Notes on:

Multivariable Calculus

David Duan

Last Updated: September 27, 2023 (Draft)

Contents

Chapter 1

Maps

Contents

MAT137 solely focused on maps from $\mathbb{R} \to \mathbb{R}$, and now we will begin the study of MAT237 with the characteristics of maps from $\mathbb{R}^n \to \mathbb{R}^m$. There are many ways of referring to them but all of these defines a function

$$\mathbf{f} : A \to B, \quad A \subseteq \mathbb{R}^n, B \subseteq \mathbb{R}^m$$

This section will refrain from being too rigour as we try to build intuitions and explore examples of multivariable functions.

1.1 Notations

In this course, we will be discussing real-value functions from $\mathbb{R}$ and also vector-valued $\mathbb{R}^n$. To help keep track of these objects, we will use different set of letters:

- lowercase Latin and Greek letters near the start of the alphabet $a, b, c, r, \alpha, \beta, \delta, \epsilon$ for real numbers;
- lower Latin near the end and other Greek letters $t, u, v, w, x, y, z, \varphi, \psi, \theta$ for variables;
- bold lowercase Latin $\mathbf{a}, \mathbf{b}, \mathbf{n}, \mathbf{u}, \mathbf{v}, \mathbf{x}, \mathbf{y}$ for vectors;
- upper or lowercase Latin letters f, g, h, F, G for real valued functions;
- vector arrow notation or γ for vector valued functions.

1.2 Curves

Maps of the form

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

are sometimes called **vector-valued functions of a real variable**, especially for $n \geq 2$, but more often they are referred to as **parametric curves** as that is what they describe physically.

Definition 1.2.1 A parametric curve

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

is the graph of the collections of points $(f_1, ..., f_n)$, obtained from the set of n continuous functions of a **parameter** t (*often time*) on an interval I

$$f_1 = f_1(t), \cdots, f_n = f_n(t)$$

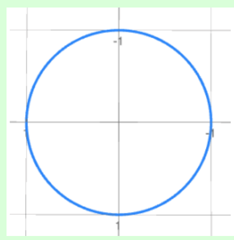
these functions are the **parametric equations**.

For maps from $\mathbb{R} \to \mathbb{R}$, we plot both the input and the output simultaneously. However, for parametric curves, we only care about the output, hence a map from $\mathbb{R} \to \mathbb{R}^n$ can be illustrated by an n dimensional graph.

Intuition: There are many benefits as to why we use parametric equations as opposed to our usual Cartesian or Polar form, some of them are:

- Allows us to graph curves that are not functions like the unit circle.
- Provides us more information with the use of the '**parameter**', such as direction and speed with respect to it (usually time).
- More often than not it is easier to differentiate and integrate a curve using its parametric equations.

Example 1.2.2 Define the map $\gamma : [0, 2\pi] \to \mathbb{R}^2$ as $\gamma(t) = (\cos(t), \sin(t))$. The image of γ is the unit circle in $\mathbb{R}^2$, namely the set $\gamma([0, 2\pi]) = \{(x, y) \in \mathbb{R}^2 : x^2 +^2 = 1\}$.



Watch this to see how the path is drawn over time.

Example 1.2.3 Define the map $\gamma_2 : [0, \infty] \to \mathbb{R}^3$ as $\gamma_2(t) = (\cos(t), \sin(t), t)$. The image of γ is a helix, where as time increases, the helix is traced out in an upward fashion. Watch this to see how the path is drawn over time.

Example 1.2.4 Let $p, q \in \mathbb{R}^n$. A straight line path $\gamma : [0, 1] \to \mathbb{R}^n$ from point p to point q is defined by $\gamma(t) = (1 - t)p + tq$. You can also view this formula $\gamma(t) = p + t(q - p)$.

1.2.1 Motion

Physically, a parametric curve $\gamma : I \to \mathbb{R}^n$ describes the motion of an object moving in $\mathbb{R}^n$ in the interval $I \subseteq \mathbb{R}$. The **position** at time t is clearly $\gamma(t)$. Well, what is its velocity at t? From MAT137, the natural suggestion should be the **velocity** at time t is given by

$$\gamma'(t) = \lim_{h \to 0} \frac{\gamma(t + h) - \gamma(t)}{h},$$

however h is scalar whereas $\frac{\gamma(t+h)-\gamma(t)}{h}$ is a vector! We have only seen limits of scalar quantities with scalar limit variables, so the above notion of a limit is not exactly as the case in MAT137. We will formally study these later but the above guess is a good try.

Example 1.2.5 An object in 3-dimensional space is moving along the path described by

$$\vec{g}(t) = \left(2t-6, 5(t-3)^2, \frac{1}{2}t^2-4\right).$$

Despite not knowing the exact notion of a limit of a vector, we can still analyze the velocity and speed of $\vec{g}'(3)$ using the tradition limit definition. Similar to the one-dimensional case, it is simply the change in position over a shorter and shorter interval, here we will use $[3, 3+h]$. Using a table of values, notice how the average velocities and average speeds approach a specific vector and scalar respectively as $h \to 0$.

h	$\frac{\vec{g}(3+h)-\vec{g}(3)}{h}$	$\left\|\frac{\vec{g}(3+h)-\vec{g}(3)}{h}\right\|$
1	$(2, 5, 3.5)$	6.42
0.1	$(2, 0.5, 3.05)$	3.68
0.01	$(2, 0.05, 3.005)$	3.61
0.001	$(2, 0.005, 3.0005)$	3.61
0.0001	$(2, 0.00005, 3.00005)$	3.60

We might guess that $\vec{g}'(3) \approx (2, 0, 3)$ and $||\vec{g}'(3)|| \approx 3.60$. This limiting process also has a nice geometric representation.

For an linear map $\gamma : \mathbb{R} \to \mathbb{R}^n$, we can write $\gamma(t) = (\gamma_1(t), ..., \gamma_n(t))$ where each $\gamma_i : \mathbb{R} \to \mathbb{R}$ are single variable functions which we call the **component functions** of γ, and we can then tools we know to compute their derivatives $\gamma_i'(t)$. This reduction to single variable calculus will be a recurring theme through multivariable calculus. And as we will later see, the derivative of γ can be written as:

$$\gamma'(t) = (\gamma_1'(t), ..., \gamma_n'(t)).$$

Recall from the last example where we've seen that the **instantaneous speed** of the object at time t is the magnitude of the instantaneous velocity $||\gamma'(t)||$, which is given by

$$||\gamma'(t)|| = \sqrt{\gamma_1'(t)^2 + \cdots + \gamma_n'(t)^2}.$$

Since it is often useful to consider just the direction of the motion at some time t without its magnitude, this leads to a definition of a special unit vector which is in the direction of motion.

Definition 1.2.6 The **unit tangent vector**, denoted $\vec{T} = \vec{T}(t)$ is the unit vector in the direction of $\gamma'(t)$:

$$\vec{T}(t) = \frac{\gamma'(t)}{||\gamma'(t)||}.$$

Example 1.2.7 Referring to example 1.2.5, we can now explicitly compute the velocity by calculating the derivative of each component function:

$$\gamma'(t) = \left(\frac{d}{dt}(2t-6), 5\frac{d}{dt}(t-3)^2, \frac{d}{dt}\left(\frac{1}{2}t^2 - 4\right)\right) = (2, 10(t-3), t).$$

Hence the velocity at $t = 3$ is given by $\gamma'(3) = (2, 0, 3)$ and speed $||\gamma'(3)|| = \sqrt{(13)} \approx 3.60$ which matches our guesses. The direction of motion at $\gamma(3)$ is given by the unit tangent vector:

$$T(3) = \frac{1}{\sqrt{(13)}}(2, 0, 3).$$

We might've also guessed that the **acceleration** of the objection is the instantaneous change in velocity which is the second derivative, so

$$\gamma''(t) = \lim_{h\to 0} \frac{\gamma'(t+h) - \gamma'(t)}{h} = (\gamma_1''(t), .., \gamma_n''(t)).$$

Example 1.2.8 One of the applications of parametric curves is trajectory. Suppose we throw a ball as far as we can, what trajectory does it follow? Let

$$\gamma : [0, \infty) \to \mathbb{R}^3$$

describe the trajectory of this ball, and say at $t = 0$, our initial velocity is $\gamma'(0) = (v_x, v_y, v_z)$ from the origin. Assuming the only force acting on the ball is the downwards acceleration due to gravity, we might guess that its velocity is given by

$$\gamma'(t) = (v_x, v_y, v_z - gt)$$

where $g = 9.81 m/s^2$. Since $\gamma'(t)$ is the derivative of $\gamma(t)$ we might as well write

$$\gamma(t) = \int_0^t \gamma'(u) du,$$

but wait, γ' outputs a vector! However due to the purpose of this chapter, we will ignore rigour for now and assume that integrating over a vector means integrating over each of its components. Then

$$\gamma(t) = \int_0^t \gamma'(u)du = \left(\int_0^t v_x du, \int_0^t v_y du, \int_0^t (v_z - gu)du\right) = \left(v_x t, v_y t, v_z t - \frac{1}{2}gt^2\right).$$

These are the classic kinematics equations for projectile motion in three dimensions! Play with this demo to test out our new expressions!

1.2.2 Frenet frame in three dimensions

Often times, it is very important to be able to describe the motion of an object *relative to its frame*. As we've seen, the direction of motion is given by the unit tangent vector $\vec{T}$. But how is the direction of motion changing? Naturally we want to compute the derivative $\vec{T}'(t)$.

Definition 1.2.9 Given a differentiable vector valued function γ and its unit tangent vector $\vec{T}$, the **principal unit normal**, denoted $\vec{N}(t)$, is given by

$$\vec{N}(t) = \frac{T'(t)}{||T'(t)||}.$$

As the name suggest, the principal unit normal is perpendicular to the unit tangent. This may come as somewhat of a surprise, but it is important to understand why so.

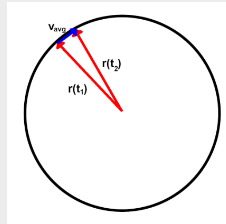
Remark 1.2.10 Algebraically, using the product rule for dot product, which states for two vectors $\vec{x}, \vec{y}$

$$(\vec{x} \cdot \vec{y})' = \vec{x} \cdot \vec{y}' + \vec{x}' \cdot \vec{y}$$

and the fact that $||\vec{T}|| = 1$, we get

$$\frac{d}{dt}(\vec{T} \cdot \vec{T}) = \frac{d}{dt}(1) \implies 2(\vec{T} \cdot \vec{T}') = 0$$

which mean $\vec{T}$ is orthogonal to its derivative. However personally, I think it is better to understand it geometrically. Since $\vec{T}$ has constant magnitude, lets of its as an object rotating on a circle around the origin where we are trying to find its instantaneous velocity, aka the derivative at that point. Using the formula for average velocity over some interval of time, we would get a secant vector, but as we shorten that interval more and more (taking a limit), that secant vector gets so small that it is close to being a vector tangent to the circle. And of course, the tangent to the circle is always perpendicular to the position vector at that point, and hence we see why $\vec{N}$ is orthogonal to $\vec{T}$



Example 1.2.11 Define $\gamma : \mathbb{R} \longrightarrow \mathbb{R}^3$ as $\gamma(t) = (\cos(t), \sin(t), t)$ for all $t \in \mathbb{R}$. We can compute $\gamma', \gamma'', \vec{T}$, and $\vec{N}$. Its derivative is calculated by taking each component function's derivative:

$$\gamma'(t) = \left(\frac{d}{dt}\cos(t), \frac{d}{dt}\sin(t), \frac{d}{dt}t\right) = (-\sin(t), \cos(t), 1)$$

Similarly to the first derivative, to find $\gamma''(t)$, take the derivative of each component of $\gamma'(t)$:

$$\gamma''(t) = \left(-\frac{d}{dt}\sin(t), \frac{d}{dt}\cos(t), \frac{d}{dt}1\right) = (-\cos(t), -\sin(t), 0)$$

Since $\|\gamma'(t)\| = \sqrt{\sin^2(t) + \cos^2(t) + 1} = \sqrt{2}$, it follows that

$$T(t) = \frac{\gamma'(t)}{\|\gamma'(t)\|} = \frac{1}{\sqrt{2}}(-\sin(t), \cos(t), 1)$$

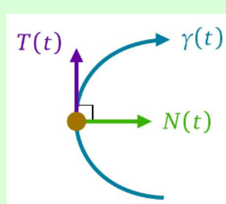
in which case

$$T'(t) = \frac{1}{\sqrt{2}}(-\cos(t), -\sin(t), 0)$$

As $\|T'\| = \frac{1}{\sqrt{2}}$, we see that N is given by

$$N(t) = \frac{T'(t)}{\|T'(t)\|} = \frac{\sqrt{2}}{\sqrt{2}}(-\cos(t), -\sin(t), 0) = (-\cos(t), -\sin(t), 0).$$

Below is an illustration of how the unit normal is oriented on the curve traced by $\gamma(t)$.



Remark 1.2.12 Note $\gamma'(t)$ is always a scalar multiple of $\vec{T}(t)$, however $\gamma''(t)$ is not necessarily a scalar multiple of $\vec{N}(t)$. The above example where $\gamma''(t)$ was indeed a scalar multiple of $\vec{N}(t)$ is just a coincidence because $||\gamma'(t)$ was just a scalar and not a function of t.

The unit tangent and unit normal span a two-dimensional plane, so these cannot be enough to represent all kinds of motion in $\mathbb{R}^3$. Since $\vec{T}$ and $\vec{N}$ are already orthogonal unit vectors, we can choose another unit vector $\vec{B}$ which is orthogonal to both of them. However, both $\vec{B}$ and $-\vec{B}$ will satisfy the requirement. To remove this ambiguity, we defined a vector called the **bi normal unit vector**.

Definition 1.2.13 The **binormal unit vector** $\vec{B}$ is the *unique* unit vector such that $\{\vec{T}, \vec{N}, \vec{B}\}$ form a positively-oriented[a] ordered orthogonal basis in $\mathbb{R}^3$, which means this set of vectors satisfies the right hand rule.

[a] A basis u, v, w in $\mathbb{R}^3$ is positively oriented if $(u \times v) \cdot w > 0$.

This ordered basis $\{\vec{T}, \vec{N}, \vec{B}\}$ forms the **Frenet frame**, or **TNB frame** which described the motion of an object in three dimensions.

Example 1.2.14 Continuing with Example 1.2.11, recall we have already calculated the unit tangent $\vec{T}$ and the unit normal $\vec{N}$ for the path $\gamma(t) = (\cos(t), \sin(t), t)$. To find $\vec{B}$ algebraically, we use the cross product

$$\vec{B} = \vec{T} \times \vec{N}.$$

The cross product $a \times b$ of two vectors $a = (a_1, a_2, a_3)$ and $b = (b_1, b_2, b_3)$ can be calculated by expressing it as a "determinant". That is, if $\{e_1, e_2, e_3\}$ is the standard basis in $\mathbb{R}^3$, then

$$a \times b = \det \begin{bmatrix} e_1 & e_2 & e_3 \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{bmatrix}$$

This is not really a determinant (since e_1, e_2, e_3 are vectors), but by naively following the rules of calculating determinants, we will end up with the correct expression. For this example:

$$\vec{B} = \vec{T} \times \vec{N} = \frac{1}{\sqrt{2}} \det \begin{bmatrix} e_1 & e_2 & e_3 \\ -\sin(t) & \cos(t) & 1 \\ -\cos(t) & -\sin(t) & 0 \end{bmatrix} = \frac{1}{\sqrt{2}} \sin(t) e_1 - \frac{1}{\sqrt{2}} \cos(t) e_2 + \frac{1}{\sqrt{2}} e_3,$$

so $\vec{B}(t) = \frac{1}{\sqrt{2}}(\sin t, -\cos t, 1)$. View this demo of the TNB frame for γ. The green arrow is $\vec{T}(t)$, the red is $\vec{N}(t)$ and the orange is $\vec{B}(t)$. Notice how they remain orthogonal to each other throughout the motion while being positively oriented.

1.2.3 Geometry of curves

We define the **trace** of a parametric curve γ as the image of γ, notice that many different parametric curves can have the same trace.

Example 1.2.15

- Define $\gamma_1 : [0, \pi] \longrightarrow \mathbb{R}^2$ by $\gamma_1(t) = (\cos(2t), \sin(2t))$. Then γ_1 traces the unit circle twice as fast.
- Define $\gamma_2 : [0, 2\pi] \longrightarrow \mathbb{R}^2$ by $\gamma_2(t) = (\cos(t - \pi), \sin(t))$. Then γ_2 traces the unit circle but starts at π instead of 0 radians of rotation.
- Define $\gamma_3 : [0, 6\pi] \longrightarrow \mathbb{R}^2$ by $\gamma_3(t) = (\cos(t), -\sin(t))$. Then γ_3 traces the unit circle three times in the opposite direction.
- Define $\gamma_4 : [0, 14.1] \longrightarrow \mathbb{R}^2$ by $\gamma_4(t) = \left(\cos\left(\frac{t}{4}\sin(t)\right), \sin\left(\frac{t}{4}\sin(t)\right)\right)$. Then γ_4 traces the unit circle in an...interesting way.

View this animation for a visual demonstration of the four parametric curves above.

Always remember that the trace of a parametric curve is simply a set, and we should not refer to this set as a "curve". We will explore these issues much later on, but for now we can take the following definition for granted.

Definition 1.2.16 Let $C \subseteq \mathbb{R}^n$ be a set. We say C is a **curve** in $\mathbb{R}^n$ if C is the trace of a continuous parametric curve $\gamma : I \longrightarrow \mathbb{R}^n$.

Often times we might be concerned with the shape of C but not how C is traced out, for example, how many times does C cross itself? How curvy is C? In other words, we may want to study curves C without worrying how it is described. Hence we also ways to describe a curve without directly using parametric curves.

Example 1.2.17 Define the set

$$C = \left\{(x, y) \in \mathbb{R}^2 : y = x^2, -2 \leq x \leq 2\right\}$$

so C describes the graph of the parabola $y = x^2$ on the domain $[-2, 2]$. Intuitively, we would consider C to be a curve and we can indeed quickly prove it. Define the parametric curve $\gamma : [-2, 2] \to \mathbb{R}^2$ as $\gamma(t) = \left(t, t^2\right)$ which yields a trace $\gamma([-2, 2]) = C$. Since γ is continuous and $\gamma([-2, 2]) = C$, it follows that C is a curve.

Hopefully this section provided a brief foray into the interesting applications of curves and whats to come!

1.3 Real-valued functions

In the previous section, we focused on maps of the form $\mathbb{R} \to \mathbb{R}^n$, this section is about maps of the opposite form

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

We often refer to these maps as **real-value functions**.

1.3.1 Scalar fields and densities

In physics, real-valued functions are called **scalar fields** or **scalar functions** or **potentials**. Below are many examples of real-value functions as it is important to know how to switch between formal and informal languages.

Example 1.3.1 Meteorologists use temperature to help predict the weather as it can allude to cloud formation and the movement of huge climate systems. Temperature itself is a scalar value (measured in Celsius) and varies depending on where you are on Earth. At any given moment, the temperature $T(x, y, z)$ depends on your position (x, y, z) on Earth. For example, suppose (x_A, y_A, z_A) is a position in the Arctic and (x_D, y_D, z_D) is a position in the Sahara Desert. You might guess that $T(x_A, y_A, z_A) < T(x_D, y_D, z_D)$ since the Sahara Desert is always much hotter than the Arctic.

Example 1.3.2 Mass is never distributed uniformly among objects because there are always more dense regions and less dense regions in a mass. For example, a block of swiss cheese has regions where there is no mass (holes) and regions where cheese is tightly packed. Since the density of the cheese depends on where in the cheese you are looking then we can describe the density with a scalar field $\varphi : C \to [0, \infty)$ where $C \subseteq \mathbb{R}^3$ is the set of points in the cheese. Our scalar field $\varphi(x, y, z)$ outputs the density in units kg/m^3 at (x, y, z). For instance, suppose $p \in C$ is a point inside a hole. Then one would expect $\varphi(p) \approx 0$ kg/m^3 as there's no mass in the hole.

Example 1.3.3 There are forces in physics that have a special connection with potentials, namely conservative forces. For example, the electrostatic force for a point charge is related to the scalar field $V : \mathbb{R}^3 \to \mathbb{R}$ defined by

$$V(x, y, z) = \frac{Q}{4\pi\varepsilon_0\sqrt{x^2 + y^2 + z^2}}.$$

The function V is often referred to as the electric potential where ε_0 is the permittivity of free space constant and Q is the charge constant. As you will see much later in vector calculus, this function can describe how the electrostatic forces of two different charges influence each other as they move nearly.

Real-valued functions arise in many other fields of study as well.

Example 1.3.4 Economists and businesses strive to maximize profit or minimize costs subject to many constraints. They must account for many parameters before making decisions resulting in real-valued functions being highly important. For example, suppose you are CEO of a company named CHAYR and must produce 20,000 chairs. The number of chairs they can produce is given by the Cobb-Douglas function:

$$P(K, L) = \frac{1}{25} K^{1/4} L^{1/3}$$

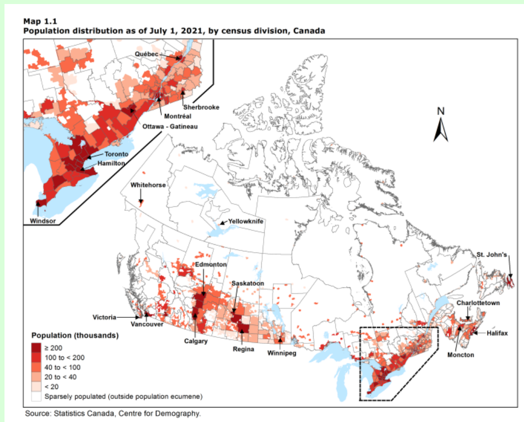
where K is their capital expenditure and L is their labour costs. The total cost is therefore $C(K, L) = K + L$. You must decide how to spend your money the minimizes costs and still produces 20,000 chairs. In other words, you must minimize $C(K, L)$ subject to the constraint that $P(K, L) = 20,000$. You will learn how to solve such multivariable optimization problems.

Example 1.3.5 A streaming site FLYX uses very complicated algorithms to find content to recommend to you. Every movie or show you watch generates data points that FLYX stores and uses to associate a rough categorization of the type of viewer you are. An example of such a data point would be how many hours you watched a specific genre/style. Then, before Netflix recommends you a movie or show, it evaluates the data, represented in n variables $x_1, \ldots, x_n$, using some function $E : \mathbb{R}^n \to \mathbb{R}$ that outputs a score $E(x_1, \ldots, x_n)$. FLYX recommends the shows with the highest scores because they believe you will enjoy the movie or show. But how did FLYX create the magical real-valued function E ? Real-valued functions are very important for data analysis.

A special kind of real-valued functions are those that are non-negative, these are often referred to as **densities**. Many densities are defined by counting a quantity and dividing by a unit of measurement.

Example 1.3.6 Population density, the number of people per unit area, of a country is valuable information for societal statistics as well as future development. Let

$C \subseteq \mathbb{R}^2$ be the set of points in Canada (ideally we would need 3 variables, but let's assume C is the set of points on a 2D map of Canada). If $\varphi : C \to [0, \infty)$ is Canada's population density function measured in persons per square kilometre, then $\varphi(x, y)$ should be approximately the number of people in a 1 km by 1 km square centred at (x, y). Below is a heatmap of φ.



As one would expect, the GTA (Greater Toronto Area) is quite red implying high population density because of the limited space and large number of people. Northern regions like Nunavut remain less shaded (low population density) due to the large amount of land, but few inhabitants.

1.3.2 Graphs, level sets, and slices

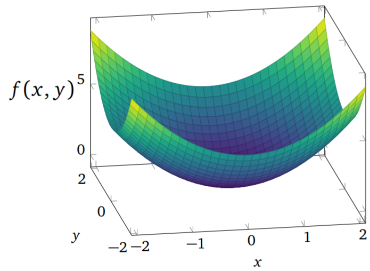
There are a few ways to visualize a multivariable function which are all defined as sets in $\mathbb{R}^n$. The most basic one is a graph.

Definition 1.3.7 Let $A \subseteq \mathbb{R}^n$. The **graph of a function** $f : A \to \mathbb{R}$ is the set in $\mathbb{R}^{n+1}$ given by

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

The graphs of function $f : \mathbb{R} \to \mathbb{R}$ have been taught to us in high school and it lies in $\mathbb{R}^2$. For a two-variable real-valued function, its graph is also called a **surface plot**.

Example 1.3.8 Define $f : \mathbb{R}^2 \to \mathbb{R}$ by $f(x, y) = x^2 + y^2$. The graph of f is the set $\{(x, y, x^2 + y^2) : (x, y) \in \mathbb{R}^2\}$, as plotted below.

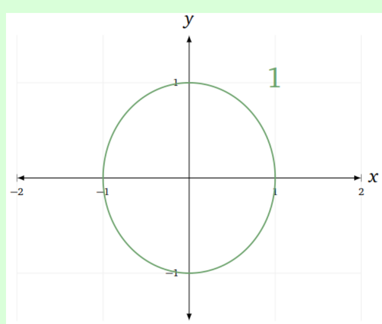


However, we cannot visualize any real-valued functions whose input space is greater or equal to three-dimension, as we cannot directly visualize four-dimensional space. So it is often helpful to "reduce dimensions". There are several ways of doing so.

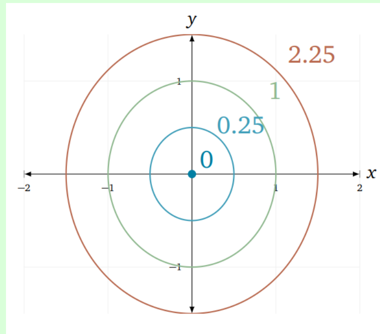
Definition 1.3.9 Let $A \subseteq \mathbb{R}^n$ and $f : A \to \mathbb{R}$ be a real-valued function. Fix $k \in \mathbb{R}$. The **level set** of f at k is the set $\{x \in \mathbb{R}^n : f(x) = k\}$. This is also referred to as the **k-level set**.

Remark 1.3.10 A level set in $\mathbb{R}^2$ is also called a **contour**, and for graph of 2-variable functions, we can create a **contour plot** by plotting the level sets for many different values.

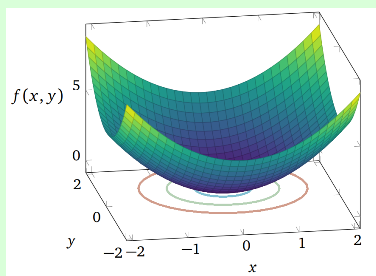
Example 1.3.11 Define the function $f : \mathbb{R}^2 \to \mathbb{R}$ as $f(x, y) = x^2 + y^2$. You can visualize f and its graph using only its level sets. Begin with the 1-level set, 0-level set and (-1)-level set. By definition, the 1-level set of f is the set $\{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$, which is the unit circle in $\mathbb{R}^2$. This single contour is plotted below.



The " 1 " indicates that for any (x, y) on this contour, $f(x, y) = 1$. The 0-level set is given by $\{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 0\} = \{(0, 0)\}$, since only the origin satisfies $x^2 + y^2 = 0$. Hence, this contour is a single point. Any k-level set for $k < 0$ contains points in the set $\{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = k\} = \varnothing$, since there are no points (x, y) satisfying $x^2 + y^2 < 0$. By plotting a few more contours, you obtain a contour plot.

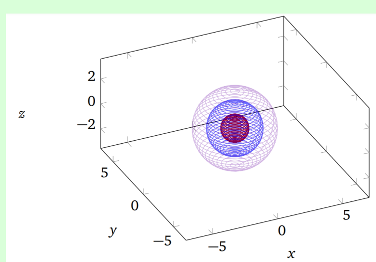


How does this contour plot correspond to the graph of f ? Imagine raising each contour by their value out of the page. For example, the 0 point remains on the page (the xy-plane) and the 0.5 circle raises out of the page by 0.5. This recreates a "skeleton" graph of f, as illustrated below.



Play with this demo to see how the contours correspond to the graph of f. Toggle switches in order. Also, watch this demo to see how the k-level set relates to the graph of f as k varies from -1 to 4 .

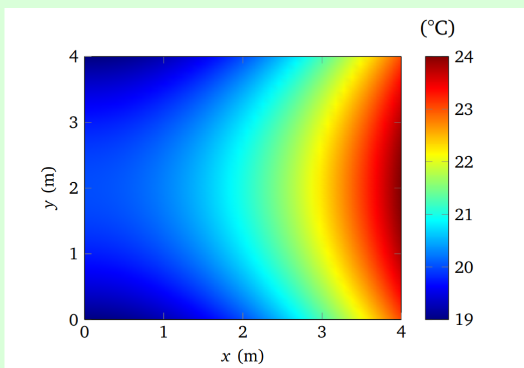
Example 1.3.12 Define the function $f(x, y, z) = x^2 + y^2 + z^2$. The graph of f lies in $\mathbb{R}^4$ so it cannot be plotted, but we can indeed plot its level sets as they lie in $\mathbb{R}^3$. Notice that the k-level set of f is just a sphere of radius $\sqrt{k}$ for $k \geq 0$.



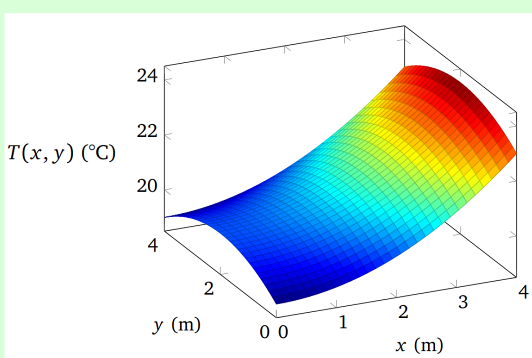
The innermost red sphere is the 1-level set, the middle blue sphere is the 4-level set, and the outermost pink sphere is the 9-level set. Three-dimensional level sets may seem abstract and not that helpful at a glance; however, they can offer valuable interpretations. For example, if f describes the magnitude of force felt by a mass at position (x, y, z), then the k-level set describes all positions in three-dimensional space that a mass would feel a force of k.

Another way of reducing dimensions we'll talk about is to use a colour gradient that corresponds to the values of the function f, which are called **heat maps**. These are like a continuous version of the contour plot.

Example 1.3.13 Suppose you are sitting in a 4 m by 4 m room with a window. The sun radiates heat through the window increasing the temperature of the room. You model the temperature in Celsius using the function $T : [0,4]^2 \longrightarrow [0,\infty)$ defined as $T(x,y) = 0.25\left(x^2 - (y-2)^2\right) + 20$. The input, x and y, are measured in meters and describe your position in the room. To visualize T on its domain, you can create a heat map:



The highest temperature, indicated by the red, is in the vicinity of the window as expected. You can see how this corresponds to the actual graph of T :



The last way we'll talk about is **slicing**, which is achieved by fixing a variable. For 2-variable functions:

Definition 1.3.14 Let $A \subseteq \mathbb{R}^2$ and $f : A \to \mathbb{R}$ be a real-valued function.

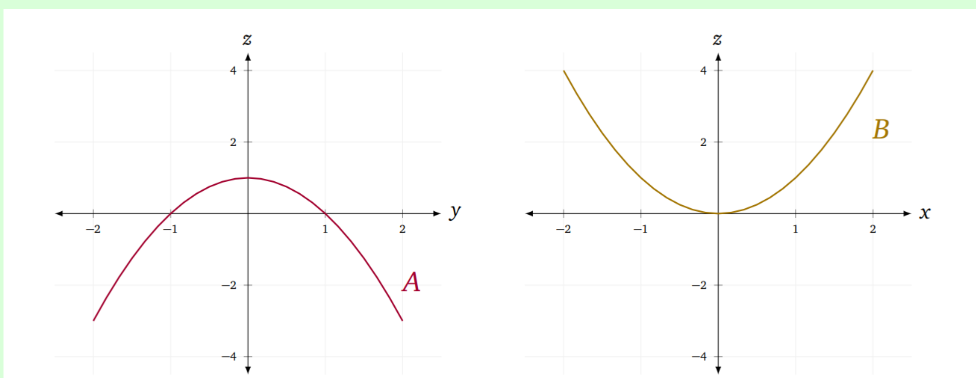
- For fixed $a \in \mathbb{R}$, the x**-slice** at a of the graph of f is the set

$$\{(y, z) \in \mathbb{R}^2 : (a, y) \in A, z = f(a, y)\}.$$

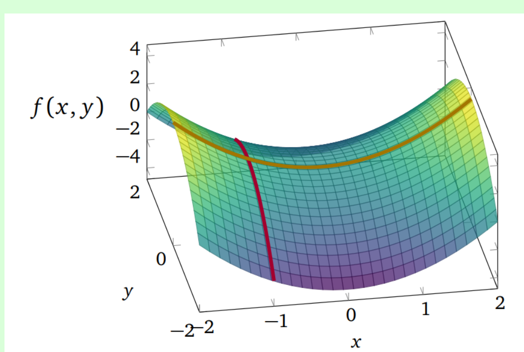
- For fixed $B \in \mathbb{R}$, the y**-slice** at B of the graph of f is the set

$$\{(x, z) \in \mathbb{R}^2 : (x, b) \in A, z = f(x, b)\}.$$

Example 1.3.15 Consider the graph of $f(x, y) = x^2 - y^2$. The x-slice at $x = -1$ is the set $A = \left\{(y, z) : z = 1 - y^2\right\}$ and the y-slice at $y = 0$ is the set $B = \left\{(x, z) : z = x^2\right\}$. Notice both A and B are sets in $\mathbb{R}^2$. These are plotted separately below.



And to visualize how these 2D slices correspond to the graph of f:



You can play with this demo to view different slices.

Similarly, we can define slices for 3-variable functions.

Definition 1.3.16 Let $A \subseteq \mathbb{R}^3$ and $f : A \to \mathbb{R}$ be a real-valued function.

- For fixed $a \in \mathbb{R}$, the x-slice at a of the graph of f is the set

$$\left\{(y, z, w) \in \mathbb{R}^3 : (a, y, z) \in A, w = f(a, y, z)\right\}.$$

- For fixed $b \in \mathbb{R}$, the y-slice at b of the graph of f is the set

$$\left\{(x, z, w) \in \mathbb{R}^3 : (x, b, z) \in A, w = f(x, b, z)\right\}.$$

- For fixed $c \in \mathbb{R}$, the z-slice at c of the graph of f is the set

$$\left\{(x, y, w) \in \mathbb{R}^3 : (x, y, c) \in A, w = f(x, y, c)\right\}.$$

These slices are sets in $\mathbb{R}^3$ so we can still plot them.

In this section, we illustrated many ways to visualize real-valued functions of two or three variables, it is good to be familiar with all of them and their relationships with the graph of the function.

1.4 Vector fields

In this section, we will talk about maps of the form

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

This is a special case where the dimension of the domain and codomain are the same. There are two major interpretations of these maps and here, we will view them as vector fields.

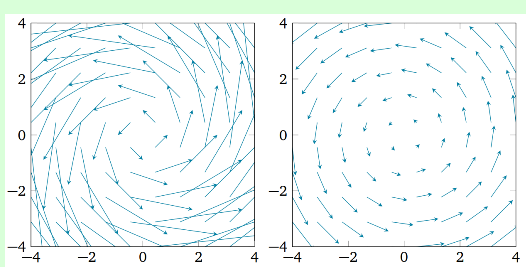
Definition 1.4.1 An n-dimensional **vector field** is a function F with domain and codomain lying in $\mathbb{R}^n$.

The name "vector field" is inspired by physical examples, and vectors $F(x)$ often represents the velocity of a fluid or a force at the point x so it is also known as **velocity field** or **force field** in those context.

Example 1.4.2 The ocean is an example of a vector field; at any given moment in time, each point x in the ocean has a velocity $F(x)$. The same is true of atmospheric winds and weather patterns, like hurricanes. Magnetic field generated by a magnet is another example of a vector field, where each point x is influenced by a force $F(x)$ imposed by the magnet. This also applys to gravitational force fields like planet Earth.

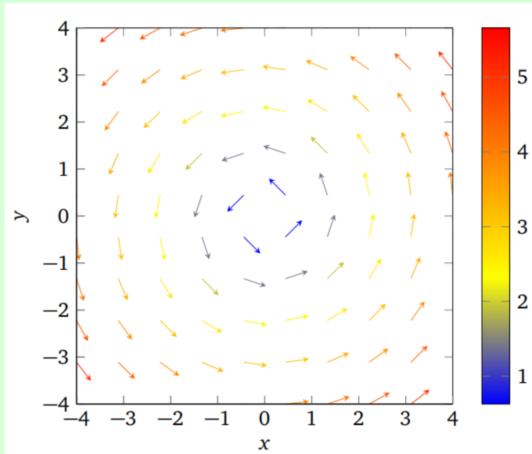
Given a two or three dimensional vector field, how would we plot it?

Example 1.4.3 Let F be the two-dimensional vector field defined by $F(x, y) = (-y, x)$. Then at each point (a, b) we would the vector $(-b, a)$. By doing this process for many points on a grid, we can produce a **vetor field plot** as shown below on the left.



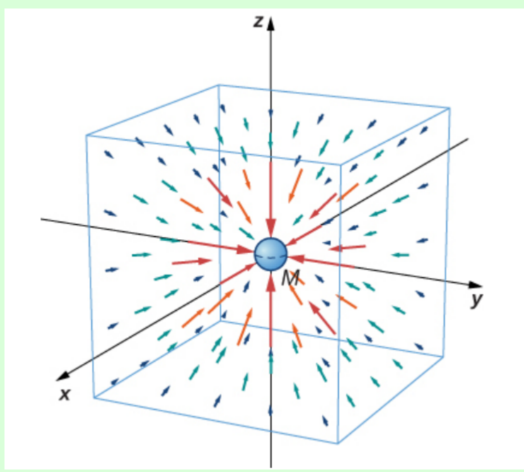
As we can see, the picture on the left is very messy as they are plotted according to scale. To avoid this, we proportionally rescale the vector so they are smaller, which here resulted in the plot on the right. Since the relative size of the vectors remain the same, we can still see where the "fluid" is moving faster.

Alternatively, we can scale all the vectors to be the same size and use a colour gradient to represent the magnitude.



Three-dimenisonal vector fields can be visualized in a similar fashion.

Example 1.4.4 The gravitation force field for earth is an example of a vector field in $\mathbb{R}^3$, notice the vectors are coloured according to the magnitude of their force.



Remark 1.4.5 It is nice to be familiar with some with equivalent notation for vector fields:

$$F = (x^2, yx, -z), \quad F = \langle x^2, yx, -z \rangle, \quad F = [x^2, yz, -z], \quad F = x^2\hat{\imath} + yx\hat{\jmath} - z\hat{k}.$$

1.5 Coordinate transformations

In this section we will study the same kind of maps as the previous section, however we will view them as transformation instead of vector field. In this text, any map with domain and codomain in $\mathbb{R}^n$ will be referred to as a **transformation**. Note that the domain and codomain should be subsets lying in the same dimension, and it is usually continuous.

Example 1.5.1 The transformation $f : \mathbb{R}^2 \to \mathbb{R}^2$ given by $f(u, v) = (u + v, u - v)$ is a linear transformation. The transformation $g : \mathbb{R}^2 \to \mathbb{R}^2$ and $h : \mathbb{R}^2 \to \mathbb{R}^2$ given by $g(u, v) = (u^2 + v^2, v)$ and $h(u, v) = (u + 1, 0)$ are not linear. We can also show that f is invertible while g and h are not.

A **coordinate transformation** $f : A \to B$ will refer to a continuous transformation that is usually bijective, and the domain A and map f create a **coordinate system** for the codomain B.
If $b = f(a)$ we can informally say any one of:

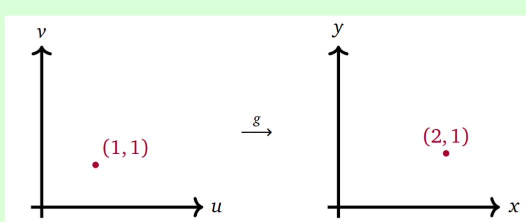
"The point b can be written as the point a in the coordinate system defined by f."
"The point b in B-space corresponds to the point a in A-space."

This is useful when we are plotting subsets of B and wish to describe these subsets using the coordinate system defined by f and A.

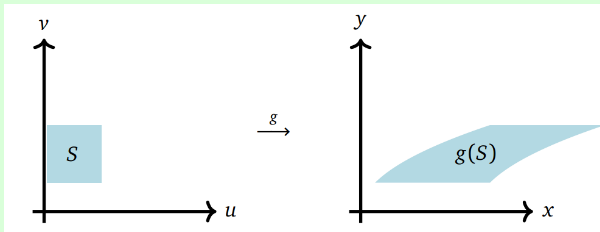
Example 1.5.2 Consider the transformation $g : \mathbb{R}^2 \to \mathbb{R}^2$ defined by $g(u, v) = \left(u^2 + v^2, v\right)$. Since the pair (u, v) denotes a point in the domain $\mathbb{R}^2$, it is a common convention to use another pair, usually (x, y), to distinguish points in the codomain $\mathbb{R}^2$. Then we may write

$$(x, y) = \left(u^2 + v^2, v\right)$$

to equivalently describe the coordinate transformation g. Using this notation, take the point $g(1, 1) = (2, 1)$ for example, this implies the point $(2, 1)$ in (x, y) space corresponds to the point $(1, 1)$ in the (u, v)-space, note $g(-1, 1) = (2, 1)$ as well hence g is not invertible.
Transformations in $\mathbb{R}^2$ are commonly visualized using two planes. For example, you can plot the point $(1, 1)$ in the (u, v)-plane and its image $g(1, 1) = (2, 1)$ in the (x, y)-plane.



This idea generalizes to any set $S \subseteq \mathbb{R}^2$ in the (u, v)-plane and its image $g(s) \subseteq \mathbb{R}^2$ in the $(x, y) - plane$. For example, you can use $S = [0, 1] \times [0.5, 1.5]$ and obtain the plot below.



Coordinate transformations are extremely important because they allow you to describe the same set of set of points in many ways, and thus can dramatically simplify many scenarios why choosing the correct coordinate system. This features mirrors "change-of-basis" in linear algebra. We will study three fundamental coordinates systems: polar coordinates in $\mathbb{R}^2$, cylindrical coordinates in $\mathbb{R}^3$, and spherical coordinates in $\mathbb{R}^3$.

1.5.1 Polar coordinates

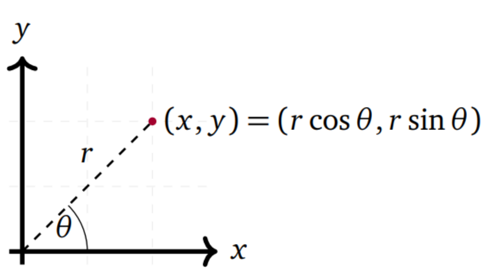
A point in the (x, y)-plane can be described using its distance from the origin and the polar angle. Formally, we define the **polar coordinate transformation** $T : \mathbb{R}^2 \to \mathbb{R}^2$ by

$$T(r, \theta) = (r \cos\theta, r \sin\theta).$$

The variable r indicates the radius and θ is the polar angle. Notice r can be negative! Informally we write

$$(x, y) = (r \cos\theta, r \sin\theta)$$

and geometrically this means:



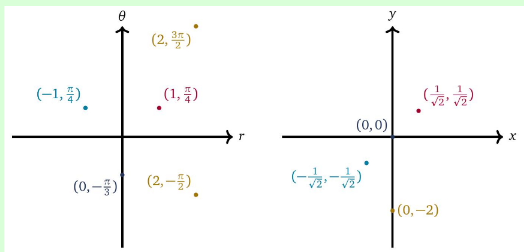
however, remember that the polar angle θ is not unique.

Example 1.5.3 The polar coordinates transformation T has a lot of symmetries. By direct calculation,

$$T\left(1, \frac{\pi}{4}\right) = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right), \quad T\left(-1, \frac{\pi}{4}\right) = \left(-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right),$$

$$T\left(2, \frac{-\pi}{2}\right) = T\left(2, \frac{3\pi}{2}\right) = (0, -2), \quad T\left(0, -\frac{\pi}{3}\right) = 0.$$

Plotting the transformation of these points:



Notice that infinitely many points in the (r, θ)-plane correspond to the same point in the (x, y)-plane.

Remark 1.5.4 Sometimes we might plot points in rectangular coordinates but *label* them in polar coordinates. For example, the point $(-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}})$ can be written as $(-1, \frac{pi}{4})$ in polar coordinates. Many authors only write $(-1, \frac{pi}{4})$ instead of $T(-1, \frac{pi}{4})$, so we have to be constantly aware on whether the labelling is in rectangular or polar.

As we will see in the next few examples, polar coordinates describes shapes like circles, hyperbola, ellipses much more simply than rectangular coordinates.

Example 1.5.5 What does the polar equation $r = 2$ represent? Informally, notice that

$$r^2 = (r\cos\theta)^2 + (r\sin\theta)^2 = x^2 + y^2.$$

Thus the equation $r = 2$ in polar coordinates should represent the circle $x^2 + y^2 = 4$. To prove this formally, $r = 2$ corresponds to the set

$$A = \{(r, \theta) : r = 2, \theta \in \mathbb{R}\} = \{(2, \theta) : \theta \in \mathbb{R}\}$$

Under the transformation T, we get

$$T(A) = \{(2\cos\theta, 2\sin\theta) : \theta \in \mathbb{R}\}$$

which is precisely the circle of radius 2 centred at $(0,0)$.

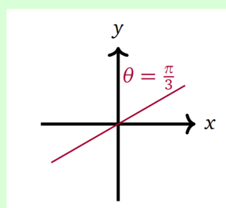
Example 1.5.6 How about the polar equation $\theta = \frac{\pi}{3}$? It is trickier to solve this equation informally as we must solve the equations $x = r\cos\theta, y = r\sin\theta$ for θ. This will involve inverse trig functions but we also don't know about the range of θ. So we will go the formal route. The polar equation $\theta = \frac{\pi}{3}$ corresponds to the set

$$B = \{(r,\theta) : r \in \mathbb{R}, \theta = \frac{\pi}{3}\} = \{(r, \frac{\pi}{3}) : r \in \mathbb{R}\}$$

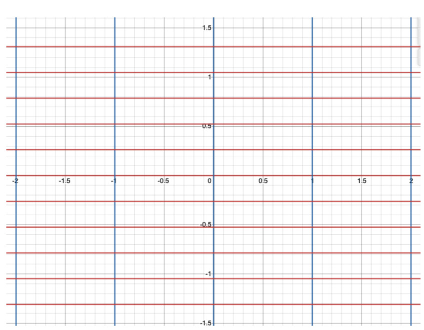
Note that r can be any real number as long as we don't specify $r \geq 0$. The set then becomes

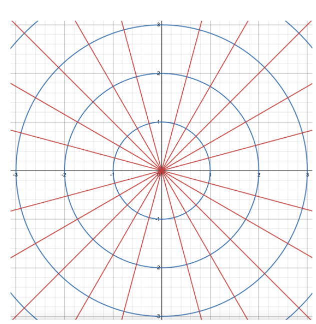
$$T(B) = \{(r\cos\frac{\pi}{3}, r\sin\frac{\pi}{3}) : r \in \mathbb{R}\} = \{\frac{r}{2}, \frac{r\sqrt{3}}{2} : r \in \mathbb{R}\}$$

under the polar coordinate transformation T, which is a line through the origin. This is plotted below:



Coordinate transformations typically preserve some kind of geometric properties or modify them in a predictable manner. This includes the *polar coordinate transformation* and we will now use two graphs to see the patterns. On the left-hand side, the grid lines are plotted in the (r,θ)2-plane. The blue vertical grid lines correspond to $r = a$ for various a and the red horizontal grid lines correspond to $\theta = b$. The picture on the right hand side in the (x,y)-plane are the grid lines after the transformation.





We have seen that the polar coordinate transformation is not bijective on its entire domain, however, restricting its domain to a subset, we can obtain a bijection.

Lemma 1.5.7 Let

$$A = (0, \infty) \times (-\pi, \pi) \text{ and } B = \mathbb{R}^2 \backslash \{(x, 0) : x \leq 0\},$$

then the polar coordinate transformation $T|_A : A \longrightarrow B$ defined by

$$T(r, \theta) = (r \cos \theta, r \sin \theta)$$

is bijective.

Proof. First, we claim that range $T|_A \subseteq B$. Suppose not, then there exist $r \cos \theta \leq 0$ and $r \sin \theta = 0$. This is only possible if θ is an odd integer multiple π, but since $\theta \in (-\pi, \pi)$, we have a contradiction, this proves the claim.
To prove f is injective, let $(r_1, \theta_1), (r_2, \theta_2) \in A$ be such that $f(r_1, \theta_1) = f(r_2, \theta_2)$. Then $r_1 \cos \theta_1 = r_2 \cos \theta_2$ and $r_1 \sin \theta_1 = r_2 \sin \theta_2$. It follows that

$$r_1^2 = (r_1 \cos \theta_1)^2 + (r_1 \sin \theta_1)^2 = (r_2 \cos \theta_2)^2 + (r_2 \sin \theta_2)^2 = r_2^2.$$

Hence $r_1 = r_2$ are they must both be positive. This implies $(\cos \theta_1, \sin \theta_1) = (\cos \theta_2, \sin \theta_2)$. As $\theta_1, \theta_2 \in (-\pi, \pi)$, we have $\cos \theta_1 = \cos \theta_2$ implies $\theta_1 = \pm \theta_2$, similarly $\sin \theta_1 = \sin \theta_2$ implies $\theta_1 = \theta_2$ or $\theta_1 = \pi - \theta_2$ or $\theta_1 = -\pi - \theta_2$. These conditions are satisfied if and only if $\theta_1 = \theta_2$, so we can conclude that $(r_1, \theta_1) = (r_2, \theta_2)$.
Finally, to prove that f is surjective. Let $(x, y) \in B$, Define $r = \sqrt{x^2 + y^2}$ and

$$\theta = \begin{cases} \arccos(\frac{x}{r}) & \text{if } y \geq 0, \\ -\arccos(\frac{x}{r}) & \text{if } y < 0, \end{cases}$$

so $r > 0$ and $\theta \in (-\pi . \pi)$ by definition.By checking the two cases depending on the sign of y, we can verify that $f(r, \theta) = (x, y)$. This proves that f is surjective and hence bijective, as required.

The proof of the above lemma shows that defining an "inverse" for the polar coordinate transformation is tricky. Many texts suggest the following relationship

$$r = \sqrt{x^2 + y^2}, \quad \theta = \arctan \frac{y}{x},$$

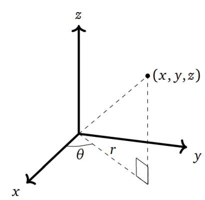
but we must be careful as this only holds for $r \in (0, \infty)$ and $\theta \in (-\frac{\pi}{2}, \frac{\pi}{2})$.

1.5.2 Cylindrical coordinates

We can extend the idea of polar coordintaes in $\mathbb{R}^2$ to a coordinate system in $\mathbb{R}^3$ in two ways. We define the **cylindrical coordinate transformation** $T : \mathbb{R}^3 \to \mathbb{R}^3$ by

$$T(r, \theta, z) = (r\cos\theta, r\sin\theta, z).$$

The variable r is the *polar radius*, θ is the *polar angle*, and the variable z is the usual rectangular coordinate. Informally, we say the point (x, y, z) can be written as (r, θ, z) in **cylindrical coordinates**.



This allows us to plot points in rectangular coordinates but label them in cylindrical coordinates. As with polar coordinates, there are infinitely many ways to write a rectangular coordinate in cylindrical coordinates. Much like the name suggests, cylindrical coordinates are useful for describing objects with rotational symmetry about the z-axis in simpler terms.

Example 1.5.8 What does the equation $r = 2$ represent in $\mathbb{R}^3$? Formally, it is the set

$$A = \{(r, \theta, z) : r = 2, \theta \in \mathbb{R}, z \in \mathbb{R}\} = \{(2, \theta, z) : \theta \in \mathbb{R}, z \in \mathbb{R}\}$$

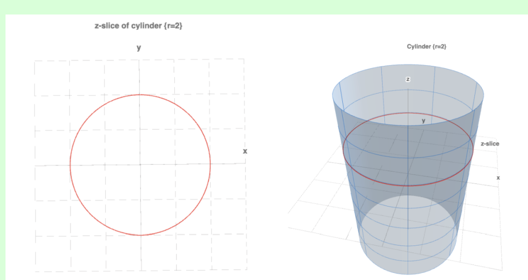
Then

$$T(A) = \{(2\cos\theta, 2\sin\theta, z) : \theta \in \mathbb{R}, z \in \mathbb{R}\},)$$

by fixing some for value for z, we see that the z-slice of $T(A)$ is the set

$$\{(2\cos\theta, 2\sin\theta) : \theta \in \mathbb{R}\} \subseteq \mathbb{R}^2,$$

which is a circle of radius 2, hence $r = 2$ creates a cylinder in $\mathbb{R}^3$!



View this demo for a visualization of this shape.

Example 1.5.9 From our example using polar coordinates in $\mathbb{R}^2$, we can guess that the cylindrical equation $\theta = \frac{\pi}{4}$ represents a plane passing through the z-axis. More formally, its image is the set

$$B = \{(r, \frac{\pi}{4}, z) : r, z \in \mathbb{R}\}$$

and so its under the transformation T, we get

$$T(B) = \{(r\cos\frac{pi}{4}, r\sin\frac{\pi}{4}) : r, z \in \mathbb{R}\} = \{(\frac{r}{\sqrt{2}}, \frac{r}{\sqrt{2}}, z) : r, z \in \mathbb{R}\}$$

View this demo for a visualization of this plane.

Example 1.5.10 The cylindrical equation $z = -1$ is just a flat plane, as its image is the set

$$C = \{(r, \theta, -1) : r, \theta \in \mathbb{R}\} \implies T(C) = \{(r\cos\theta, r\sin\theta, -1) : r, \theta \in R\}$$

The cylindrical coordinate transformation is also a bijection once we restrict its domain.

Lemma 1.5.11 Let

$$A = (0, \infty) \times (-\pi, \pi) \times \mathbb{R} \text{ and } B = \mathbb{R}^3 \backslash \{(x, 0, z) : x \leq 0, z \in \mathbb{R}\},$$

then the cylindrical coordinate transformation $T|_A : A \to B$ defined by

$$T(r, \theta, z) = (r\cos\theta, r\sin\theta, z)$$

is bijective.

Proof. An analogous proof can be given using the arguments for polar coordinates.

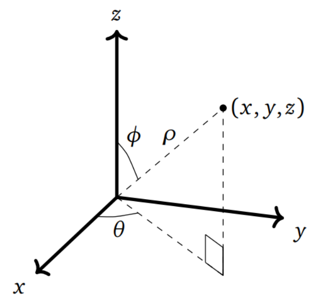
There are many other ways to restrict the domain to obtain a bijection with cylindrical coordinates, but for now, these examples will be enough. The key takeaway is that this coordinate system can nicely describe objects with rotational symmetry about eh z-axis.

1.5.3 Spherical coordinates

Another way of extending polar coordinates in $\mathbb{R}^2$ is through spherical coordinates in $\mathbb{R}^3$. We define the **spherical coordinate transformation** $T : \mathbb{R}^3 \to \mathbb{R}^3$ by

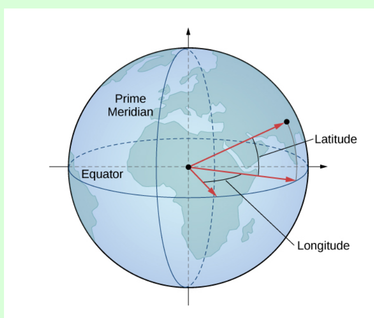
$$T(\rho, \theta, \phi) = (\rho\cos\theta\sin\phi, \rho\sin\theta\sin\phi, \rho\cos\phi).$$

The variable ρ is the *(spherical) radius*, the variable θ is the *polar angle (or azimuthal angle)*, and the variable ϕ is the *inclination angle (or zenith angle)*. Similarly to polar coordinates, ρ, θ, ϕ can be any real numbers, despite many resources restricting their values. Informally, we say the point (x, y, z) can be written as (ρ, θ, ϕ) in **spherical coordinates**. The geometry of this coordinates system is shown below.



Play around with this demo to get a feeling of how spherical coordinates work.

Example 1.5.12 Locations on the Earth's surface are described with latitude and longitude, this is an example of spherical coordinates. Roughly speaking, longitude relates to the polar angle θ and latitude relates to the azimuthal angle ϕ. Although the Earth is not a perfect sphere, it spherical radius ρ is $\approx 6,378km$.



As usual, there are infinitely many ways to write some point in (x, y, z)-space as a spherical coordinates.

Example 1.5.13 Lets look at what the equation $\rho = 2$ represent in $\mathbb{R}^3$. Since ρ is the distance from the origin, we might guess that we should get a sphere with radius 2. This is indeed correct and we can verify it in many ways. Informally, notice

$$\begin{aligned} x^2 + y^2 + z^2 &= \rho^2 \cos^2\theta \sin^2\phi + \rho^2 \sin^2\theta \sin^2\phi + \rho^2 \cos^2\phi \\ &= \rho^2 \sin^2\phi + \rho^2 \cos^2\phi \\ &= \rho^2 = 4, \end{aligned}$$

Formally, the spherical equation correspond to the set of points

$$A = \{(\rho, \theta, \phi) : \rho = 2, \theta \in \mathbb{R}, \phi \in \mathbb{R}\} = \{(2, \theta, \phi) : \theta \in \mathbb{R}, \phi \in \mathbb{R}\}$$

Applying T, we get

$$T(A) = \{(2\cos\theta \sin\phi, 2\sin\theta \sin\phi, 2\cos\phi) : \theta \in \mathbb{R}, \phi \in \mathbb{R}\}$$

but what is the shape of this set? From the equation above, we can show every point in $T(A)$ satisfies $x^2 + y^2 + z^2 =$ so $T(A)$ is a subset of the sphere of radius 2, but to show $T(A)$ is equal to the sphere of radius 2, we need to show that opposite direction as well. This requires some extra work and is left as an exercise for the reader. (Apply the same idea with how you proved Lemma below)

Example 1.5.14 The spherical equation $\theta = \frac{\pi}{4}$ represents the same object as it does in cylindrical coordinates. This should not be a surprise the polar angle is fixed in both cases, while the other two variable can span the whole 2d plane. Formally this conclusion is not as obvious. If $B = \left\{\left(\rho, \frac{\pi}{4}, \phi\right) : \rho, \phi \in \mathbb{R}\right\}$ then its image under the spherical coordinate transformation is

$$\begin{aligned} T(B) &= \left\{\left(\rho\cos\frac{\pi}{4}\sin\phi, \rho\sin\frac{\pi}{4}\sin\phi, \rho\cos\phi\right) : \rho, \phi \in \mathbb{R}\right\} \\ &= \left\{\left(\frac{1}{\sqrt{2}}\rho\sin\phi, \frac{1}{\sqrt{2}}\rho\sin\phi, \rho\cos\phi\right) : \rho, \phi \in \mathbb{R}\right\}. \end{aligned}$$

Indeed it is not obvious that this matches 1.5.9. The proof of this is left as an exercise.

We can as well restrict the domain to make the spherical coordinate transformation a bijection.

Lemma 1.5.15 Let

$$A = (0, \infty) \times (-\pi, \pi) \times (0, \pi) \text{ and } B = \mathbb{R}^3 \backslash \{(x, 0, z) : x \leq 0, z \in \mathbb{R}\},$$

then the spherical coordinate transformation $T|_A : A \to B$ defined by

$$T(\rho, \theta, \phi) = (\rho \cos\theta \sin\phi, \rho \sin\theta \sin\phi, \rho \cos\phi)$$

is bijective.

Proof. A proof can be given using similar arguments as polar coordinates.

1.6 Manifolds

In linear algebra, we learned how to calculate the dimension of a set described by some linear equations, which is calculated by finding the size of a basis, this means we can study "flat" object lines, planes, and subspaces. However, many geometric objects are not defined this way, thus inducing a important question

*If a set $S \subseteq \mathbb{R}^n$ is described by a **nonlinear** equation, then what is its "dimension"? How do we even define "dimension"?*

This question will be a central problem in multivariable calculus, and we will need to develop some significant theory before even approaching it. At the moment, we will use low dimensional cases to get some intuition.
The first case is the idea of a "curve", which can be through as bending a straight line. Since a straight line is a 1-dimensional linear object in some higher dimensional space, a curve should probably be thought of as a 1-dimensional nonlinear object living in higher dimensions.
The second case is the idea of a "surface". This will capture the core issues of the question above. intuitively speaking, a surface is presumable created by bending a piece of a plane in $\mathbb{R}^3$, so a surface can probably be thought of as a 2-dimensional nonlinear object living in 3-dimensional space.
The general case defines the idea of a "**manifold"**. Fix $k, n \in \mathbb{N}^+$ with $k < n$, intuitively a k-dimensional manifold in $\mathbb{R}^n$ should presumably be created by bending a piece of a k-dimensional plane in $\mathbb{R}^n$. But to be able to rigorously define manifolds, it will take many chapters of preparation and right now, we will take a step back and look at a more foundational question:

How can a set $S \subseteq \mathbb{R}^n$ be described by nonlinear equations? is there more than one way?

In this section, we will explore the three fundamental forms for describing sets: parametric form, explicit form, and implicit form.

1.6.1 Parametric form

One natural way to describe sets with nonlinear equations is using maps of the form

$$\mathbb{R}^n \to \mathbb{R}^m$$

where $n < m$. We have already seen the case $n = 1$ but these are very special, so we will focus on when $n \geq 2$. The visual focus here will be on the case $n = 2$ and $m = 3$. These maps will presumably create surfaces, i.e. 2-dimensional manifolds in $\mathbb{R}^3$.

Example 1.6.1 The **unit sphere** is the sphere of radius 1 centred at the origin. Let S be the unit sphere in $\mathbb{R}^3$, we can define $g : \mathbb{R}^2 \to \mathbb{R}^3$ by

$$g(\theta, \phi) = (\cos\theta \sin\phi, \sin\theta \sin\phi, \cos\phi).$$

This is the spherical coordinate transformation with the spherical radius fixed, so

$$S = \{g(\theta, \phi) : (\theta, \phi) \in \mathbb{R}^2\} = img(g).$$

Hence it seems reasonable to guess that S will be a "2-dimensional manifold in $\mathbb{R}^3$ as it can be described as the image of a map with domain $\mathbb{R}^2$.
Unfortunately, this calculation is not good enough for a definition of a surface. For example, we could be silly and define $h : \mathbb{R}^2 \to \mathbb{R}^3$ by

$$h(\theta, \phi) = (\cos\theta, \sin\theta, 0),$$

the image of h is a unit circle lying in the $z = 0$ plane. We see that despite being described by a function from $\mathbb{R}^2 \to \mathbb{R}^3$, we have a 1-dimensional manifold! How can we find a formal way of distinguishing these two scenarios? That is a tough question for much later.

Despite the challenges presented by this example, we can extend this idea to any dimension.

Definition 1.6.2 Let $m, n \in \mathbb{N}^+$ with $n < m$. A set $S \subseteq \mathbb{R}^m$ can be written in **parametric form** (with n-variables) if there exists a set $A \subseteq \mathbb{R}^n$ and a continuous map $g : A \to \mathbb{R}^m$ such that

$$S = \{g(x) : x \in A\} = img(g).$$

Equivalently, we say the set S is **parametrized by** g.

If a set in $\mathbb{R}^m$ is parametrized by a map with n inputs, then we might guess that it should be an n-dimensional manifold in $\mathbb{R}^m$. However, as the above example illustrated, the set could be anything! Thus, the definition above is only a starting point and does not fully match our intuitive understanding of an n-dimensional manifold. The definition simply gives one way to describe sets with nonlinear equations.

1.6.2 Explicit form

Another way to describe sets is a special case of parametric form. In particular, we have already encountered a large class of sets which are easy to parametrize: graphs.

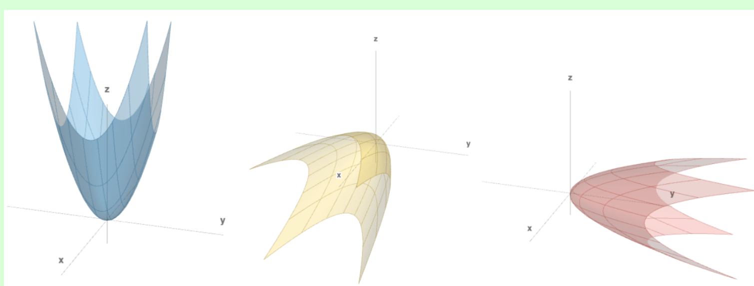
Example 1.6.3 The graph of the real-valued function $f(x, y) = x^2 + y^2$ is the set

$$S = \{(x, y, z) \in \mathbb{R}^3 : z = f(x, y)\} = \{(x, y, z) \in \mathbb{R}^3 : z = x^2 + y^2\}$$

in $\mathbb{R}^3$. To parametrize this set, simply define $g : \mathbb{R}^2 \to \mathbb{R}^3$ by $g(x, y) = (x, y, x^2 + y^2)$ so $\operatorname{img}(g) = S$. Now, by convention, the graph of f is always defined using the z-coordinate but as long as we can express one coordinate as a function of the others, the set will be a graph. For example, sets

$$S_1 = \{(x, y, z) \in \mathbb{R}^3 : x = y^2 + z^2\} \quad S_2 = \{(x, y, z) \in \mathbb{R}^3 : y = x^2 + z^2\}$$

are also graphs of f which can be parametrized as well. From left to right, the plots of S, S_1, S_2 are below.



These sets are all called **paraboloids**.

The above example illustrates that if $f : \mathbb{R}^2 \to \mathbb{R}$ is a real-valued function, then any of the following sets

$$\{x, y, z) \in \mathbb{R}^3 : x = f(y, z)\}, \{x, y, z) \in \mathbb{R}^3 : y = f(x, z)\}, \{x, y, z) \in \mathbb{R}^3 : z = f(x, y)\}$$

are graphs of f, with the third set being the one commonly referred to.

Example 1.6.4 Real-valued functions are not the only maps that create graphs. We can use vector-valued functions, too. For example, the graph of the map $\gamma : \mathbb{R} \to \mathbb{R}^2$

defined by $\gamma(t) = (\cos t, \sin t)$ is the set

$$S = \{(t, \cos t, \sin t) : t \in \mathbb{R}\}$$

lying in $\mathbb{R} \times \mathbb{R}^2 = \mathbb{R}^3$. This produces a helix along the x-axis. Notice the y and z variables are functions of the x variable.

Graphs can therefore be formalized to any dimension.

Definition 1.6.5 Let $m, n \in \mathbb{N}^+$ with $n < m$. Let $A \subseteq \mathbb{R}^n$. The **graph** of a function $f : A \to \mathbb{R}^{m-n}$ is the set

$$S = \{(x, f(x)) : x \in A\} \subseteq \mathbb{R}^n \times \mathbb{R}^{m-n} = \mathbb{R}^m.$$

A set $S' \subseteq \mathbb{R}^m$ is a graph of f if S' is the same as S up to reordering the variables.

Remark 1.6.6 "Reordering variables" can be formally expressed as $S' = \pi(S)$ for a linear transformation $\pi : \mathbb{R}^m \to \mathbb{R}^m$ given by $\pi(x) = Px$ where P is a $m \times m$ permutation matrix.

As the examples demonstrate, graphs can always be written in parametric form. This gives a new definition:

Definition 1.6.7 Let $m, n \in \mathbb{N}^+$ with $n < m$. A set $S \subseteq \mathbb{R}^m$ can be written in **explicit form (in n variables)** if S if S is a graph of a continuous function $f : A \to \mathbb{R}^{m-n}$ where $A \subseteq R^n$.

Example 1.6.8 In the last subsection, we've shown that the unit sphere $S \in \mathbb{R}^3$ can be written in parametric form. However it cannot be represented in explicit, here is a sketch of the proof:

Proof. We prove by contradiction,so suppose that S can be written in explicit form. Then there are 6 cases to consider by 1.6.6. Lets just consider the case

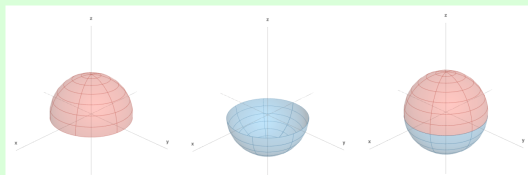
$$S = \{(x, y, z) \in \mathbb{R}^3 : z = f(x, y)\}$$

for some continuous real-valued two variable function f. Then notice that $(0, 0, \pm 1) \in S$ which implies f is not a function, a contradiction. We can then extend this to the other five cases $x = f(y, z), y = f(x, z)$,etc.. using an analogous argument, as $(0, \pm 1, 0) \in S$ and $(\pm 1, 0, 0) \in S$.

Although S cannot be written in explicit form, it can be written as a finite union of sets in explicit form! In particular,

$$S = \left\{(x, y, z) \in \mathbb{R}^3 : z = \sqrt{1 - x^2 - y^2}\right\} \cup \left\{(x, y, z) \in \mathbb{R}^3 : z = -\sqrt{1 - x^2 - y^2}\right\}$$

so S is the union of the graphs of $f(x, y) = \sqrt{1 - x^2 - y^2}$ and of $g(x, y) = -\sqrt{1 - x^2 - y^2}$, which each represents the upper and lower hemisphere.



Informally, we can see that each hemisphere passes the vertical line tests while the whole sphere does not.

Thus, sets in explicit form can always be written in parametric form, but the converse is not necessarily true, as sets in explicit form is a very special form of parametric form.

1.6.3 Implicit form

Sets can also be naturally described by nonlinear equations using maps of the form

$$\mathbb{R}^n \to \mathbb{R}^m$$

when $n > m$. For instance, a real-valued function $\mathbb{R}^3 \to \mathbb{R}$ can create a surface via its level sets.

Example 1.6.9 The unit sphere S in $\mathbb{R}^3$ is defined to be the set of points that are distance 1 away from the origin, i.t.

$$S = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 = 1\},$$

which is the 1-level set of the function $f(x, y, z) = x^2 + y^2 + z^2$. The implicit equation

$$x^2 + y^2 + z^2 = 1$$

does not *explicitly* express one variable in terms of the others.

Other sets can also be created by looking at maps $\mathbb{R}^n \to \mathbb{R}^m$ where $n > m \geq 2$.

Example 1.6.10 Define $f : \mathbb{R}^3 \to \mathbb{R}^2$ by $g(x, y, z) = (z - x^2 - y^2, x^2 + y^2 + z^2)$. Consider the set C in $\mathbb{R}^3$ defined by

$$C = \{(x, y, z) \in \mathbb{R}^3 : g(x, y, z) = (0, 9)\}.$$

In other words, C is the set of points satisfying the two non-linear equations

$$z - x^2 - y^2 = 0, \quad x^2 + y^2 + z^2 = 9.$$

Lucikly, we can solve to find that C is the circle $x62 + y^2 = a$ lying in the plane $z = a$ where $a = \frac{1}{2}(\sqrt{37} - 1)$. In other words, C is a curve implicitly defined by 2 non-linear equations. View this demo to see how C is the intersection of two surfaces.

These examples suggest a third way to describe sets by nonlinear equations.

Definition 1.6.11 Let $m, n \in \mathbb{N}^+$ with $n > m$. A set $S \subseteq \mathbb{R}^n$ can be written in **implicit form (with m equations)** if there exists a constant $c \in \mathbb{R}^m$, a set $A \subseteq \mathbb{R}^n$, and a continuous function $f : A \to \mathbb{R}^m$ such that

$$S = f^{-1}(\{c\}) = \{x \in \mathbb{R}^n : f(x) = c\}.$$

Remark 1.6.12 the notation f^{-1} is not the inverse function of f in this case, but the preimage of the set $\{c\}$

Example 1.6.13 The unit sphere S in $\mathbb{R}^3$ can be written in implicit form, because it is the 1-level set of the continuous function $f(x, y, z) = x^2 + y^2 + z^2$, that is $S = f^{-1}(\{1\})$. Similarly, the set C in Example 1.6.10 is $C = g^{-1}(\{(0, 9)\})$.

Example 1.6.14 The paraboloid $\{(x, y, z) \in \mathbb{R}^3 : z = x^2 + y^2\}$ is written in implicit form, isnce it is the 0 level set of $f(x, y, z) = z - x^2 - y^2$. The same rick can be used for any graph, so we see that any sets in explicit form can always be written in implicit form.

If a set in $\mathbb{R}^n$ is written in implicit form with m nonlinear equations, then what would you guess to be its "dimension"?We can use linear algebra to formulate an educated guess: A system with n variables and m linear equations can be represented by an equation of the form $Ax = 0$, where A is an $m \times n$ matrix and $x \in \mathbb{R}^n$ is unknown. The set of solutions to this linear system is the null space of A, if m, n and the m linear equations are linearly independent, then the null space is

$(n - m)$-dimensional. While sets in implicit form are defined using nonlinear equations, a similar principle appears to hold true in some sense.

Example 1.6.15 The unit sphere S in $\mathbb{R}^3$ is written in implicit form with 1 nonlinear equation $x^2 + y^2 + z^2 = 1$ in 3 variables x, y, z. This means the sphere should presumably be $3 - 1 = 2$ dimensional, which it is! Similarly, this also holds for Example 1.6.10 which is 1 dimensional. This acts as some evidence that the principles of linear algebra may carry over to nonlinear systems.

This investigation is the beginning of something much greater, namely the implicit function theorem. We will explore that in depth later but for now, the key takeaway is that sets can have three different descriptions: parametric, explicit, and implicit. Each with its own advantages and disadvantages.

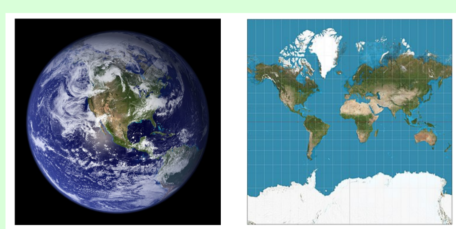
1.7 Projections

Lastly, we will talk about maps of the form

$$R^n \longrightarrow \mathbb{R}^m$$

where $n > m$. These maps are sometimes referred to as **projections** as they push higher dimensional objects into a lower dimensional space. As usual, we will focus on the special case $\mathbb{R}^3 \longrightarrow \mathbb{R}^2$ to provide some visual intuitions.

Example 1.7.1 Creating a map of the Earth is a classical projection problem. You must push a sphere, an object in $\mathbb{R}^3$, into a rectangle, an object in $\mathbb{R}^2$. The Mercator projection, invented in 1569 , is the modern standard, which is still used in Google Maps to this day.



This non-linear projection preserves local directions so its popularity rose due to its use for ocean navigation. However, by reducing dimensions, it loses some geometric information. The Mercator projection distorts distances dramatically. Greenland and the African continent appear to be the same size on the map, but Africa is actually 14 times larger!

There are some simple and common examples of projections.

Example 1.7.2 For $i \in \{1, ..., n\}$, the map $\pi_i : \mathbb{R}^n \to \mathbb{R}$ given by

$$\pi_i (x_1, \ldots, x_n) = x_i$$

is called the i^{th} coordinate map. These are convenient for proofs when you are trying to express some quantities in terms of continuous maps. Notice π_i is a linear transformation.

Remark 1.7.3 The Greek letter π is often used for projections. Since π is used for other reasons, this is an abuse of notation. You are permitted to abuse notation provided the context makes your notation unambiguous.

Example 1.7.4 Example 1.6.5 For $i \in \{1, \ldots, n\}$, the map $\Pi_i : \mathbb{R}^n \to \mathbb{R}^{n-1}$ given by

$$\Pi_i (x_1, \ldots, x_n) = (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$$

is the i^{th} **coordinate plane** projection. These are also linear transformations which are convenient for proofs. They also have a natural interpretation as the "shadow" of an object.
In particular, for $n = 3$, the map $\Pi_3(x, y, z) = (x, y)$ is also referred to as the projection into the xy-plane. The image of the paraboloid $P = \left\{(x, y, z) : z = x^2 + y^2 \leq 1\right\}$ under Π_3 is the unit disk $\left\{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\right\}$ in $\mathbb{R}^2$. Informally speaking, Π_3 produces the "shadow" of an object in $\mathbb{R}^3$ when viewed from above.

Chapter 2

Topology

Contents

2.1 Balls, spheres, rectangles, and cubes

2.1.1 Balls and spheres

In this course, we will use the Euclidean inner product on $\mathbb{R}^n$, and the notion of distance is defined by the norm of a point $x = (x_1, ..., x_n) \in \mathbb{R}^n$ as

$$||x|| = \sqrt{x_1^2 + \cdots + x_n^2}.$$

This function defines the most fundamental sets in $\mathbb{R}^n$: balls and spheres.

Definition 2.1.1

- The **open ball of radius** r **centred at** a is the set $B_r(a) = \{x \in \mathbb{R}^n : ||x - a|| < r\}$.
- The **closed ball of radius** r **centred at** a is the set $\{x \in \mathbb{R}^n : ||x - a|| \leq r\}$.
- The **sphere of radius** r **centred at** a is the set $\{x \in \mathbb{R}^n : ||x - a|| = r\}$.

Remark 2.1.2 Other notations for the open ball include $B(a, r)$, $B(a; r)$, or $B_a(r)$, but none of them will appear in this text.

Remark 2.1.3 An open or closed ball is **puncture** if it excludes the centre. For example, $B_r(a) \backslash \{a\}$ is a punctured open ball.

The word "ball" and "sphere" are inspired by the three-dimensional case, where balls are always *solid* and spheres are always *hollow*. In the 1-dimensional case, "balls' are reduced to intervals, an open ball of radius r at a would be $(a - r, a + r)$, while a closed ball would be $[a - r, a + r]$. In two-dimensional, "balls" are **disks** and "spheres" are **circles**.

Remark 2.1.4 Notice that an open ball of radius 0 is empty, a closed ball of radius 0 centred at $a \in \mathbb{R}^n$ is the singleton $\{a\}$. It is silly to refer to balls of radius 0 but sometimes we may want to include this degenerate case in a formal proof or statement.

Definition 2.1.5 The $(n-1)$-dimensional **unit sphere** in $\mathbb{R}^n$ is the sphere of radius 1 centred at the origin and is denoted S^{n-1}. In other words, $S^{n-1} = \{x \in \mathbb{R}^n : ||x|| = 1\}$.

Example 2.1.6 S^1 is the unit circle $\{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$ in $\mathbb{R}^2$. S^2 is the unique sphere $\{(x, y, z) \in \mathbb{R}^2 : x^2 + y^2 + z^2 = 1\}$ in $\mathbb{R}^3$.

2.2 Rectangles and cubes

We can also generalize intervals in $\mathbb{R}^n$ using Cartesian products.

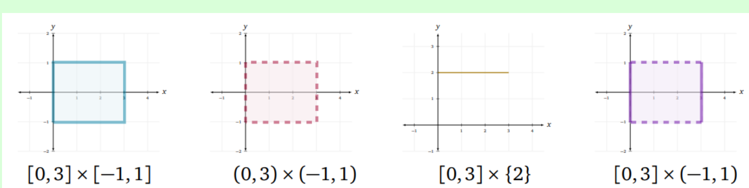
Definition 2.2.1 A **closed rectangle** in $\mathbb{R}^n$ is a set R of the form

$$R = [a_1, b_1] \times \cdots \times [a_n, b_n] = \{(x_1, ..., x_n) : x_i \in [a_i, b_i], 1 \leq i \leq n, \}$$

where $a_1, b_1, ..., a_n, b_n \in \mathbb{R}$ and $a_i < b_i$ for all $1 \leq i \leq n$.

Remark 2.2.2 The set $(a_1, b_1) \times \cdots \times (a_n, b_n)$ is an open rectangle. In general, a "rectangle" refers to a closed rectangle unless specified otherwise.

Example 2.2.3 1-dimensional rectangles are closed intervals, like $[a, b]$ where $a < b$, while open rectangle are (a, b). The singleton $\{a\}$ is not a rectangle. In two dimension, rectangle are what they are in the usual colloquial sense. The set $[0, 3] \times [-1, 1]$ is a rectangle in $\mathbb{R}^2$. The set $(0, 3) \times (-1, 1)$ is an open rectangle. The set $[0, 3] \times \{2\}$ is not a rectangle. The set $[0, 3] \times (-1, 1)$ is neither an open rectangle nor a closed rectangle.



Definition 2.2.4 An n-dimensional **hypercube** is a set in $\mathbb{R}^n$ of the form

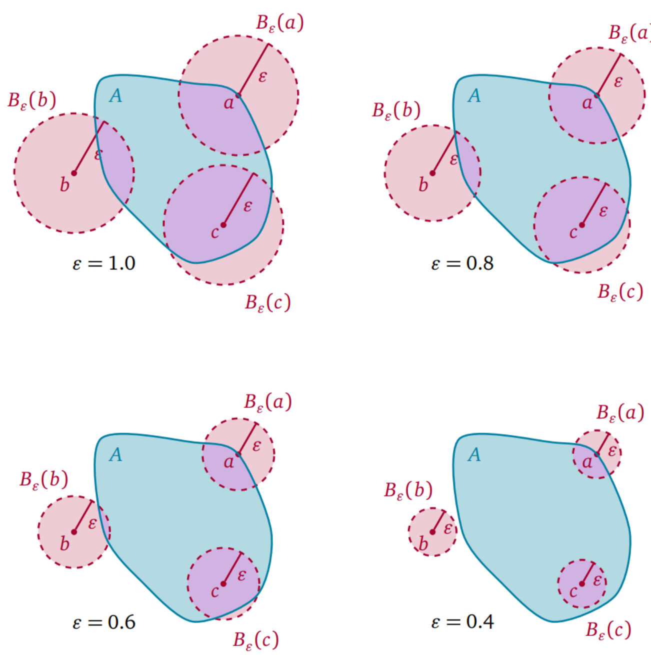
$$[a,b]^n = [a,b] \times \cdots \times [a,b]$$

The unit hypercube is the set $[0,1]^n$.

Note that a 2-dim hypercube is just the **square** and $[0,1]^2$ is the **unit square**. A 3-dim hypercube is a **cube** and $[0,1]^3$ is the **unit cube**. They are all "solid"

2.2.1 Interior, boundary, and closure

Given a solid object in 3-dimensions, we hvae an intuitive physical understanding of its inside, outside, and edge. But how to we extend this idea to $\mathbb{R}^n$? A simple yet brilliant solution is to "zoom in" with respect to a point.



The four pictures above illustrates a set $A \subseteq \mathbb{R}^2$ and three different points $a, b, c \in \mathbb{R}^2$, with the radii of each ball getting smaller and smaller each time.

2.2.2 Interior

With the intuition above, we can define an interior point as

Definition 2.2.5 Let $A \subseteq \mathbb{R}^n$ be a set. A point $p \in \mathbb{R}^n$ is an **interior point** of A if there exists $\epsilon > 0$ such that $B_\epsilon(p) \subseteq A$.

Example 2.2.6 The point $p = 2$ is an interior point of the interval $A = [1, 4)$ since $B_{1/2}(2) = (1.5, 2.5)$ is a subset of A. To show that some point is not an interior point of A, we need to prove the negation:

$$\forall \epsilon > 0, B_\epsilon(5) = (5 - \epsilon, 5 + \epsilon) \nsubseteq [1, 4) = A$$

This is easy as for any $\epsilon > 0$, $x \in (5 - \epsilon, 5 + \epsilon)$ yet $x \notin [1, 4)$, hence $B_\epsilon(5) \nsubseteq [1, 4)$ so 5 is not an interior point of A. Now to check that an endpoint is not an interior point, we can follow the same argument but our choice of x will be dependent on ϵ.

Proving an open ball is a subset of another set is a bit more trickier in higher dimensions. Suppose we are given the set

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

How do we prove that the point $p = (1, 0)$ is an interior point of A? Using the definition, it suffices to show that the open ball $B_{1/2}((1, 0))$ is a subset of A. Formally, this means all points in $B_{1/2}((1, 0))$ are in A. So let $(x, y) \in B_{1/2}((1, 0))$, then $(x - 1)^2 + y^2 < (\frac{1}{2})^2$, since $y^2 \geq 0$, this implies

$$(x - 1)^2 < (\frac{1}{2})^2 \implies x - 1 < \frac{1}{2} \implies x \leq 2.$$

Thus $(x, y) \in A$ as required, which means $(1, 0)$ is an interior point of A.

Definition 2.2.7 Let $A \subseteq \mathbb{R}^n$ be a set. The **interior** of A, denoted A^o or $\text{int}(A)$, is the set of interior points of A.

Example 2.2.8 The interval of $A = [1, 4)$ is the open interval $A^o = (1, 4)$.For $(1, 4) \subseteq A^o$, we need to prove

$$\forall x \in (1, 4), \exists \epsilon > 0, s.t.\ (x - \epsilon, x + \epsilon) \subseteq [1, 4).$$

Let $x \in (1, 4)$, set $\epsilon = \min(\frac{x-1}{2}, \frac{4-x}{2}) > 0$ and we can verify that this ϵ indeed satisfy the equation above, implying that $A^o = (1, 4)$. Conversely, to prove that $A^o \subseteq (1, 4)$, we must show that is $p \notin (1, 4) \implies p \notin A^o$. Take any $p \in A^c$, so $p < 1$ or $p \geq 4$. For any $\epsilon > 0$, the ball $B_\epsilon(p) = (p - \epsilon, p + \epsilon)$ is not a subset of A as p itself is not in A, this shows $p \notin A^o$. Finally to show that $p = 1$ is not an interior point of A, for any ϵ simply take $x = 1 - \frac{\epsilon}{2}$ which is in $B_\epsilon(1)$ but not in A. This completes the proof.

There are some important examples which we should remember, let $a \in \mathbb{R}^n$ and $r > 0$.

- The interior of $\mathbb{R}^n$ is $\mathbb{R}^n$.
- The interior of any finite set is empty. (Any open ball contains an infinite number of points)

- The interior of the open ball $B_r(a)$ is the open ball itself.
- The interior of the closed ball $\{x \in \mathbb{R}^n : ||x - a|| \leq r\}$ is the open ball $B_r(a)$
- The interior of any sphere is empty.
- The interior of the hypercube $[c, d]^n$ is the open hypercube $(c, d)^n$, where $c < d$.

Not all sets have nice geometric interpretations. The set of rational numbers $\mathbb{Q}$ is an example. Its interior is empty because any interval of rational numbers must contain irrational numbers, so $B_\epsilon(q)$ cannot be a subset of $\mathbb{Q}$ for any $q \in \mathbb{Q}$. The interior satisfies natural properties with respect to other set operations.

Lemma 2.2.9 Let A and B be sets of $\mathbb{R}^n$, then

1. $A^o \subseteq A$
2. $A^o \cup B^o \subseteq (A \cup B)^o$
3. $A^o \cap B^o = (A \cap B)^o$
4. $A^o \times B^o = (A \times B)^o$

Proof.
1. Trivial from definition.
2. Suppose $x \in A^o \cup B^o$, without loss of generality suppose $x \in A^o$, then there exist an open ball $B_\epsilon(x) \subseteq A$ for some $\epsilon > 0$, clearly $B_\epsilon(x) \subseteq A \cup B$ hence $x \in (A \cup B)^o$.
3. First suppose $x \in A^o \cap B^o$, then there exist open ball $B_\epsilon(x) \subseteq A$ and $B_\phi(x) \subseteq B$ for some $\epsilon > 0$ and $\phi > 0$. Take $r = \min(\epsilon, \phi)$, then $B_r(x) \subseteq A \cap B$, so x is an interior point of $A \cap B$, as required. Conversely, if $x \in (A \cap B)^o$, by definition there is an open ball $B_r(x) \subseteq A \cap B$, hence this open ball is an contained entirely in A and B, so it is an interior point of A and B, completing the proof.

4. First suppose $(a,b) \in A^o \times B^o$, then by definition we have two open balls of radius ϵ, ϕ such that $B_\epsilon(a) \subseteq A$ and $B_\phi(b) \subseteq B$. Let $r = \min(\epsilon, \phi)$ and consider the open ball $B_r((a,b))$. For any $(x,y) \in B_r((a,b))$, we have

$$||(x,y)-(a,b)|| = \sqrt{(x-a)^2 + (y-b)^2} < r \implies x-a < r \text{ and } y-b < r,$$

so $(x,y) \in B_\epsilon(a) \times B_\phi(b) \subseteq A \times B$. Since x, y was arbitrary we have that $B_r((a,b))$ is entirely contained in $A \times B$, so $A^o \times B^o \subseteq (A \times B)^o$.
For the other way, suppose $(a,b) \in (A \times B)^o$, then there exist open ball $B_r((a,b)) \subseteq A \times B$ for positive r. Consider the open balls $B_\epsilon(a)$ and $B_\phi(b)$, where $\epsilon = \phi = \frac{r}{\sqrt{2}}$, we want to show that they are subsets of A and B, respectively. Take any $(x,y) \in B_\epsilon(a) \times B_\phi(b)$, $||(x,y)-(a,b)|| = \sqrt{(x-a)^2 + (y-b)^2}$, since $x - a < \frac{r}{\sqrt{2}}$ and $y - b < \frac{r}{\sqrt{2}}$, this implies

$$||(x,y)-(a,b)|| < \sqrt{\frac{r^2}{2} + \frac{r^2}{2}} = r,$$

so $(x,y) \in B_r((a,b))$, we can then conclude $B_\epsilon(a) \times B_\phi(b) \subseteq B_r((a,b))$, hence $(A \times B)^o \subseteq A^o \times B^o$, completing the proof.

2.2.3 Boundary

A point p should be on the "edge" of a region A if no matter how close we zoom into p, we can see points inside A and outside A. This leads to the definition of a **boundary point**.

Definition 2.2.10 let $A \subseteq \mathbb{R}^n$ be a set. A point $p \in \mathbb{R}^n$ is a **boundary point** of A if for every $\epsilon > 0$, the sets $B_\epsilon(p) \cap A$ and $B_\epsilon(p) \cap A^c$ are both non-empty.

The collection of boundary points produces the concept of an "edge" of some set.

Definition 2.2.11 Let $A \subseteq \mathbb{R}^n$ be a set. The **boundary** of A, denoted ∂A, is the set of boundary points of A.

Here is a list of some examples of boundaries of sets. Let $a \in \mathbb{R}^n$ and $r > 0$.

- The boundary of $\mathbb{R}^n$ is empty.
- The boundary of any finite set A is A itself.
- The boundary of the closed interval $[c,d]$ is the finite set$\{c,d\}$.
- The boundary of the open ball $B_r(a)$ is the sphere $\partial B_r(a) = \{x \in \mathbb{R}^n : ||x-a|| = r\}$.
- The boundary of $B_1(0)$ is the unit sphere $\partial B_1(0) = S^{n-1}$.
- The boundary of any sphere is the sphere itself.

Lemma 2.2.12 For any set $A \subseteq \mathbb{R}^n$,

$$A^o \cap \partial A = \emptyset$$

Proof. Suppose not, then there exist $x \in A$ such that $x \in A^o$ and $x \in \partial A$, by definition, $x \in A^o$ implies that there exist a open ball $B_r(x) \subseteq A$, and $x \in \partial A$ implies that for all $r' > 0$, $B_{r'}(x) \cap A^c \neq \emptyset$. Clearly these two contradicts each other, hence the two sets are disjoint.

2.2.4 Closure

Defining limits on $\mathbb{R}^n$ produces a new issue. Over $\mathbb{R}$, when we try to find the limit of a point x, we were satisfied by simply look at some open interval $(x-\epsilon, x+\epsilon)$ that belonged to that function. However, this simplicity exists because there are only two ways of approaching a point on the real number line: from the left or from the right.

This changes completely when we work with $\mathbb{R}^n$ for $n \geq 2$,as we can approach a point in infinitely many way. But first, how do we even know where we can take limits of some function with domain A? As a result, we must determine which points in $\mathbb{R}^n$ can be approached by points only from A. Intuitively, this can be thought of as no matter how close we zoom into a point x, we can see points in A which are not x. Formally:

Definition 2.2.13 Let $A \subseteq \mathbb{R}^n$ be a set. A point $p \in \mathbb{R}^n$ is a **limit point** of A if for every $\epsilon > 0$, the set $B_\epsilon(p) \backslash \{p\} \cap A \neq \emptyset$.

Here is a list of examples of sets of limit points. Let $a \in \mathbb{R}^n$ and $r > 0$.

- The set of limit points of R^n is R^n.
- The set of limit points of any finite set is empty.
- The set of limit points of the open ball $B_r(a)$ is the closed ball $\{x \in \mathbb{R}^n : ||x-a|| \leq r\}$.
- The set of limit points of any sphere is the sphere itself.
- The set of limit points of any closed rectangle is the rectangle itself.

It is a good exercise to also prove that every interior point of A is also a limit point of A.

It should not be a surprise that the set of limits points of A might not contain A itself, but sometimes we may be interested in a new set which contains all points of A along with its limit points, as it has some nice properties, this leads to a new definition:

Definition 2.2.14 Let $A \subseteq \mathbb{R}^n$. The **closure** of A, denoted $\overline{A}$ or cl(A), is the union of the set A and the set of limit points of A.

Here is a list of examples of closures. Let $a \in \mathbb{R}^n$ and $r > 0$:

- The closure of $\mathbb{R}^n$ is $\overline{\mathbb{R}^n} = \mathbb{R}^n$.
- The closure of any finite set is the set itself.
- The closure of the open ball $B_r(a)$ is the closed ball $\overline{B_r(a)} = \{x \in \mathbb{R}^n : ||x - a|| \leq r\}$.
- The closure of any sphere is the sphere itself.
- The closure of any closed rectangle is the rectangle itself.
- The closure of $\mathbb{Q}$ is $\overline{\mathbb{Q}} = \mathbb{R}$.

Again, closure respects natural set operations.

Lemma 2.2.15 Let A and B be sets in $\mathbb{R}^n$. Then

- $A \subseteq \overline{A}$.
- $\overline{A \cup B} = \overline{A} \cup \overline{B}$.
- $\overline{A \cap B} \subseteq \overline{A} \cap \overline{B}$.
- $\overline{A \times B} = \overline{A} \times \overline{B}$.

Proof. The proof will be left as an exercise. These facts may be easier to prove once a alternate definition of limit point is given in the next section.

Now that we have gone through the three important definitions of interior, boundary, and closure of set. It should not be hard to see how they are related.

Lemma 2.2.16 Let $A \subseteq \mathbb{R}^n$ be a set, then

- $\overline{A} = A^o \cup \partial A$.
- $\partial A = \overline{A} \backslash A^o$.

Proof. To prove the first statement, we first show $A^o \cup \partial A \subseteq A^o$. Since $A^o \subseteq A \subseteq \overline{A}$, it suffices to show that $\partial A \backslash A \subseteq \overline{A}$. Take $p \in \partial A$ with $p \notin A$. By definition, for all $\epsilon > 0$, the set $B_\epsilon(p) \backslash \{p\}$ contains points in A, hence $p \in \overline{A}$. This proves $A^o \cup \partial A \subseteq \overline{A}$.
For the other direction, we prove by showing $A \subseteq A^o \cup \partial A$ and $\overline{A} \backslash A \subseteq \partial A$. Fix $p \in A$, if there exist $\epsilon > 0$ such that $B_\epsilon(p) \subseteq A$ then $p \in A^o$. Otherwise, for every $\epsilon > 0$, the ball $B_\epsilon(p)$ is not a subset of A, however $B_\epsilon(p) \cap A$ is not empty as it contains $p \in A$. Therefore, for every $\epsilon > 0$, both $B_\epsilon(p) \cap A$ and $B_\epsilon(p) \cap A^c$ so $p \in \partial A$. Now fix $q \in \overline{A}$ with $q \notin A$. By definition of a limit point, for every $\epsilon > 0$, the punctured open ball $B_\epsilon(q) \backslash \{q\}$ contains points in A, moreover $B_\epsilon(q) \cap A^c$ is non-empty for every ϵ since $q \in A^c$. Therefore, $q \in \partial A$ which proves that $\overline{A} \backslash A \subseteq \partial A$, finishing the proof. The second statement follows quickly from the first.

2.3 Sequences

We have seen a description of limit points in terms of open balls, there are also equivalent definitions using sequences which some might find more intuitive.

Definition 2.3.1 A **sequence** in $\mathbb{R}^n$ is a function with domain $\{k \in \mathbb{Z} : k \geq k_0\}$ for some fixed $k_0 \in \mathbb{Z}$ and codomain $\mathbb{R}^n$.

As with sequences in $\mathbb{R}$, we can specify what it means to "pick terms from a sequence".

Definition 2.3.2 Let $x : \mathbb{N}^+ \to \mathbb{R}^n$ be a sequence and let $m : \mathbb{N}^+ \to \mathbb{N}^+$ be s strictly increasing function. The sequence $\{x(m(k))\}_{k=1}^{\infty}$ is a **subsequence** of the sequence $\{x(k)\}_{k=1}^{\infty}$.

Remark 2.3.3 With this definition, the domains of x and m can be modified to allow for subsets of $\mathbb{Z}$ which are not $\mathbb{N}^+$, however the codomain of m must be a subset of the domain of x.

2.3.1 Convergence of sequences

We can view sequences in a way where each term $x(k)$ represents a finite approximation to some desired value, and the next term $x(k+1)$ is usually a refined version of the previous term, so intuitively, the limit of the sequence exists if and only we can refine these approximations to any arbitrary accuracy.

Definition 2.3.4 Let $\{x(k)\}_k$ be a sequence in $\mathbb{R}^n$. Then $\{x(k)\}$ **converges** if there exist $p \in \mathbb{R}^n$ for which

$$\forall \epsilon > 0, \exists K \in \mathbb{N}\ s.t.\ \forall k \in \mathbb{N}, k \geq K \implies ||x(k) - p|| < \epsilon.$$

If the above holds, we write $lim_{k\to\infty} x(k) = p$ or equivalently $x(k) \to p$. Otherwise, the sequence **diverges**.

Example 2.3.5 The sequence $\{x(k)\}_{k=1}^{\infty}$ given by $x(k) = \left(2 + \frac{1}{k}, \frac{\sin k}{k}\right)$ converges to $(2, 0)$.
Proof. Let $\varepsilon > 0$ be arbitrary. Take K to be any natural number greater than $\sqrt{2}/\varepsilon$. For any $k \in \mathbb{N}$, if we assume $k \geq K$, then:

$$\|x(k) - p\| = \left\| \left(2 + \frac{1}{k}, \frac{\sin k}{k}\right) - (2, 0) \right\| = \sqrt{\left(2 + \frac{1}{k} - 2\right)^2 + \left(\frac{\sin k}{k}\right)^2}$$
$$= \sqrt{\frac{1 + \sin^2 k}{k^2}} \leq \sqrt{\frac{2}{k^2}} < \epsilon$$

so $\|x(k) - p\| < \varepsilon$. By the definition, $\lim_{k\to\infty} x(k) = p$ as desired.

Convergence of sequences has a geometric interpretation with shrinking balls in $\mathbb{R}^n$.

A sequence $\{x(k)\}_k$ in $\mathbb{R}^n$ converges to x if every open ball centred at x contains infinitely many points of the sequence $\{x(k)\}_k$.

We have had many scenarios where we split up vectors in $\mathbb{R}^n$ and instead look at each component, this is no different for sequences. A sequence in $\mathbb{R}^n$ can be thought of as n sequences in $\mathbb{R}$, where we write

$$x(k) = (x_1(k), x_2(k), ..., x_n(k)) \in \mathbb{R}^n$$

so the n sequences in $\mathbb{R}$ are given by $\{x_1(k)_k, ..., \{x_n(k)\}_k\}$. This allows you to connect the notions of convergence between $\mathbb{R}$ and $\mathbb{R}^n$.

Lemma 2.3.6 Let $\{x(k)\}_k$ be a sequence in $\mathbb{R}^n$ with $x(k) = (x_1(k), ..., x_n(k))$. Fix $p = (p_1, ..., p_n) \in \mathbb{R}^n$. The sequences $\{x(k)\}_k$ converges to p if and only if $\{x_i(k)\}_k$ converges to p_i for all $i = 1, 2, ..., n$.

Proof. $\Rightarrow$ Assume $\{x(k)\}_k$ converges to $a = (a_1, ..., a_n) \in \mathbb{R}^n$. Fix $i \in \{1, ..., n\}$ and fix $\epsilon > 0$. It suffices to show that $x_i(k) \to a_i$ as $k \to \infty$. Since $x(k) \to a$, there exists $K \in \mathbb{N}$ such that

$$k \geq K \to ||x(k) - a|| < \epsilon.$$

Take this K and let $k \in N$ satisfy $k \geq K$. Then

$$|x_i(k) - a_i| = \sqrt{|x_i(k) - a_i|^2} \leq \sqrt{|x_1(k) - a_1|^2 + \cdots + |x_n(k) - a_n|^2} = ||x(k) - a|| < \epsilon$$

Since ϵ and i was arbitrary, this proves the desired implication.
$\Leftarrow$ Assume $\{x_i(k)\}$ converges for all $i = 1, 2, ..., n$. For each $i \in \{1, 2, ..., n\}$, let $a_i \in \mathbb{R}$ be the limit of $\{x_i(k)\}_k$. Let $a = (a_1, ..., a_n) \in \mathbb{R}^n$. It suffices to prove that $x(k) \to a$. Let $\epsilon > 0$. For $i \in \{1, ..., n\}$, since $x_i(k) \to a_i$, there exists $K_i \in \mathbb{N}$ such that

$$k \geq K_i \implies |x_i(k) - a_i| < \frac{\epsilon}{\sqrt{n}}$$

Take $K = \max\{K_1, ..., K_n\}$, then

$$||x(k) - a|| = \sqrt{|x_1(k) - a_i|^2 + \cdots + |x_n(k) - a_n|^2} \leq \sqrt{n \max_{1 \leq i \leq n} |x_i(k) - a_i|^2} < \epsilon,$$

as $k \geq K \geq K_i$ for all $i = 1, ..., n$. This completes the proof.

Using this lemma, limit laws in $\mathbb{R}^n$ can be proved by appealing to the limit laws in $\mathbb{R}$ instead of directly through the formal definition of the limit.

2.3.2 Limit points, boundary points, and interior points

A sequence of points $\{x(k)\}_k \in \mathbb{R}^n$ often arises in optimization or search algorithms. In these cases, the approximations may only naturally lie in a particular set $A \subseteq \mathbb{R}^n$ such as the domain of the optimizing function. Sometimes these approximations will converge to a point inside your set A, i.e. the interior of A. Sometimes the limit of this sequence may 'fall outside' of the set A, i.e. the boundary of A. This perspective leads to an equivalent formulation for limit points.

Lemma 2.3.7 Let $A \subseteq \mathbb{R}^n$ be a set. A point $p \in \mathbb{R}^n$ is a limit point of A if and only if there exists a sequence of points in $A \backslash \{p\}$ which converges to p.

This lemma makes it quite easy to exhibit limit points in explicit examples.

Example 2.3.8 Let $A = \left\{(x, y) \in \mathbb{R}^2 : 1 \leq x^2 + y^2 < 2\right\} \cup \{(0, 0)\}$. You can quickly check that $(1, 1)$ is a limit point of A using the sequential definition. Define the sequence

$$x(k) = \left(1 - \frac{1}{4k}, 1 - \frac{1}{4k}\right)$$

for $k \geq 1$ so $x(k) \to (1, 1)$ as $k \to \infty$ by Lemma 2.3.14. Moreover, $x(k) \in A \backslash \{(1, 1)\}$ since for $k \geq 1$,

$$2 > \left(1 - \frac{1}{4k}\right)^2 + \left(1 - \frac{1}{4k}\right)^2 \geq 2\left(\frac{3}{4}\right)^2 > 1.$$

Thus, $(1, 1)$ is a limit point of A by Lemma 2.3.17.
A similar argument can work for the limit point $(-1, 0)$. On the other hand, $(0, 0)$ and $(3, 0)$ are not limit points of S but how would you prove it? Try the definition with open balls instead.

Proof. $\Leftarrow$ Let $\{x(k)\}_n$ be a sequence of points in $A \backslash \{p\}$ which converge to p. Fix $\epsilon > 0$. By definition of convergence, there exists $K \in \mathbb{N}$ such that

$$\forall k \in \mathbb{N}, k \geq K \implies ||x(k) - p|| < \epsilon$$

Then $x(k) \in B_\epsilon(p)$, and by assumption $x(k) \in A \backslash \{p\}$. Thus $B_\epsilon(p) \backslash \{p\}$ contains points of A, and so p is a limit point of A.
$\Rightarrow$ Assume $p \in \mathbb{R}^n$ is a limit point of A. It suffices to construct a sequence $\{x(k)\}_k$ such that $\forall k \in \mathbb{N}, x(k) \in A \backslash \{p\}$ and $x(k) \to P$. For each $k \in \mathbb{N}^+$, choose a point $x(k) \in B_{1/k}(p) \cap A$ with $x(k) \neq p$, such a point exists by assumption. It remains to show that $x(k) \to p$ as $k \to \infty$. Fix $\epsilon > 0$. Take $K = \lceil \frac{1}{\epsilon} \rceil$. Let $k \in \mathbb{N}$ satisfy $k \geq K$. Then

$$||x(k) - p|| < \frac{1}{k} \geq \frac{1}{K} leq \epsilon.$$

This completes the proof.

Lemma 2.3.9 Let $A \subseteq \mathbb{R}^n$ be a set. Let $p \in \mathbb{R}^n$ be a point.

- The point p is an interior point of A if and only if for every sequence $\{x(k)\}_k$ of points converging to p, there exists $K \in \mathbb{N}^+$ such that $\{x(k)\}_{k=K}^\infty \subseteq A$.
- The point p is a boundary point of A if and only if there exists a sequence of points in A converging to p and there exists a sequence of points in A^c converging to p.

Proof. This is left as an exercise. The ideas are similar to the proof of the last lemma.

In summary, these definitions are just another tool for verifying whether a point lies on the boundary or interior. Sometimes the sequential definitions are easier, and sometimes the open ball definitions are easier.

2.4 Open sets and closed sets

Given a set $A \subseteq \mathbb{R}^n$, we will often only consider sequences converging to points inside A. For example, if the set A is the domain of a map, we may want a sequence of approximations $\{x(k)\}_k$ lying inside A to converge to a points p within the domain A. Otherwise, we cannot necessarily evaluate the map at the point p. We can ensure this feature in two different ways:

1. *If a sequence in $\mathbb{R}^n$ converges to $a \in A$, then the tail of the sequence belongs to A.*
2. *If a sequence in A converges to $a \in \mathbb{R}^n$, then a must belong to A.*

Many sets A will not satisfy either property, so each of these two properties of sets in $\mathbb{R}^n$ warrants their own definitions.

2.4.1 Open sets

First, lets define a set which satisfy the following.

If a sequence in $\mathbb{R}^n$ converges to $a \in A$, then the tail of the sequence belongs to A.

In other words, no matter how we approach $a \in A$ we must eventually lie inside A. That's precisely the sequential definition of an interior point! This suggests a definition.

Definition 2.4.1 A set $A \subseteq \mathbb{R}^n$ is **open** if every point of A is an interior point of A.

If a is an interior point of A, then we can approach a however we want. On the other hand, if a is a boundary point of A then we can only approach a along certain sequences while still lying inside A, that is, the set A restrict on how the sequence can approach a boundary point a. Some examples to remember:

- The empty set $\emptyset$ is vacuously open.
- The set $\mathbb{R}^n$ is open.
- For $a, b \in \mathbb{R}$ with $a < b$, the open interval (a, b) in $\mathbb{R}$ is open.
- For $p \in \mathbb{R}^n$ and $r > 0$, the open ball $\{x \in \mathbb{R}^n : ||x - p|| < r\}$ in $\mathbb{R}^n$ is open.

Example 2.4.2 The set $A = \{(x, y) \in \mathbb{R}^2 : y > 1\}$ is open.

Proof. Let $(a, b) \in A$ be given. Fix $r = \frac{b-1}{2}$ so $(a, b) \in A$ implies $r > 0$. It suffices to show that $B_r((a, b)) \subseteq A$. For $(x, y) \in B_r((a, b))$, it follows that

$$\begin{aligned} |y - b| \leq ||(x, y) - (a, b)|| < r &\implies |y - b| < \frac{b-1}{2} \\ &\implies b - \frac{b-1}{2} < y < b + \frac{b-1}{2} \\ &\implies y > \frac{b+1}{2} > 1 \end{aligned}$$

Therefore, $(x, y) \in A$, which proves that $B_r((a, b))$ is contained in A, as desired.

Lemma 2.4.3 The interior of a set $A \subseteq \mathbb{R}^n$ is open.

Proof. Let $a \in A^o$. By definition, there exists $\epsilon > 0$ such that $B_\epsilon(a) \subseteq A$. Fix $x \in B_\epsilon(a)$, it suffices to show that $x \in A^o$. Since the open ball $B_\epsilon(a)$ is open and $x \in B_\epsilon(a)$, there exists $\delta > 0$ such that $B_\delta(x) \subseteq B_\epsilon(a)$ As $B_\epsilon(a) \subseteq A$, it follows that $B_\delta(x) \subseteq A$ so $x \in A^o$.

This gives rise to equivalent definitions of an open set.

Lemma 2.4.4 Let $A \subseteq \mathbb{R}^n$. The following are equivalent:

1. A is open.
2. $A = A^o$.
3. $A \cap \partial A = \emptyset$.

Proof. Suppose first A is open, then by definition every point of A is an interior point of A, so $A \subseteq A^o$, clearly $A^o \subseteq A$ hence $A = A^o$. Now assume $A = A^o$, then in previous sections we've shown that the interior and boundary of a set is disjoint, hence $A \cap \partial A = \emptyset$, as required. Lastly, assume $A \cap \partial A = \emptyset$, since every point in A is either a boundary point or an interior point, this implies that A does not contain any boundary points, hence A is open.

Example 2.4.5 Consider the interval $I = [137, 237)$ in $\mathbb{R}$. The point 137 is a boundary point of I because the sequence $\left\{137 + \frac{1}{k}\right\}_{k=1}^{\infty}$ in I, and the sequence $\left\{137 - \frac{1}{k}\right\}_{k=1}^{\infty}$ in I^c both converge to 137. Therefore, $137 \in I \cap \partial I$ so $I \cap \partial I$ is non-empty. By Lemma 2.4.5, I is not open.

2.4.2 Closed sets

If we want to define a property of a set $A \subseteq \mathbb{R}^n$ satisfying the following.

If a sequence in A converges to a point $a \in \mathbb{R}^n$, then $a \in A$.

In other words, A must contain all of its limit points. This suggest the following definition:

Definition 2.4.6 A set $A \subseteq \mathbb{R}^n$ is **closed** if every limit point of A belongs to A.

Some common examples of closed sets are:

1. The empty set $\emptyset$ is vacuously closed.
2. The set $\mathbb{R}^n$ is closed.
3. For $a, b \in \mathbb{R}$ with $a < b$, the closed interval $[a, b]$ in $\mathbb{R}$ is closed.
4. For $p \in \mathbb{R}^n$ and $r > 0$, the closed ball $\{x \in \mathbb{R}^n : ||x - p|| \leq r\}$ is closed.

Example 2.4.7 The set $A = \{(x, y) \in \mathbb{R}^2 : y \geq 1\}$ is closed. To prove this, let (a, b) be a limit point of A, then there exist a sequence $\{x(k), y(k)\}_k$ in $A \backslash \{(a, b)\}$ converging to (a, b), so $x(k) \to a$ and $y(k) \to b$. As $y(k) \geq 1$ for all $k \in \mathbb{N}$ and $y(k) \to b$, it follows by a limit law over $\mathbb{R}$ that $b = \lim_{x\to\infty} y(k) \geq 1$. Thus $(a, b) \in A$.

Closed sets are often produced via the closure.

Lemma 2.4.8 The closure of a set A is closed.

Proof. Let p be a limit point of $\overline{A}$. Then there exist a sequence $\{x(k)\}_{k=1}^{\infty}$ in $\overline{A}\backslash\{p\}$ converging to p. For each $k \in \mathbb{N}^+$, $x(k) \in \overline{A}$ implies that there exists $y(k) \in A$ satisfying $||x(k) - y(k)|| < \frac{1}{k}$. Thus, it suffices to show that the sequence $\{y(k)\}_{k=1}^{\infty}$ in A converges to p, which implies that p is a limit of A, which implies $p \in \overline{A}$.
Fix $\epsilon > 0$, as $x(k) \to p$, there exists $K \in \mathbb{N}$ such that

$$\forall k \in \mathbb{N}, k \leq K \implies ||x(k) - p|| < \frac{\epsilon}{2}.$$

Set $K' = \max\{K, \lceil \frac{2}{\epsilon} \rceil\}$. By the triangle inequality, for $k \in \mathbb{N}$ with $k \geq K'$,

$$\begin{aligned} ||y(k) - p|| = ||y(k) - x(k) + x(k) - p|| &\leq ||y(k) - x(k)|| + ||x(k) - p|| \\ &< \frac{1}{k} + \frac{\epsilon}{2} \\ &\leq \frac{1}{K'} + \frac{\epsilon}{2} \leq \epsilon. \end{aligned}$$

Hence, $y(k) \to p$ as required.

Lemma 2.4.9 Let $A \subseteq \mathbb{R}^n$, then the following are equivalent:

1. A is closed.
2. $A = \overline{A}$.
3. $\partial A \subseteq A$.

Proof. This is left as an exercise.

2.5 Set operations

Lemma 2.5.1 A set $A \subseteq \mathbb{R}^n$ is open if and only if its complement $A^c = \mathbb{R}^n \backslash A$ is closed.

Proof. Assume $A \subseteq \mathbb{R}^n$ is an open set. let p be a limit point of the complement A^c, so there exists a sequence $\{x(k)\}_{k=1}^{\infty}$ in $A^c\backslash\{p\}$ satisfying $x(k) \to p$. Since A is open, if $p \in A$ then the tail of this sequence must lie in A which is impossible since $x(k) \in A^c$ for all $k \geq 1$ for all $k \in \mathbb{N}$. Therefore, p must lie in the complement $\mathbb{R}^n\backslash A = A^c$. This shows A^c is closed.

Conversely, assume $B = A^c$ is a closed set, so we must show $B^c = A$ is open. Let $p \in B^c$ be arbitrary. Since B is closed, the point p cannot be a limit point of B. Therefore, there exists $\epsilon > 0$ such that $B_\epsilon(p) \backslash \{p\}$ does not contain any points of B. In other words, $B_\epsilon(p) \backslash p \subseteq B^c$, since B is closed, we can deduce $B_\epsilon(p) \subseteq B^c$. This shows that p is an interior point of B^c and so B^c is open.

Example 2.5.2 The empty set $\emptyset$ and the set $\mathbb{R}^n$ are both open and closed in $\mathbb{R}^n$, which is sometimes called **clopen**. They are also the only two clopen sets in $4\mathbb{R}^n$.

Example 2.5.3 There are examples of sets which are neither open nor closed:

- The interval $[13, 237)$ is neither open nor closed as shown in previous examples.
- The set $A = \{\frac{1}{n} : n \in \mathbb{N}\}$ is neither open nor closed. It is not open because any open ball around any point is not contained in the set, it is not closed because 0 is a limit point and does not belong to A.
- The set $B = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 < 1, y \geq 0\}$ is the upper half of an open disk, including the points along the x-axis. It is not open because $(0, 0)$ is not an interior point, and it is not closed because $(0, 1)$ is a boundary point that does not belong to B.
- The set of rationals $\mathbb{Q}$ is neither open nor closed since its interior $\mathbb{Q}^o = \emptyset$ and its closure $\overline{\mathbb{Q}} = \mathbb{R}$ are not equal to $\mathbb{Q}$.

Open sets and closed sets respect other basic set operations, too.

Lemma 2.5.4 Lemma 2.4.18 All of the following are true for sets in $\mathbb{R}^n$:

1. A finite intersection of open sets is open.
2. A finite or infinite union of open sets is open.
3. A finite union of closed sets is closed.
4. A finite or infinite intersection of closed sets is closed.
5. A finite Cartesian product of open sets is open.
6. A finite Cartesian product of closed sets is closed.

Proof. TBD

Example 2.5.5 An infinite intersection of open sets may not necessarily be open, such as:
- $\bigcap_{\varepsilon>0}(-\varepsilon,\varepsilon)$. This is equal to the singleton $\{0\}$, which is not open.
- $\bigcap_{n=1}^{\infty}\left(0,1+\frac{1}{n}\right)$. This is equal to $(0,1]$, which is not open.
An infinite union of closed sets may not necessarily be closed, such as:
- $\bigcup_{0<\varepsilon<1}[-\varepsilon,\varepsilon]$. This is equal to the interval $(-1,1)$, which is not closed.
- $\bigcup_{n=1}^{\infty}\left[0,1-\frac{1}{n}\right]$. This is equal to $[0,1)$, which is not closed.

2.6 Compact sets

Suppose a set $A\subseteq\mathbb{R}^n$ is the domain of a real-valued function f and we want to determine its maximum value, but how do we find it? We can try constructing a sequence of points $\{x(k)\}_{k=1}^{\infty}$ in A, where each value $f(x(k))$ is attempting to approximate the maximum value of f. Ideally, we want the sequence to converge to some point p inside A, but there are two critical issues:

- *How can we ensure that the sequence in A converges to a point $p\in\mathbb{R}^n$?*
- *How can we ensure that $p\in A$?*

From last section, we have successfully addressed the second question by assuming the set A is closed, but it is unfortunately not enough for the first point.

Example 2.6.1 Consider the closed set $A=[-1,1]$. The sequence $x(k)=(-1)^k$ for $k\in\mathbb{N}^+$ lying in A does not converge. On the other hand, notice that $x(2k)=1$ for all $k\in\mathbb{N}^+$ so the subsequence of even terms $x(2),x(4),...$ converges to $1\in A$.
Consider the closed set $B=[0,\infty)$. The sequence $y(k)=2^k$ for $k\in\mathbb{N}^+$ lying in B does not converge and, no matter how we pick terms from this sequence, we cannot form a subsequence which will converge. This is because B is not bounded, so the terms of the sequence $x(k)$ can always stay far apart.

As seen above, we can always construct an "alternating" sequence like the example above which will not converge in a closed set. However, we can indeed pick infinitely terms from a sequence and form a new subsequence which will converge. If our set A has the property that ew can always do this for any sequence in A, then we will have addressed the first question!

2.6.1 Definitions of compactness

Definition 2.6.2 A set $A\subseteq\mathbb{R}^n$ is **compact** if every sequence of A has a subsequence which converges to a point lying inside A

Example 2.6.3 Here are some common examples of non-compact sets.

- The interval $A = (0,1]$ in $\mathbb{R}$ is not compact. Verify this with the sequence $\left\{\frac{1}{k}\right\}_{k=1}^{\infty}$ since it (and any subsequence) converges to 0 which does not belong to $A = (0,1]$.
- The set $B = \left\{\frac{1}{n} : n \in \mathbb{N}\right\}$ in $\mathbb{R}$ is not compact for the same reason as A.
- The set $\mathbb{R}^n$ is not compact since, for example, the sequence $x(k) = (k, k, \ldots, k)$ and any subsequence of it does not converge.
- The unit open ball $B_1(0)$ in $\mathbb{R}^n$ is not compact because you can construct a sequence which converges to a point on the unit sphere $\partial B_1(0) = S^{n-1}$.

These non-examples illustrate two necessary conditions for compactness: closed and bounded.

Definition 2.6.4 A set $A \subseteq \mathbb{R}^n$ is **bounded** if $\exists R > 0$ such that $A \subseteq \{x \in \mathbb{R}^n : ||x|| < R\}$. A set that is not bounded is called **unbounded**.

If a set A is unbounded, then it cannot be compact as we can construct a sequence $\{x(k)\}_k$ lying in A such that $||x(k)|| \to \infty$, which implies that the terms will never get close to each other, and so no subsequence can converge.

Example 2.6.5 Here are some examples of compact sets. r

- The empty set $\emptyset$ is vacuously compact.
- Any finite set A is compact since some elements in A must repeat infinitely often, and we can simply choose this constant subsequence.
- The close interval $[a, b]$ in $\mathbb{R}$ is compact.

These examples illustrate the difficult of verifying whether a set A is compact by definition, bu luckily, there is a natural equivalent definition of compactness that is much easier to see.

Theorem 2.6.6 (Bolzano-Weierstrass) A set in $\mathbb{R}^n$ is compact if an only if it is both closed and bounded.

Proof. $\Rightarrow$ First assume $A \subseteq \mathbb{R}^n$ is compact. To show A is closed, let p be a limit point of A, there exists a sequence $\{x(k)_k\}$ lying in A such that $x(k) \to p$, then any subsequence of $\{x(k)_k\}$ must therefore converge to p as well. Since A is compact, this implies $p \in A$ and hence A is closed.
Second, we show A is bounded by contrapositive. If A is not bounded, then $\forall k \in \mathbb{N}^+$, the set A is not a subset of $B_k(0)$. Using this, we can construct a sequence so that each $x(k) \in A$ has the property $||x(k)|| > k$. Let $\{x(m(k))\}_k$ be any subsequence of $\{x(k)\}_k$ with m a strictly increasing function. It suffices to show that $\{x(m(k))\}$ does not converge to any point in $\mathbb{R}^n$. Fix $p \in \mathbb{R}^n$, we

need to prove that

$$\exists \epsilon > 0 \ s.t. \ \forall K \in \mathbb{N}^+, \exists k \in \mathbb{N}^+ \ s.t. \ k \geq K \text{ and } ||x(m(k)) - p|| \geq \epsilon.$$

Take $\epsilon = 1$ and let $K \in \mathbb{N}^+$. Since $m(k)$ is a strictly increasing function, set k to be large enough so that $m(k) > K$ and $m(k) - ||p|| \geq 1$, then $||x(m(k))|| > m(k)$ by construction of x. By rearranging the triangle inequality, we see by our choice of the sequence x that

$$||x(m(k)) - p|| \geq ||x(m(k))|| - ||p|| > m(k) - ||p|| \geq 1 = \epsilon$$

This shows that $x(m(k))$ does not converge to p, as required. Since the subsequence and point p were arbitrary, the set A is not compact. This completes the proof of this direction.
$\Leftarrow$ Assume $A \subseteq \mathbb{R}^n$ is closed and bounded. Let $\{x(k)\}_k$ be a sequence lying in A. We claim that $\{x(k)\}_k$ has a convergent subsequence, which means we must prove that this subsequence converges to a point p in A, and since A is closed, this would prove A is compact.
Express the sequence $x(k) = (x_1(k), ..., x_n(k))$ in terms of its components, so $x_i(k) \in \mathbb{R}$ for all $k \in \mathbb{N}^+$ and all $i \in \{1, ..., n\}$. For $i \in \{1, ..., n\}$, we show that each sequence $\{x_i(k)\}_k$ of real numbers is bounded in $\mathbb{R}$. Since A is bounded, there exists $R > 0$ such that $A \subseteq B_R(0)$, so $\forall a \in A$, we have $||a|| < R$. Writing $a = (a_1, ..., a_n)$, it follows that $|a_i| \leq ||a|| < R$ for all $i \in \{1, ..., n\}$. This implies that $A \subseteq (-R, R)^n$ so $x(k) \in (-R, R)^n$ for every $k \in \mathbb{N}^+$. Thus each coordinate sequence is bounded within $(-R, R)$. Since ever bounded sequence in $\mathbb{R}$ has a convergent subsequence, we get that the first coordinate sequence $\{x_1(k)\}_k$ in $\mathbb{R}$ has a convergent subsequence, say $\{x_1(m_1(k)))\}_k$ for some strictly increasing function m_1,converging to some real number $p_1 \in \mathbb{R}$. Then the second coordinate sequence also has subsequence $\{x_2(m_1(k))\}_k$ in $\mathbb{R}$ which is bounded. Thus this subsequence has a convergent subsequence $\{x_2(m_2(k))\}_k$ where m_2 is a strictly increasing function whose range lies in side the range of m_1. Sine any subsequence of a convergent sequence still converges, so $x_1(m_2(k)) \to p_1$ as $k \to \infty$.
Repeat this process for all the coordinates, which ultimately gives us a strictly increasing function m_n and real numbers $p_1, ..., p_n \in \mathbb{R}$ such that

$$\forall i \in \{1, ..., n\}, x_i(m_n(k)) \to p_i \text{ as } k \to \infty.$$

This implies that the subsequence $\{x(m_n(k))\}_k$ converges to $p = (p_1, ..., p_n) \in \mathbb{R}^n$, which completes the proof.

Example 2.6.7 Going back to the interval $[a, b]$ with $a < b$. Clearly it is both bounded and closed, so it is compact. $[a, \infty), (a, b), (-\infty, b)$ are all not compact as it is either not closed or not bounded.
Fix $r > 0$ and $a \in \mathbb{R}$. The open ball $B_r(a)$ is not closed so it is not compact. The closed ball $\overline{B_r(a)}$ and the sphere $\partial B_r(a)$ are both closed anc bounded, so they are compact.

2.6.2 Set operations and subsets

Like open and closed sets, compact sets respect set operations.

Lemma 2.6.8 All of the following are true for sets in $\mathbb{R}^n$.

1. A finite union of compact sets is compact.
2. A finite or infinite intersection of compact sets is compact.
3. A finite Cartesian product of compact sets is compact.

Proof. This is left as an exercise, we have to verify that each statement is true if we replace every instance of "compact" with "bounded". Then this follows from previous lemmas.

Lemma 2.6.9 Let A be a compact set in $\mathbb{R}^n$. If $B \subseteq A$ and B is closed then B is compact.

Proof. This follows directly from the Bolzano-Weierstrass theorem but we can also give a proof from the definition of compactness. Let $\{x(k)\}_k$ be a sequence in B. Since $B \subseteq A$, the sequence also lies in A. By the compactness of A, it has a subsequence $\{x(m(k))\}_k$ converging to some $p \in A$, now this subsequence must also lies in B, and since B closed and $x(m(k)) \to p$, this implies $p \in B$. Hence, B is compact.

Example 2.6.10 Consider the following subset of $\mathbb{R}^3$:

$$S = \left\{(x, y, z) \in \mathbb{R}^3 : 1 \le x^2 + y^2 \le 2, -1 \le z \le 1\right\}$$

Intuitively, this is a hollowed out cylinder in $\mathbb{R}^3$. You can prove that this is a compact set using the above two lemmas. First, the set

$$A = \left\{(x, y) \in \mathbb{R}^2 : 1 \le x^2 + y^2 \le 2\right\}$$

is a closed subset of the closed ball $\left\{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \le 2\right\}$, hence A is compact in $\mathbb{R}^2$. The closed interval $B = [-1, 1]$ is compact in $\mathbb{R}$. Thus, since a finite Cartesian product of compact sets is compact, we have

$$A \times B = \left\{(x, y) \in \mathbb{R}^2 : 1 \le x^2 + y^2 \le 2\right\} \times [-1, 1] = S$$

is compact in $\mathbb{R}^3$, as desired.

2.7 Limits

All of our labour in the previous section will help generalize the most fundamental notion of calculus: *limits*. The underlying formalities are very similar to single variable calculus, but there are some new subtleties in higher dimensions.
In $\mathbb{R}^n$, we can only approach a real number from either the left or the right, and limits on the endpoints of an interval are often separately defined a one-sided limits. However in $\mathbb{R}^n$, limits need to be able to approach a point from any possible direction and in any weird way. It is therefore not reasonable to use a separate definition for limits at boundary points. Hence, a good definition for limit in $\mathbb{R}^n$ should be indistinguishable between boundary points versus interior points of a set A.

2.7.1 Formal definitions

The formal definition of a limit in $\mathbb{R}$ can be generalized to $\mathbb{R}^n$ by simply replacing the absolute value with the corresponding norm.

Definition 2.7.1 Let $f : A \to \mathbb{R}^m$ be a function with $A \subseteq \mathbb{R}^n$. Let $a \in \mathbb{R}^n$ be a limit point of A and let $b \in \mathbb{R}^m$. Then b is the **limit of f at a** if

$$\forall \epsilon > 0, \exists \delta > 0 \; s.t. \; \forall x \in A, 0 < ||x - a|| < \delta \implies ||f(x) - b|| < \epsilon.$$

If the above holds, we write $\lim_{x \to a} f(x) = b$ or write $f(x) \to b$ as $x \to a$. Otherwise, the limit does not exist.

Example 2.7.2 Here is a proof that $\lim_{(x,y)\to(2,3)}(x + y) = 5$ using the formal definition of a limit.

Proof. Fix $\epsilon > 0$. Take $\delta = \epsilon/2$. let $(x, y) \in \mathbb{R}^2$. Assume $0 < ||(x, y) - (2, 3)|| < \delta$. This implies that

$$|x - 2| \leq \sqrt{(x-2)^2 + (y-3)^2} = ||(x, y) - (2, 3)|| < \delta$$

and similarly $|y - 3| < \delta$. Then

$$||(x + y) - t|| = |x + y - 5| = |(x - 2) + (y - 3)| \leq |x - 2| + |y - 3| < 2\delta = \epsilon.$$

This completes the proof.

There are a few trick we can use to estimate the norms that arises in the proof. The first is to estimate the distance between a fixed coordinate of two points by the distance between two points which was used in the example above. The other is to use an intermediate approximation by "fixing one variable at a time", this occurs in the previous example with the triangle inequality. The next example shows both ideas as well.

Example 2.7.3 Here is a proof that $\lim_{(x,y)\to(2,3)} xy = 6$ using the formal definition of a limit.

Proof. Fix $\varepsilon > 0$. Take $\delta = \min\{\varepsilon/6, 1\}$. Let $(x,y) \in \mathbb{R}^2$. Assume $0 < \|(x,y) - (2,3)\| < \delta$. Note that

$$|x-2| \leq \|(x,y)-(2,3)\| < \delta \quad \text{and} \quad |y-3| \leq \|(x,y)-(2,3)\| < \delta$$

Moreover, as $\delta \leq 1$, this implies that $2 < y < 4$. Then

$$\begin{aligned}
\|xy-6\| = |xy-6| &= |xy-2y+2y-6| \\
&\leq |xy-2y| + |2y-6| && \text{by the triangle inequality} \\
&\leq |x-2|\cdot|y| + 2|y-3| \\
&\leq 4|x-2| + 2|y-3| \\
&< 6\delta && \text{as } |x-2| < \delta \text{ and } |y-3| < \delta \\
&\leq \varepsilon && \text{as } \delta \leq \varepsilon/6.
\end{aligned}$$

This completes the proof.

Notice that we fixed one variable at a time by introducing a function $f(x,y) = xy$, hence the first bracket is $f(x,y) - f(2,y)$ where only x-coordinate is different, and $f(2,y) - f(2,3)$ has only the x coordinate different. Another idea is to use the existence of single variable limits as an input.

Example 2.7.4 Here is a proof that $\lim_{(x,y)\to(0,0)} \cos(x+y) = 1$.

Proof. Fix $\varepsilon > 0$. As $\cos(t)$ is continuous at $t = 0$ and $\cos 0 = 1$, there exists $\delta_1 > 0$ such that

$$\forall t \in \mathbb{R}, |t| < \delta_1 \implies |\cos(t) - 1| < \varepsilon$$

Take $\delta = \delta_1/2$. Let $(x,y) \in \mathbb{R}^2$. Assume $0 < \|(x,y)\| < \delta$ so this implies, by the triangle inequality, that

$$|x+y| \leq |x| + |y| \leq 2\|(x,y)\| < 2\delta = \delta_1.$$

Hence, it follow from the first equation that $|\cos(x+y) - 1| < \varepsilon$ as required.

Now that we have some foundations for limits of maps $\mathbb{R}^n \to \mathbb{R}^m$, naturally we will explore

continuity in high dimensions next. The definitions and basic properties parallel each other in almost every way, the key difference is that in high dimension we can approach a point in countless many ways. This feature leads to many possible ways that a function can be discontinuous.
Looking back at our definition of limit, notice that a limit is only defined at limit points of the domain. This means the limit at any interior point is well defined, but not necessarily at all boundary points. In particular, a limit at an isolated point of a function's domain is not defined.

Example 2.7.5 Let $A = [1,3) \cup \{7\}$ in $\mathbb{R}$. Define the function $f : A \to \mathbb{R}$ by

$$f(x) = \begin{cases} x & 1 \le x < 2, \\ 100 & 2 \le x < 3, \\ e^{\pi} & x = 7. \end{cases}$$

The limit of f at a for any $a \in [1,3)$ is defined even though it might not exist, but the limit of f at 7 is not defined.

Moreover, we only consider points belonging to the puncture ball and the function's domain, as we cannot evaluate the function at points outside its domain.

Example 2.7.6 Define $f(x,y) = \log(1 - x62 - y^2)$. Although f is only defined inside the open ball $B_1(0,0)$, the limit of f as$(x,y) \to (1,0)$ is still defined according to our definition, despite the fact that it doesn't exist.

To prove that a limit does not exist using our current definition is possible but rather tedious, as you can verify with the negation of this formal open ball definition. Instead, we can use an equivalent sequential definition of the limit.

Lemma 2.7.7 Let $A \subseteq \mathbb{R}^n$ be a set and let $f : A \to \mathbb{R}^m$ be a function. Let $a \in \mathbb{R}^n$ be a limit point of A and let $b \in \mathbb{R}^m$. Then $\lim_{x \to a} f(x) = b$ if and only if for every sequence of points $\{x(k)\}_k$ in $A \backslash \{a\}$ with $x(k) \to a$, the sequence of point $\{f(x(k))\}_k$ in $\mathbb{R}^m$ converges to b, that is, $f(x(k)) \to b$.

Proof. First suppose that a is a limit point of A and $\lim_{x \to a} f(x) = b$,let $\{x(k)\}_k$ be a sequence in $A \backslash \{a\}$ which converges $\to a$. Now for this sequence, since it converges, there exist K such that for all $k \ge K$, $||x(k) - a|| < \delta$. Then by the definition of the limit, it follows that for all $k \ge K$, $||f(x(k)) - b|| < \epsilon$, which shows that this arbitrary sequence converges to b.
Conversely, suppose that every sequence in $A \backslash \{a\}$ with $x(k) \to a$ satisfies $f(x(k)) \to b$. We prove by contradiction and suppose that $\lim_{x \to a} f(x) \neq b$. Then by definition, there exist an $\epsilon_0 > 0$ such that for every $\delta > 0$, there exists x such that $||x - a|| < \delta$ but $||f(x) - b|| \ge \epsilon_0$. Consider a sequence $x(k)$ such that each term satisfies $||x(k) - a|| < \delta = \frac{1}{k}$, then $x(k)$ approaches a as $k \to \infty$ and

$||f(x(k)) - b|| \geq \epsilon_0$. Clearly this is a contradiction as we constructed a sequence which tends to a as $k \to \infty$, but also $f(x(k))$ does not approach b no matter the value of k. Completing the proof.

Example 2.7.8 Here is a proof that $\lim_{(x,y)\to(0,0)} \frac{xy}{x^2+y^2}$ does not exist.
Proof. Take the first sequence to be $x(k) = (0, 1/k)$ for $k \in \mathbb{N}^+$. By Lemma 2.3.14, this converges to $(0, 0)$. Notice that for $k \in \mathbb{N}^+$,

$$f(x(k)) = \frac{0}{0^2 + 1/k^2} = 0$$

so $\lim_{k\to\infty} f(x(k)) = 0$ by Lemma 2.3.14. Take the second sequence to be $y(k) = (1/k, 1/k)$ for $k \in \mathbb{N}^+$. By Lemma 2.3.14, this converges to $(0, 0)$. Notice that for $k \in \mathbb{N}^+$,

$$f(y(k)) = \frac{1/k^2}{1/k^2 + 1/k^2} = \frac{1}{2}$$

so $\lim_{k\to\infty} f(y(k)) = 1/2$ by Lemma 2.3.14. Since $\lim_{k\to\infty} f(x(k)) \neq \lim_{k\to\infty} f(y(k))$ where $\{x(k)\}_k$ and $\{y(k)\}_k$ both converge to $(0, 0)$, the desired limit does not exist by Lemma 2.6.8.

Like single variable calculus, these definitions capture the essence of limits but are not efficient for computation.

2.7.2 Basic properties

The most important and fundamental property allows you to reduce limits of vector valued functions to limits of real valued functions.

Theorem 2.7.9 Let $f : A \to \mathbb{R}^m$ where $A \subseteq \mathbb{R}^n$. Let a be a limit point of A and let $b = (b_1, \ldots, b_m) \in \mathbb{R}^m$. Let $f_1, \ldots, f_m$ be the coordinate functions of f so $f = (f_1, \ldots, f_m)$. Then

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

if and only if for all $i \in \{1, \ldots, m\}$

$$\lim_{x\to a} f_i(x) = b_i$$

Proof. Proof. ($\implies$) Assume $\lim_{x\to a} f(x) = b$. Fix $i \in \{1, \dots, m\}$. Let $\varepsilon > 0$. By assumption, there exists $\delta > 0$ such that

$$\forall x \in A, 0 < \|x - a\| < \delta \implies \|f(x) - b\| < \varepsilon.$$

For any $x \in A$, we have that $|f_i(x) - b_i| \leq \|f(x) - b\|$ so the above implies with the same δ that

$$\forall x \in A, 0 < \|x - a\| < \delta \implies |f_i(x) - b_i| < \varepsilon$$

Since $\varepsilon > 0$ was arbitrary, this shows that $f_i(x) \to b_i$ as $x \to a$.
($\impliedby$) Assume $\lim_{x\to a} f_i(x) = b_i$ for every $i \in \{1, \dots, m\}$. Let $\varepsilon > 0$. By assumption, for each $i \in \{1, \dots, m\}$, there exists $\delta_i > 0$ such that

$$\forall x \in A, 0 < \|x - a\| < \delta_i \implies |f_i(x) - b_i| < \frac{\varepsilon}{\sqrt{m}}$$

Set $\delta = \min\{\delta_1, \dots, \delta_m\}$. Let $x \in A$. Assume $0 < \|x - a\| < \delta$. For each $i \in \{1, \dots, m\}$, we have $\delta \leq \delta_i$ so (2.6.3) implies that $|f_i(x) - b_i| < \varepsilon/\sqrt{m}$. It follows that

$$\begin{aligned} \|f(x) - b\| = \sqrt{|f_1(x) - b_1|^2 + \cdots + |f_m(x) - b_m|^2} &\leq \sqrt{m \max_{1\leq i\leq m} |f_i(x) - b_i|^2} \\ &< \sqrt{m \cdot \frac{\varepsilon^2}{m}} \\ &= \varepsilon. \end{aligned}$$

Since $\varepsilon > 0$ was arbitrary, this shows that $f(x) \to b$ as $x \to a$.

As you can see, this makes finding limits much easier!

Example 2.7.10 Define $f(x, y) = (x + y, xy)$ so $f = (f_1, f_2)$ can be written using its coordinate functions $f_1(x, y) = x + y$ and $f_2(x, y) = xy$. By Examples 2.6.3 and 2.6.4, it follows that

$$\lim_{(x,y)\to(2,3)} f_1(x, y) = 5 \quad \text{and} \quad \lim_{(x,y)\to(2,3)} f_2(x, y) = 6$$

Therefore, by Theorem 2.6.10,

$$\lim_{(x,y)\to(2,3)} f(x, y) = \left(\lim_{(x,y)\to(2,3)} f_1(x, y), \lim_{(x,y)\to(2,3)} f_2(x, y)\right) = (5, 6).$$

Theorem 2.7.11 Let $A \subseteq \mathbb{R}^n$ be a set and let a be a limit point of A. Let f and g be $\mathbb{R}^m$-valued functions defined on A. Let ϕ be a real-valued function defined on A. Let $\lambda \in \mathbb{R}$ and $b \in \mathbb{R}^m$ be constants. All of the following hold:

(a) (Constants) $\lim_{x\to a} b = b$ and $\lim_{x\to a} x = a$.
(b) (Linearity) If $\lim_{x\to a} f(x)$ and $\lim_{x\to a} g(x)$ exist then $\lim_{x\to a}(f(x)+\lambda g(x))$ exists and

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

(c) (Scalar product) If $\lim_{x\to a} f(x)$ and $\lim_{x\to a} \phi(x)$ exist then $\lim_{x\to a}(\phi(x)f(x))$ exists and

$$\lim_{x\to a}(\phi(x)f(x)) = \left(\lim_{x\to a} \phi(x)\right)\left(\lim_{x\to a} f(x)\right)$$

(d) (Dot product) If $\lim_{x\to a} f(x)$ and $\lim_{x\to a} g(x)$ exist then $\lim_{x\to a}(f(x)\cdot g(x))$ exists and

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

Proof. These are all left as exercises, for a) and b), use the formal definition, for c), use the theorem above to reduce to the case $m = 1$ and prove using the formal definition. For d), use b) and c).

Finally, the squeeze theorem also exists for real-valued functions.

Theorem 2.7.12 (Squeeze theorem)
Let $A \subseteq \mathbb{R}^n$ be a set and let a be a limit point of A. Let f, g, h be real-valued functions with domain A. Assume there exists $\delta > 0$ such that

$$\forall x \in A, 0 < \|x - a\| < \delta \implies f(x) \leq g(x) \leq h(x)$$

If $\lim_{x\to a} f(x) = \lim_{x\to a} h(x) = b$ for some $b \in \mathbb{R}$ then $\lim_{x\to a} g(x) = b$

Proof. Exercise.

2.7.3 Limits with infinity

In high dimensions, limits with infinity are still defined using a limit, but since there are countlessly many ways to diver outward in $\mathbb{R}^n$ for $n \geq 2$, the notion of $\pm\infty$ no longer makes any sense. Instead, we only consider whether the norm of a sequence grows arbitrarily large.

Definition 2.7.13 Let $A \subseteq \mathbb{R}^n$ be unbounded. Let $f : A \to \mathbb{R}^m$ and let $b \in \mathbb{R}^m$. Define b to be the limit of $f(x)$ as $\|x\| \to \infty$ provided

$$\forall \varepsilon > 0, \exists M > 0 \text{ s.t. } \forall x \in A, \|x\| > M \implies \|f(x) - b\| < \varepsilon.$$

If the above holds, then write $\lim_{\|x\|\to\infty} f(x) = b$ or write $f(x) \to b$ as $\|x\| \to \infty$. Otherwise, the limit $\lim_{\|x\|\to\infty} f(x)$ does not exist.

As we will discover, limits involving infinity are often pretty tricky to deal with.

Example 2.7.14 In the two-dimensional plane, you can check that

$$\lim_{\|(x,y)\|\to\infty} \frac{1}{x^2+y^2} = 0 \quad \text{yet} \quad \lim_{\|(x,y)\|\to\infty} \frac{1}{x^2} \text{ does not exist.}$$

This may seem strange at first but notice that the latter function $\frac{1}{x^2}$ does not tend to zero along all possible sequences $\{(x(k), y(k))\}_k$ with $\|(x(k), y(k))\| \to \infty$.

We can similarly define limits of real-valued functions diverging to infinity.

Definition 2.7.15 Definition 2.6.16 Let $A \subseteq \mathbb{R}^n$ be a set. Let a be a limit point of A. Let $f : A \to \mathbb{R}$ be a real-valued function. The limit of f at a diverges to $+\infty$ provided

$$\forall M > 0, \exists \delta > 0 \text{ s.t. } \forall x \in A, 0 < \|x - a\| < \delta \implies f(x) > M$$

If the above holds, then write $\lim_{x\to a} f(x) = +\infty$ or write $f(x) \to +\infty$ as $x \to a$. The definition for $f(x) \to -\infty$ as $x \to a$ is similar.

Example 2.7.16 Example 2.6.17 In the two-dimensional plane, you can check that

$$\lim_{(x,y)\to(0,0)} \frac{1}{x^2+y^2} = +\infty \quad \text{yet} \quad \lim_{(x,y)\to(0,0)} \frac{x}{x^2+y^2} \text{ does not exist and is not } \pm\infty.$$

Again, you can check the latter limit by using the sequential definition of the limit.

2.8 Continuity

2.8.1 Formal definitions

Coming from single variable calculus, we might think that a function $f : A \to \mathbb{R}^m$ is continuous at a point $a \in A$ whenever

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

This is mostly correct except that the limit is only if a is a limit point of A, but not if a is an isolated point. Hence this motivates us to come up with a better definition:

Definition 2.8.1 Let $f : A \to \mathbb{R}^m$ be a function with domain $A \subseteq R^n$. Let $a \in A$ be a point. The function f is **continuous at** a provided

$$\forall \epsilon > 0, \exists \delta > 0, s.t. \forall x \in A, ||x - a|| < \delta \implies ||f(x) - f(a)|| < \epsilon.$$

With this, notice that all isolated points are continuous since there exist a small enough δ such that $A \cap B_\delta(a) = \{a\}$, so it is vacuously continuous. And for all non-isolated point, we regain our old definition that f is continuous at a if and only if $\lim_{x \to a} f(x) = f(a)$.

Example 2.8.2 Set $f(x, y) = x + y$ and $g(x, y) = xy$. From Examples 2.6.3 and 2.6.4, it follows that

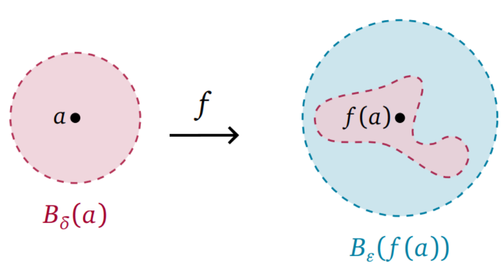
$$\lim_{(x,y) \to (2,3)} f(x, y) = 5 = f(2, 3), \quad \text{and} \quad \lim_{(x,y) \to (2,3)} g(x, y) = 6 = g(2, 3).$$

Thus, both f and g are continuous at $(2, 3)$.

Geometrically, we can using subsets and open balls to create a more intuitive definition. Notice f is continuous at a is equivalent to

$$\begin{aligned} &\forall \varepsilon > 0, \exists \delta > 0 \text{ s.t. } \forall x \in A, x \in B_\delta(a) \implies f(x) \in B_\varepsilon(f(a)) \\ \iff &\forall \varepsilon > 0, \exists \delta > 0 \text{ s.t. } \forall x \in A \cap B_\delta(a), f(x) \in B_\varepsilon(f(a)) \\ \iff &\forall \varepsilon > 0, \exists \delta > 0 \text{ s.t. } f(A \cap B_\delta(a)) \subseteq B_\varepsilon(f(a)). \end{aligned}$$

It can be illustrated with the following diagram.



as with limits, there is also a sequential definition of continuity.

Lemma 2.8.3 Let $f : A \to \mathbb{R}^m$ be a function with domain $A \subseteq \mathbb{R}^n$. Let $a \in A$ be a point. Then f is continuous at a if and only if for every sequence $\{x(k)\}_k$ in A converging to a, the sequence $\{f(x(k))\}_k$ in $\mathbb{R}^m$ converges to $f(a)$.

This lemma is very useful in verifying a discontinuity, and is left as an exercise, note that isolated points must be considered separately.

Definition 2.8.4 Let $f : A \to \mathbb{R}^m$ be a function with domain $A \subseteq \mathbb{R}^n$. For a subset $S \subseteq A$, the function f is continuous on S if f is continuous at a for every $a \in S$. The function f is continuous if f is continuous on its domain A.

Example 2.8.5 The function $f(x, y) = x + y$ is continuous on its domain $\mathbb{R}^2$.

Proof. Fix $(a, b) \in \mathbb{R}^2$. Let $\varepsilon > 0$. Take $\delta = \frac{\varepsilon}{2}$. Let $(x, y) \in \mathbb{R}^2$. Assume

$$\|(x, y) - (a, b)\| < \delta$$

so, as usual, $|x - a| < \delta$ and $|y - b| < \delta$. Then

$$\begin{aligned} \|f(x,y) - f(a,b)\| = |(x+y) - (a+b)| &= |x - a + y - b| \\ &\leq |x-a| + |y-b| \quad \text{by triangle inequality} \\ &< \delta + \delta \quad \text{as } |x-a| < \delta \text{ and } |y-b| < \delta \\ &= 2\delta = \varepsilon, \end{aligned}$$

since $\delta = \varepsilon/2$. This completes the proof.

Recall in single variable calculus, there were 3 common types of discontinuities: removable, jump, or infinite. Discontinuities in higher dimensions are much more diverse. However **removable discontinuity** still remains and is straightforward to describe.

Example 2.8.6 Define

$$F(x, y) = \begin{cases} x + y & \text{if } (x, y) \neq (2, 3) \\ 237 & \text{otherwise.} \end{cases}$$

By Example above, F is continuous on $\mathbb{R}^2 \backslash \{(2, 3)\}$, but F is not continuous at $(2, 3)$

since

$$\lim_{(x,y)\to(2,3)} F(x,y) = 5 \neq 237 = F(2,3)$$

It becomes much more complicated for every other cases, as there are infinitely many ways for a function's limit to not exist. Notice that these example below are only about maps $\mathbb{R}^2 \to \mathbb{R}$ yet they are already so unpredictable.

Example 2.8.7 For $(x, y) \neq (0, 0)$, define

$$f(x,y) = \frac{1}{x^2+y^2}, \quad g(x,y) = \frac{x}{x^2+y^2}, \quad h(x,y) = \frac{x}{x+y}$$

and set $f(0,0) = g(0,0) = h(0,0) = 0$. The functions f, g, and h are all discontinuous for the same reason as the function H in Example 2.7.10, namely, their limits at $(0,0)$ do not exist. However, the discontinuities all look completely different! View the graphs of f, g, h in this Math3D demo by toggling each surface one at a time. Notice $f(x,y) \to +\infty$ as $(x,y) \to (0,0)$ but the same is not at all true for g or h.

As we see, even just trying to identify whether or not a function is continuous at a point is not so easy.

2.8.2 Basic properties

Like limits, the most fundamental property allows us to reduce checking continuity of vector-valued functions to continuity of real-valued functions.

Theorem 2.8.8 The map $f = (f_1, ..., f_m) : A \to \mathbb{R}^m$ is continuous at $a \in A$ if and only if for each $i \in \{1, ..., m\}$, the component function f_i is continuous at a.

Proof. This follows almost immediately from 2.7.9.

This naturally leads to the following lemma:

Lemma 2.8.9 Every linear transformation $\mathbb{R}^n \to \mathbb{R}^m$ is continuous.

Proof. By theorem above, it suffices to check linear maps of the form $\mathbb{R}^n \to \mathbb{R}$. That is, for fixed $c_1, ..., c_n \in \mathbb{R}$, we must show $f(x_1, ..., x_n) = c_1x_1 + \cdots + c_nx_n$ is continuous on $\mathbb{R}^n$. Fix $(a_1, ..., a_n)$ in $\mathbb{R}^n$ and $\epsilon > 0$, let $c' = max(c_1, ..., c_n)$ and take $\delta = \frac{\epsilon}{|c'|n}$. Let $(x_1, ..., x_n) \in \mathbb{R}^n$ such that

$||(x_1, ..., x_n) - (a_1, ..., a_n)|| < \delta$, then each $|x_i - a_i| < ||(x_1, ..., x_n) - (a_1, ..., a_n)|| < \delta$ as usual. Then

$$\begin{aligned} ||f(x_1, ..., x_n) - f(a_1, ..., n)|| &= |c_1x_1 + \cdots + c_nx_n - (c_1a_1 + \cdots + c_na_n)| \\ &\leq |c_1||x_1 - a_1| + \cdots + |c_n||x_n - a_n| \\ &\leq |c'|(n\delta) \\ &= \epsilon \end{aligned}$$

Hence every linear transformation $\mathbb{R}^n to \mathbb{R}$ is continuous, this completes the proof.

Example 2.8.10 The lemma above shows all of the following maps are continuous.

- The identity map $\text{id} : \mathbb{R}^n \to \mathbb{R}^n$ defined by $\text{id}(x) = x$ is linear and hence continuous.
- Fix $i \in \{1, 2, \ldots, n\}$. The i th coordinate projection map $\pi_i : \mathbb{R}^n \to \mathbb{R}$ defined by $\pi_i(x) = x_i$ is linear and hence continuous.
- For any $m \times n$ matrix A, the linear map $f : \mathbb{R}^n \to \mathbb{R}^m$ defined by $f(x) = Ax$ is continuous. By a theorem in linear algebra, this actually describes all possible linear maps.

There is also the standard list of properties that follows from limits.

Theorem 2.8.11 Theorem 2.7.16 Let $A \subseteq \mathbb{R}^n$ and let $a \in A$. Let f and g be $\mathbb{R}^m$-valued functions defined on A. Let ϕ be a real-valued function defined on A. Let $\lambda \in \mathbb{R}$. All of the following hold:
(a) If f and g are continuous at a then the function $f + \lambda g$ is continuous at a.
(b) If f and g are continuous at a then their dot product $f \cdot g$ is continuous at a.
(c) If f and ϕ are continuous at a then their scalar product ϕf is continuous at a.

Proof. Again by theorem 2.8.8, it suffices to check all of these properties for real-valued functions. Notice b) and c) are the same in this case. Each statement can then be shown directly by the definition of continuity.

This allows us to create more continuous functions.

Example 2.8.12 Define the map $f : \mathbb{R}^n \to \mathbb{R}$ by $N(x) = \|x\|^2 = x_1^2 + \cdots + x_n^2$. Notice f is the dot product of the identity map $\text{id} : \mathbb{R}^n \to \mathbb{R}^n$ with itself. That is, for $x \in \mathbb{R}^n$,

$$\mathrm{id}(x) \cdot \mathrm{id}(x) = x \cdot x = \|x\|^2 = f(x)$$

Since id is a linear map and hence continuous, the map f is continuous by Theorem 2.7.16(b).

Another fundamental property of continuity relates to composition of functions.

Corollary 2.8.13 Let $f : A \to B$ where $A \subseteq \mathbb{R}^n$ and $B \subseteq \mathbb{R}^m$. Let $g : B \to \mathbb{R}^k$. Let $a \in A$. If f is continuous at a and g is continuous at $f(a)$ then $g \circ f$ is continuous at a.

This corollary follows immediately from the following more general theorem.

Theorem 2.8.14 Let $f : A \to B$ where $A \subseteq \mathbb{R}^n$ and $B \subseteq \mathbb{R}^m$. Let $g : B \to \mathbb{R}^k$. Let a be a limit point of A and let $b \in B$. If $\lim_{x\to a} f(x) = b$ and g is continuous at b then $\lim_{x\to a} g \circ f(x) = g(b)$.

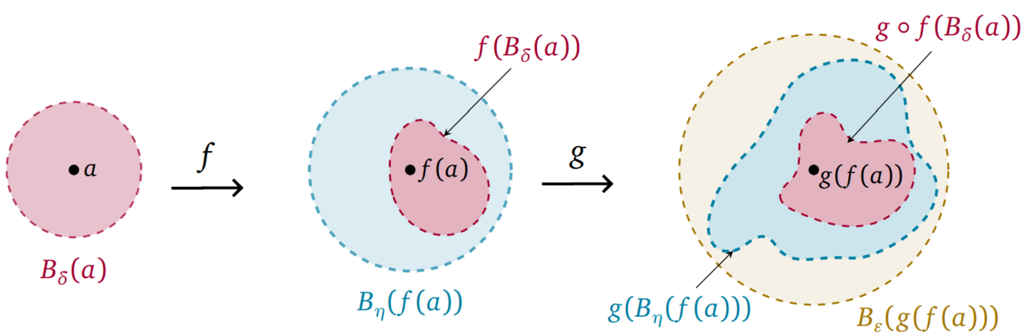
Proof. Fix $\epsilon > 0$. Since g is continuous at b, there exists $\delta' > 0$ such that

$$\forall y \in B, ||y - b|| < \delta' \implies ||g(y) - g(b)|| < \epsilon.$$

Since $f(x) \to b$ as $x \to a$, there exists $\delta > 0$ such that

$$\forall x \in A, 0 < ||x - a|| < \delta \implies ||f(x) - b|| < \delta'.$$

Fix $x \in A$. Assume $0 < ||x - a|| < \delta$. By above, $||f(x) - b|| < \delta'$, which implies $g(f(x)) - g(b)|| < \epsilon$, as required.



Equipped with the theorem and corollary above, we can construct even more continuous functions.

Example 2.8.15 We know $g(t) = \cos(t)$ is continuous on $\mathbb{R}$ and $f(x,y) = (x+y)$ is continuous on $\mathbb{R}^2$. Therefore, $g \circ f(x,y) = g(x+y) = \cos(x+y)$ is continuous on $\mathbb{R}^2$.

Another class of continuous functions are multivariable polynomials.

Definition 2.8.16 Definition 2.7.22 A monomial in the n variables $x_1, \ldots, x_n$ is a function of the form $x_1^{\alpha_1} \cdots x_n^{\alpha_n}$ for some $\alpha_1, \ldots, \alpha_n \in \mathbb{N}$. A polynomial in the n variables $x_1, \ldots, x_n$ is a linear combination of monomials in n variables with real coefficients.

Example 2.8.17 The function $p(x,y,z) = xy + 3z^4$ is a polynomial in the 3 variables x, y, z. It is a linear combination of the monomials xy and z^4 which respectively correspond to $(\alpha_1, \alpha_2, \alpha_3) = (1,1,0)$ and $(0,0,4)$.To show p is continuous, we write it as a composition of functions. Namely, define

- $f : \mathbb{R}^2 \to \mathbb{R}$ by $f(x,y) = xy$, and $g : \mathbb{R} \to \mathbb{R}$ by $g(x) = x^4$.
- $\pi_3 : \mathbb{R}^3 \to \mathbb{R}$ by $\pi_3(x,y,z) = z$, and $\pi_{1,2} : \mathbb{R}^3 \to \mathbb{R}^2$ by $\pi_{1,2}(x,y,z) = (x,y)$

We can check f is continuous by definition. From single-variable calculus, the function g is continuous.Since the projections π_3 and $\pi_{1,2}$ are linear transformations, they are also continuous. Notice that

$$\forall (x,y,z) \in \mathbb{R}^3, \quad p(x,y,z) = xy + 3z^4 = (f \circ \pi_{1,2})(x,y,z) + 3\,(g \circ \pi_3)(x,y,z).$$

From 2.8.14 and our previous observations, $f \circ \pi_{1,2}$ and $g \circ \pi_3$ are continuous. Thus, the polynomial p is continuous as a linear combination of continuous functions are continuous.

Lemma 2.8.18 All polynomials in n variables are continuous on $\mathbb{R}^n$.

Proof. It suffices to show all monomials in n variables are continuous. This is left as an exercise, you only need to know the fact that all linear transformations are continuous, continuity of $f(x,y) = xy$, and continuity of single-variable powers x^n. Use induction on the degree of th monomial.

2.8.3 Topological properties

Topological properties of sets like open, closed, compact, etc. help characterize the behaviour of functions on sets. But there is a dual perspective: how do functions affect the properties of these sets? Without any restriction, anything can happen. But if we only look at continuous function, we get many useful properties.

Continuous functions preserve topological properties of set (under image or preimage).

This can be translated into the following.

Theorem 2.8.19 Theorem 2.7.25 Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be a function. The following are equivalent:
(a) f is continuous on $\mathbb{R}^n$
(b) The preimage $f^{-1}(U)$ is open for every open set $U \subseteq \mathbb{R}^m$
(c) The preimage $f^{-1}(V)$ is closed for every closed set $V \subseteq \mathbb{R}^m$.

Proof. The proofs that (a) implies (b) and (a) implies (c) are below. The converse statements are left as exercises. Assume f is continuous on $\mathbb{R}^n$. By definition,

$$\begin{aligned} &\forall a \in \mathbb{R}^n, \forall \varepsilon > 0, \exists \delta > 0 \text{ s.t. } f\left(B_\delta(a)\right) \subseteq B_\varepsilon(f(a)) \\ \iff &\forall a \in \mathbb{R}^n, \forall \varepsilon > 0, \exists \delta > 0 \text{ s.t. } B_\delta(a) \subseteq f^{-1}\left(B_\varepsilon(f(a))\right). \end{aligned}$$

To prove (b), fix an open set $U \subseteq \mathbb{R}^m$. Let $a \in f^{-1}(U)$ so $f(a) \in U$. Since U is open, there exists $\varepsilon > 0$ such that $B_\varepsilon(f(a)) \subseteq U$. By (2.7.2), there exists $\delta > 0$ such that

$$B_\delta(a) \subseteq f^{-1}\left(B_\varepsilon(f(a))\right) \subseteq f^{-1}(U)$$

as required. This establishes that $f^{-1}(U)$ is open and hence (b) holds. To prove (c), fix a closed set $V \subseteq \mathbb{R}^m$ so V^c is open. Note $f^{-1}\left(V^c\right) = \left(f^{-1}(V)\right)^c$ is open since (b) already follows from (a). You can therefore conclude that $f^{-1}(V)$ is closed, so (c) also holds.

This solves many problems involving whether a set is open or closed, as well as the continuity of a function as shown below.

Example 2.8.20 Example 2.7.27 Let $A = \left\{(x,y,z) \in \mathbb{R}^3 : xy + 3z^4 \leq 8\right\}$. The polynomial $f : \mathbb{R}^3 \to \mathbb{R}$ given by $f(x,y,z) = xy + 3z^4$ is continuous by Lemma 2.7.24. Notice that

$$f^{-1}((-\infty, 8]) = \left\{(x,y,z) \in \mathbb{R}^3 : f(x,y,z) \leq 8\right\} = A$$

Since the interval $(-\infty, 8]$ is closed, it follows that the set A is closed.

Example 2.8.21 Let $f : \mathbb{R}^2 \to \mathbb{R}$ be the function defined by $f(x,y) = 237$ for $x^2 + y^2 < 1$ and $f(x,y) = 0$ otherwise. By definition, you have that

$$f^{-1}(\{237\}) = \left\{(x,y) \in \mathbb{R}^2 : x^2 + y^2 < 1\right\} = B_1((0,0))$$

which is an open ball. Thus, $f^{-1}(\{237\})$ is open, yet the set $\{237\}$ is closed. hence, f cannot be continuous on $\mathbb{R}^n$.

Furthermore, continuity also preserves compactness under image.

Theorem 2.8.22 If A is a compact subset of $\mathbb{R}^n$, and f is an $\mathbb{R}^m$-valued function that is continuous on A, then $f(A)$ is a compact subset of $\mathbb{R}^m$.

Proof. If A is empty, then $f(A)$ is empty and hence compact. Assume A is not empty. Let $\{y(k)\}_k$ be a sequence in $f(A)$, we want to show there exist a subsequence which converges to a point inside $f(A)$. For each $k \in \mathbb{N}^+$, the set $f^{-1}(y(k))$ is non-empty since $y(k) \in f(A)$ and A is non-empty. Thus, we can choose some $x(k) \in f^{-1}(y(k))$ for each $k \in \mathbb{N}^+$. The sequence $\{x(k)\}_k$ lies in A, so since A is compact, it follows by definition that there exists some strictly increasing $m : \mathbb{N}^+ \to \mathbb{N}^+$, such that the subsequence $\{x(m(k))\}_k$ converges to some $a \in A$. By continuity of f

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

Thus, the subsequence $\{y(m(k))\}_k$ converges to $f(a) \in f(A)$. hence $f(A)$ is compact.

This theorem will be decisive in the proof of the extreme value theorem. For now, we can use it to show seemingly complicated sets are compact or that a function is not continuous.

Example 2.8.23 Consider the set $B = \{(xy, yz, xz) : 0 \le x, y, z \le 1\}$. The function $f : \mathbb{R}^3 \to \mathbb{R}^3$ given by $f(x,y,z) = (xy, yz, xz)$ satisfies

$$f\left([0,1]^3\right) = \{f(x,y,z) : 0 \le x, y, z \le 1\} = B$$

Moreover, f is a polynomial in each component, so f is continuous as each component is continuous. Since the cube $[0,1]^3$ is compact, the set $B = f\left([0,1]^3\right)$ is therefore compact.

Example 2.8.24 Define the function $f : \mathbb{R}^2 \to \mathbb{R}$ by $f(x) = 1/x$ if $x \neq 0$ and $f(0) = 0$. For any $0 < \varepsilon < 1$, the image of the set $A = [\varepsilon, 1]$ is the set $f(A) = [1, 1/\varepsilon]$, which is compact. On the other hand, the image of the set $B = [0,1]$ is the set $\{0\} \cup [1, \infty)$, which is not compact. Hence, f is not continuous.

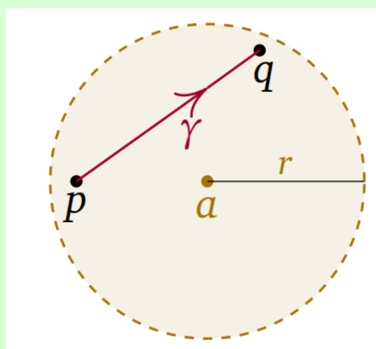
2.9 Path-connected sets

We have investigated some key topological properties of set. They each have their own intuitive concept, but none of them ensure that a set is "connected" in a natural sense. There are examples of open, closed, and compact sets which are disjoint union of sets. So how can we capture the concept of "connected" for a set $S \subseteq \mathbb{R}^n$. Informally, this is the same as being able to walk from one point of S to any other point in S without leaving S.

Definition 2.9.1 A set $S \subseteq \mathbb{R}^n$ is **path-connected** if for every pair of points $p, q \in S$ three exists a continuous function $\gamma : [a, b] \in \mathbb{R}^n$ such that $\gamma(a) = p$ and $\gamma(b) = q$ and $\text{img}(\gamma) \subseteq S$.

A few simple examples can demonstrate that this definition is sensible.

Example 2.9.2 Fix $a \in \mathbb{R}^n$ and $r > 0$. The open ball $B_r(a)$ is path-connected.



The "picture proof" is illustrated above and the formal proof is below.
Proof. Fix $p, q \in B_r(a)$. Define $\gamma : [0, 1] \to \mathbb{R}^n$ by

$$\gamma(t) = (1 - t)p + tq$$

so $\gamma(0) = p$ and $\gamma(1) = q$. Note γ is the straight line segment from p to q. Each component of γ is a linear single-variable polynomial t, so it is continuous. It remains to check $\text{img}(\gamma)$ is contained in the open ball. Fix $t \in [0, 1]$. Observe that

$$\begin{aligned} \|\gamma(t) - a\| &= \|(1-t)p + tq - a\| \\ &= \|(1-t)(p-a) + t(q-a)\| \\ &\leq \|(1-t)(p-a)\| + \|t(q-a)\| && \text{by the triangle inequality} \\ &= |1-t|\|p-a\| + |t|\|q-a\| && \text{as } t \in \mathbb{R} \\ &< |1-t|r + |t|r = r && \text{as } p, q \in B_r(a). \end{aligned}$$

This implies that $\gamma(t) \in B_r(a)$ so $\text{img}(\gamma)$ is in the open ball.

The open ball is a special example of a path-connected set since we can connect any two points by a line segment. This common property warrants its own terminology.

Definition 2.9.3 A set $S \subseteq \mathbb{R}^n$ is **convex** if the line segment between any two points $p, q \in S$ lies inside S.

Convex sets are an especially nice case of path-connected sets, but we will not delve into them too deeply in this course. Balls, cubes, planes, and regular polygons are all convex, however anything that has an indent or a part jutting out is not convex. It is obvious that all convex sets are path-connected but not vice versa. To prove a set is not path-connected requires the use of the intermediate value theorem.

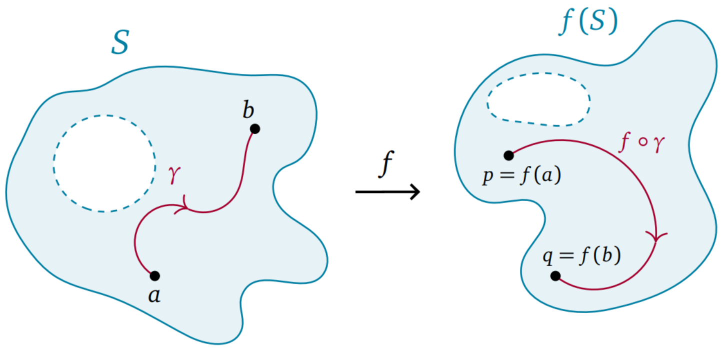
Example 2.9.4 The set $S = \left\{(x, y) \in \mathbb{R}^2 : x \neq 0\right\}$ is the 2-dimensional plane minus the vertical axis. It is not path-connected. You can illustrate the proof with a picture.

Proof. Take $p = (-1, 0)$ and $q = (1, 0)$. Let $\gamma : [a, b] \to \mathbb{R}^2$ be any continuous function such that $\gamma(a) = p = (-1, 0)$ and $\gamma(b) = q = (1, 0)$. Write $\gamma = (\gamma_1, \gamma_2)$ in terms of its component functions, each of which are continuous by Theorem 2.7.12. Note $\gamma_1(a) = -1$ and $\gamma_1(b) = 1$ and $\gamma_1 : [a, b] \to \mathbb{R}$ is continuous. By the intermediate value theorem, there exists $c \in (a, b)$ such that $\gamma_1(c) = 0$. This implies that $\gamma(c)$ lies on the vertical axis, so $\gamma(c) \notin S$. This proves $\text{img}(\gamma) \nsubseteq S$ and hence S is not path-connected.

Now that we've understood path-connectedness, we can add yet another topological property of continuous functions, namely, they preserve path-connectedness sets under image.

Theorem 2.9.5 Let $S \subseteq \mathbb{R}^n$ be a path-connected set. Let f be a $\mathbb{R}^m$-valued function defined on S, if f is continuous on S then $f(S)$ is path-connected.

Proof. Let $p, q \in f(S)$ be arbitrary. By definition, there exists $a, b \in S$ such that $f(a) = p$ and $f(b) = q$. Since S is path-connected, there exists a continuous map $\gamma : [0, 1] \to \mathbb{R}^n$ such that $\gamma(0) = a$ and $\gamma(1) = b$ and the range of γ lies inside S. Since the range of γ lies inside the domain of f, we may define the map $f \circ \gamma : [0, 1] \to \mathbb{R}^m$. An illustration is below.



It suffices to show that the parametric curve $f \circ \gamma$ is continuous, has range lying in $f(S)$, starts at p, and ends at q. Since f and γ are continuous, it follows that $f \circ \gamma$ is continuous. The range of γ lying in S implies that the range of $f \circ \gamma$ lies in $f(S)$. Finally, $f \circ \gamma(0) = f(a) = p$ and $f \circ \gamma(1) = f(b) = q$, so we have indeed found a continuous function from p to q.

You can see this theorem in action with a difficult-to-describe set.

Example 2.9.6 Recall the set $B = \{(xy, yz, xz) : 0 \le x, y, z \le 1\}$ and the continuous function $f : \mathbb{R}^3 \to \mathbb{R}$ given by $f(x, y, z) = (xy, yz, xz)$. Since the cube $[0, 1]^3$ is path-connected, it follows that $B = f\left([0, 1]^3\right)$ is also path-connected.

By taking $S = [a, b]$ and $m = n = 1$ in 2.9.5, we have actually recovered the intermediate value theorem over $\mathbb{R}$.

Corollary 2.9.7 Let f be a real-valued function defined on $[a, b]$, if f is continuous on $[a, b]$, then $f([a, b])$ is parth-connected.

Remark 2.9.8 As $[a, b]$ is compact, $f([a, b])$ is a compact path-connected set. You can verify that the only compact path-connected sets in $\mathbb{R}$ are closed intervals, so $f([a, b]) = [c, d]$ for some $c, d \in \mathbb{R}$ with $c \le d$.

Generalizing a core theorem from calculus like the IVT is a major step towards understanding multivariable calculus. In the next section, we will talk about another generalization from single variable calculus: the extreme value theorem.

2.10 Global extrema

Multivariable optimization is one of the most fundamental applications of calculus. Over several chapters of this textbook, we will develop basic techniques for solving these optimization problems. You must first address a simple yet deceptively challenging question:

For a given optimization problem, does a solution exist?

Of course, one way is to actually find the solution itself and prove it is optimal, but this is not a good method to execute in general. Even worse, we might be searching for a solution which doesn't exist. In this section, we will characterize a broad class of optimization problems where we are guaranteed the existence of a solution

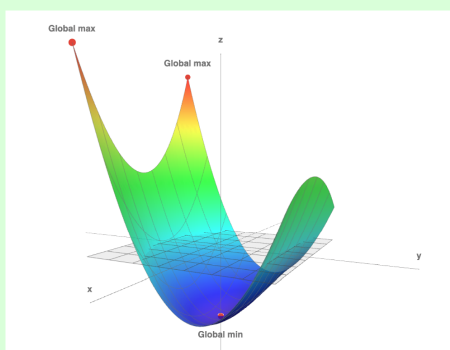
2.10.1 definitions of global extreme

Definition 2.10.1 Let $A \subseteq \mathbb{R}^N$ and let f be a real-valued function defined on A.

- A point $p \in A$ is a **global maximum point of f on** A if $f(p) \geq f(x)$ for all $x \in A$.
- If a maximum point of f on A exists, then f **attains a global maximum on** A

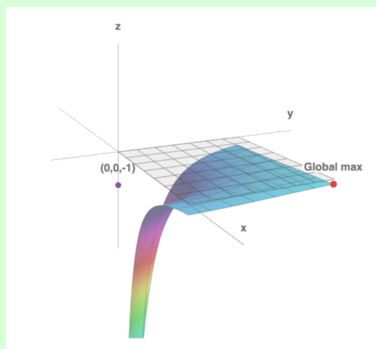
The definitions of minimum point, minimum value, and attaining a minimum are similar.

Example 2.10.2 Consider $f(x, y) = x^2 + 4y^2 - 2x^2y - 2$ on the rectangle $A = [-1, 1] \times [-1, 1]$. The graph of f is below which you can also view on Math3D.



From looking at the graph, you can see that f attains its minimum value at $(0, 0)$ and its maximum values at $(1, -1)$ and $(-1, -1)$. Note that there may be multiple points of A at which f attains its minimum or maximum. Moreover, notice that the minimum is attained on the interior of A, whereas the maxima are attained on the boundary of A. If you consider f on the subset $B = (-1, 1)^2$ then f still attains its minimum on B since $(0, 0) \in B$ but it does not attain a maximum on B. You can get arbitrarily close to $(-1,1)$ or $(-1,-1)$ on B but you cannot reach either point. While f is continuous on A and continuous on B, the issue is that B is not closed whereas A is closed (and hence contains its boundary).

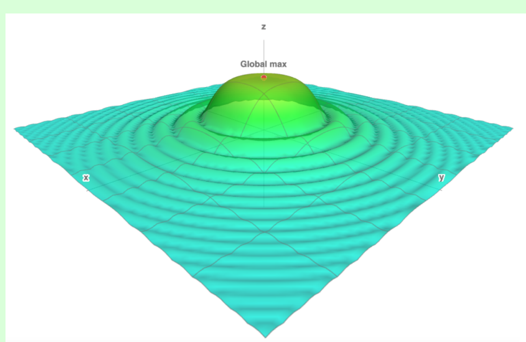
Example 2.10.3 Let $A = [0,2] \times [0,2]$. Define $f : A \to \mathbb{R}$ by $f(0,0) = -1$ and $f(x,y) = \frac{-1}{x^2+y^2}$ otherwise. While A is closed, notice f is not continuous on A. In fact, as $(x,y) \to (0,0)$, $f(x,y) \to -\infty$ so f does not attain a minimum on A. It does, however, attain its maximum at $(2,2)$. You can view these features in its graph, which can you also view on Math3d.



Example 2.10.4 Let $A = \mathbb{R}^2$. Define $f : A \to \mathbb{R}$ by

$$f(x,y) = \frac{3\left|\sin\left(x^2+y^2\right)\right| + 3}{x^2+y^2+1}$$

The domain $A = \mathbb{R}^2$ is unbounded and you can check that f is continuous. It attains its maximum on A at $(0,0)$. You can check that $f(x,y) > 0$ always yet $f(x,y) \to 0$ as $\|(x,y)\| \to \infty$. You can prove that this implies that f has no minimum on A. These features can be seen in its graph which you can view on Math3D.



With these examples, we observe that the key ingredients to guarantee the existence of extreme is that the function should be continuous and the domain should be closed and bounded.

2.10.2 Extreme value theorem

Collecting all of these observations, it is quite easy to formulate the generalization of the extreme value theorem to $\mathbb{R}^n$.

Theorem 2.10.5 (Extreme value theorem)
If $A \subseteq \mathbb{R}^n$ is a non-empty compact set and the map $f : A \to \mathbb{R}$ is continuous, then f attains it maximum and minimum values at points of A.

Proof. It suffices to show a maximum is attained since the proof for the minimum is similar. Since f is continuous on A and A is compact, it follows that $f(A)$ is compact. By Bolzano-Weierstrass, this implies $f(A)$ is a bounded subset of $\mathbb{R}$ so the quantity

$$M = \sup f(A) = \sup y : y \in f(A)$$

is finite. It remains to show there exists $p \in A$ such that $f(p) = M$ By definition of the supremum, for each $k \in \mathbb{N}^+$, there exists $y(k) \in f(A)$ such that

$$M - \frac{1}{k} < y(k) \leq M.$$

As $k \to \infty$, this implies that $y(k) \to M$ so M is a limit point of $f(A)$. Since $f(A)$ is compact and hence closed, it follows that $M \in f(A)$. Therefore, there exists $p \in A$ such that $f(p) = M$.

Now, equipped with the extreme value theorem,we can use it to guarantee the existence of a solution to an optimization problem.

Example 2.10.6 At a given moment, is there a hottest point on the Earth? The Earth can be roughly viewed as the unit sphere S^2 in $\mathbb{R}^3$. Let $T(x, y, z)$ be the temperature in Celsius at a given point $(x, y, z) \in S^2$, so $T : S^2 \to \mathbb{R}$. Presumably, temperature should vary continuously across the Earth's surface so T is continuous. The sphere S^2 is compact so by the extreme value theorem T attains its maximum and minimum on S^2. A maximum point $p \in S^2$ of T corresponds to the hottest point on Earth.

Example 2.10.7 Can you find extrema for f on the open rectangle $B = (-1, 1)^2$? The extreme value theorem guarantees you nothing here since B is not compact. Extrema may or may not exist; in this case, a minimum exists and a maximum does not but anything can happen in general.

Example 2.10.8 Recall Example 2.10.3 where $A = [0,2]^2$ and $f : A \to \mathbb{R}$ is defined by $f(0,0) =$ -1 and $f(x,y) = -\frac{1}{x^2+y^2}$ otherwise. Can you find extrema for f on A ? The extreme value theorem again gives no direct information since f is not continuous on A. Extrema may or may not exist; in this case, a maximum exists but a minimum does not.

Despite these setbacks, the extreme value theorem can still be used to prove the existence of extrema for non-compact sets. Here is one such result.

Lemma 2.10.9 Let $A \subseteq \mathbb{R}^n$ be closed and unbounded. Let f be a continuous real-valued function on A. If $f(x) \to \infty$ as $||x|| \to \infty$ in A, then f attains a maximum on A.

Proof. Fix $k \in \mathbb{R}^n$

There are many more result of this type which we will not list, rather you can formulate them yourself and prove them with a clever application of the extreme value theorem for compact sets. This concludes our discussion on topology, and we are now ready to build differential calculus in higher dimensions!

Chapter 3

Derivatives

Contents

In this chapter, we will devote our time in trying to unravel the definition of the derivatives for maps $\mathbb{R}^n \to \mathbb{R}^m$, and it will take a long and winding road. In the beginning, we will try to reduce the definition to differentiation of single variable maps $\mathbb{R}$ to $\mathbb{R}$. This will be quite successful for the special case of parametric curves and somewhat successful for real-valued functions.
However, this attempt to generalize "rates" and "slopes" to higher dimension will not be enough, because those notions are in the end, a one-dimensional intuition. The crucial insight from calculus will be to interpret single variable derivatives via linear approximations.
The ultimate definition of derivative for any map $\mathbb{R}^n \to \mathbb{R}^m$ will require both linear algebra and calculus, and following the underlying philosophy of

Nonlinear maps are well approximated by linear maps.

This will be discovered by viewing derivatives from four different perspective: physical, geometric, analytic, and algebraic. By translating between these overarching viewpoints, you will hopefully truly understand derivatives.

3.1 Derivatives of one variable

In this section, we will look at the very special case of derivatives for maps of one variables, that is, derivatives of parametric curves which are of the form

$$\mathbb{R} \to \mathbb{R}^m.$$

These are a very special case as we can quickly reduces to single variable calculus, and we can use it to build familiarity with the four viewpoints of the derivative.

3.1.1 Definition

Definition 3.1.1 Let $A \subseteq \mathbb{R}$ and let $f : A \to \mathbb{R}^m$ be a funiton. Let a be an interior point of A. The **derivative of f at a** is defined to be

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

provided the limit exists. If it does, then f is **differentiable at a**.

This definition is visually identical to what was introduced in single variable calculus. however, notice that $h \in \mathbb{R}$ is a scalar which approaches 0, while the limit quantity $\frac{f(a+h)-f(a)}{h}$ is a scalar multiplied by a vector. This implies that $f'(a) \in \mathbb{R}^m$. More, the limit is equivalent to

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

by using composition. Lastly, the derivative is only defined at interior points of A for simplicity.

Example 3.1.2 Let $f : \mathbb{R} \to \mathbb{R}^2$ be defined by $f(x) = (\cos x, \sin x)$. Notice

$$\begin{aligned} f'(\pi) &= \lim_{h \to 0} \frac{(\cos(\pi + h), \sin(\pi + h)) - (\cos \pi, \sin \pi)}{h} \\ &= \lim_{h \to 0} \left(\frac{\cos(\pi + h) - \cos \pi}{h}, \frac{\sin(\pi + h) - \sin \pi}{h} \right) \\ &= \left(\lim_{h \to 0} \frac{\cos(\pi + h) - \cos \pi}{h}, \lim_{h \to 0} \frac{\sin(\pi + h) - \sin \pi}{h} \right) \end{aligned}$$

These are the derivatives of $\cos x$ and $\sin x$ at $x = \pi$, so the above is $(-\sin \pi, \cos \pi) = (0, -1)$. Hence, f is differentiable at π and $f'(\pi) = (0, -1)$.

3.1.2 Basic properties

The prior example shows that derivatives of parametric curves can be calculated very easily.

Lemma 3.1.3 Let $A \subseteq \mathbb{R}$ and let $f = (f_1, ..., f_m) : A \to \mathbb{R}^m$. Let a be an interior point of A. The function is differentiable at a if and only if for every $i \in \{1, ..., m\}$, the component function f_i s differentiable at a. If so,

$$f'(a) = (f_1'(a), \dots, f_m'(a)) = \begin{bmatrix} f_1'(a) \\ \vdots \\ f_m'(a) \end{bmatrix}.$$

Proof. This follows immediately from the fact that the limit of a function can be calculated component wise.

Chapter 4

Improper Integrals

Contents

4.1 TODO

4.2 TODO

4.3 Convergence tests

Now that we have learnt the basic of calculating improper integral and the monotone convergence theorem, the question becomes how do we deal with them without direct computation and analyze functions that are both positive and negative.
With the tool given in single variable calculus. The many comparison tests for integrals are quite easy to establish in multivariable calculus as the ideas are almost identical.

4.3.1 Basic comparison test

The simplest comparison test follows exactly as in single variable calculus, and can be proved with the monotone convergence theorem.

Theorem 4.3.1 Let $\Omega \subseteq \mathbb{R}^n$ be a set with an exhaustion by compact Jordan measurable sets. Let f and g be real-valued locally integrable functions on Ω.

1. If $0 \leq f \leq g$ on Ω and $\int_\Omega g dV$ converges, then $\int_\Omega f dV$ converges.
2. If $0 \leq f \leq g$ on Ω and $\int_\Omega f dV$ diverges, then $\int_\Omega g dV$ diverges.

For functions which may be positive or negative, recall that with infinite series of real numbers back in first year. We can the notion of absolute convergence. That is,

Let $\{a_n\}_n \subseteq \mathbb{R}$. If $\sum_n |a_n|$ converges, then $\sum_n a_n$ converges.