of the single variable x and look at the process that defined the derivative for $f(x,b)$ at a

$$\lim_{h\to 0}\frac{f(a+h,b)-f(a,b)}{h}. \tag{295}$$

If this limit exists we have a type of derivative of a multivariable function known as the partial derivative of f with respect to x at (a,b). One writes:

$$\frac{\partial f}{\partial x}(a,b)=\lim_{h\to 0}\frac{f(a+h,b)-f(a,b)}{h}. \tag{296}$$

Similarly,

$$\frac{\partial f}{\partial y}(a,b)=\lim_{h\to 0}\frac{f(a,b+h)-f(a,b)}{h}. \tag{297}$$

would be the partial derivative of f with respect to y at (a,b) if the limit on the right-hand side exists. In general, one has the following definition for partial derivatives:

Definition 10.2. *Let f be a function of n-variables, U be a subset of $\mathrm{dom}(f)$. Suppose $\mathbf{a}\in U$. The partial derivative of f with respect to the i^{th}-variable x_i at $\mathbf{a}$ is defined as*

$$\frac{\partial f}{\partial x_i}(\mathbf{a})=\lim_{h\to 0}\frac{f(a_1,...,a_i+h,...,a_n)-f(a_1,...,a_i,...,a_n)}{h} \tag{298}$$

if the limit on the right-hand side exists.

One could let $\mathbf{a}$ vary in U. If the limits exist then one obtains functions of $\mathbf{x}$, i.e. partial derivative functions. These are often denoted many notations listed below:

- Leibniz: $\frac{\partial f}{\partial x_i}$, $\partial_{x_i}f$ and $\partial_i f$, where the latter is often used by lazy relativists.
- f_{x_i}
- Euler: $D_i f$.

Example 10.3. *Find the partial derivative functions of*

1. $f(x,y)=x^n y^m$ *for* $m,n\geq 1$.
2. $f(x,y)=\frac{y^m}{x^n}$ *for* $m\geq 1$, $n\geq 1$ *and* $x\neq 0$.
3. $f(x,y,z)=\ln(xyz)$ *for* $xyz>0$.
4. $f(x,y,z)=\ln(x+y+z)$ *for* $x+y+z>0$.
5. $f(r,\theta)=r\cos\theta+\sin\theta$.
6. $f(x,y)=\arctan(y/x)$ *for* $x\neq 0$.
7. *For* $z^3=1-6xyz-x^3-y^3$, *find* $\partial_x z$.

- $\partial_x(x^n y^m)=nx^{n-1}y^m$ *and* $\partial_y(x^n y^m)=mx^n y^{m-1}$.
- $\partial_x\frac{y^m}{x^n}=-n\frac{y^m}{x^{n+1}}$ *and* $\partial_y\frac{y^m}{x^n}=m\frac{y^{m-1}}{x^n}$.
- $\partial_x\ln(xyz)=\frac{1}{x}$, $\partial_y\ln(xyz)=\frac{1}{y}$ *and* $\partial_z\ln(xyz)=\frac{1}{z}$.
- $\partial_{x_i}\ln(x+y+z)=\frac{1}{x+y+z}$ *for* $x_i=1,2,3$.
- $\partial_r(r\cos\theta+\sin\theta)=\cos\theta$, $\partial_\theta(r\cos\theta+\sin\theta)=\cos\theta-r\sin\theta$.
- $\partial_x\arctan(y/x)=-\frac{y}{x^2+y^2}$ *and* $\partial_x\arctan(y/x)=\frac{x}{x^2+y^2}$.
- $\partial_x z^3=3z^2\partial_x z=-6yz-3x^2$. *Therefore,* $\partial_x z=\frac{-6yz-3x^2}{3z^2}$.

10.3 Higher Derivatives

Let f be a function of two variables and that the partial derivatives of f, $\partial_x f$ and $\partial_y f$, exist in some region D of $\mathbb{R}^2$. We can at this point treat $\partial_x f$ and $\partial_y f$ as functions on D and compute their partial derivatives (if limit above exists). In other words we can consider,

$$\partial_x\partial_x f = \partial_x^2 f = (f_x)_x = f_{xx} = D_1D_1 f = \frac{\partial}{\partial x}\Big(\frac{\partial f}{\partial x}\Big) = \frac{\partial^2 f}{\partial x^2} \tag{299}$$

$$\partial_y\partial_y f = \partial_y^2 f = (f_y)_y = f_{yy} = D_2D_2 f = \frac{\partial}{\partial y}\Big(\frac{\partial f}{\partial y}\Big) = \frac{\partial^2 f}{\partial y^2} \tag{300}$$

$$\partial_x\partial_y f = \partial_{xy}^2 f = (f_y)_x = f_{yx} = D_1D_2 f = \frac{\partial}{\partial x}\Big(\frac{\partial f}{\partial y}\Big) = \frac{\partial^2 f}{\partial x\partial y} \tag{301}$$

$$\partial_y\partial_x f = \partial_{yx}^2 f = (f_x)_y = f_{xy} = D_2D_1 f = \frac{\partial}{\partial y}\Big(\frac{\partial f}{\partial x}\Big) = \frac{\partial^2 f}{\partial y\partial x} \tag{302}$$

Note that we have separated the cases $\partial_y\partial_x f$ and $\partial_x\partial_y f$. The former means differentiate with respect to x first and then with respect to y, whilst the latter means differentiate with respect to y first and then with respect to x. These **may not** be equal; we will state a theorem below about when these operations commute below. First let's do an example:

Example 10.4. *Let*

$$f(x,y) \doteq x^7 - x^2y^3 + yx + 2. \tag{303}$$

Then,

$$\partial_x\partial_y f = -6xy^2 + 1 = \partial_y\partial_x f. \tag{304}$$

Example 10.5. *Let*

$$f(x,y) \doteq \begin{cases} 0 & \mathbf{x} = \mathbf{0} \\ \frac{xy(x^2-y^2)}{x^2+y^2} & \mathbf{x} \neq \mathbf{0} \end{cases} \tag{305}$$

For $\mathbf{x} \neq \mathbf{0}$, via the quotient rule, we have

$$\partial_x f = \frac{y(x^2-y^2)+xy(2x)}{x^2+y^2} - \frac{xy(x^2-y^2)(2x)}{(x^2+y^2)^2} = \frac{x^4y+4x^2y^3-y^5}{(x^2+y^2)^2}. \tag{306}$$

and

$$\partial_y f = \frac{y(x^2-y^2)+xy(2x)}{x^2+y^2} - \frac{xy(x^2-y^2)(2x)}{(x^2+y^2)^2} = -\frac{y^4x+4y^2x^3-x^5}{(x^2+y^2)^2}. \tag{307}$$

For $\mathbf{x} = \mathbf{0}$, we have

$$\partial_x f(0,0) = \lim_{h\to 0}\frac{f(h,0)-f(0,0)}{h} = \lim_{h\to 0}\frac{0}{h} = 0. \tag{308}$$

Similarly, we have

$$\partial_x f(0,0) = \lim_{h\to 0}\frac{f(0,h)-f(0,0)}{h} = \lim_{h\to 0}\frac{0}{h} = 0. \tag{309}$$

Lets look at $\partial_x\partial_y f(0,0)$ and $\partial_y\partial_x f(0,0)$:

$$\partial_x\partial_y f = \lim_{h\to 0}\frac{\partial_y f(h,0)-\partial_y f(0,0)}{h} = \lim_{h\to 0}\frac{h-0}{h} = 1 \tag{310}$$

and

$$\partial_y\partial_x f = \lim_{h\to 0}\frac{\partial_x f(0,h)-\partial_y f(0,0)}{h} = \lim_{h\to 0}\frac{-h-0}{h} = -1. \tag{311}$$

Therefore, $\partial_y\partial_x f \neq \partial_x\partial_y f$.

This last example illustrates the failure of the following theorem:

Theorem 10.1. *Let D be some disk in $\mathbb{R}^2$ containing (a, b). Suppose $f : D \to \mathbb{R}$. If $\partial_x\partial_y f$ and $\partial_y\partial_x f$ exists and are continuous on D, then 'partial derivatives commute':*

$$\partial_x\partial_y f(a, b) = \partial_y\partial_x f(a, b). \tag{312}$$

Remark 10.2. *This theorem generalises to more derivatives and to functions of more variables. For example,*

$$\partial_x\partial_y\partial_z f = \partial_y\partial_z\partial_x\partial_z f = \partial_z\partial_x\partial_y f = \partial_y\partial_x\partial_z f = \partial_x\partial_z\partial_y f = \partial_z\partial_y\partial_x f \tag{313}$$

if these functions are continuous.

10.4 Partial Differential Equations

A partial differential equation (PDE) is a relation between partial derivatives of a multivariable function $f(x_1, ..., x_n)$. For example,

$$\partial_x f = \partial_y f, \qquad \partial_x\partial_y f = 0, \qquad f\partial_z f = 2(\partial_x f)^2. \tag{314}$$

Here f would be 'the unknown' of the equation for which (in an ideal world) we'd like to solve for. Be warned this is not always possible; a solution may not even exist. If it does exist, it may not be unique.

Some famous examples of PDE are listed below (**the last four examples are simply included for interest, you're not expected to fully understand the notation or all comments related to them**):

- **Laplace's equation**:

$$\Delta u \doteq \partial_x^2 u + \partial_y^2 u = 0, \qquad \Delta u \doteq \partial_x^2 u + \partial_y^2 + \partial_z^2 u = 0, \qquad \Delta u \doteq \partial_{x_1}^2 u + ... + \partial_{x_n}^2 u = 0 \tag{315}$$

  This first is Laplace's equation in 2-dimensions, the second in 3-dimensions and the last in n-dimensions. Note that in 1-dimension it $u(x)$ must satisfy

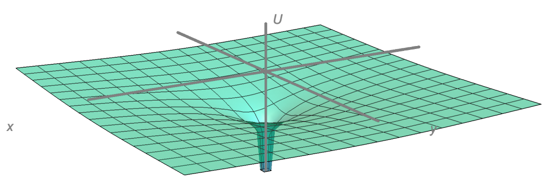
$$\Delta u = \partial_x^2 u = \frac{d^2 u}{dx^2} = 0 \implies u = ax + b \tag{316}$$

  for a, b constants in $\mathbb{R}$. The Δ often gets called 'the Laplacian'. Solutions to this equation are called harmonic functions, which crop up everywhere in physics. For example, in gravitation, fluid dynamics, heat conduction, electrostatics.

  Lets look at $u(x, y) \doteq \ln(x^2 + y^2)$ for Laplace's equation in 2-dimensions. Now,

$$\partial_x^2 u = -\frac{2(x^2 - y^2)}{(x^2 + y^2)^2}, \quad \partial_y^2 u = \frac{2(x^2 - y^2)}{(x^2 + y^2)^2} \implies \partial_x^2 u + \partial_y^2 u = 0. \tag{317}$$

  So $u(x, y)$ is a harmonic function and is plotted below:



  This function could describe the electric potential due to a line of unit charge entire z-axis.

- The **heat equation**:

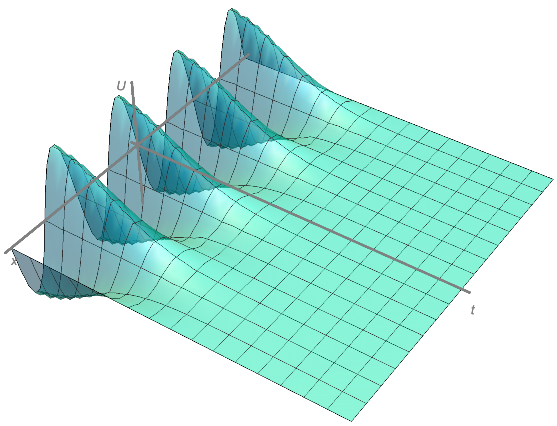
$$\partial_t u - \Delta u = 0 \tag{318}$$

As the name suggests this equation describes how heat is conducted or diffused through a given material. Solutions are called caloric functions. Note that, if u is independent of time, we have Laplace's equation, i.e. if the temperature distribution in some material is not evolving it must be a harmonic function.

Let's look at $u(t, x) \doteq e^{-\kappa^2 t} \sin(\kappa x)$ for the heat equation with the 1-dimensional Laplacian. We have,

$$\partial_t u = -\kappa^2 u, \quad \partial_x^2 u = -\kappa^2 u \implies \partial_t u - \partial_x^2 u = 0. \tag{319}$$

So, $u(t, x)$ is a caloric function and is plotted below:



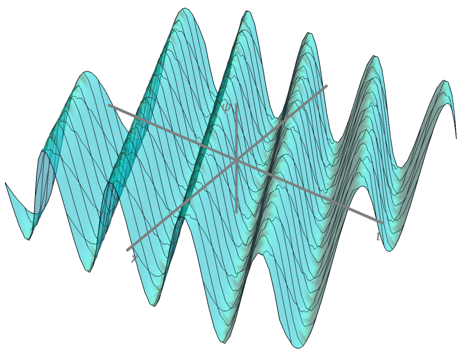
- The **wave equation**:

$$\Box \Psi \doteq -\partial_t^2 \Psi + \Delta \Psi = 0. \tag{320}$$

Suppose we take $\Delta\Psi$ as the 1-dimensional Laplacian of $\Psi(t, x)$. Lets look at the function $\Psi(t, x) \doteq \sin(x - t)$. Computing partial derivatives gives

$$\partial_t^2 \Psi = -\sin(x - t), \qquad \partial_x^2 \Psi = -\sin(x - t) \implies -\partial_t^2 \Psi + \partial_x^2 \Psi = 0. \tag{321}$$

The function $\Psi(t, x) \doteq \sin(x - t)$ is plotted below:



This is wave type behaviour, hence the name for the equation. This function Ψ could be the a sound wave, a light wave, an ocean wave etc.

- **Maxwell's equations** for electromagnetism. In the absence of charges and currents these have the form:

$$\mathrm{div}\mathbf{E} = 0 = \mathrm{div}\mathbf{B} \tag{322}$$
$$\mathrm{curl}\mathbf{E} = -\partial_t \mathbf{B} \tag{323}$$
$$\mathrm{curl}\mathbf{B} = \partial_t \mathbf{E}. \tag{324}$$

  Here $\mathbf{E}$ and $\mathbf{B}$ are vectors in $\mathbb{R}^3$ representing the electric field and magnetic fields respectively. One can show that these equations imply that

$$-\partial_t^2 \mathbf{E} + \Delta \mathbf{E} = 0, \tag{325}$$
$$-\partial_t^2 \mathbf{B} + \Delta \mathbf{B} = 0. \tag{326}$$

  In otherwords, the electric and magnetic fields satisfy the wave equation and therefore, admit wave-type solutions. This is often why it is said that light (which is electromagnetic radiation) is a wave.

- The **vaccuum Einstein equation** for general relativity,

$$\mathrm{Ric}(g) = 0, \tag{327}$$

  which can be written in the form

$$\Box_g(g_{ij}) = N(g, \partial g). \tag{328}$$

  Here, $\Box_g$ is an altered version of the $\Box$ for the wave equation above. Additionally, g_{ij} can be thought of as a collection of functions $g_{ij}(x_0, x_1, x_2, x_3)$ with $1 \leq i, j \leq 4$ and $N(g, \partial g)$ denotes a term that involves g_{ij} and $\partial g_{ij}/\partial x_k$ and no higher derivatives.

  The functions g_{ij} can be thought of as encoding the gravitational field. There are g_{ij} that model the gravitational fields of stars, black holes, galaxies, the universe...

  When the gravitational field is 'weak' the term $N(g, \partial g)$ can be neglected and $\Box_g$ replaced with the wave equation $\Box$. The functions g_{ij} then satisfy the wave equation, which gave rise to Einsteins famous prediction of gravitational waves in 1916. Direct experimental confirmation of their existence was produced 2016 by the LIGO observations.

- The **Navier-Stokes equation** for fluid dynamics.

$$\partial_t \mathbf{v} + (\mathbf{v} \cdot \nabla)\mathbf{v} = -\frac{1}{\rho}\nabla p + \nu \Delta \mathbf{v} + \mathbf{f}. \tag{329}$$

  Here, $\mathbf{v}$ is the velocity of the fluid. It is a vector in $\mathbb{R}^3$ that depends on time, t and space, $\mathbf{x}$. The function p is the pressure of the fluid and ν is something known as the viscosity (related to thickness of the fluid), $\mathbf{f}$ is some external force. There is a 1 million dollar prize to be awarded to the person who resolves an open problem about the existence of solutions (in some catagory) to this equation. This is one of 7 Millennium Prize Problems put forward by the Clay Mathematics Institute.

- The **Black–Scholes equation** in mathematical finance:

$$\partial_t V + \frac{1}{2}\sigma^2 S^2 \partial_S^2 V + rS\partial_S V - rV = 0. \tag{330}$$

  Here, V is the price of an option (which is right–or 'option'–to buy a stock at a given predetermined price) as a function of stock price S and time t. The constant r is the risk-free interest rate, and σ is the volatility of the stock. For example, you could have a friend that potentially wants to buy your car at the price of $20k$. They could ask you to sell them the right (or option) to buy your car in the next month for 500. This could vary in time dependent on how desperate your friend is for the car or how much you like your car.

The field of analysis of PDE has been and continues to be a very active field of research in mathematics currently.

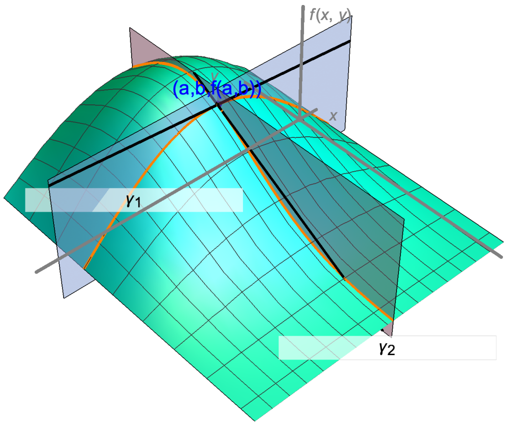
11 Tangent Planes and Linear Approximations

This section is concerned with the local properties of multivariable functions. In particular, how to partial derivatives characterise the behaviour of a multivariable function close to a particular point.

11.1 Tangent Planes

For a function of a single variable $f(x)$, $^{df}/_{dx}$ is interpreted a rate of change with respect to x. Similarly, for a multivariable function $f(x_1, ..., x_n)$, partial derivatives of f can be interpreted as rates of change when all but one variable is fixed. Effectively, this means we're looking at rates of change in certain directions. For example, suppose we have a function f of two variables (x, y). Further suppose the partial derivatives of f exists at (a, b), i.e. $\partial_x f(a, b)$ and $\partial_y f(a, b)$ exist.

The surface $\{(x, y, z) : z = f(x, y)\}$ near (a, b) is plotted below in cyan.



Note that $y = b$ is a plane in $\mathbb{R}^3$ which intersects the surface in a curve, denoted in the diagram with γ_1. This curve has the equation $g(x) = f(x, b)$ and therefore it's gradient at $x = a$ is

$$\frac{dg}{dx}(a) = \partial_x f(a, b). \tag{331}$$

So, there is a line tangent to the curve γ_1 with gradient $m_1 = \partial_x f(a, b)$ in the plane determined by $y = b$, i.e.

$$z = m_1(x - a) + f(a, b), \tag{332}$$

in $\{(x, y, z) : y = b\}$. Supposing that $m_1 \neq 0$, i.e. $\partial_x f(a, b) \neq 0$ then the symmetric equation of the line is

$$\frac{z - f(a, b)}{\partial_x f(a, b)} = \frac{(x - a)}{1}, \qquad y = b. \tag{333}$$

We can put this into the form for the vector equation of a line as:

$$\mathbf{x}(\lambda) = (a, b, f(a, b)) + \lambda \mathbf{v}_1, \tag{334}$$

with $\mathbf{v}_1 = (1, 0, \partial_x f(a, b))$. This is plotted in black above γ_1. Similarly, $x = a$ is a plane in $\mathbb{R}^3$ which intersects the surface in a curve, denoted in the diagram with γ_2. This curve has the equation $h(y) = f(a, y)$ and therefore it's gradient at $y = b$ is

$$\frac{dh}{dy}(b) = \partial_y f(a, b). \tag{335}$$

So, there is a line tangent to the curve γ_2 with gradient $m_2 = \partial_y f(a,b)$ in the plane determined by $x = a$, i.e.

$$z = m_2(y - b) + f(a,b), \tag{336}$$

We can put this into the form for the vector equation of a line as:

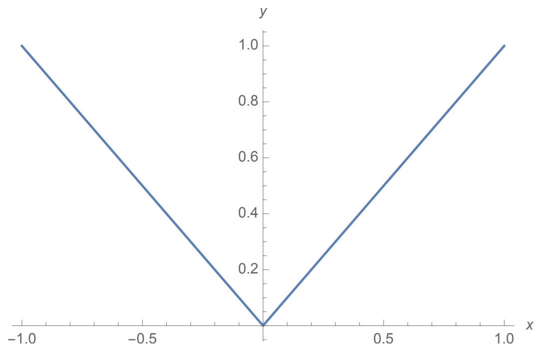
$$\mathbf{x}(\lambda) = (a, b, f(a,b)) + \lambda \mathbf{v}_2. \tag{337}$$

with $\mathbf{v}_2 = (0, 1, \partial_y f(a,b))$. This is plotted in black above γ_2.

What's so special about curves resulting from intersecting with planes parallel to the zy and zx-planes? Well, both nothing and something.

- Nothing: We could have any curve on the surface through $(a, b, f(a,b))$ and look at its tangent line at $(a, b, f(a,b))$, **if** that line exists, which leads to the something...
- Something: The only stipulation about the curves resulting from $\{x = a\}$ and $\{y = b\}$ intersection was that the partial derivatives existed.

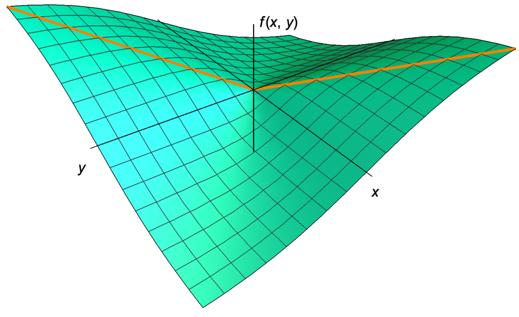
This latter point means that the curves that result from cross-sections $x = a$ and $y = b$ cannot break at (a,b), i.e. be discontinuous, or have corners like $y = |x|$ (which is continuous but not differentiable at $x = 0$) plotted below:



So, the stipulation that $\partial_x f(a,b)$ and $\partial_y f(a,b)$ exist, for the curves resulting from $\{x = a\}$ and $\{y = b\}$ intersection, means that the tangent line is well-defined. In the case of other curves, the tangent line may not be well defined. Consider the function,

$$f(x,y) \doteq \begin{cases} \frac{xy}{\sqrt{x^2+y^2}}, & (x,y) \neq (0,0) \\ 0 & (x,y) = (0,0). \end{cases} \tag{338}$$

This is plotted below:



Now one can check that

$$\partial_x f(0,0) = 0 = \partial_y f(0,0), \tag{339}$$

Hence, the partial derivatives exist at $(0,0)$, and, therefore, the tangent lines to $(0,0,0)$ in the $\mathbf{i}$ and $\mathbf{j}$ directions exist. However, consider the curve in orange above, which is given by $y = x$. Therefore,

$$\left(x, x, f(x,x) = \frac{|x|}{\sqrt{2}}\right), \tag{340}$$

defines a curve in the surface. This has no tangent line at $x = 0$ since $|x|$ is not differentiable at $x = 0$. What's gone wrong here? Well, note that

$$\partial_x f = \frac{y^3}{(x^2+y^2)^{\frac{3}{2}}}, \qquad \partial_y f = \frac{x^3}{(x^2+y^2)^{\frac{3}{2}}} \tag{341}$$

for $(x,y) \neq (0,0)$. Lets focus on $\partial_x f(x,y)$. Take $x = 0$, then

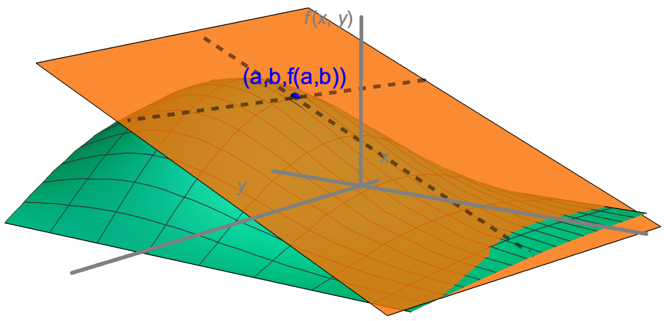
$$\partial_x f(0,y) = \frac{y^3}{(y^2)^{\frac{3}{2}}} = \frac{y^3}{|y|^3} = \begin{cases} 1 & y > 0 \\ -1 & y < 0. \end{cases} \tag{342}$$

Therefore, $\lim_{(x,y)\to(0,0)} \partial_x f(x,y)$ does not exist. In particular, $\partial_x f(x,y)$ is not continuous at $(0,0)$.

This example illustrates the following fact: if the partial derivatives are continuous, then all the tangent lines exist. So when the partial derivatives are continuous, these tangent lines form the **tangent plane** at $(a,b,f(a,b))$, which we will define as the plane which contains the two tangent lines resulting from the intersection with $\{x = a\}$ and $\{y = b\}$, i.e. the plane containing the point $\mathbf{x}_0 = (a,b,f(a,b))$ with normal

$$\mathbf{n} = \mathbf{v}_1 \times \mathbf{v}_2 = (-\partial_x f(a,b), -\partial_y f(a,b), 1) \tag{343}$$

This is plotted below:



Recall that our equation for the plane was

$$\langle \mathbf{n}, \mathbf{x} - \mathbf{x}_0 \rangle = 0. \tag{344}$$

Expanding gives

$$\partial_x f(a,b)(x-a) + \partial_y f(a,b)(y-b) - (z - f(a,b)) = 0, \tag{345}$$

which is the equation of the tangent plane at $(a,b,f(a,b))$.

Let's do an example:

Example 11.1. *Take the ellipsoid*

$$\frac{x^2}{4} + \frac{y^2}{16} + \frac{z^2}{49} = 1. \tag{346}$$

Recall that we can solve for z in terms of two functions

$$z = f_{\pm}(x, y) = \pm 7\sqrt{1 - \frac{x^2}{4} - \frac{y^2}{16}}. \tag{347}$$

Lets find the tangent plane at $(1, 1, \frac{7}{4}\sqrt{11})$. To easy notation let $f = f_+$ and let's find $\partial_x f$ and $\partial_y f$:

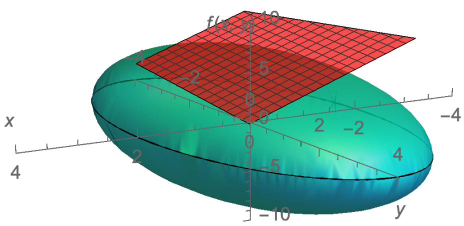
$$\partial_x f = -\frac{7x}{4\sqrt{1 - \frac{x^2}{4} - \frac{y^2}{16}}}, \qquad \partial_y f = -\frac{7y}{16\sqrt{1 - \frac{x^2}{4} - \frac{y^2}{16}}}. \tag{348}$$

So,

$$\partial_x f(1, 1) = -\frac{7}{\sqrt{11}}, \qquad \partial_y f(1, 1) = -\frac{7}{4\sqrt{11}}. \tag{349}$$

Therefore, the equation of the plane at $(1, 1, \frac{7}{4}\sqrt{11})$ tangent to the ellipsoid is

$$\frac{7}{\sqrt{11}}(x - 1) + \frac{7}{4\sqrt{11}}(y - 1) + \left(z - \frac{7}{4}\sqrt{11}\right) = 0. \tag{350}$$



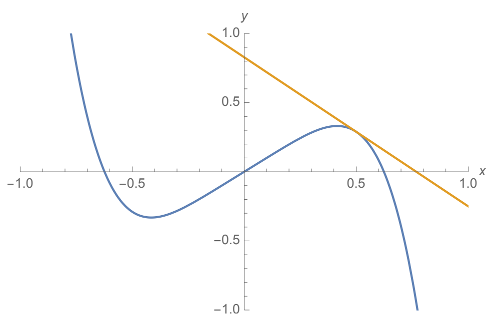
11.2 Linear Approximations of Multivariable Functions

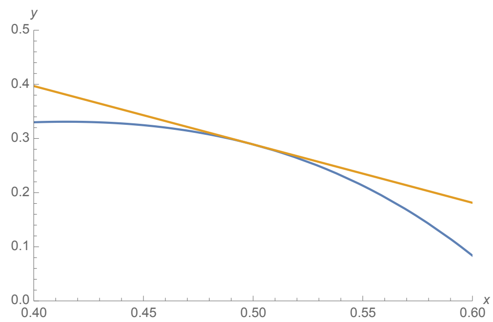
Definition 11.1. *A function f from D (a region in $\mathbb{R}^n$) to $\mathbb{R}$ is called* **linear** *if*

$$f(\mathbf{x}) = b + a_1 x_1 + ... + a_n x_n, \tag{351}$$

where $\mathbf{x} = (x_1, ..., x_n)$ and a_i are real constants for $1 \le i \le n$.

Suppose we have a single variable function $f(x)$ which has derivative $f'(a)$ near some point $x = a$. This could look something like the following:





Here we have the function $f(x) = x^7 - 7x^5 + x$ and the line in orange is the tangent line at $(1/2, 37/128)$. Clearly when we have a global plot of the graph this line doesn't approximate $f(x)$ at all well. However, as we 'zoom' in on the graph (i.e. we look locally) the tangent line approximates the function better and better. See the right-hand diagram. Therefore, one would expect that we can approximate the function near a point by the tangent line at that point.

This is the intuition behind **linear approximation** but let's make this a bit more precise. Consider the definition of the derivative

$$\frac{df}{dx}(a) = \lim_{h \to 0} \frac{f(a+h) - f(a)}{h} \tag{352}$$

and suppose the limit on the right-hand side exists. Let $h = x - a$ then this can be rewritten as

$$\frac{df}{dx}(a) = \lim_{x \to a} \frac{f(x) - f(a)}{x - a}. \tag{353}$$

From the definition of the limit one knows that for all $\epsilon > 0$ that there exists a $\delta > 0$ such that if $0 < |x - a| < \delta$ then

$$\left| \frac{f(x) - f(a)}{x - a} - \frac{df}{dx}(a) \right| < \epsilon. \tag{354}$$

For x close to a (ϵ small and δ small), this means that

$$\frac{f(x) - f(a)}{x - a} = \frac{df}{dx}(a) + \epsilon(x) \tag{355}$$

where $\epsilon(x) \to 0$ as $x \to a$. Rearranging gives

$$f(x) = f(a) + \frac{df}{dx}(a)(x - a) + R(x) \tag{356}$$

where $R(x) = \epsilon(x)(x - a)$, called the remainder, goes to 0 faster than $x - a$ as $x \to a$. This means that for x close to a one can indeed approximate $f(x)$ as

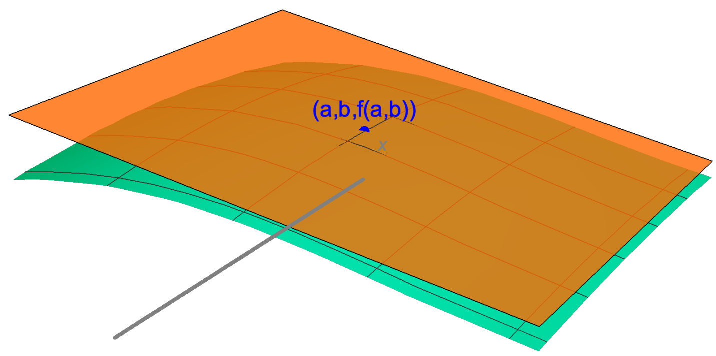
$$f(x) \approx L(x) \doteq f(a) + \frac{df}{dx}(a)(x - a). \tag{357}$$

The function $L(x)$ is a linear function on $\mathbb{R}$ and is known as the **linear approximation** of $f(x)$ at a. Notice that for functions of one variable, the linear approximation only exists if $f'(a)$ exists, i.e. the tangent line exists at $x = a$. In this case it is always a valid approximation.

For a function f of two variables the intuition is similar: if f has **continuous partial derivatives**, the tangent plane at a point $(a, b, f(a, b))$ with equation (from (345))

$$z = f(a, b) + (\partial_x f)(a, b)(x - a) + (\partial_y f)(a, b)(y - b) \tag{358}$$

is well-defined since it contains all tangent lines to the surface at $(a, b, f(a, b))$. In this case, the tangent plane becomes a good approximation of the function near $(a, b, f(a, b))$ see the following diagram:



This means that in analogy with the argument about tangent lines for functions of a single variable we can approximate $f(x,y)$ using (358), i.e.

$$f(x,y) \approx L(x,y) \doteq f(a,b) + \partial_x f(a,b)(x-a) + \partial_y f(a,b)(y-b). \tag{359}$$

The function $L(x,y)$ is known as the **linear approximation** of $f(x,y)$ at (a,b). It is important to stress that even though one can write down the linear approximation in equation (359) when the partial derivatives exist, it is not always a 'good' approximation. It is a **valid approximation** when the partial derivatives are **continuous** at (a,b).

One can generalise the linear approximation to a function f of n-variables $\mathbf{x} = (x_1, ..., x_n)$. In this case, the linear approximation f at $\mathbf{a} = (a_1, ..., a_n)$ is

$$L(\mathbf{x}) = f(\mathbf{a}) + (\partial_{x_1} f)(\mathbf{a})(x_1 - a_1) + ... + (\partial_{x_n} f)(\mathbf{a})(x_n - a_n). \tag{360}$$

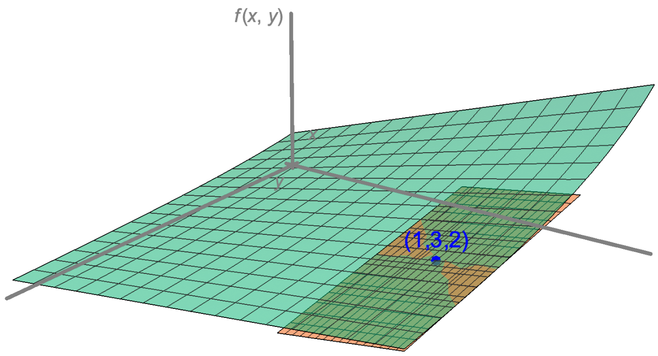
This is a **valid approximation** for $\mathbf{x}$ close to $\mathbf{a}$ when the partial derivatives $\partial_{x_i} f$ are **continuous** at $\mathbf{a}$ for $1 \leq i \leq n$.

Example 11.2. *Lets find the linear approximation of $f(x,y) = \frac{1+y}{1+x}$ at $(1,3)$.*

Lets compute the first partial derivatives for $x > -1$

$$\partial_x f = -\frac{1+y}{(1+x)^2}, \qquad \partial_y f = \frac{1}{1+x}. \tag{361}$$

Note that the partial derivatives in equation (361) are both rational functions and, therefore, continuous on their domain of definition. In particular at $(1,3)$. Hence, the tangent plane at $(1,3)$ is well-defined and the linear approximation is valid as shown by the following plot:



At $(1,3)$,

$$\partial_x f(1,3) = -1, \quad \partial_y f(1,3) = \frac{1}{2}. \tag{362}$$

Therefore, the linear approximation of $f(x,y)$ at $(1,3)$ is

$$L(x,y) = f(1,3) + \partial_x f(1,3)(x-1) + \partial_y f(1,3)(y-3) = 2 - (x-1) + \frac{1}{2}(y-3). \tag{363}$$

Example 11.3. *Lets find the linear approximation of $f(x,y) = xe^{xy}$ at $(1,0)$.*

Note that $f(x,y)$ is well defined on all $\mathbb{R}^2$ and therefore, we can freely compute the partial derivatives

$$\partial_x f = (1+xy)e^{xy}, \qquad \partial_y f = x^2 e^{xy} \tag{364}$$

which are continuous everywhere on $\mathbb{R}^2$ since these functions are products of polynomials and the exponential. Therefore, the linear approximation is good at $(1,0)$:

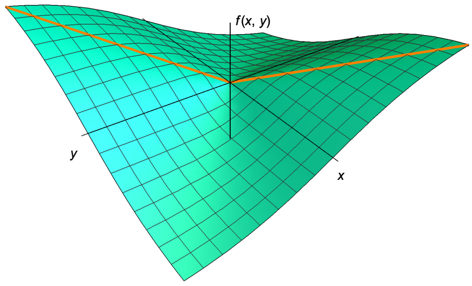
$$L(x,y) = f(1,0) + \partial_x f(1,0)(x-1) + \partial_y f(1,0)(y-1) = 1 + (x-1) + y = x + y. \tag{365}$$

Let's illustrate how the linear approximation fails for non-continuous partial derivatives with the following example:

Example 11.4. *Consider again the function,*

$$f(x,y) \doteq \begin{cases} \frac{xy}{\sqrt{x^2+y^2}}, & (x,y) \neq (0,0) \\ 0 & (x,y) = (0,0). \end{cases} \tag{366}$$

This is plotted below:



As above

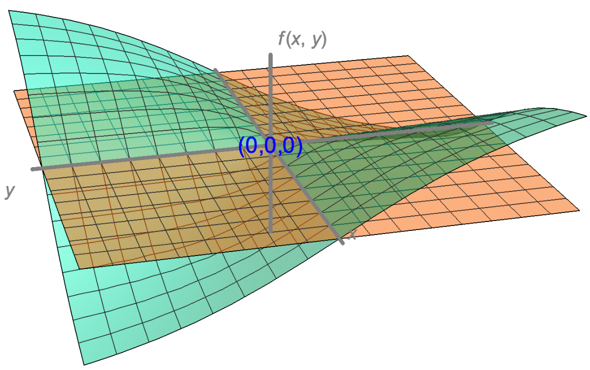
$$\partial_x f = \begin{cases} 0 & \mathbf{x} = (0,0) \\ \frac{y^3}{(x^2+y^2)^{\frac{3}{2}}} & \mathbf{x} \neq (0,0) \end{cases} \tag{367}$$

$$\partial_y f = \begin{cases} 0 & \mathbf{x} = (0,0) \\ \frac{x^3}{(x^2+y^2)^{\frac{3}{2}}} & \mathbf{x} \neq (0,0) \end{cases} \tag{368}$$

Since the partial derivatives exist at $(0,0)$ *we can use equation* (359) *to write down a linear approximation for* $f(x,y)$ *at* $(0,0)$*:*

$$L(x,y) = f(0,0) + \partial_x f(0,0)(x-0) + \partial_y f(0,0)(y-0) = 0. \tag{369}$$

However, recall also that the partial derivatives are not continuous and therefore, the tangent plane was not well defined. One can see this from the following plot:



12 Differentiability for Multivariable Functions

This section is about what it means for a multivariable to be differentiable, not just *partial* differentiable.

12.1 Motivation and Definition

Lets try to reverse engineer the computation above where we derived the linear approximation for a function $f(x)$ of a single variable from the definition of differentiability. In other words, let's derive another condition that a function of two variables must satisfy for the linear approximation to be valid. This will lead us to a definition of what it means for a function of many variables to be differentiable, **not** just partial differentiable.

To this end, suppose $x = a + h_1$ and $y = b + h_2$ for h_1, h_2 small. So, we have $h_1 = x - a$ and $h_2 = y - b$. To ease notation define

$$\mathbf{h} = (h_1, h_2), \tag{370}$$

which has the property $||\mathbf{h}|| \to 0$ as $(x, y) \to (a, b)$. Suppose that the linear approximation in equation (359) is valid for (x, y) close to (a, b). Therefore, one has

$$f(a + h_1, b + h_2) = f(a, b) + \partial_x f(a, b)h_1 + \partial_y f(a, b)h_2 + R(h_1, h_2) \tag{371}$$

where our remainder $R(h_1, h_2) = R(x - a, y - b)$ goes to zero faster than h_1 or h_2, i.e.

$$f(a + h_1, b + h_2) = f(a, b) + \partial_x f(a, b)h_1 + \partial_y f(a, b)h_2 + \varepsilon(h_1, h_2)||\mathbf{h}|| \tag{372}$$

where $\varepsilon(h_1, h_2) \to 0$ as $||\mathbf{h}|| \to 0$. Therefore,

$$\frac{|f(a + h_1, b + h_2) - f(a, b) - \langle \nabla f(a, b), \mathbf{h}\rangle|}{||\mathbf{h}||} \leq |\varepsilon(h_1, h_2)|. \tag{373}$$

where we've introduced the **vector** of partial derivatives

$$\nabla f \doteq (\partial_x f, \partial_y f) = \partial_x f\mathbf{i} + \partial_y f\mathbf{j}, \tag{374}$$

which is called the **gradient** of f.

Remark 12.1. *In* $\mathbb{R}^n$,

$$\nabla f \doteq (\partial_{x_1} f, \partial_{x_2} f, ..., \partial_{x_n} f) = \partial_{x_1} f\mathbf{e}_1 + \partial_{x_2} f\mathbf{e}_2 + ... + \partial_{x_n} f\mathbf{e}_n. \tag{375}$$

One makes the following definition about differentiability for multivariable functions,

Definition 12.1 (Differentiability)**.** *Let* $f : \mathbb{R}^n \to \mathbb{R}$*. The function* f *is said to be differentiable at* $\mathbf{x}_0 \in \mathbb{R}^n$ *if there exists a* $\mathbf{v} \in \mathbb{R}^n$ *such that*

$$\lim_{\mathbf{h}\to\mathbf{0}} \frac{|f(\mathbf{x}_0 + \mathbf{h}) - f(\mathbf{x}_0) - \langle \mathbf{v}, \mathbf{h}\rangle|}{||\mathbf{h}||} = 0. \tag{376}$$

The condition that a multivariable function is differentiable at (a, b) implies that the linear approximation of a the function at (a, b) is valid, just like the single variable case! Note that previously we said that the linear approximation at (a, b) was valid/good if the partial derivatives of the function at (a, b) were continuous. Therefore, we expect some relation between (continuous) partial differentiability and differentiability.

12.2 Relation to Partial Differentiability

If f is differentiable at $\mathbf{x}_0$ then this implies all partial derivatives exist at $\mathbf{x}_0$ and $\mathbf{v} = \nabla f$ since one can look at $\mathbf{h} = (h, 0)$ and $\mathbf{h} = (0, h)$ respectively for $\partial_x f$ and $\partial_y f$. Explicitly, suppose

$$\lim_{\mathbf{h}\to\mathbf{0}} \frac{|f(\mathbf{x}_0+\mathbf{h}) - f(\mathbf{x}_0) - \langle \mathbf{v}, \mathbf{h}\rangle|}{||\mathbf{h}||} = 0. \tag{377}$$

then since this is a multivariable limit, the limit along all paths through $\mathbf{h} = \mathbf{0}$ must exist and be equal. Therefore, the limit along the path $\mathbf{h} = (h, 0)$ must exist and be equal to zero:

$$\lim_{h\to 0} \frac{|f(x_0+h, y_0) - f(x_0, y_0) - v_1 h|}{|h|} = 0. \tag{378}$$

We can rewrite this as

$$\lim_{h\to 0} \left| \frac{f(x_0+h, y_0) - f(x_0, y_0)}{h} - v_1 \right| = 0. \tag{379}$$

Unpacking the definition of the limit this says that for any number $\epsilon > 0$ there exists another number $\delta > 0$ such that if $0 < |h - 0| < \delta$ then

$$\left| \left| \frac{f(x_0+h, y_0) - f(x_0, y_0)}{h} - v_1 \right| - 0 \right| < \epsilon \tag{380}$$

or equivalently

$$\left| \frac{f(x_0+h, y_0) - f(x_0, y_0)}{h} - v_1 \right| < \epsilon. \tag{381}$$

This says that

$$\lim_{h\to 0} \frac{f(x_0+h, y_0) - f(x_0, y_0)}{h} = v_1, \tag{382}$$

i.e. the limit on the left-hand side exists and is v_1. At this point recall that the limit on the left-hand side is the definition of $\partial_x f(x_0, y_0)$:

$$\partial_x f(x_0, y_0) = \lim_{h\to 0} \frac{f(x_0+h, y_0) - f(x_0, y_0)}{h}. \tag{383}$$

Therefore, $\partial_x f(x_0, y_0)$ exists and $v_1 = \partial_x f(x_0, y_0)$. A similar argument shows that $\partial_x f(x_0, y_0)$ and $v_2 = \partial_x f(x_0, y_0)$.

The converse is **not** true: all partial derivatives can exist but f may not be differentiable. For example

Example 12.1. *Let $f : \mathbb{R}^2 \to \mathbb{R}$ be defined by*

$$f(x, y) \doteq \begin{cases} \frac{y^3}{x^2+y^2} & (x, y) \neq (0, 0) \\ 0 & (x, y) = (0, 0). \end{cases} \tag{384}$$

One can check that the gradient at $(0, 0)$ is

$$\nabla f|_{(0,0)} = (0, 1). \tag{385}$$

If f is differentiable then the limit of

$$g(\mathbf{h}) \doteq \frac{|f(\mathbf{h}) - f(\mathbf{0}) - \langle \mathbf{v}, \mathbf{h}\rangle|}{||\mathbf{h}||} \tag{386}$$

should vanish as $\mathbf{h} \to 0$ for $\mathbf{v} = \nabla f(0, 0)$. One can compute that

$$g(\mathbf{h}) = \frac{\left|\frac{h_2^3}{h_1^2+h_2^2} - h_2\right|}{\sqrt{h_1^2+h_2^2}} = \frac{|h_2| h_1^2}{(h_1^2+h_2^2)^{\frac{3}{2}}}. \tag{387}$$

Take $h_2 = h_1$ then

$$g(h_1, h_1) = \frac{|h_1|^3}{2\sqrt{2}|h_1|^3} = \frac{1}{2\sqrt{2}} \neq 0. \tag{388}$$

So, we've argued that for the linear approximation to be valid our function must be differentiable and we've showed that partial differentiability does not imply differentiability. So what use is partial differentiability then? Well, as you might expect, if the partial derivatives are continuous then partial differentiability implies differentiability. We simply state the following theorem, which is probably the most useful theorem to know regarding differentiability of multivariable functions:

Theorem 12.1. *If the partial derivatives $\partial_x f$ and $\partial_y f$ exist near (a, b) and are continuous at (a, b), then f is differentiable at (a, b).*

Remark 12.2. *In practise you're unlikely to ever prove that a function is differentiable directly using definition 12.1. You will use theorem 12.1.*

We will not prove theorem 12.1 in general but we will illustrate it with an example:

Example 12.2. *Show that $f(x, y) = \frac{1+y}{1+x}$ is differentiable at $(1, 3)$.*

So the partial derivatives are for $x > -1$

$$\partial_x f = -\frac{1+y}{(1+x)^2}, \qquad \partial_y f = \frac{1}{1+x}. \tag{389}$$

Note that these partial derivatives are continuous on their domains of definition since they are rational functions. Further $(1, 3)$ is in the domain of definition so the partial derivatives are continuous at $(1, 3)$, hence the function is differentiable at $(1, 3)$ by theorem 12.1.

Let's show directly that f is differentiable at $(1, 3)$. For f to be differentiable at $(1, 3)$,

$$\lim_{\mathbf{h}\to 0} \frac{|f(1+h_1, 3+h_2) - f(1,3) - \langle \mathbf{v}, \mathbf{h}\rangle|}{||\mathbf{h}||} = 0 \tag{390}$$

for $\mathbf{v} = \nabla f(1, 3)$. Evalutating the partial derivatives at $(1, 3)$ gives:

$$\partial_x f(1,3) = -1, \quad \partial_y f(1,3) = \frac{1}{2} \implies \nabla f = (-1, 1/2). \tag{391}$$

Define for notational simplicity:

$$g(\mathbf{h}) \doteq \frac{|f(1+h_1, 3+h_2) - f(1,3) - \langle \nabla f(1,3), \mathbf{h}\rangle|}{||\mathbf{h}||}. \tag{392}$$

Evaluating:

$$g(\mathbf{h}) = \frac{\left|\frac{4+h_2}{2+h_1} - 2 + h_1 - \frac{1}{2}h_2\right|}{\sqrt{h_1^2 + h_2^2}} = \frac{|h_1||h_1 - \frac{1}{2}h_2|}{|2+h_1|\sqrt{h_1^2+h_2^2}}. \tag{393}$$

For any $\epsilon > 0$, we want to show that there is a $\delta > 0$ such that if $||\mathbf{h}|| < \delta$ then

$$|g(\mathbf{h})| < \epsilon. \tag{394}$$

So, note that (the triangle inequality on $\mathbb{R}$ tells us)

$$\left|h_1 - \frac{1}{2}h_2\right| \leq |h_1| + \frac{1}{2}|h_2|. \tag{395}$$

Now $|h_1|, |h_2| \leq ||\mathbf{h}||$, so

$$\left|h_1 - \frac{1}{2}h_2\right| \leq \frac{3}{2}||\mathbf{h}||. \tag{396}$$

Hence,

$$|g(\mathbf{h})| \leq \frac{|h_1|}{|2+h_1|}. \tag{397}$$

You can see that this goes to zero for $h_1 \to 0$ in the region

$$\{(h_1, h_2) \in \mathbb{R}^2 : h_2 > -2\}. \tag{398}$$

We have that $|h_1| \leq ||\mathbf{h}||$, so pick $\delta = \min(1, \epsilon)$, therefore,

$$|h_1| < ||\mathbf{h}|| < \min(\epsilon, 1) \leq 1. \tag{399}$$

So,

$$|2 + h_1| > |2 - |h_1|| \geq 1 \implies \frac{1}{|2 + h_1|} < \frac{1}{|2 - |h_1||} \leq 1. \tag{400}$$

Therefore,

$$|g(\mathbf{h})| \leq \frac{|h_1|}{|2 + h_1|} < |h_1| < \min(\epsilon, 1) \leq \epsilon. \tag{401}$$

13 The Chain Rule

The chain rule provides a way to differentiate a composite function. You are probably familiar with this for functions of a single variable. This section is about how to generalise this to multivariable functions.

13.1 Single Variable Functions

Suppose f is a differentiable function of a single variable x and suppose $x = g(t)$ for a variable t and g is differentiable. Then f is the composite function $f(g(t))$ or $(f \circ g)(t)$. To differentiate with respect to t one uses the chain rule:

$$\frac{df}{dt} = \frac{df}{dg}\frac{dg}{dt}, \qquad \frac{df}{dt} = \frac{df}{dx}\frac{dx}{dt}, \qquad \frac{df}{dt} = \frac{df}{dx}\frac{dg}{dt} \tag{402}$$

or if you want to be explicit in your arguments

$$\left.\frac{df}{dt}\right|_{t=a} = \frac{df}{dx}(g(a))\frac{dx}{dt}(a). \tag{403}$$

Example 13.1. *Suppose $f(x) = 3x^2$ and $x = g(t) = 2t^2 - 1$. Find $\frac{df}{dt}$.*

We compute using the chain rule

$$\frac{df}{dt} = \frac{df}{dx}\frac{dx}{dt} = (6x)(4t) = 24t(2t^2 - 1). \tag{404}$$

Example 13.2. *Suppose $f(x) = e^{7x}$ and $x = g(t) = \frac{1}{7}\ln(t)$ for $t > 0$. Find $\frac{df}{dt}$.*

We compute using the chain rule

$$\frac{df}{dt} = \frac{df}{dx}\frac{dx}{dt} = (7e^{7x})\Big(\frac{1}{7t}\Big) = 1. \tag{405}$$

13.2 Multivariable Functions: A First Step

Suppose we have a differentiable function of two variables, $f : D \to \mathbb{R}$ where D is a subset of $\mathbb{R}^2$. So f takes (x, y) and evalutates $f(x, y)$. Suppose $x = g(t)$ and $y = h(t)$ are differentiable, so that one has a single variable function in reality. How does one compute df/dt?

Proposition 13.1. *Let f be a differentiable function of two variables (x, y) such that $x = g(t)$ and $y = h(t)$ and g and h are differentiable. Then*

$$\frac{df}{dt} = \partial_x f\frac{dx}{dt} + \partial_y f\frac{dy}{dt}, \qquad \frac{df}{dt} = \partial_g f\frac{dg}{dt} + \partial_h f\frac{dh}{dt}. \tag{406}$$

or if you wish to be explicit in your arguments

$$\left.\frac{df}{dt}\right|_{t=a} = \partial_x f(g(a), h(a))\frac{dx}{dt}(a) + \partial_y f(g(a), h(a))\frac{dy}{dt}(a). \tag{407}$$

Example 13.3. *Let*

$$f(x, y) = xy \tag{408}$$

with $x = \sin(t)$ and $y = \cos(t)$. We can compute

$$\frac{df}{dt} = (y)(\cos t) + (x)(-\sin(t)) = \cos^2(t) - \sin^2(t). \tag{409}$$

Example 13.4. *Let*

$$f(x,y) = x^2y + 3xy^4 \tag{410}$$

with $x = \sin(2t)$ *and* $y = \cos(t)$. *Find* $\frac{df}{dt}(t=0)$. *We can compute*

$$\frac{df}{dt} = (2xy + 3y^4)(2\cos(2t)) + (x^2 + 12xy^3)(-sin(t)). \tag{411}$$

Let's be lazy and not simplify. We can just find $(x(0), y(0)) = (0,1)$ *to give*

$$\left.\frac{df}{dt}\right|_{t=0} = 6. \tag{412}$$

13.3 Multivariable Functions: Adding Complexity

One could complicate things further and have $x = g(s,t)$ and $y = h(s,t)$. In this case we have the following chain rule for the partial derivatives of f:

Proposition 13.2. *Let* f *be a differentiable function of two variables* (x,y) *such that* $x = g(s,t)$ *and* $y = h(s,t)$ *and* g *and* h *are differentiable. Then*

$$\frac{\partial f}{ds} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial s}, \qquad \frac{\partial f}{dt} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial t}, \tag{413}$$

or if you want to be explicit in arguments

$$\left.\frac{\partial f}{ds}\right|_{(s,t)=(a,b)} = \left.\frac{\partial f}{\partial x}\right|_{(g(a),h(b))}\left.\frac{\partial x}{\partial s}\right|_{(a,b)} + \left.\frac{\partial f}{\partial y}\right|_{(g(a),h(b))}\left.\frac{\partial y}{\partial s}\right|_{(a,b)}. \tag{414}$$

Example 13.5. *Suppose* $F(x,y) = xy$ *and* $x = st$ *and* $y = \ln(st)$ *with* $st > 0$ *then*

$$\partial_s F = yt + \frac{x}{st}t = yt + \frac{x}{s} = t\ln(st) + t, \qquad \partial_t F = s\ln(st) + s \tag{415}$$

Example 13.6. *Let* $f(x,y) = e^x\sin(y)$ *and* $x = st^2$, $y = s^2t$.

$$\partial_s f = e^x\sin(y)t^2 + e^x\cos(y)2st = e^{st^2}t^2\sin(s^2t) + 2e^{st^2}st\sin(s^2t) \tag{416}$$

$$\partial_t f = 2ste^x\sin(y) + s^2e^x\cos(y) = 2ste^{st^2}\sin(s^2t) + s^2e^{t^2s}\cos(s^2t). \tag{417}$$

Example 13.7. *Let* $R(s,t) = G(u(s,t), v(s,t))$ *and*

$$u(1,2) = 5, \qquad v(1,2) = 7 \tag{418}$$

$$\partial_s u(1,2) = 4, \qquad \partial_t u(1,2) = -3 \tag{419}$$

$$\partial_s v(1,2) = 2, \qquad \partial_t v(1,2) = 6 \tag{420}$$

$$\partial_u G(5,7) = 9, \qquad \partial_v G(5,7) = -2. \tag{421}$$

Compute $\partial_s R(1,2)$ *and* $\partial_t R(1,2)$.

13.4 Multivariable Functions: Generality

Let's do the general case:

Proposition 13.3. *Let* f *be a differentiable function of* n*-variables* $x_1, ..., x_n$. *Suppose that* $x_1, ..., x_n$ *are given by differentiable functions of* m*-variables* $y_1, ..., y_m$, *i.e.* $x_i = g_i(y_1, ..., y_m)$. *Then*

$$\frac{\partial f}{\partial y_i} = \sum_{k=1}^{n}\frac{\partial f}{\partial x_k}\frac{\partial x_k}{\partial y_i}, \qquad \frac{\partial f}{\partial y_i} = \sum_{k=1}^{n}\frac{\partial f}{\partial g_k}\frac{\partial g_k}{\partial y_i}, \tag{422}$$

for all $i = 1, ..., m$.

Example 13.8. *Suppose* $u(x,y,z) = x^4y + y^2z^3$, *where* $x = rse^t$ *and* $y = rs^2e^{-t}$ *and* $z = r^2s\sin(t)$. *Find* $\partial_s u$ *at* $(r,s,t) = (2,1,0)$.

Using the chain rule one has

$$\begin{aligned}\partial_s u &= \partial_x u \partial_s x + \partial_y u \partial_s y + \partial_z u \partial_s z && (423)\\ &= 4x^3yre^t + (x^4 + 2yz^3)2rse^{-t} + 3yz^2r^2\sin(t). && (424)\end{aligned}$$

Evaluating at $(r,s,t) = (2,1,0)$ *or equivalently* $(x,y,z) = (2,2,0)$ *one has*

$$\partial_s u(2,1,0) = 192. \tag{425}$$

Example 13.9. *Suppose* $u(t,r) = \sin(t - kr)$ *for a constant* k, $r = ||\mathbf{x}||$ *where* $\mathbf{x} = (x,y,z)$ *and* $t = t$. *Compute:*

$$k^2\partial_t^2 u - \partial_z^2 u - \partial_y^2 u - \partial_x^2 u. \tag{426}$$

Computing directly gives

$$\partial_z u = \partial_t u \partial_z t + \partial_r u \partial_z r = -k\cos(t - kr)\frac{z}{||\mathbf{x}||}. \tag{427}$$

Taking a second derivative using the product rule gives

$$\partial_z^2 u = -k(\partial_z \cos(t - kr))\frac{z}{||\mathbf{x}||} - k\cos(t - kr)\partial_z\Big(\frac{z}{||\mathbf{x}||}\Big). \tag{428}$$

Using the chain rule for multivariable functions on the first term and the chain rule for functions of a single variable on the second term gives,

$$\partial_z^2 u = -\frac{k^2z^2}{||\mathbf{x}||^2} + \frac{k(z^2 - ||\mathbf{x}||^2)\cos(t - kr)}{||\mathbf{x}||^3} \tag{429}$$

The function is symmetric under $z \leftrightarrow x$ *and* $z \leftrightarrow y$, *so*

$$\begin{aligned}\partial_x^2 u &= -\frac{k^2x^2}{||\mathbf{x}||^2} + \frac{k(x^2 - ||\mathbf{x}||^2)\cos(t - kr)}{||\mathbf{x}||^3} && (430)\\ \partial_y^2 u &= -\frac{k^2y^2}{||\mathbf{x}||^2} + \frac{k(y^2 - ||\mathbf{x}||^2)\cos(t - kr)}{||\mathbf{x}||^3}. && (431)\end{aligned}$$

Additionally,

$$\partial_t^2 u = -\sin(t - kr). \tag{432}$$

Therefore,

$$k^2\partial_t^2 u - \partial_z^2 u - \partial_y^2 u - \partial_x^2 u = 0. \tag{433}$$

13.5 Implicit Functions

One cannot always solve $F(x,y) = 0$ for y. We then view F as an implicit definition of y as a function of x, i.e.

$$F(x, y(x)) = 0. \tag{434}$$

How do we find $\frac{dy}{dx}$?

Proposition 13.4. *Let y be an implicitly defined function of x by*

$$F(x, y) = 0. \tag{435}$$

Suppose F and y are differentiable and $\partial_y F \neq 0$ then

$$\frac{dy}{dx} = -\frac{\partial_x F}{\partial_y F}. \tag{436}$$

Proof. Writing $F(x, y(x)) = 0$ and using the chain rule gives

$$\frac{dF}{dx} = 0 = \partial_x F + \partial_y F \frac{dy}{dx}. \tag{437}$$

□

This generalises to the case where z is an implicitly defined function of (x, y) through

$$F(x, y, z) = 0. \tag{438}$$

Suppose F and z are differentiable and $\partial_z F \neq 0$ then

$$\frac{\partial z}{\partial x} = -\frac{\partial_x F}{\partial_z F}, \qquad \frac{\partial z}{\partial y} = -\frac{\partial_y F}{\partial_z F}. \tag{439}$$

Example 13.10. *Suppose $F(x, y, z) = x^3 + y^3 + z^3 + 6xyz - 1 = 0$. Find $\partial_x z$ and $\partial_y z$.*

$$\partial_x F = 3x^2 + 6yz, \quad \partial_y F = 3y^2 + 6yz, \quad \partial_z F = 6xy + 3z^2. \tag{440}$$

Therefore,

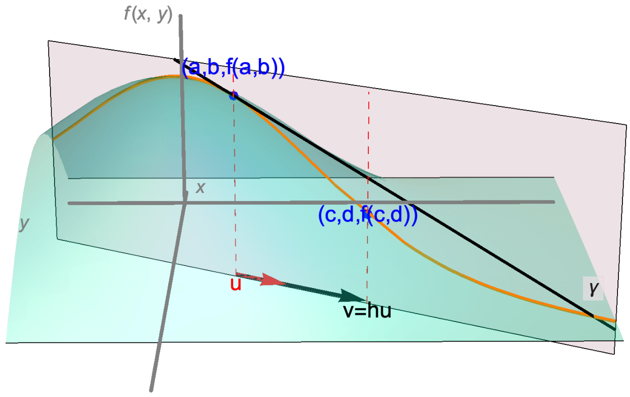
$$\partial_x z = -\frac{x^2 + 2yz}{2xy + z^2}, \quad \partial_x y = -\frac{y^2 + 2xz}{2xy + z^2}. \tag{441}$$

14 Directional Derivatives and the Gradient Vector

As the name suggests directional derivatives are derivatives of a function f in a direction specified by a vector $\mathbf{u}$.

14.1 Introduction and Definition

Suppse f is a function of two variables (x, y) and $\mathbf{u} = (u_1, u_2)$ is a **unit** vector in the xy-plane. The surface defined by the equation $z = f(x, y)$ is plotted below:



Let p be a point on this surface with Cartesian coordinates $\mathbf{x}_0 = (a, b, f(a, b))$. Consider intersecting the surface with the plane containing $\mathbf{x}_0$, $\mathbf{u}$ and $\mathbf{k}$ (the unit vector in the z-direction), i.e. the plane with equation,

$$\langle \mathbf{n}, \mathbf{x} - \mathbf{x}_0 \rangle, \quad \mathbf{n} = \mathbf{u} \times \mathbf{k}. \tag{442}$$

This produces a curve γ lying in the surface defined by $z = f(x, y)$ along which $\mathbf{u}$ points. Let q be a points on this curve with Cartesian coordinates $(c, d, f(c, d))$. The projections of p and q to the xy-plane have Cartesian coordinates (a, b) and (c, d) respectively. The displacement vector $\mathbf{v}$ (the black arrow on the above diagram) from the projection of p to the projection of q is proportional to $\mathbf{u}$ by a positive scalar multiple, i.e.

$$\mathbf{v} = h\mathbf{u} = (hu_1, hu_2). \tag{443}$$

Consider the new function

$$g(h) = \frac{f(a + hu_1, b + hu_2) - f(a, b)}{h}. \tag{444}$$

This ratio is the change in height in z as you move along γ over the change in length in the xy-plane. Taking the limit of g as $h \to 0$ gives the rate of change of f in the direction of $\mathbf{u}$. We make the following definition:

Definition 14.1. *Let f be a function of two vartiables (x, y). Then the directional derivative of f at (a, b) in the direction of the unit vector $\mathbf{u} = (u_1, u_2)$ is*

$$D_{\mathbf{u}} f(a, b) = \lim_{h \to 0} \frac{f(a + hu_1, b + hu_2) - f(a, b)}{h}, \tag{445}$$

if the limit on the right-hand side exists.

Remark 14.1. *This generalises to n-variables*

$$D_{\mathbf{u}} f(a_1, ..., a_n) = \lim_{h \to 0} \frac{f(a_1 + hu_1, a_2 + hu_2, ..., a_n + hu_n) - f(a_1, a_2, ..., a_n)}{h}, \tag{446}$$

Example 14.1. *Note that even all directional derivatives existing does not imply differentiablility. For example, let*

$$f(x,y) \doteq \begin{cases} \frac{x^3}{x^2+y^2} & (x,y) \neq (0,0) \\ 0 & (x,y) = (0,0). \end{cases} \tag{447}$$

Now,

$$D_{\mathbf{u}}f(0,0) = \lim_{h\to 0} \frac{f(hu_1,hu_2) - f(0,0)}{h} = \lim_{h\to 0} \frac{(hu_1)^3}{h[(hu_1)^2 + (hu_2)^2]} = \frac{u_1^3}{u_1^2 + u_2^2}. \tag{448}$$

So all directional derivatives exist at $(0,0)$. In particular, by take $\mathbf{u} = (1,0)$ and $\mathbf{u} = (0,1)$, one finds $\nabla f = (1,0)$.

So, to check the differentiablility of f one has to check the following limit vanishes:

$$\lim_{\mathbf{h}\to\mathbf{0}} \frac{|f(h_1,h_2) - f(0,0) - \langle \nabla f(0,0), \mathbf{h}\rangle|}{||\mathbf{h}||} = \lim_{\mathbf{h}\to\mathbf{0}} \frac{|h_1|h_2^2}{(h_1^2+h_2^2)^{\frac{3}{2}}}. \tag{449}$$

Take $h_1 = mh_2$ then

$$\frac{|h_1|h_2^2}{(h_1^2+h_2^2)^{\frac{3}{2}}} = \frac{|m||h_2|^3}{((m^2+1)h_2^2)^{\frac{3}{2}}} = \frac{|m|}{(m^2+1)^{\frac{3}{2}}} \neq 0. \tag{450}$$

Therefore, the limit in definition differentiablility does not vanish and hence, f is not differentiable.

14.2 The Relation Between Directional Derivatives and the Gradient Vector

If f is differentiable then we have a nice relation to the gradient vector above:

Proposition 14.1. *Suppose f is a differentiable function of n-variables $(x_1, ..., x_n)$ and $\mathbf{u}$ is a unit vector in $\mathbb{R}^n$. Then all directional derivatives exist and*

$$D_{\mathbf{u}}f = \langle \nabla f, \mathbf{u}\rangle. \tag{451}$$

Proof. (Non-Examinable) We will prove this in $\mathbb{R}^2$. If f is differentiable one has

$$\lim_{\mathbf{h}\to 0} \frac{|f(a_1+h_1, a_2+h_2) - f(a_1,a_2) - \langle \nabla f, \mathbf{h}\rangle|}{||\mathbf{h}||} = 0. \tag{452}$$

Set $\mathbf{h} = h\mathbf{u}$ where $\mathbf{u}$ is unit. Then one has

$$\lim_{h\to 0} \left| \frac{f(a_1+hu_1, a_2+hu_2) - f(a_1,a_2)}{h} - \langle \nabla f, \mathbf{u}\rangle \right| = 0, \tag{453}$$

since $||\mathbf{h}|| = h||\mathbf{u}|| = h$. Now this say that for all $\epsilon > 0$, there exists a $\delta > 0$ such that if $|h - 0| < \delta$ then

$$\left| \left| \frac{f(a_1+hu_1, a_2+hu_2) - f(a_1,a_2)}{h} - \langle \nabla f, \mathbf{u}\rangle \right| - 0 \right| < \epsilon, \tag{454}$$

which is equivalent to

$$\left| \frac{f(a_1+hu_1, a_2+hu_2) - f(a_1,a_2)}{h} - \langle \nabla f, \mathbf{u}\rangle \right| < \epsilon, \tag{455}$$

which by definition says that

$$\lim_{h\to 0} \frac{f(a_1+hu_1, a_2+hu_2) - f(a_1,a_2)}{h} = \langle \nabla f, \mathbf{u}\rangle. \tag{456}$$

$\square$

Example 14.2. *Find the directional derivative of the function $f(x, y) = x^3 - 3xy + 4y^2$ in the direction $\mathbf{u} = (1, \frac{1}{2})$. Evaluate $D_{\hat{\mathbf{u}}}f(1, 2)$.*

Let's now compute the partial derivatives:

$$\partial_x f = 3x^2 - 3y \tag{457}$$
$$\partial_y f = -3x + 8y. \tag{458}$$

These are continuous and therefore f is differentiable.

Since f is differentiable one has from proposition 14.1,

$$D_{\hat{\mathbf{u}}}f = \langle \nabla f, \hat{\mathbf{u}} \rangle \tag{459}$$

So let's make $\mathbf{u}$ a unit vector. Its length is $||\mathbf{u}|| = \frac{\sqrt{5}}{2}$. Therefore, $\hat{\mathbf{u}} = (2/\sqrt{5}, 1/\sqrt{5})$ and

$$D_{\hat{\mathbf{u}}}f = \langle (3x^2 - 3y, -3x + 8y), (2/\sqrt{5}, 1/\sqrt{5}) \rangle = 2/\sqrt{5}(3x^2 - 3y) + 1/\sqrt{5}(-3x + 8y) \tag{460}$$
$$= \frac{6}{\sqrt{5}}x^2 - \frac{3}{\sqrt{5}}x + \frac{2}{\sqrt{5}}y. \tag{461}$$

Evaluating at $(1, 2)$ gives

$$D_{\mathbf{u}}f(1, 2) = \frac{7}{\sqrt{5}}. \tag{462}$$

14.3 Properties of the Gradient Vector

We have already seen one interesting property of the gradient vector. Namely for differentiable functions,

$$D_{\mathbf{u}}f = \langle \nabla f, \mathbf{u} \rangle. \tag{463}$$

Another interesting property follows from considering level sets/curves/surfaces. Suppose we consider a level curve γ of a continuous function of two variables $f(x, y)$, i.e. we look at

$$\{(x, y) \in \mathbb{R}^2 : f(x, y) = k, k = \text{const}\}. \tag{464}$$

The level curve γ can expressed as a continuous vector-valued function $\mathbf{r}(t) = (x = r_1(t), y = r_2(t))$. So we have

$$g(t) = f(r_1(t), r_2(t)) = k. \tag{465}$$

We note that $dg/dt = 0$ and suppose $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$ exist. We now compute $df(r_1(t), r_2(t))/dt$ using the chain rule:

$$0 = \frac{\partial f}{\partial x}\frac{dr_1}{dt} + \frac{\partial f}{\partial y}\frac{dr_2}{dt} = \langle \nabla f, \mathbf{r}'(t) \rangle. \tag{466}$$

The vector $\mathbf{r}'(t)$ is tangent to the curve γ. Hence ∇f must be orthogonal to γ, i.e. **the gradient vector is orthogonal to the level curves**. Note that this generalised to level surfaces and general level sets.

The final interesting property of the gradient vector that we will mention here is that $\nabla f(\mathbf{x})$ points in the direction of maximum rate of increase of f at $\mathbf{x}$, and that maximum rate of change is $||\nabla f(\mathbf{x})||$. Precisely:

Proposition 14.2. *Suppose f is a differentiable of n-variables. The maximum value of the directional derivative $D_{\mathbf{u}}f(\mathbf{x})$ is $||\nabla f(\mathbf{x})||$ and it occurs when the unit vector $\mathbf{u}$ is in the same direction as ∇f.*

Proof. From propositions 14.1 and 3.2 one has

$$D_{\mathbf{u}}f = \langle \nabla f, \mathbf{u} \rangle = ||\mathbf{u}||||\nabla f|| \cos\theta = ||\nabla f|| \cos\theta. \tag{467}$$

The left-hand side is maximised when $\theta = 0$. Hence, $D_{\mathbf{u}}f$ has maximimum $||\nabla f||$ and this occurs when $\theta = 0$, i.e. when $\mathbf{u}$ and ∇f are in the same direction. □

15 Extrema: Maxima and Minima

15.1 Review of Single Variables

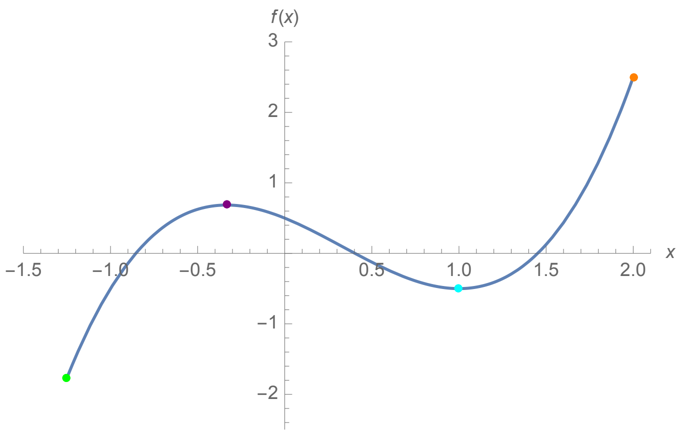
Here we briefly recall the notions associated to extrema of functions of a single variable.

Definition 15.1 (Maxima/Minima/Extrema)**.** *Let $f : D \to \mathbb{R}$ where D is a subset of $\mathbb{R}$. We have the following notions of maxima and minima (collectively known as extrema):*

- *The point $a \in D$ is said to be a local maximum for f if there exists an $\epsilon > 0$ such that $f(x) \leq f(a)$ for all $x \in (a - \epsilon, a + \epsilon)$.*
- *The point $a \in D$ is said to be a local minimum for f if there exists an $\epsilon > 0$ such that $f(x) \geq f(a)$ for all $x \in (a - \epsilon, a + \epsilon)$.*
- *The point $a \in D$ is said to be a global maximum for f if $f(x) \leq f(a)$ for all $x \in D$.*
- *The point $a \in D$ is said to be a global minimum for f if $f(x) \geq f(a)$ for all $x \in D$.*

One can place the word strict in front of these notions if the inequality is strict.

This definition is illustrated with the following picture:



Here we have a function on the interval $D = [-\frac{5}{4}, 2]$.

- The point in purple is a local max.
- The point in cyan is a local min.
- The point in orange is a global max.
- The point in green is a global min.

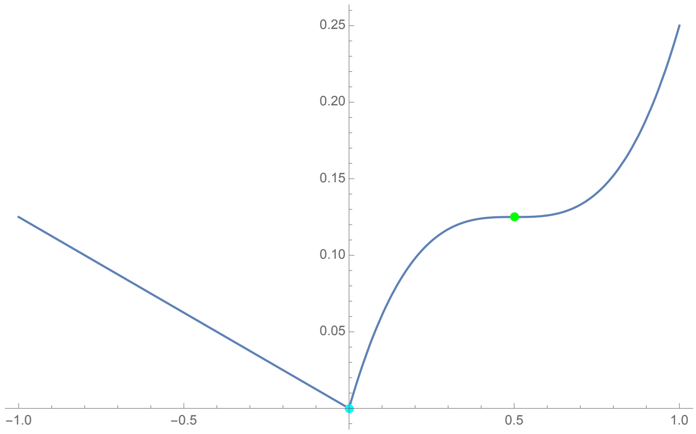
Recall also the following definition:

Definition 15.2 (Critical/Stationary Point)**.** *Let $\mathrm{dom}(f)$ be a subset of $\mathbb{R}$. A stationary point of a function $f : \mathrm{dom}(f) \to \mathbb{R}$ is a number $a \in \mathbb{R}$ such that $f'(a) = 0$. A critical point $a \in \mathbb{R}$ of f is a stationary point or a point where $f'(a)$ does not exist.*

Example 15.1. *Let*

$$f(x) = \begin{cases} -\frac{1}{8}x & x < 0 \\ \left(x - \frac{1}{2}\right)^3 + \frac{1}{8} & x \geq 0. \end{cases} \tag{468}$$

This is plotted below:



This function is continuous on $\mathbb{R}$: observe that $\lim_{x\to 0} f(x) = 0 = f(0)$. For $x \neq 0$, the functions derivative is the following

$$f'(x) = \begin{cases} -\frac{1}{8} & x < 0 \\ 3\left(x - \frac{1}{2}\right)^2 & x > 0. \end{cases} \tag{469}$$

Therefore is has a stationary point at $x = \frac{1}{2}$ (plotted in green). For $x = 0$ then

$$\lim_{h\to 0^+} \frac{f(h) - f(0)}{h} = \lim_{h\to 0^+} \frac{(h - \frac{1}{2})^3 + \frac{1}{8} - 0}{h} = \frac{3}{4}. \tag{470}$$

$$\lim_{h\to 0^-} \frac{f(h) - f(0)}{h} = \lim_{h\to 0^-} \frac{-\frac{1}{8}h - 0}{h} = -\frac{1}{8}. \tag{471}$$

Therefore, $f'(0)$ doesn't exist, which means its a critical point, shown in cyan in the figure above.

15.1.1 Local Extrema

The following theorem of Fermat gives a nessecary condition for finding a local min/max for differentiable functions via stationary points:

Theorem 15.1 (Interior Extremum Theorem)**.** *Suppose f has a local minimum or maximum at a. If f is differentiable at a then $f'(a) = 0$, i.e. a is a stationary point.*

This can be used to attempt to identify local maximums and minimums for functions that are differentiable. You should think that it narrows down the points we need to examine as potential local min/max. For example,

Example 15.2. *Let $f(x) = x^3 - 2x^2 + 4$. The function f is differentiable everywhere with*

$$f'(x) = 3x^2 - 4x. \tag{472}$$

To find the stationary points we consider $3x^2 - 4x = 0$, which gives the stationary points $x = 0, x = \frac{4}{3}$.

What this doesn't tell us is whether the stationary points are local maximum/minimums or something else. There are a few ways to check, the most common is the second derivative test:

Proposition 15.1 (Second Derivative Test)**.** *Suppose f'' is continuous near a. If a is a stationary point*

- *and $f''(a) > 0$, then f has a local min at a.*
- *and $f''(a) < 0$, then f has a local max at a.*
- *and if $f''(a) = 0$, then the test is inconclusive.*

Let's return to our example

Example 15.3. *Let* $f(x) = x^3 - 2x^2 + 4$*. We've found*

$$f'(x) = 3x^2 - 4x. \tag{473}$$

You can compute that

$$f''(x) = 6x - 4, \tag{474}$$

which is continuous. So, evaluting at the stationary points $x = 0, x = \frac{4}{3}$ *gives*

$$f''(0) = -4 < 0 \implies (0, f(0)) \text{ is a max.} \tag{475}$$

$$f''(\frac{4}{3}) = 4 > 0 \implies (\frac{4}{3}, f(\frac{4}{3})) \text{ is a min..} \tag{476}$$

What's so inconclusive about the case $f''(0) = 0$? Here's a little bit of context:

Definition 15.3. *Suppose* f *is a function of a single variable* x *which is twice differentiable with continuous second derivative on its domain* $\mathrm{dom}(f)$*. A point* $a \in \mathrm{dom}(f)$ *is said to be a point of inflection of* $f''(x)$ *changes sign at* a*, i.e.* $f''(a) = 0$*,* $f''(x) > 0$ *(*$f''(x) < 0$*) for* $x < a$ *and* $f''(x) < 0$ *(*$f''(x) > 0$*) for* $x > a$*.*

Example 15.4. *Let's look at three examples:*

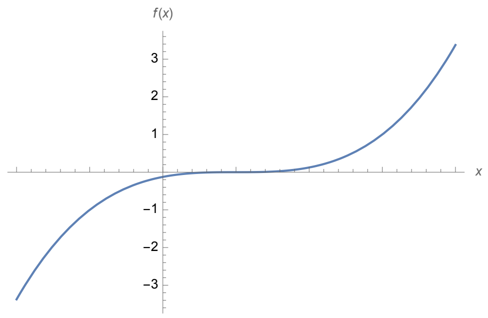
1. *Let*

$$f(x) = (x - \frac{1}{2})^3. \tag{477}$$

Then

$$f'(x) = 3(x - \frac{1}{2})^2, \qquad f''(x) = 6(x - \frac{1}{2}). \tag{478}$$

So, $f'(x) = 0$ *implies* $x = \frac{1}{2}$ *is a stationary point, with* $f''(\frac{1}{2}) = 0$*. This is in fact a point of inflection (and* **not** *a maxima or minima) as shown in the following plot:*



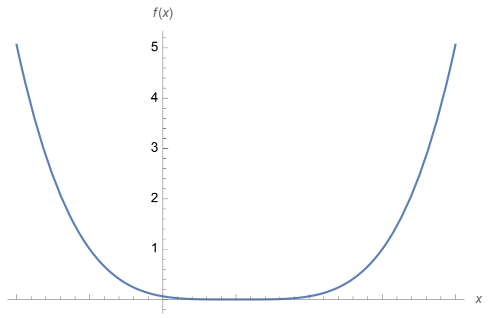
2. *Let*

$$f(x) = (x - \frac{1}{2})^4. \tag{479}$$

Then

$$f'(x) = 4(x - \frac{1}{2})^3, \qquad f''(x) = 12(x - \frac{1}{2})^2. \tag{480}$$

So, $f'(x) = 0$ *implies* $x = \frac{1}{2}$ *is a stationary point, with* $f''(\frac{1}{2}) = 0$*. However this is not a point of inflection since* $f''(x) \geq 0$ *for all* x*. In fact* $x = \frac{1}{2}$ *is a minima:*



3. *Finally, a point where $f'' = 0$ does not need to be a stationary point. For example, take*

$$f(x) = -\sin(x). \tag{481}$$

This gives

$$f'(x) = -\cos(x), \qquad f''(x) = \sin(x). \tag{482}$$

We have $f''(n\pi) = 0$ for n integer and $f''(x)$ changes sign at $n\pi$, i.e. $n\pi$ is a point of inflection. However, $f'(n\pi) = -1(\pm 1)^n \neq 0$. So, it is not a stationary point and certainly not a maxima or minima of $-\sin(x)$.

15.1.2 Global Extrema

The following theorem characterises when one can find an absolute maxima or minima for a function:

Theorem 15.2 (Extreme Value Theorem). *Suppose $f : [a,b] \to \mathbb{R}$ is continuous. Then f attains an absolute maximum $f(c)$ and absolute minimum $f(d)$ for some $c, d \in [a,b]$.*

Remark 15.1. *Relax closedness of the interval or continuity and this theorem is not true.*

To go from local maximum/minimums to absolute maximum/minimums of a continuous function f on a closed interval $[a,b]$ we have the following method:

- Find the critical points of f in (a,b) by considering f': call them $c_1, ..., c_m \in \mathbb{R}$ and evaluate f at $c_1, ..., c_m$, i.e. compute $f(c_1), ..., f(c_m)$.
- Evaluate $f(a)$ and $f(b)$.
- The absolute maximum of f on $[a,b]$ is then

$$\max(f(a), f(b), f(c_1), ..., f(c_m)). \tag{483}$$

- The absolute minimum of f on $[a,b]$ is then

$$\min(f(a), f(b), f(c_1), ..., f(c_m)). \tag{484}$$

Example 15.5. *Take $f(x) = |x|$ on $[-1,1]$. One has the following derivative*

$$f'(x) = \begin{cases} 1 & x > 0 \\ -1 & x < 0 \\ \text{undefined} & x = 0. \end{cases} \tag{485}$$

Hence, there are no stationary points but there is a critical point at $x = 0$, for which $f(0) = 0$. We can evaluate at the end points $f(-1) = 1 = f(1)$. Therefore, the absolute max of f is

$$\max(1,0) = 1, \tag{486}$$

and the absolute min of f is

$$\max(1,0) = 0. \tag{487}$$

15.2 Local Extrema of Functions of Two Variables

We can generalise our notion of local max/min to 2-variables. Effectively, a local max is a point $\mathbf{a}$ in $\mathbb{R}^2$ such that $f(\mathbf{a})$ is greater than f as any point 'near' $\mathbf{a}$. Similarly, a local min is a point $\mathbf{a}$ in $\mathbb{R}^2$ such that $f(\mathbf{a})$ is less than f as any point 'near' $\mathbf{a}$. We formalise this as follows:

Definition 15.4 (Local Maxima/Minima)**.** *Let $f : D \to \mathbb{R}$ where D is a subset of $\mathbb{R}^2$. We have the following notions of maxima and minima:*

- *The point $\mathbf{a} \in D$ is said to be a local maximum for f if there exists an $\epsilon > 0$ such that $f(\mathbf{x}) \leq f(\mathbf{a})$ for all $\mathbf{x} \in \{\mathbf{x} \in \mathbb{R}^2 : ||\mathbf{x} - \mathbf{a}|| < \epsilon\}$.*
- *The point $\mathbf{a} \in D$ is said to be a local minimum for f if there exists an $\epsilon > 0$ such that $f(\mathbf{x}) \geq f(\mathbf{a})$ for all $\mathbf{x} \in \{\mathbf{x} \in \mathbb{R}^2 : ||\mathbf{x} - \mathbf{a}|| < \epsilon\}$.*

One can place the word strict in front of these notions if the inequality is strict.

The notions of critical and stationary points generalise as:

Definition 15.5 (Critical/Stationary Point)**.** *Let $\mathrm{dom}(f)$ be a subset of $\mathbb{R}^2$. A stationary point of a function $f : \mathrm{dom}(f) \to \mathbb{R}$ is a point $\mathbf{a} \in \mathbb{R}$ such that $\partial_x f(\mathbf{a}) = 0$ and $\partial_y f(\mathbf{a}) = 0$. A critical point $\mathbf{a} \in \mathbb{R}^2$ of f is a stationary point or a point where $\partial_x f(\mathbf{a})$ and/or $\partial_y f(\mathbf{a})$ does not exist.*

We have the generalisation of theorem 15.1:

Theorem 15.3. *Suppose f has a local max/min at (a, b) and $\partial_x f(a, b)$ and $\partial_y f(a, b)$ exist. Then (a, b) is a stationary point.*

Proof. If f has a local max/min at (a, b) then $g(x) = f(x, b)$ has a local max/min at a. By theorem 15.1 $g'(a) = 0$, which implies $\partial_x f(a, b) = 0$. Similarly, if f has a local max/min at (a, b) then $h(y) = f(a, y)$ has a local max/min at b then by theorem 15.1 $h'(b) = 0$, which implies $\partial_y f(a, b) = 0$. □

The generalisation of the second derivative test for single variable functions is

Proposition 15.2. *Suppose the second partial derivatives of f are continuous on a disk with center (a, b), and suppose that (a, b) is a stationary point of f. Further, let*

$$\mathcal{D}(a,b) = \left| \begin{pmatrix} \partial_x^2 f(a,b) & \partial_{xy}^2 f(a,b) \\ \partial_{xy}^2 f(a,b) & \partial_y^2 f(a,b) \end{pmatrix} \right| = \partial_x^2 f(a,b)\partial_y^2 f(a,b) - [\partial_{xy}^2 f(a,b)]^2. \tag{488}$$

The we have the following cases:

- *If $\mathcal{D}(a, b) > 0$ and $\partial_x^2 f(a, b) > 0$, then $f(a, b)$ is a local min.*
- *If $\mathcal{D}(a, b) > 0$ and $\partial_x^2 f(a, b) < 0$, then $f(a, b)$ is a local max.*
- *If $\mathcal{D}(a, b) < 0$, then $f(a, b)$ is a saddle point.*
- *If $\mathcal{D}(a, b) = 0$, then the test is inconclusive.*

Example 15.6. *Suppose $f(x, y) = axy - bx^2 - cy^2$ with $a \neq 0$ and $b, c > 0$. This function has partial derivatives*

$$\partial_x f = ay - 2bx, \qquad \partial_y f = ax - 2cy. \tag{489}$$

To find the stationary points we look for $\partial_x f = 0$ and $\partial_y f = 0$. One can solve the first of these for $x = \frac{a}{2b}y$ which gives

$$\left(\frac{a^2}{2b} - 2c\right)y = 0. \tag{490}$$

So either $a^2 - 4bc = 0$ or $y = 0$. So one has two cases:

1. *If $a^2 - 4bc \neq 0$ then $(x, y) = (0, 0)$ is a stationary point.*
2. *If $a^2 - 4bc = 0$ then $(x, y) = (\frac{a}{2b}y, y)$ is a line of stationary points.*

Let's find the second deritivatives:

$$\partial_x^2 f = -2b, \qquad \partial_y^2 f = -2c, \qquad \partial_{xy}^2 f = a. \tag{491}$$

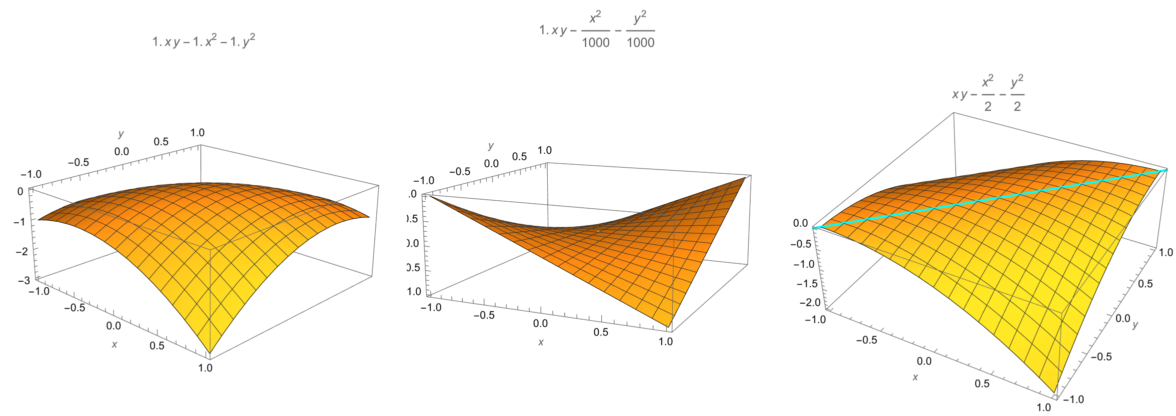
So

$$\mathcal{D} = \partial_x^2 f \partial_y^2 f - (\partial_{xy}^2 f)^2 = 4bc - a^2. \tag{492}$$

Let's compute our cases:

1. *If $a^2 - 4bc < 0$ then $(x, y) = (0, 0)$ is a stationary point, $\mathcal{D}(0, 0) > 0$ and $\partial_x^2 f(0, 0) < 0$ so we have a maximum at $(0, 0)$ by the second derivative test.*
2. *If $a^2 - 4bc > 0$ then $(x, y) = (0, 0)$ is a stationary point, $\mathcal{D}(0, 0) < 0$, so we have a saddle point at $(0, 0)$ by the second derivative test.*
3. *If $a^2 - 4bc = 0$ then $(x, y) = (\frac{a}{2b}y, y)$ is a line of stationary points, but the second derivative test is inconclusive. One can show that this is a line of maximums.*

Here is the plots of the three cases (1,2,3 from left to right):



Example 15.7. *Let*

$$f(x, y) = \frac{x^2 + y^2 + xy - \frac{1}{2}x}{x^2 + y^2}. \tag{493}$$

Note that its domain of definition is $\{(x, y) : x, y \neq 0\}$. To look for stationary points in its domain we compute:

$$\partial_x f = \frac{(2y-1)(y^2 - x^2)}{(x^2 + y^2)^2}, \qquad \partial_y f = \frac{x(x^2 + y - y^2)}{(x^2 + y^2)^2} \tag{494}$$

So,

$$\partial_x f = 0 \implies y = \frac{1}{2} \qquad y = \pm x. \tag{495}$$

Therefore,

$$\partial_y f(x, \frac{1}{2}) = \frac{4x}{1 + x^2}, \implies x = 0 \tag{496}$$

and

$$\partial_y f(x, \pm x) = \frac{x(x^2 \pm x - x^2)}{2x^2} = \pm\frac{1}{4x^2} \neq 0. \tag{497}$$

So, $(x, y) = (0, \frac{1}{2})$ is a stationary point of f.

Lets check its second derivative

$$\partial_x^2 f = \frac{x(2y-1)(x^2-3y^2)}{(x^2+y^2)^3}, \qquad \partial_y^2 f = \frac{x^3(1-6y)+xy^2(2y-3)}{(x^2+y^2)^3}, \tag{498}$$

and

$$\partial_{xy}^2 f = -\frac{x^4+3x^2(1-2y)y+(y-1)y^3}{(x^2+y^2)^3}. \tag{499}$$

So,

$$\mathcal{D}\Big(0, \frac{1}{2}\Big) = -4 < 0. \tag{500}$$

Therefore, $(x, y) = (0, \frac{1}{2})$ is a saddle point.

Example 15.8. *Suppose*

$$f(x, y) = \arctan(xy). \tag{501}$$

Note the following trick if you forget the derivative of $\arctan$. Let $w(z) = \arctan(z)$. Therefore, from the chain rule

$$z = \tan(w(z)) \implies 1 = \frac{dz}{dz} = \frac{d\tan(w)}{dw}\frac{dw}{dz}. \tag{502}$$

Now,[9]

$$\tan(w) = \frac{\sin(w)}{\cos(w)}. \tag{503}$$

So, from the product rule

$$\frac{d}{dw}\tan(w) = \frac{\frac{d}{dw}\sin(w)}{\cos(w)} + \sin(w)\frac{d}{dw}\frac{1}{\cos(w)} = 1 + \frac{\sin^2(w)}{\cos^2(w)} = 1 + \tan^2(w). \tag{504}$$

Therefore,

$$\frac{dz}{dz} = 1 = (1+\tan^2(z))\frac{dw}{dz} \implies \frac{dw}{dz} = \frac{1}{1+z^2}. \tag{505}$$

So,

$$\frac{d}{dz}\arctan(z) = \frac{1}{1+z^2}. \tag{506}$$

So, returning to the maxima and minima of $\arctan(xy)$ we can search for stationary points by considering

$$\partial_x f = \frac{y}{1+(xy)^2}, \qquad \partial_y f = \frac{x}{1+(xy)^2} \tag{507}$$

using the chain rule. The only stationary point is $(x, y) = (0, 0)$ and there are no critical points. Let's attempt the second derivative test,

$$\partial_x^2 f = -\frac{2y^3x}{(1+(xy)^2)^2}, \qquad \partial_y^2 f = -\frac{2x^3y}{(1+(xy)^2)^2}, \qquad \partial_{xy}^2 f = \frac{1-x^2y^2}{(1+x^2y^2)^2}. \tag{508}$$

So,

$$\mathcal{D}(0, 0) = -1 < 0. \tag{509}$$

Therefore, we have a saddle point at $(x, y) = (0, 0)$.

[9] I remember this with a very stupid rule: $\tan = \frac{\text{sheep}}{\text{cows}}$ because you could happily stack a sheep on a cow since it's smaller but a cow would crush the sheep.

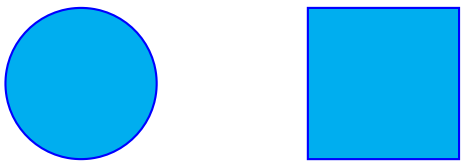
15.3 Global Extrema of Functions of Two Variables

Definition 15.6 (Global Maxima/Minima). *Let $f : D \to \mathbb{R}$ where D is a subset of $\mathbb{R}^2$. We have the following notions of maxima and minima:*

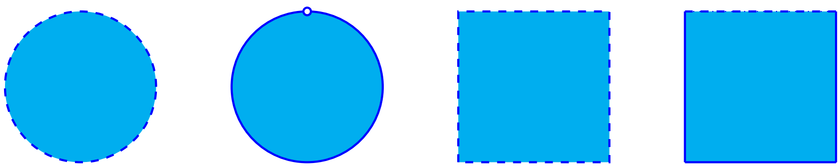
- *The point $\mathbf{a} \in D$ is said to be a global maximum for f if $f(\mathbf{x}) \leq f(\mathbf{a})$ for all $\mathbf{x} \in D$.*
- *The point $\mathbf{a} \in D$ is said to be a global minimum for f if $f(\mathbf{x}) \geq f(\mathbf{a})$ for all $\mathbf{x} \in D$..*

One can place the word strict in front of these notions if the inequality is strict.

What happens to the extreme value theorem 15.2 for function of two variables. We need a notion of closedness for sets in $\mathbb{R}^2$ to replace the closed interval $[a, b]$. Effectively, what one requires is that the set in $\mathbb{R}^2$ contains all points on its boundary. For example,



would be closed where as,



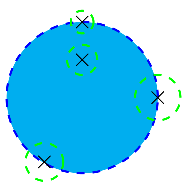
would be not be closed. For this course you can associated the term **closed** with **contains all boundary points**.

Non-Examinable: If you're interested in making this precise read on. The relevant definition to capture the idea of boundary points is this:

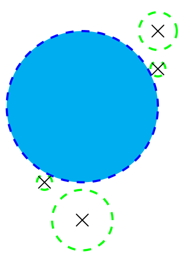
Definition 15.7. *Let S be a set in $\mathbb{R}^2$, $\mathbf{x} \in \mathbb{R}^2$ is a* **limit point** *of S if for all $\epsilon > 0$ there exists $\mathbf{y} \in S$ such that*

$$\mathbf{y} \in \{(x, y) \in \mathbb{R}^2 : ||\mathbf{x} - \mathbf{y}|| < \epsilon\}. \tag{510}$$

The following would be a limit points of S



The following would **not** be a limit point of S:



We then characterise closedness with the condition that all boundary points must be included:

Definition 15.8. *A set S is* **closed** *if S contains all its limit points.*

Something slightly problematic is that $\mathbb{R}$ as a subset of $\mathbb{R}^2$ (i.e. the x-axis say) is closed. The property we need to capture from the extreme value theorem for single variable functions is the fact we were using an **interval**, which is finite or 'bounded'.

Definition 15.9. *Let S be a set in $\mathbb{R}^2$. The set S is bounded if S is contained in a disk/ball of radius $R < \infty$:*

$$B_R = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq R^2\}. \tag{511}$$

Theorem 15.4. *If f is continuous on a closed and bounded set D in $\mathbb{R}^2$, then f attains an absolute maximum value $f(x_1, y_1)$ and an absolute minimum $f(x_2, y_2)$ at some points (x_1, y_1) and (x_2, y_2) in D.*

Why is this theorem important you may ask? Well, if we have a continuous function on a closed and bounded set D we know that the global max/min are attained. If the set was not closed or bounded the function could **approach** a maximum but never attain it. So we know the function attains its global max/min in the set D. At this point we know the local max/min are contained in the stationary points of the function, if those lie in D then these are good candidates for global max/min. We have two other places to check: 1. the critical points, since the function can have odd behaviour there (think $|x|$ has a minimum at $x = 0$ but not stationary points), 2. the boundary, the value on the boundary could be greater than all these points.

So, to go from local maximum/minimums to absolute maximum/minimums of a continuous function f on a closed and bounded set D we have the following method:

- Find the critical points of f in D by considering $\partial_x f, \partial_y f$: call them $\mathbf{c}_1, ..., \mathbf{c}_m \in \mathbb{R}^2$ and evaluate f at $\mathbf{c}_1, ..., \mathbf{c}_m$, i.e. compute $f(\mathbf{c}_1), ..., f(\mathbf{c}_m)$.
- Evaluate f on the boundary of D.
- The absolute maximum of f on D is then the maximum value from steps 1 and 2.
- The absolute minimum of f on D is then the minimum value from steps 1 and 2.

Example 15.9. *Find the absolute maximum and minimum values of the function*

$$f(x, y) = -x^2 + 7xy + 1 + 3y \tag{512}$$

on the square $\{(x, y) \in \mathbb{R}^2 : -1 \leq x \leq 1, -1 \leq y \leq 1\}$.

Note that we have a polynomial, so we have continuity everywhere. Let's find the critical points:

$$\partial_x f = -2x + 7y, \qquad \partial_y f = 7x + 3. \tag{513}$$

Therefore, $x = -\frac{3}{7}$ for $\partial_y f = 0$ and $y = -\frac{6}{49}$ for $\partial_x f = 0$. Now one can evaluate $f(-\frac{3}{7}, -\frac{6}{49}) = \frac{40}{49}$.

Let's evaluate on the boundary which is given by the four sets

$$\{(x, y) \in \mathbb{R}^2 : x = 1 \text{ and } -1 \leq y \leq 1\} \tag{514}$$
$$\{(x, y) \in \mathbb{R}^2 : x = -1 \text{ and } -1 \leq y \leq 1\} \tag{515}$$
$$\{(x, y) \in \mathbb{R}^2 : y = 1 \text{ and } -1 \leq x \leq 1\} \tag{516}$$
$$\{(x, y) \in \mathbb{R}^2 : y = -1 \text{ and } -1 \leq x \leq 1\}. \tag{517}$$

1. *$x = 1$ then $f(1, y) = 10y$, which is an increasing function from -10 at $y = -1$ to 10 at $y = 1$.*
2. *$x = -1$ then $f(-1, y) = -4y$, which is an decreasing function from 4 at $y = -1$ to -4 at $y = 1$.*

3. $y = 1$ *then* $f(x, 1) = -x^2 + 7x + 4$, *which is* 10 *at* $x = 1$ *and* -4 *at* $x = -1$ *and has a maximum of at* $x = \frac{7}{2} > 1$ *(so is not in our domain).*

4. $y = -1$ *then* $f(x, 1) = -x^2 - 7x - 2$, *which is* 4 *at* $x = -1$ *and* -10 *at* $x = 1$ *and has a maximum at* $x = -\frac{7}{2} < -1$ *(so is not in our domain).*

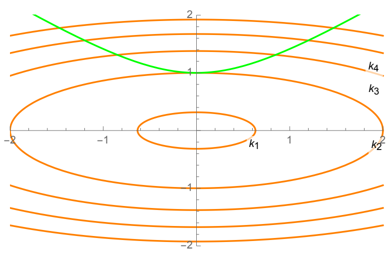
Therefore, our absolute maximum is 10 *at* $(x, y) = (1, 1)$ *and absolute minimum is* -10 *at* $(-1, -1)$.

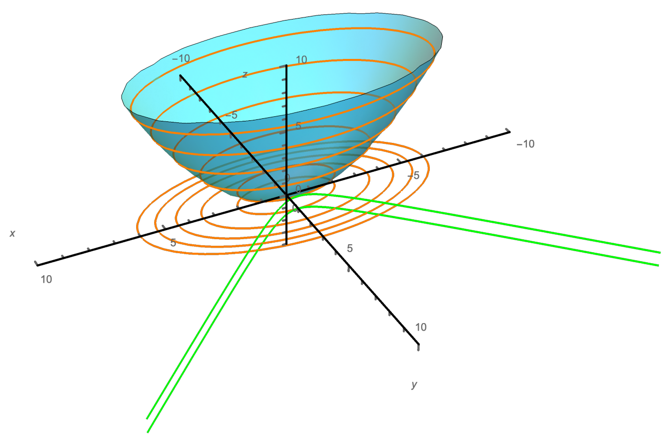
16 Optimization: Lagrange Multipliers

In the previous section 15, we studied maximums and minimums of functions. If the function represents some physical quantity, say a heat distribution in a room, then the maximum and minimum come with a physical interpretation also, the point with highest and lowest temperture respectively. You could imagine that you want to know where these maximums and minimums are in a physical problem to help you with a decision, i.e. if its a cold day you want to sit at the hottest place in the room. On a less personal level: suppose you work for a renewable energy company and you want to build a new wind farm. You could collect data on where the wind speeds around the world. This could be modeled with a function s which depends on the longditude and lattitude (x, y). Finding the maximum is then important to know where to place your windmills. However, you may not want to place them anywhere in the world. You could have a constraint on where they can be coming from where the energy needs to be, the cost of transport of that energy to the consumer and perhaps where people live. This would be a **contrained** optimisation problem: find the maximum (or minimum) of a function with a constraint. This is where the method of Lagrange multipliers comes into play: find the maximum or minimum of a function $f(x, y, z)$ given a constraint that $g(x, y, z) = k$ where k is a constant.

16.1 Illustration of the Idea

Let's think about finding extreme values of a function of two variables $f(x, y)$ given a constraint $g(x, y) = k$. So $(x, y, f(x, y))$ determines a surface in space $\mathbb{R}^3$. The constaint is a level curve of g, i.e. a curve in the plane. We can plot these together in the plane $\mathbb{R}^2$, i.e. we can plot the level curve of g along with a collection of level curves of f, $f(x, y) = k_1, ..., k_n$. This is plotted below for $f(x, y) = \frac{x^2}{2} + y^2$ and $g(x, y) = x^2 - y^2 = 1$ and $k_1 = \frac{1}{10}, k_2 = 1, k_3 = \frac{19}{10}, k_4 = \frac{14}{5}....$





We can see here that we've rephrased the problem of minimisation under the constraint as: which is the lowest level curve of $f(x, y)$ which touches the level curve given by the constraint? We can see this is the level curve of f that is tangent to the level curve $g(x, y) = k$, i.e. the ellipse that touches the hyoerbola on the y-axis. You can also see this from the neighbouring picture in $3D$.

16.2 Derivation of the Method

It turns out that above is the general principle: the extreme values of $f(x, y)$ under the constraint $g(x, y) = k$ occur when the level set of $f(x, y)$ is **tangent** to the level curve $g(x, y) = k$. This can be argued as follows. Suppose we want to find the extreme values of $f(x, y, z)$ subject to the constraint $g(x, y, z) = k$.

- An extreme point for $f(x, y, z)$ subject to the constraint $g(x, y, z) = k$ must lie of the surface S defined by $g(x, y, z) = k$. Let this point be denoted $\mathbf{x}_0 = (x_0, y_0, z_0)$.
- Suppose we take a curve γ in S with equation $\mathbf{r}(t) = (r_1(t), r_2(t), r_3(t))$ lies on S and at t_0, $\mathbf{r}(t_0) = (x_0, y_0, z_0)$. So it passes through the extreme point of f subject to the constraint $g(x, y, z) = k$.
- Denote $h(t) = f(\mathbf{r}(t)) = f(r_1(t), r_2(t), r_3(t))$. Note that since $\mathbf{r}(t)$ constrains the inputs of f to the surface S, then the extreme value of f subject to the constraint $g(x, y, z) = k$ occur at the extreme values of h, i.e. when the derivative of h vanishes. We know that the extreme value of f subject to the constraint $g(x, y, z) = k$ occurs at (x_0, y_0, z_0) so

$$h'(t_0) = 0. \tag{518}$$

- Using the chain rule gives

$$\begin{aligned} 0 = h'(t_0) &= \partial_x f(x_0, y_0, z_0)r_1'(t_0) + \partial_y f(x_0, y_0, z_0)r_2'(t_0) + \partial_z f(x_0, y_0, z_0)r_3'(t_0) && (519)\\ &= \langle \nabla f(\mathbf{x}_0), \mathbf{r}'(t_0)\rangle. && (520) \end{aligned}$$

  So $\nabla f(\mathbf{x}_0)$ is orthogonal to $\mathbf{r}'(t_0)$. Note that the curve determined by $\mathbf{r}$ was an arbitrary curve in S that passes through $\mathbf{x}_0$. So this has to be true for any such curve and, therefore, $\nabla f(\mathbf{x}_0)$ is orthogonal to the level surface S determined by the constraint $g(\mathbf{x}) = k$.
- As shown in section 14.3, the gradient of g is orthogonal to the level surfaces of g.
- This shows that $\nabla f(\mathbf{x}_0)$ and $\nabla g(\mathbf{x}_0)$ must be parallel, i.e.

$$\nabla f(\mathbf{x}_0) = \lambda \nabla g(\mathbf{x}_0) \tag{521}$$

  for $\lambda \in \mathbb{R}$, which gets called the Lagrange Multiplier, and $\nabla g(\mathbf{x}_0) \neq \mathbf{0}$. This is the equation we want to solve to find the extreme values of f under the constraint $g(x, y, z) = k$.

This is a sketch of the derivation of the equation. In practise what you need to do is the following:

The method of Lagrange Multipliers: Suppose you wish to find the maximum and minimum values of $f(x, y, z)$ subject to the constraint $g(x, y, z) = k$ under the assumption that these extreme values exist and $\nabla g \neq \mathbf{0}$ on the surface $\{g(x, y, z) = k\}$. Then one executes the following algorithm:

1. Find all values of x, y, z and λ such that

$$\nabla f = \lambda \nabla g \quad \text{and} \quad g(x, y, z) = k. \tag{522}$$

2. Evaluate f at all points (x, y, z, λ) that result from the first step. The largest of these is the maximum of f subject to the constraint and the smallest is the minimum of f subject to the constraint.

16.3 Examples

Example 16.1. *Find the extreme values of*

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

subject to the constraint

$$g(x, y) = x^2 + y^2 = 1$$

.

Let's compute the gradient of f and g. We have

$$\nabla f = (2x, 4y), \qquad \nabla g = (2x, 2y). \tag{523}$$

So we want to solve the system of equations

$$2x = 2\lambda x \tag{524}$$
$$4y = 2\lambda y \tag{525}$$
$$g(x, y) = x^2 + y^2 = 1. \tag{526}$$

So the first has solutions $x = 0$ or $\lambda = 1$.

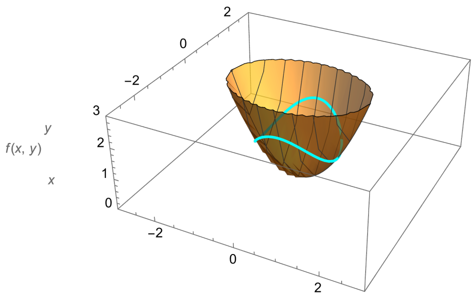
1. *If $x = 0$ then $y = \pm 1$ from the constaint $y^2 = 1 - x^2$. The second equation then enforces $\lambda = 2$.*

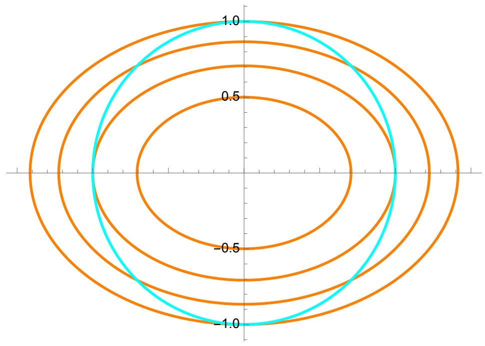
2. *If $\lambda = 1$ then $y = 0$ from the second equation. This enforces $x = \pm 1$ from the constraint.*

Therefore we have 4 solutions $(x, y, \lambda) = (0, 1, 2)$, $(x, y, \lambda) = (0, -1, 2)$, $(x, y, \lambda) = (1, 0, 1)$ and $(x, y, \lambda) = (-1, 0, 1)$. Evaluating f at these points gives

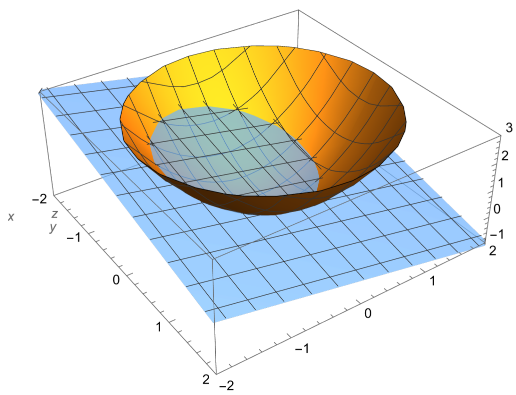
$$f(0, 1) = 2, \quad f(0, -1) = 2, \quad f(1, 0) = 1, \quad f(-1, 0) = 1. \tag{527}$$

So we have the maximum of f subject to the constraint g is 2 at $(0, \pm 1)$ and the minimum of f subject to the constraint g is 1 at $(\pm 1, 0)$.





Example 16.2. *The plane $x + y + 2z = 2$ intersects the elliptic paraboloid $z = x^2 + y^2$ in the an ellipse. Find the points on this ellipse nearest and farthest the from origin $(0, 0, 0)$.*



First, notice that the distance from the origin is given by

$$d(x, y, z) = \sqrt{x^2 + y^2 + z^2} \tag{528}$$

Note that minimising and maximising the distance from the origin is equivalent to maximising/minimising the distance squared. Second, note that we can solve for z here using the plane equation:

$$z = 1 - \frac{1}{2}x - \frac{1}{2}y. \tag{529}$$

Therefore, we want to minimise/maximise

$$f(x,y) = d\big(x,y,1-\frac{1}{2}x-\frac{1}{2}y\big)^2 = x^2+y^2+\big(1-\frac{1}{2}x-\frac{1}{2}y\big)^2. \tag{530}$$

We need to find an equation for our constraint, which is to lie on the ellipse determined by the intersection of the paraboloid and the plane. To do this we can substitute in the plane equation solved for z*:*

$$1-\frac{1}{2}x-\frac{1}{2}y = x^2+y^2 \implies g(x,y) = x^2+y^2+\frac{1}{2}x+\frac{1}{2}y = 1. \tag{531}$$

Therefore, by the method of Lagrange multipliers, we need to find (x,y,λ) *such that*

$$\nabla f = \lambda\nabla g \tag{532}$$
$$g(x,y) = 1 \tag{533}$$

Computing the gradients we have

$$\nabla f = \Big(-1+\frac{5}{2}x+\frac{1}{2}y, -1+\frac{1}{2}x+\frac{5}{2}y\Big), \tag{534}$$
$$\nabla g = \Big(2x+\frac{1}{2}, 2y+\frac{1}{2}\Big). \tag{535}$$

So, equating $\nabla f = \lambda\nabla g$ *gives*

$$-1+\frac{5}{2}x+\frac{1}{2}y = \lambda\Big(2x+\frac{1}{2}\Big) \tag{536}$$
$$-1+\frac{1}{2}x+\frac{5}{2}y = \lambda\Big(2y+\frac{1}{2}\Big). \tag{537}$$

Solving the second for x *gives*

$$x = 2+\lambda+(4\lambda-5)y. \tag{538}$$

Substituting this into the first gives

$$(\lambda-1)(2+\lambda+(4\lambda-6)y) = 0. \tag{539}$$

So either $\lambda = 1$ *or* $2+\lambda+(4\lambda-6)y = 0$*. If* $\lambda \neq \frac{3}{2}$ *then the second can be solved for* y

$$y = -\frac{2+\lambda}{4\lambda-6}. \tag{540}$$

If $\lambda = \frac{3}{2}$ *then* $2+\lambda = 0$*, i.e.* $\lambda = -2$ *which is a contradiction. So either*

$$y = -\frac{2+\lambda}{4\lambda-6} \quad \textit{or} \quad \lambda = 1. \tag{541}$$

If $\lambda = 1$ *then from* (538) $x = 3-y$*. Therefore, from the constraint*

$$g(3-y,y) = (3-y)^2+y^2+\frac{1}{2}(3-y)+\frac{1}{2}y = 1. \tag{542}$$

So, y *has to be a root of the polynomial:*

$$2y^2-6y+\frac{19}{2} = 0 \implies 2\Big(y-\frac{3}{2}\Big)^2+\frac{10}{2} = 0 \tag{543}$$

which has no solution.

If $y = -\frac{2+\lambda}{4\lambda-6}$ then

$$x = -\frac{2+\lambda}{4\lambda-6} \tag{544}$$

from (538). *Then plugging x, y into the constraint gives that λ must satisfy:*

$$\frac{9\lambda^2 - 27\lambda + 8}{4(3-2\lambda)^2} = 0 \implies \lambda = \frac{1}{3}, \quad \text{or} \quad \lambda = \frac{8}{3}. \tag{545}$$

This then gives $x = y = \frac{1}{2}$ or $x = y = -1$. We can then solve for z via the plane equation this gives two points in $\mathbb{R}^3$:

$$(x, y, z) = (\frac{1}{2}, \frac{1}{2}, \frac{1}{2}), \quad \text{or} \quad (x, y, z) = (-1, -1, 2). \tag{546}$$

To figure out the max/min we can simply plug into the distance:

$$d(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})^2 = \frac{3}{4}, \qquad d(-1, -1, 2)^2 = 6. \tag{547}$$

Therefore, distance is maximise at $(-1, -1, 2)$ and minimised $(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$.

Example 16.3. *Find the max/min of $f(x, y) = x^2 + y^2 + 4x - 4y$ subject to the constraint $x^2 + y^2 \leq 9$.*

This problem can be dealt with using Lagrange multipliers by introducting a dummy variable to encode the constraint with an equality:

$$x^2 + y^2 \leq 9 \implies x^2 + y^2 - 9 \leq 0. \tag{548}$$

We can write this as

$$9 - x^2 - y^2 = z^2 \tag{549}$$

since, for $x^2+y^2-9 \leq 0$, $9-x^2-y^2$ has to be some positive number. Therefore, we want to max/minimise

$$f(x, y, z) = x^2 + y^2 + 4x - 4y, \tag{550}$$

subject to the constraint

$$g(x, y, z) = 9 - x^2 - y^2 - z^2 = 0. \tag{551}$$

Computing gradients gives

$$\nabla f = (2x + 4, 2y - 4, 0), \qquad \nabla g = (-2x, -2y, -2z). \tag{552}$$

Therefore, the system resulting from the Lagrange multiplier method gives

$$2x + 4 = -2\lambda x \tag{553}$$

$$2y - 4 = -2\lambda y \tag{554}$$

$$0 = -2\lambda z \tag{555}$$

Therefore, $\lambda = 0$ or $z = 0$. If $\lambda = 0$ then $x = -2$ and $y = 2$. If $z = 0$ then

$$x = \frac{-2}{1+\lambda}, \qquad y = \frac{2}{1+\lambda}. \tag{556}$$

Substituting into our constraint gives

$$9 - \frac{8}{(1+\lambda)^2} = 0 \implies \lambda = -1 \pm \frac{2\sqrt{2}}{3}, \tag{557}$$

and, therefore,

$$(x, y, z) = (-\frac{3}{\sqrt{2}}, \frac{3}{\sqrt{2}}, 0), \qquad (x, y, z) = (\frac{3}{\sqrt{2}}, -\frac{3}{\sqrt{2}}, 0). \tag{558}$$

Evalutating f at these points gives a maximum of $9 + 12\sqrt{2}$ at $(x, y) = (\frac{3}{\sqrt{2}}, -\frac{3}{\sqrt{2}})$ and a minimum at $(x, y) = (-2, 2)$.

17 Complex Numbers

What x solves the equation

$$x^2 = -1? \tag{559}$$

Clearly, no real number can satisfy this equation. However, we could define a new 'number' $i = \sqrt{-1}$ to be a solution to this equation. This is how imaginary and complex numbers arise.

17.1 Introduction: Definition and Operations

Definition 17.1. *The complex numbers, denoted $\mathbb{C}$, is the set or collection of all numbers of the form*

$$z = a + bi \tag{560}$$

where $a, b \in \mathbb{R}$ and i is a specific element, called the imaginary unit, which satisfies $i^2 = 1$.

One can introduce some terminology to discuss complex numbers:

- The **real part** of $z = a + bi$ is a, this is sometimes written $\mathrm{Re}(z) = a$.
- The **imaginary part** of $z = a + bi$ is b, this is sometimes written $\mathrm{Im}(z) = b$.
- The **complex conjugate** of a complex number $z = a + bi$, denoted $\bar{z}$, is given by,

$$\bar{z} = a - bi. \tag{561}$$

- The **absolute value or modulus** of a complex number $z = a + bi$ is denoted $|z|$ and is given by

$$|z| = \sqrt{a^2 + b^2}. \tag{562}$$

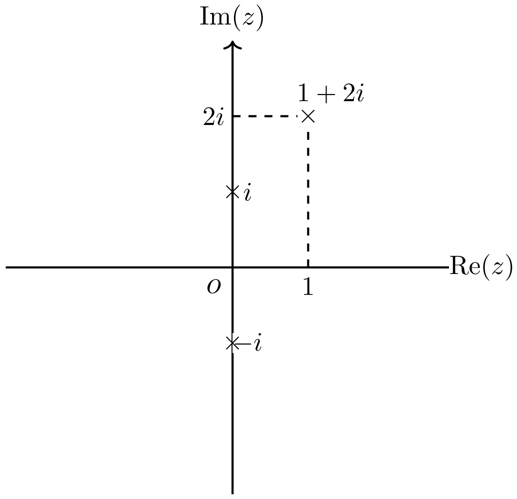
- Two complex numbers $z_1, z_2 \in \mathbb{C}$ are equal if their real parts are equal and their imaginary parts are equal,

$$\mathrm{Re}(z_1) = \mathrm{Re}(z_2), \qquad \mathrm{Im}(z_1) = \mathrm{Im}(z_2). \tag{563}$$

We can identify $\mathbb{C}$ with $\mathbb{R}^2$ via

$$z = a + bi \mapsto (a, b) \in \mathbb{R}^2, \qquad (a, b) \mapsto a + bi \in \mathbb{C}, \tag{564}$$

which means that we can draw $\mathbb{C}$ as we drew $\mathbb{R}^2$. We associate the x-axis with the real part of the complex numbers and the y-axis with the imaginary part of the complex numbers. In this context the plane is called the Argand or complex plane and the x-axis is called the real axis and the y-axis is called the imaginary axis. This is plotted below:



Note that the distance of $z \in \mathbb{C}$ from the origin is given by the absolute value of z, i.e. $|z| = \sqrt{a^2 + b^2}$.

One can define addition and subtraction of imaginary numbers. Suppose

$$z_1 = a + bi \qquad z_2 = c + di. \tag{565}$$

Then,

$$z_1 + z_2 = (a + c) + (b + d)i, \qquad z_1 - z_2 = (a - c) + (b - d)i. \tag{566}$$

Further multiplication is defined with the usual commutative and distributive laws holding:

$$z_1 z_2 = (a + bi)(c + di) = (ac - bd) + (ad + cb)i \tag{567}$$

where we've used $i^2 = -1$.

Complex conjugates satisfy some very nice relations:

Proposition 17.1. *For $z, w \in \mathbb{C}$ one has*

$$\overline{z + w} = \bar{z} + \bar{w}, \qquad \overline{zw} = \bar{z}\bar{w}, \qquad \overline{z^n} = \bar{z}^n, \qquad z\bar{z} = |z|^2 \tag{568}$$

Proof. Problem Sheet 10. □

Division of two complex numbers is slightly more tricky. Suppose one has the quotient $\frac{z}{w}$ with $z = a+bi$ and $w = c + di$. One can use the complex conjugate of w and the very useful trick of multiplying by 1 to simplify a quotient:

$$\frac{z}{w} = \frac{z}{w}\frac{\overline{w}}{\overline{w}} = \frac{z\overline{w}}{|w|^2} = \frac{(a + bi)(c - di)}{|w|^2} = \frac{ac + bd}{|w|^2} + \frac{bc - ad}{|w|^2} i \tag{569}$$

17.2 Polar Form, DeMoivre's Theorem, Roots

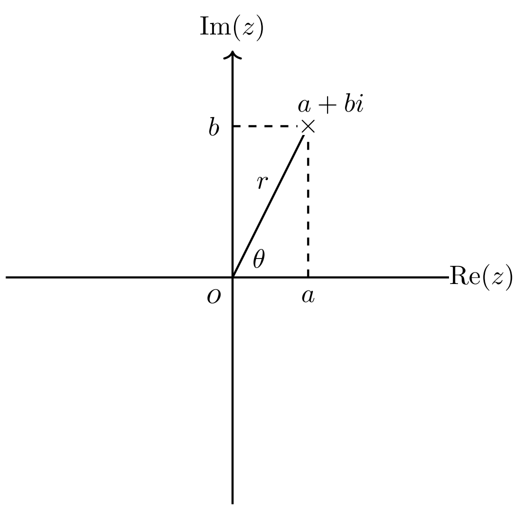
As we've discussed above $\mathbb{C}$ can be identified with the plane $\mathbb{R}^2$. If you recall section 1, we discussed polar coordinates (r, θ). Where we had

$$x = r\cos\theta, \qquad y = r\sin\theta. \tag{570}$$

We can simply translate now to write any complex number z as

$$z = a + bi = r(\cos\theta + i\sin\theta). \tag{571}$$

This is the polar form of a complex number. The following picture should help with visualisation:



From our discussion on polar coordinates in section 1, we have that

$$r = |z| = \sqrt{a^2 + b^2}, \qquad \tan\theta = \frac{b}{a}. \tag{572}$$

One often calls the angle θ the **argument or phase** of z, denoted $\theta = \arg(z)$ which is usually restricted to $\theta \in [0, 2\pi)$ (or often $\theta \in (-\pi, \pi]$).

Example 17.1. *Write the following complex numbers in polar form:*

$$z_1 = 1 + i, \qquad z_2 = 1 - 7i, \qquad z_3 = \sqrt{2} - 5i. \tag{573}$$

Let's find the absolute value $|z_1| = \sqrt{2}$, $|z_2| = 5\sqrt{2}$ *and* $|z_3| = 3\sqrt{3}$. *Now*

$$1 + i = \sqrt{2}(\cos\theta + i\sin\theta) \implies \cos\theta = \frac{1}{\sqrt{2}} = \sin\theta \implies \theta = \frac{\pi}{4}. \tag{574}$$

Similarly

$$1 - 7i = 5\sqrt{2}(\cos\theta + i\sin\theta) \implies \cos\theta = \frac{1}{5\sqrt{2}}, \sin\theta = -\frac{7}{5\sqrt{2}} \implies \theta = 2\pi - \arccos(\frac{1}{5\sqrt{2}}), \tag{575}$$

or $\theta \approx 1.545\pi$. *Finally,*

$$\sqrt{2} - 5i = 3\sqrt{3}(\cos\theta + i\sin\theta) \implies \cos\theta = \frac{\sqrt{2}}{3\sqrt{3}}, \sin\theta = -\frac{5}{3\sqrt{3}} \implies \theta = 2\pi - \arccos(\frac{\sqrt{2}}{3\sqrt{3}}), \tag{576}$$

or $\theta \approx 1.588\pi$.

The polar form lets us gain new insight on mulitplication and division. Recall the trigonometic formulas

$$\cos\theta_1 \sin\theta_2 = \frac{1}{2}\Big[\sin(\theta_1 + \theta_2) - \sin(\theta_1 - \theta_2)\Big] \tag{577}$$

$$\cos\theta_1 \cos\theta_2 = \frac{1}{2}\Big[\cos(\theta_1 - \theta_2) + \cos(\theta_1 + \theta_2)\Big] \tag{578}$$

$$\sin\theta_1 \sin\theta_2 = \frac{1}{2}\Big[\cos(\theta_1 - \theta_2) - \cos(\theta_1 + \theta_2)\Big] \tag{579}$$

This means

$$z_1 z_2 = r_1 r_2\Big[\cos\theta_1\cos\theta_2 - \sin\theta_1\sin\theta_2 + i\cos\theta_1\sin\theta_2 + i\sin\theta_2\cos\theta_1\Big] \tag{580}$$

$$= r_1 r_2\Big[\cos(\theta_1 + \theta_2) + i\sin(\theta_1 + \theta_2)\Big] \tag{581}$$

and

$$\frac{z_1}{z_2} = \frac{r_1}{r_2}\Big[\cos(\theta_1 - \theta_2) + i\sin(\theta_1 - \theta_2)\Big], \qquad z_2 \neq 0. \tag{582}$$

We are about to state a theorem about the polar form of z^n. The proof of this is by a method known as induction. Theorems that can be proved in this way depend on a natural number n. The idea is to prove the **base** case $n = 0$ or $n = 1$. Then we **assume** that the statement holds for some $n = k > 1$ and prove that it holds for $n = k + 1$. Since k is arbitrary this then must hold for all n.

Theorem 17.1 (De Moivre's Theorem)**.** *If* $z = r(\cos\theta + i\sin\theta)$ *and* $n \in \{1, 2, 3, ...\}$ *then*

$$z^n = (r\cos\theta + ir\sin\theta)^n = r^n(\cos(n\theta) + i\sin(n\theta)). \tag{583}$$

Proof. (Non-examinable). This is clear for $n = 1$. So lets prove the base case $n = 2$

$$z^2 = r^2(\cos^2\theta - \sin^2\theta) + 2r^2 i\cos\theta\sin\theta \tag{584}$$

From our trignometic formulas in equations (577)-(579) one has

$$z^2 = r^2\cos(2\theta) + r^2 i\sin(2\theta). \tag{585}$$

Lets assume this holds for $n-1$ and prove it for n. Under this assumption

$$\begin{aligned} z^n &= r^n(\cos((n-1)\theta) + i\sin((n-1)\theta))(\cos(\theta) + i\sin(\theta)) \\ &= r^n\Big[\cos((n-1)\theta)\cos\theta - \sin((n-1)\theta)\sin\theta\Big] + ir^n\Big[\cos((n-1)\theta)\sin\theta + \sin((n-1)\theta)\cos\theta\Big]. \end{aligned} \tag{586}$$

Using the trig. identities (equations (577)-(579)) once more gives the result. Therefore, by induction we have the result. □

De Moivre's theorem can be use to find the nth root of a complex number z, i.e. find w such that

$$w^n = z. \tag{587}$$

Writing w and z in polar form as

$$w = s\big(\cos\varphi + i\sin\varphi\big), \qquad z = r\big(\cos\theta + i\sin\theta\big) \tag{588}$$

gives, via DeMoivre's theorem,

$$s^n\big(\cos(n\varphi) + i\sin(n\varphi)\big) = r\big(\cos\theta + i\sin\theta\big). \tag{589}$$

This requires,

$$s = r^{\frac{1}{n}}, \quad \cos(n\varphi) = \cos\theta, \quad \sin(n\varphi) = \sin\theta. \tag{590}$$

This can be simplified to

$$s = r^{\frac{1}{n}}, \qquad \varphi = \frac{\theta + 2k\pi}{n}, \tag{591}$$

where k is an integer. This gives distinct solutions for $k \in \{0, 1, ..., n-1\}$. So we have the following result:

Proposition 17.2. *Let $z = r(\cos\theta + i\sin\theta)$ and $n \in \{1, 2, 3, ...\}$. Then z has n distinct roots given by*

$$w_k = r^{\frac{1}{n}}\Big[\cos\Big(\frac{\theta + 2k\pi}{n}\Big) + i\sin\Big(\frac{\theta + 2k\pi}{n}\Big)\Big], \tag{592}$$

for $k = 0, ..., n-1$.

Remark 17.1. *All n roots of z have modulus $r^{\frac{1}{n}}$. Hence, they lie on a circle. Moreover, successive roots are spaced by $\frac{2\pi}{n}$, i.e. they are spaced equally on the circle.*

Example 17.2. *Find all distinct w such that $w^4 = 1$.*

The above proposition gives us 4 distinct roots:

$$w_k = r^{\frac{1}{4}}\Big[\cos\Big(\frac{\theta + 2k\pi}{4}\Big) + i\sin\Big(\frac{\theta + 2k\pi}{4}\Big)\Big] \tag{593}$$

for $k = 0, 1, 2, 3$. The absolute value of z is

$$r = |z| = \sqrt{1^2} = 1 \tag{594}$$

and $\theta = \arctan(0) = 0$. So,

$$w_k = \Big[\cos\Big(\frac{k\pi}{2}\Big) + i\sin\Big(\frac{k\pi}{2}\Big)\Big]. \tag{595}$$

for $k = 0, 1, 2, 3$. Evaluating gives

$$w_0 = 1, \quad w_1 = i, \quad w_2 = -1, \quad w_3 = -i. \tag{596}$$

17.3 Complex Functions and the Fundamental Theorem of Algebra

So far in this course, we've studied functions that take real numbers as inputs and give out real numbers as outputs. What about functions that take complex numbers as inputs and give out complex numbers as outputs? We would denote this notationally,

$$f : \mathbb{C} \to \mathbb{C}. \tag{597}$$

If $z = x + iy$ then

$$f(z) = u(x, y) + iv(x, y) \tag{598}$$

for $u, v : \mathbb{R}^2 \to \mathbb{R}$.

17.3.1 Polynomials

Let's start with a univariate polynomial of a complex number, z. This is an expression of the form

$$P(z) = a_n z^n + a_{n-1} z^{n-1} ... + a_1 z^1 + a_0, \qquad a_0, ..., a_n \in \mathbb{C}. \tag{599}$$

The **degree** of the polynomial is the largest power appearing in it's equation, i.e. n. The **coefficients** of the polynomial are $a_0, ..., a_n$. A **root** of a polynomial is a complex number z_0 such that $P(z_0) = 0$. A root's **multiplicity** is the number of times that root appears in the factorisation, i.e.

$$P(x) = (x - 1)^2 (x - 7) \tag{600}$$

has roots $x = 1$ with multiplicity 2 and $x = 7$ with multiplicity 1.

For univarite polynomial's one has a very beautiful theorem about the existence of roots called the Fundamental Theorem of Algebra.[10]

Theorem 17.2. *Every non-zero, univariate polynomial of degree n with complex coefficients has, counted with multiplicity, exactly n complex roots.*

We do not give the proof here but we will content ourselves with a proof for quadratic equations

$$P(z) = az^2 + bz + c. \tag{601}$$

To study the roots sets $P(z) = 0$. Then

$$az^2 + bz + c = 0. \tag{602}$$

One can rewrite this as

$$a\left(z + \frac{b}{2a}\right)^2 - \frac{b^2}{4a} + c = 0. \tag{603}$$

So,

$$\left(z + \frac{b}{2a}\right)^2 = \frac{b^2 - 4ac}{4a^2}. \tag{604}$$

Now when we consider a, b, c as complex numbers and allow for complex roots as above, $\sqrt{\frac{b^2-4ac}{a^2}}$ is well-defined. Therefore,

$$z = \frac{b \pm \sqrt{b^2 - 4ac}}{2a}, \tag{605}$$

which is two roots unless $b^2 - 4ac = 0$, in which case $\frac{b}{2a}$ appears as a root with multiplicity 2.

[10]This theorem is usually attributed to Gauss in 1799. However, as is usual in mathematics there is a complicated history here. If you're interested in the history Wikipedia lays it out well.

17.3.2 Complex Exponentials

The complex exponential is the extension of e^x to the complex plane. We denote this with e^z. How is this defined? Well, there's a couple of ways but one is included here for completeness. The most common is through an infinite series[11]

$$e^z \doteq \sum_{n=0}^{\infty} \frac{z^n}{n!} = 1 + z + \frac{z^2}{2} + \frac{z^3}{6} + ... \tag{606}$$

It turns out that the complex exponential has the same property as the real exponential, namely

$$e^{z_1+z_2} = e^{z_1} e^{z_2}. \tag{607}$$

There is a further property in relation to the trignometric functions $\sin$ and $\cos$. Let's do some rough manipulation with this formula.[12] Consider, e^{ix} for $x \in \mathbb{R}$. By the definition via the series one has

$$e^{ix} = 1 + ix + \frac{(ix)^2}{2!} + \frac{(ix)^3}{3!} + \frac{(ix)^4}{4!} + \frac{(ix)^5}{5!} + ... \tag{608}$$

Noting that $i^{2n} = (-1)^n$ for $n = 1, 2, 3, ...$ and $i^{2n+1} = (-1)^n i$ for $n = 1, 2, 3, ...$ one can collect real and imaginary parts of e^{ix} as

$$e^{ix} = \left(1 - \frac{x^2}{2!} + \frac{x^4}{4!} + ...\right) + i\left(x - \frac{x^3}{3!} + \frac{x^5}{5!} + ...\right). \tag{609}$$

The real part is the series expansion of $\cos(x)$ and the imaginary part is the series expansion of $\sin(x)$. We have arrived at Euler's formula

$$e^{ix} = \cos x + i \sin x. \tag{610}$$

Let's compute using this formula $e^{i\pi}$:

$$e^{i\pi} = \cos \pi + i \sin \pi = 1 \implies e^{i\pi} = -1. \tag{611}$$

This is often viewed as one of the most remarkable relations in mathematics. On the left-hand side you have two irrational numbers, e and π and the imaginary unit. Via Euler's formula, these combine to give the integer number -1!

Let's note one last thing about the complex exponential. From Euler's formula one has

$$e^{ix} = \cos(x) + i \sin(x), \qquad e^{-ix} = \cos(-x) + i \sin(-x) = \cos(x) - i \sin(x). \tag{612}$$

Therefore,

$$\cos(x) = \frac{1}{2}\left(e^{ix} + e^{-ix}\right), \qquad \sin(x) = \frac{1}{2i}\left(e^{ix} - e^{-ix}\right), \tag{613}$$

which are very useful identities to know if you forget your trignometric identities.

17.3.3 A Couple of Interesting Uses/Properties Associated to Complex Numbers

This section is non-examinable.

Definition 17.2 (Complex Differentiable/Holomorphic Functions). *A complex-valued function f of a single complex variable z is* **complex differentiable/holomorphic** *at z_0 if the following limit exists*

$$f'(z_0) = \lim_{z \to z_0} \frac{f(z) - f(z_0)}{z - z_0}. \tag{614}$$

(Note this is defined in the complex plane the limit has to agree on any path.)

[11]For those of you concerned with convergence of this series one can check via the ratio test that this convergence for all $z \in \mathbb{C}$ and is therefore a well-defined object.

[12]Note: the following manipulation can be made rigorous by noting the absolute convergence of the series. If this doesn't mean anything to you do not worry about it.

Theorem 17.3. *Let $z = x + iy$ and let f be a complex-valued function of z, i.e. $f : \mathbb{C} \to \mathbb{C}$ and*

$$f(z) = u(x, y) + iv(x, y). \tag{615}$$

If the function f is **complex-differential/holomorphic** *it satisfies the Cauchy–Riemann equations*

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}, \qquad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}. \tag{616}$$

If f satisfies the Cauchy–Riemann equations and is continuous and the first partial derivatives of u and v exist then f is holomorphic.

Example 17.3. *The function $f(z) = z^2$ is holomorphic:*

$$f(z) = (x + iy)^2 = x^2 - y^2 + 2ixy, \implies u(x, y) = x^2 - y^2, \qquad v(x, y) = 2xy. \tag{617}$$

$$\partial_x u = 2x, \qquad \partial_y v = 2x \tag{618}$$

$$\partial_y u = -2y, \qquad \partial_x v = 2y. \tag{619}$$

So, the Cauchy–Riemann equations are satisfied.

Why are holomorphic functions useful you may ask? Well they have some very nice properties:

- They are infinitely differentiable.
- At any point z_0 in it's domain you can find a small disk where the function coincides with its Taylor series

$$f(z) = f(z_0) + f'(z_0)(z - z_0) + \frac{1}{2} f''(z_0)(z - z_0)^2 + \frac{1}{3} f'''(z_0)(z - z_0)^3 + ... \tag{620}$$

  This is extremely useful for approximating functions beyond the linear approximation.
- You can differentiate this series freely, just like a polynomial.
- The whole behaviour of the function can be reconstructed from knowledge of the function in the neighbourhood of a single point.

Let's end the section on complex numbers with a result that allows us to solve an ordinary differential equation.

Example 17.4. *Let $u(x)$ satisfy the ordinary differential equation*

$$\frac{d^2 u}{dx^2} + ku = 0, \tag{621}$$

for k constant. This is the equation modelling simple harmonic motion, which arises everywhere in physics. A simple example is a mass on the end of a spring.

We want to solve for u. You may or may not of seen how to solve this equation before. Here we use a trick: let's write

$$\Big(\frac{d^2}{dx^2} + k\Big)u = 0, \tag{622}$$

which we can manipulate to

$$\Big(\frac{d}{dx} + ik\Big)\Big(\frac{d}{dx} - ik\Big)u = 0. \tag{623}$$

So, this tells us that

$$\Big(\frac{d}{dx} + ik\Big)f = 0 \tag{624}$$

where

$$f = \Big(\frac{d}{dx} - ik\Big)u. \tag{625}$$

Lets assume $f \neq 0$, then the first equation (624) *can be rewritten as*

$$\frac{1}{-ikf}\frac{d}{dx}f = 1 \implies \frac{d}{dx}\Big(\frac{i}{k}\log(f)\Big) = 1 \tag{626}$$

where we've used $\frac{1}{i} = -i$ and $\frac{d}{dx}\log f = \frac{1}{f}\frac{d}{dx}f$ by the chain rule. We can now integrate both sides to obtain

$$\frac{i}{k}\log(f) = x + c_1 \implies f(x) = e^{-ik(c_1+x)} = c_2 e^{-ikx}, \tag{627}$$

where $c_2 = e^{-ikc_1}$.

We can return to equation 625 and substitute in f to find

$$\Big(\frac{d}{dx} - ik\Big)u = c_2 e^{-ikx}. \tag{628}$$

Multiply both sides by e^{-ikx} then we find

$$e^{-ikx}\frac{d}{dx}u - ike^{-ikx}u = c_2 e^{-2ikx}. \tag{629}$$

The left hand side can be written as

$$\frac{d}{dx}(e^{-ikx}u). \tag{630}$$

So, our equation becomes

$$\frac{d}{dx}(e^{-ikx}u) = c_2 e^{-2ikx}. \tag{631}$$

Both sides can be integrated to find

$$e^{-ikx}u = c_2 e^{-2ikx} + c_3. \tag{632}$$

Therefore,

$$u(x) = c_3 e^{ikx} + c_2 e^{-ikx}. \tag{633}$$

You can invert the relations:

$$\cos(kx) = \frac{1}{2}\Big(e^{ikx} + e^{-ikx}\Big), \qquad \sin(kx) = \frac{1}{2i}\Big(e^{ikx} - e^{-ikx}\Big), \tag{634}$$

to find

$$u(x) = c_4\cos(kx) + ic_5\sin(kx). \tag{635}$$

where c_4 and c_5 are related to c_3, c_4.

Appendices

A Abbreviations

- $\therefore$: therefore, or 'it follows that'.
- $\text{Prop}^{\underline{\text{n}}}$: proposition. This is a statement of a result, like a theorem only smaller and less important.
- $\text{f}^{\underline{\text{n}}}$: function
- $\text{def}^{\underline{\text{n}}}$: definition
- $\text{not}^{\underline{\text{n}}}$: notation
- Rem. or Rmk: remark/a comment.
- Cor.: a corollary. a statement of a result that follows from a proposition or theorem above it.
- pt: shorthand for 'point'.
- $\#$: shorthand for 'number'.
- w/: shorthand for 'with'.
- s.t.: shorthand for 'such that'.
- w.r.t: shorthand for 'with respect to'.
- std: shorthand for 'standard'.
- resp.: shorthand for 'respectively'
- $//$: shorthand for 'parallel'.
- $\perp$: shorthand for 'perpendicular' or orthogonal.
- $(\dagger)$, $(\ddagger)$, $(\dagger\dagger)$ $(\star)$, $(\star\star)$, $(\circ)$: said 'dagger', 'double dagger', 'star', 'double star', 'circle'. Labelling of equations to refer back to in lectures. Does not hold over multiple lectures, only for that lecture.
- RHS: right-hand side
- LHS: right-hand side