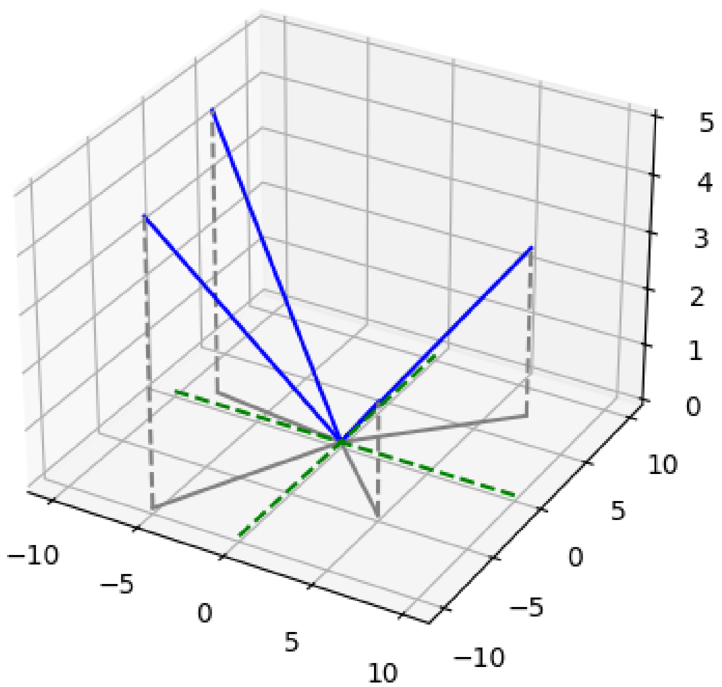
# The Notes and Assignments Sets

Introduction to Mathematics for Data Science

**Zehao Qian**

Data Science (Social Analytics)

”The mind prevails over matter,
transcending the ordinary to reach the divine.”



# Introduction to Math for DS Group Task 0
# The Solutions of Three Problems via LLM

IMDS Group 24
ChatGPT (Zehao Qian) and Bard (Mohammad Jamshaid Iqbal)
Bing (Chloe Mendez) annd Answer Analysis (Zeyue Song)

October 24, 2023

Sometimes, a little pre-setting could make GPT's performance better. Below is the first sentence I spoke to ChatGPT:

I'm a data scientist, you're my assistant, I'm having some problems that need mathematical modeling right now, and you're here to assist me.

## 1 Question 1

The air temperature just above the ground at a particular point on the Earth often varies in a periodic manner over a 24 hour cycle. The daily mean value also varies with the seasons over an annual cycle. If we measure time t in hours, what would be an appropriate mathematical model for the temperature as a function of t?

**ChatGPT:** establish a mathematical model for the temperature as a function of time, you can consider using trigonometric functions to describe both daily and seasonal variations. Typically, temperature fluctuates in a sinusoidal or cosine-like manner, so you can employ a sine function to model it. Here is a potential model:

$$T(t) = A \sin\left(\frac{2\pi}{24}t + \phi\right) + B$$

Where: - $T(t)$ represents the temperature at time $t$. - $A$ is the amplitude, representing the maximum temperature variation. - $t$ is the time in hours. - $\frac{2\pi}{24}$ controls the periodicity, with $2\pi$ corresponding to a full day (24 hours). - $\phi$ is the phase, indicating the starting point of the temperature within a day. - $B$ is the average temperature, which can vary with the seasons.

You would need to determine the values of the parameters $A$, $\phi$, and $B$ based on actual data. Typically, these parameters can be obtained by fitting the model to historical temperature data. Once you have the appropriate parameters, you can use the model to predict the temperature at any given time.

Please note that this is a simplified model example, and actual situations may be more complex, requiring more data and consideration of factors such as geographical location, elevation, and so on."

**Bard:** An appropriate mathematical model for the air temperature just above the ground at a particular point on the Earth as a function of time t in hours is:

$$T(t) = T_a + T_d \sin(\frac{2\pi t}{24}) + T_y \sin(\frac{2\pi t}{365.25})$$

where:

- T(t) is the air temperature at time t in degrees Celsius
- $T_a$ is the annual mean temperature in degrees Celsius
- $T_d$ is the amplitude of the diurnal temperature variation in degrees Celsius
- $T_y$ is the amplitude of the annual temperature variation in degrees Celsius

The diurnal temperature variation can be modeled using a sinusoidal function with a period of 24 hours. The annual temperature variation can be modeled using a sinusoidal function with a period of 365.25 days. The total temperature is the sum of the annual mean temperature, the diurnal temperature variation, and the annual temperature variation.

The values of $T_a$, $T_d$, and $T_y$ will vary depending on the location on Earth. For example, the annual mean temperature will be higher at the equator than at the poles. The amplitude of the diurnal temperature variation will be greater on clear days than on cloudy days. The amplitude of the annual temperature variation will be greater at higher latitudes than at lower latitudes. This model can be used to predict the air temperature at a particular location and time, given the values of $T_a$, $T_d$, and $T_y$. It can also be used to study the effects of climate change on air temperature.

**New Bing:** "The temperature variation over a 24-hour cycle and an annual cycle can be modeled using a trigonometric function. A sinusoidal function can be used to model the temperature variation over a 24-hour cycle, while a cosine function can be used to model the temperature variation over an annual cycle 1. A possible mathematical model for the temperature as a function of time t in hours is:

$$T(t) = Asin(2\pi t/24) + Bcos(2\pi t/365) + C$$

where A, B, and C are constants that depend on the location and other factors 1. This model assumes that the temperature variation over a 24-hour cycle is sinusoidal, with a period of 24 hours, and that the temperature variation over an annual cycle is cosine, with a period of 365 days 1."

Why use the trigonometric function to model?

**ChatGPT:** The use of a trigonometric function like the sine function in the mathematical model for temperature variations is based on the observation that many natural phenomena, including temperature changes, exhibit periodic behavior that can be described using trigonometric functions. Here's why we use a sine function (which is a type of trigonometric function) in this context:

1. **Periodicity:** Trigonometric functions, such as sine and cosine, are naturally periodic. They repeat themselves over a regular interval, making them suitable for modeling phenomena that exhibit periodic behavior. In the case of temperature variations, we often see daily and seasonal patterns, both of which are inherently periodic.

2. **Smoothness:** Trigonometric functions are smooth and continuous, which makes them a good choice for modeling continuous physical phenomena like temperature changes. They provide a smooth curve that can capture gradual variations in temperature.

3. **Amplitude and Phase:** Trigonometric functions have parameters like amplitude (A) and phase ($\phi$), which allow us to adjust the height and phase shift of the periodic function. This flexibility enables us to fine-tune the model to fit the specific data we have.

4. **Widely Applicable:** Trigonometric functions are widely used in physics and engineering to model periodic phenomena. They have a strong theoretical foundation and have been successfully applied to describe various natural phenomena, including oscillations, waves, and cyclical variations.

5. **Empirical Evidence:** Through empirical observation and data analysis, it has been found that temperature variations often follow sinusoidal patterns. Therefore, using a sine function is often a reasonable first approximation.

6. **Widely Applicable:** Trigonometric functions are widely used in physics and engineering to model periodic phenomena. They have a strong theoretical foundation and have been successfully applied to describe various natural phenomena, including oscillations, waves, and cyclical variations.

While a sine function is a common choice for modeling temperature variations, it's important to note that the actual behavior of temperature can be influenced by numerous factors, and more complex models may be needed to account for additional variables and irregularities in the data. The choice of using a sine function is often a simplification that provides a reasonable starting point for analysis.

Analysis: After analyzing the three answers of Chatgpt, Bard, and Bing, we found that all three models have some things in common. All three answers use trigonometric functions to describe changes in the mathematical model of temperature as a function of time. These three answers are simulated by sine functions. In addition, the three measure " t " in hours, so it makes sense to multiply the number of days by the number of hours in a day to keep the graph even. Admittedly, the existence of all three is reasonable, However, when compared, Bard and Bing's answer is better, because ChatGPT's answer does not give a yearly model.

## 2 Question 2

We have a 20 litre bucket that has a hole in the bottom. Every 10 minutes one litre of water flows out of the hole. There is a tap filling the bucket that adds 1 litre of water every 5 minutes. If the bucket starts empty, how long does it take for the bucket to fill?

**ChatGPT:** To solve this problem, we can set up an equation to track the net rate at which water enters or leaves the bucket. We know that water flows out of the hole at a rate of 1 liter every 10 minutes and water is added to the bucket at a rate of 1 liter every 5 minutes. Therefore, the net rate of water being added to the bucket is:

Net rate = (rate of filling) - (rate of draining)
Net rate = (1 liter per 5 minutes) - (1 liter per 10 minutes)
Net rate = 0.2 liters per minute - 0.1 liters per minute
Net rate = 0.1 liters per minute
This means that every minute, the bucket gains 0.1 liters of water.

To find out how long it takes for the bucket to fill, we need to figure out how many minutes it will take for the bucket to accumulate 20 liters of water (its capacity).

Time = (Volume of the bucket) / (Net rate)
Time = 20 liters / 0.1 liters per minute
Time = 200 minutes
So, it will take 200 minutes for the bucket to fill completely.

**Bard:** Let's assume that the amount of water in the bucket at time t is denoted by V(t). We can then model the rate of change of V(t) using the following equation: dV/dt = inflow - outflow

where: inflow is the rate at which water is flowing into the bucket outflow is the rate at which water is flowing out of the bucket In this case, the inflow is 1 litre per 5 minutes, so we can write: inflow = 1/5 litres per minute

The outflow is 1 litre per 10 minutes, so we can write: outflow = 1/10 litres per minute

Substituting these values into the equation above, we get: dV/dt = 1/5 - 1/10 = 1/10 litres per minute

This equation tells us that the amount of water in the bucket is increasing at a rate of 1/10 litres per minute. We can now use this equation to solve for the time it takes for the bucket to fill. To do this, we need to integrate the equation above. This gives us:

$$V(t) = V_0 + (1/10) * t$$

where $V_0$ is the amount of water in the bucket at time t = 0. In this case, the bucket is empty at time t = 0, so $V_0 = 0$. We can now set V(t) = 20 litres (the capacity of the bucket) and solve for t: 20 = 0 + (1/10) * t

t = 200 minutes

Therefore, it takes 200 minutes for the bucket to fill.

**New Bing:** "Let's first calculate the rate at which water is flowing out of the bucket. Since 1 litre of water flows out every 10 minutes, the rate of water flowing out is 0.1 litres per minute. Now, let's calculate the rate at which water is flowing into the bucket. Since 1 litre of water is added every 5 minutes, the rate of water flowing in is 0.2 litres per minute. Since the rate of water flowing in is greater than the rate of water flowing out, the bucket will eventually fill up. We can use the following formula to calculate how long it takes for the bucket to fill up: time = volume / (rate in - rate out) where volume is the volume of the bucket, rate in is the rate at which water flows into the bucket, and rate out is the rate at which water flows out of the bucket. Substituting the values we calculated earlier, we get: time = 20 / (0.2 - 0.1) = 200 minutes Therefore, it will take 200 minutes for the bucket to fill up."

Analysis: ChatGpt, Bard, and Bing give the same answer to this question, and are quite right. But mathematical formulas can never fully reflect the complex reality of the situation, and we need to consider them in terms of real life. For example, we need to take into account that the rate at which water flows out of a bucket is unlikely to be consistent throughout the filling, and that the pressure will increase accordingly. Moreover, the bucket will only fill temporarily, and after this brief equilibrium, it will either overflow if the tap remains open, or empty if the tap is turned off.

## 3 Question 3

What is the derivative of $f(x) = 5x(x-1) - 5x^2$?

**ChatGPT:** To find the derivative of the function $f(x) = 5x(x-1) - 5x^2$, you can apply the rules of differentiation. Here's how to do it step by step:

1. Expand the terms within the parentheses: $f(x) = 5x^2 - 5x - 5x^2$
2. Now, differentiate each term separately with respect to $x$: $\frac{d}{dx}(5x^2) - \frac{d}{dx}(5x) - \frac{d}{dx}(5x^2)$
3. Apply the power rule of differentiation, which states that $\frac{d}{dx}(x^n) = nx^{n-1}$, where $n$ is a constant:
4. Simplify the expression by combining like terms: $10x - 10x - 5$
5. The $10x$ and $-10x$ terms cancel each other out, leaving: $-5$

So, the derivative of the function $f(x) = 5x(x-1) - 5x^2$ is $-5$.

**Bard:** To find the derivative of $f(x) = 5x(x-1) - 5x^2$, we can use the product rule and the power rule. The product rule states that the derivative of the product of two functions u(x) and v(x) is: d/dx[u(x)v(x)] = u'(x)v(x) + u(x)v'(x)

The power rule states that the derivative of $x^n$ is $nx^{n-1}$. Let's differentiate f(x) using these rules:

$$f'(x) = \frac{d}{dx}[5x(x-1) - 5x^2]$$

Therefore, the derivative of f(x) is $10x - 5$.

**New Bing:** "The derivative of the function $f(x) = 5x(x-1) - 5x^2$ can be calculated using the product rule and the power rule of differentiation. The product rule states that if f(x) = u(x) * v(x), then f'(x) = u'(x) * v(x) + u(x) * v'(x). Using the product rule, we get: $f'(x) = (5x)'(x-1) + 5x(x-1)' - (5x^2)$' Applying the power rule, we get: f'(x) = (5(x-1)) + 5x(1) - (10x) Simplifying, we get: f'(x) = 5x - 5 Therefore, the derivative of $f(x)$ is $f'(x) = 5x - 5$."

Analysis: ChatGPT's answer is correct, Bard and Bing both give the wrong answer. We think that the product rule should not be used in this problem, but the rules of differentiation should be applied. Because in this case the expression is not expanded, and there is an addition after the multiplication. Therefore, ChatGpt's answer is the correct one.

**DIFFERENTIATION, OPTIMISATION AND INTEGRATION**

INTRODUCTION TO MATHEMATICS IN DATA SCIENCE

## 1. Differentiation

The main purpose of the derivative is to understand *the (instantaneous) rate of change* of a given function with respect to its variable. How would one define and compute such a thing?
A good example to have in mind when thinking of the derivative is *velocity*. Simply put, the velocity is defined as

$$\text{velocity} = \frac{\text{distance traversed}}{\text{time it took us to traverse that distance}}.$$

For example: If it took us 2 hours to traverse 100ml we would say that we have driven *with an average velocity* of 50mph. That, however, doesn't tell us the velocity at each given point of our journey, which might be very important if we want to make sure that we didn't break any speed limits. How then, could we compute our velocity at a certain given time?
Say that $x(t)$ is our position at time $t$, and we're interested in computing the velocity[1] at a given time $t_0$. Assuming that we are at times that are *very close to* $t_0$ we "expect" that our velocity is "almost constant" and equals the velocity at time $t_0$, $v(t_0)$. This means that if $t$ is very close to $t_0$

$$x(t) \approx \underbrace{x(t_0)}_{\text{position at time } t_0} + \underbrace{v(t_0)\,(t-t_0)}_{\text{additional distance traversed in time } t-t_0}$$

or in other words

$$\frac{x(t)-x(t_0)}{t-t_0} \approx v(t_0).$$

The above approximation becomes more and more accurate as $t$ gets closer and closer to $t_0$ *but does not touch it.* Thus, if we want to be able to define the derivative precisely we must *take $t$ to $t_0$*, or in other words: Use the notion of a limit for the appropriate "average rate".
This is exactly the idea of how to define the rate of change of a given function (position or not).

[1]The velocity can be negative, indicating that we go backwards. The speed, which is the absolute value of the velocity, is always non-negative.

**Definition 1.** Let $f : (a,b) \to \mathbb{R}$ be a given function, and let $x_0$ be a point in $(a,b)$. We say that $f$ is differentiable at the point $x_0$ if

$$\lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0}$$

exists and is finite. If that happens, we denote the limit by $f'(x_0)$ and call it *the derivative of* $f$ at the point $x_0$.

*Remark* 1. Looking at the definition, and how we conceived it, we can actually give a geometric interpretation to the derivative. Much like with velocity, a function that is differentiable at $x_0$ satisfies

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

The above means that we can approximate the function $f$ at the point $x_0$ with the line

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

which passes through the point $(x_0, f(x_0))$. such a line is called *the tangent line* of $f$ at the point $x_0$, and the derivative, $f'(x_0)$, is exactly its slope!

Now that we have defined the derivative, we ask ourselves if we can find techniques that will help us compute it without using the definition. The next theorem, and the table that follows it, will be our main tools to compute derivatives in this module.

**Theorem 1.1** (Rules of Differentiation)**.** *Let $f$ and $g$ be real valued functions. Then*

*(i) (Linearity) If $f$ and $g$ are differentiable at $x_0$, then for any $\alpha, \beta \in \mathbb{R}$ the function $\alpha f(x) + \beta g(x)$ is also differentiable at $x_0$ and*

$$(\alpha f + \beta g)'(x_0) = \alpha f'(x_0) + \beta g'(x_0).$$

*(ii) (Product Rule) If $f$ and $g$ are differentiable at $x_0$, then the function $h(x) = f(x) \cdot g(x)$ is also differentiable at $x_0$ and*

$$h'(x_0) = (f \cdot g)'(x_0) = f'(x_0)g(x_0) + g'(x_0)f(x_0).$$

*(iii) (Quotient Rule) If $f$ and $g$ are differentiable at $x_0$ and $g(x_0) \neq 0$, then the function $h(x) = \frac{f(x)}{g(x)}$ is also differentiable at $x_0$ and*

$$h'(x_0) = \left(\frac{f}{g}\right)'(x_0) = \frac{f'(x_0)g(x_0) - g'(x_0)f(x_0)}{\left(g(x_0)\right)^2}.$$

*(iv) (Chain Rule) If $g$ is differentiable at $x_0$ and $f$ is differentiable at $g(x_0)$, then the function $h(x) = (f \circ g)(x) = f(g(x))$ is also differentiable at $x_0$ and*

$$h'(x_0) = (f \circ g)'(x_0) = f'(g(x_0))\, g'(x_0).$$

Table of Derivatives

| Function | Derivative |
|---|---|
| $f(x) = C,\ C \in \mathbb{R}$ | $f'(x) = 0$ |
| $f(x) = x$ | $f'(x) = 1$ |
| $f(x) = x^n,\ n \in \mathbb{N}$ | $f'(x) = nx^{n-1}$ |
| $f(x) = x^\alpha,\ \alpha \in \mathbb{R}$ | $\alpha x^{\alpha-1}$ |
| $f(x) = \sin(x)$ | $f'(x) = \cos(x)$ |
| $f(x) = \cos(x)$ | $f'(x) = -\sin(x)$ |
| $f(x) = \tan(x)$ | $f'(x) = \frac{1}{\cos^2(x)}$ |
| $f(x) = \cot(x)$ | $f'(x) = -\frac{1}{\sin^2(x)}$ |
| $f(x) = e^x$ | $f'(x) = e^x$ |
| $f(x) = a^x,\ a > 0$ | $f'(x) = a^x \ln(a)$ |
| $f(x) = \ln(x)$ | $f'(x) = \frac{1}{x}$ |

**Example 1.** Find the derivatives of the following functions:

a) $f(x) = x^2 + 3x + 2$.
b) $f(x) = \frac{x}{x-1}$.
c) $f(x) = x^4 + \sin(x)$.
d) $f(x) = \frac{1}{2+\cos(x)}$.
e) $f(x) = x^4 \sin(x)$.
f) $f(x) = \frac{2-\sin(x)}{2+\sin(x)}$.

*Solution.* a) Using the linearity of the derivative we find that

$$f'(x) = 2x + 3 + 0 = 2x + 3.$$

b) Using the quotient rule we find that

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

c) Using the linearity of the derivative we find that

$$f'(x) = 4x^3 + \cos(x) + 3 + 0 = 2x + 3.$$

d) Using the quotient rule we find that

$$f'(x) = \frac{0 \cdot (2 + \cos(x)) - 1 \cdot (0 - \sin(x))}{(2 + \cos(x))^2} = \frac{\sin(x)}{(2 + \cos(x))^2}.$$

e) using the product rule we find that

$$f'(x) = 4x^3 \cdot \sin(x) + x^4 \cdot \cos(x).$$

f) Using the quotient rule we find that

$$f'(x) = \frac{(2 - \cos(x))(2 + \sin(x)) - (2 - \sin(x))(2 + \cos(x))}{(2 + \sin(x))^2}$$

$$= \frac{4-2\cos(x)+2\sin(x)-\sin(x)\cos(x)-(4-2\sin(x)+2\cos(x)-\sin(x)\cos(x))}{(2+\sin(x))^2}$$

$$= \frac{4\left(\sin(x)-\cos(x)\right)}{(2+\sin(x))^2}$$

□

**Example 2.** Find the derivatives of the following functions:

a) $f(x) = \cos(2x) - \sin x$.
b) $f(x) = \sqrt{1+x^2}$.
c) $f(x) = \left(2-x^2\right)\cos\left(x^2\right)$.
d) $f(x) = \sin(\sin(x))$.
e) $f(x) = \frac{x}{\sqrt{4-x^2}}$.

*Solution.* a) We start by differentiating $\cos(2x)$. Denoting by $h(x) = \cos(x)$ and $g(x) = 2x$ we see that

$$\cos(2x) = h\left(g(x)\right)$$

and as such, using the chain rule, we find that

$$(\cos(2x))' = h'(g(x))g'(x) = -\sin(2x)\cdot 2 = -2\sin(2x).$$

Thus, with the linearity of the derivative we find that

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

b) Denoting by $h(x) = \sqrt{x} = x^{\frac{1}{2}}$ and $g(x) = 1+x^2$ we see that

$$f(x) = h\left(g(x)\right)$$

and as such, using the chain rule, we find that

$$f'(x) = h'(g(x))g'(x) = \frac{1}{2}\left(1+x^2\right)^{-\frac{1}{2}}\cdot 2x = \frac{x}{\sqrt{1+x^2}}.$$

c) We start by differentiating $\cos\left(x^2\right)$. Denoting by $h(x) = \cos(x)$ and $g(x) = x^2$ we see that

$$\cos\left(x^2\right) = h\left(g(x)\right)$$

and as such, using the chain rule, we find that

$$\left(\cos\left(x^2\right)\right)' = h'(g(x))g'(x) = -\sin\left(x^2\right)\cdot 2x = -2x\sin\left(x^2\right).$$

Using the product rule of we find that

$$f'(x) = -2x\cdot\cos\left(x^2\right)+\left(2-x^2\right)\cdot\left(-2x\sin\left(x^2\right)\right) = 2x^3\sin\left(x^2\right)-4x\sin\left(x^2\right)-2x\cos\left(x^2\right).$$

d) Denoting by $h(x) = \sin(x)$ and $g(x) = \sin(x)$ we see that

$$f(x) = h\big(g(x)\big)$$

and as such, using the chain rule, we find that

$$f'(x) = h'(g(x))g'(x) = \cos(\sin(x))\cos(x).$$

e) We start by differentiating $\sqrt{4-x^2}$. Denoting by $h(x) = \sqrt{x} = x^{\frac{1}{2}}$ and $g(x) = 4 - x^2$ we see that

$$\sqrt{4-x^2} = h\big(g(x)\big)$$

and as such, using the chain rule, we find that

$$\left(\sqrt{4-x^2}\right)' = h'(g(x))g'(x) = \frac{1}{2}\left(4-x^2\right)^{-\frac{1}{2}} \cdot (-2x) = -\frac{x}{\sqrt{4-x^2}}.$$

Using the quotient rule we find that

$$f'(x) = \frac{1 \cdot \sqrt{4-x^2} - x \cdot \left(-\frac{x}{\sqrt{4-x^2}}\right)}{4-x^2} = \frac{\left(4-x^2\right)+x^2}{\sqrt{4-x^2}\left(4-x^2\right)} = \frac{4}{\left(4-x^2\right)^{\frac{3}{2}}}.$$

□

**Example 3.** Find the derivatives of the following functions:

a) $f(x) = e^{3x-1}$.
b) $f(x) = e^{4x^2}$.
c) $f(x) = e^{\sqrt{x}}$.
d) $f(x) = \ln(1+x^2)$.
e) $f(x) = \ln\left(\sqrt{1+x^2}\right)$.
f) $f(x) = \ln(\ln(x))$.

a) Using the chain rule we have that

$$f'(x) = e^{3x-1}(3x-1)' = 3e^{3x-1}.$$

b) Using the chain rule we have that

$$f'(x) = e^{4x^2}\left(4x^2\right)' = 8xe^{4x^2}.$$

c) Using the chain rule we have that

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

d) Using the chain rule we have that

$$f'(x) = \frac{1}{1+x^2} \cdot \left(1+x^2\right)' = \frac{2x}{1+x^2}.$$

e) We can use the chain rule with the functions $h(x) = \ln(x)$ and $g(x) = \sqrt{1+x^2}$ but remembering that

$$\ln\left(a^b\right) = b\ln a$$

we see that

$$f(x) = \ln\left(\left(1+x^2\right)^{\frac{1}{2}}\right) = \frac{\ln\left(1+x^2\right)}{2}.$$

Thus, using the previous part of the question and the linearity of the derivative we find that

$$f'(x) = \frac{1}{2}\cdot\frac{2x}{1+x^2} = \frac{x}{1+x^2}.$$

f) Using the chain rule we have that

$$f'(x) = \frac{1}{\ln(x)}\left(\ln(x)\right)' = \frac{1}{\ln(x)}\cdot\frac{1}{x} = \frac{1}{x\ln(x)}.$$

## 2. Optimisation

The derivative is an essential tool in investigating the local, and global, minimum and maximum of a given function, defined as follows:

**Definition 2.** Let $f$ be a given function on a interval (possibly infinite) $I$. We say that $x_0$ is a local minimum of $f$ if there exists a neighbourhood of $x_0$, $U \subset I$ where

$$f(x_0) \leq f(x) \quad \text{for all } x \in U.$$

$x_0$ is called a global minimum of $f$ on $I$ if

$$f(x_0) \leq f(x) \quad \text{for all } x \in I.$$

Similarly, one can define a local and global maximum by demanding that $f(x_0) \geq f(x)$ in an appropriate neighbourhood $U$ (local) or the entire interval $I$ (global).

The key connections between the derivative and the investigation of the etrema of the function is given in the following theorem:

**Theorem 2.1.** *Let $f: I \to \mathbb{R}$ be differentiable. Then*

*(i) If $x_0$ is an local etremum (minimum or maximum) then $f'(x_0) = 0$.*
*(ii) If $f'(x) > 0$ in an interval $(a,b)$ then $f$ is strictly increasing in $(a,b)$, i.e. if $x_1 > x_2$ then $f(x_1) > f(x_2)$.*
*(iii) If $f'(x) \geq 0$ in an interval $(a,b)$ then $f$ is increasing (or non-decreasing) in $(a,b)$, i.e. if $x_1 > x_2$ then $f(x_1) \geq f(x_2)$.*
*(iv) If $f'(x) < 0$ in an interval $(a,b)$ then $f$ is strictly decreasing in $(a,b)$, i.e. if $x_1 > x_2$ then $f(x_1) < f(x_2)$.*
*(v) If $f'(x) \leq 0$ in an interval $(a,b)$ then $f$ is decreasing (or non-increasing) in $(a,b)$, i.e. if $x_1 > x_2$ then $f(x_1) \leq f(x_2)$.*

**Corollary 2.2.** When searching for a local extrema for a differentiable function $f$, we must solve the equation $f'(x) = 0$.

How do we search for a local extrema?

**Step 1:** Find all the candidates for local extrema by solving the equation $f'(x) = 0$.
**Step 2:** Draw the interval $I$ as a line, and add to it the points you've found in the previous step and points where one can't differentiate (we won't see such points in out module).
**Step 3:** Choose a random point in between two points on the plotted line and check the sign of the derivative in it. Mark $+$ or $-$ on the line to indicate it sign. A $+$ sign implies that the function is strictly increasing in the interval between these points, while a $-$ sign implies that the function is strictly decreasing between them.
**Step 4:** A point is a local minimum if the sign on the plotted line switches from $-$ to $+$, and a local maximum if the sign on the plotted line switches from $+$ to $-$.

The study of global extrema is a bit more involved, as we may need to consider the boundaries of the interval $I$. The process is as follows:

- Find all local extrema.
- Add the boundary points of the interval that are in the interval.
- Compare the values of the function on the aforementioned points, and the limit of the functions on the boundaries of the interval that are *not* in the interval (including infinity). If the minimal or maximal of these values is attained in a local extrema or a boundary point that is in the interval - it is a global extrema. Else, there is no global minimum or maximum, respectively.

*Remark* 2. A function that is differentiable on $[a, b]$, where $a < b$ are finite real numbers, will always have a global minimum and maximum. This is a corollary of an important theorem in Calculus. As such you know that in order to find global extrema on a closed bounded interval all you need to do is to

- Solve the equation $f'(x) = 0$.
- Compare the values of $f$ on these points to $f(a)$ and $f(b)$, and find the points that give the maximum and minimum.

**Example 4.** Consider the function $f(x) = x^4 - 2x^2$.

a) Find all local minimum and maximum of $f$ on $\mathbb{R}$.
b) Find the global minimum and maximum of $f$ on $[-2, 2]$.

*Solution.* We won't draw the points on a line in these notes, but instead will investigate the appropriate intervals between the suspected points.

a) We start by solving the equation $f'(x) = 0$. Since
$$f'(x) = 4x^3 - 4x = 4x\left(x^2 - 1\right)$$
we see that
$$f'(x) = 0 \quad \Leftrightarrow \quad x = 0, \text{ or } x^2 = 1.$$
Thus, our candidates for local extrema are $x = 0, -1, 1$.
Let us investigate the intervals between these points:
   - $x < -1$: For $x = -2$ we have that $f'(-2) = -8(4-1) < 0$. The associated sign is $-$, and we know that $f$ is strictly decreasing in this domain.
   - $-1 < x < 0$: For $x = -\frac{1}{2}$ we have that $f'\left(-\frac{1}{2}\right) = -2\left(\frac{1}{4} - 1\right) > 0$. The associated sign is $+$, and we know that $f$ is strictly increasing in this domain.
   - $0 < x < 1$: For $x = \frac{1}{2}$ we have that $f'\left(\frac{1}{2}\right) = 2\left(\frac{1}{4} - 1\right) < 0$. The associated sign is $-$, and we know that $f$ is strictly decreasing in this domain.
   - $x > 1$: For $x = 2$ we have that $f'(2) = 8(4-1) > 0$. The associated sign is $+$, and we know that $f$ is strictly increasing in this domain.

   From the above we conclude that $x = -1$ and $x = 1$ are a local minimums (the sign changes from $-$ to $+$) and $x = 0$ is a local maximum (the sign changes from $+$ to $-$).

b) To find the global extrema in the interval we compare the values of $f$ on $0, -1$ and $1$ with the values of $f$ on the boundary points $-2$ and $2$.
$$f(-2) = 8, \quad f(-1) = -1, \quad f(0) = 0, \quad f(1) = -1, \quad f(2) = 8.$$
Thus our global maximum is attained at $x = -2$ and $x = 2$, and equals 8, and our global minimum is attained at $x = -1$ and $x = 1$ and equals $-1$.

□

**Example 5.** A group of miners are trapped underground at a depth of 300 metres. A rescue team starts at the bottom of an abandoned mine shaft that is 600 metres West of the trapped miners and has a depth of 100 metres. The rescue team decides to start digging horizontally towards the East, and then to dig directly towards the trapped miners (potentially in diagonally). At a depth of 100 metres the rock is soft and it takes only 5 minutes to dig one horizontal metre. However, at any depth below this, the rock is hard and it takes 13 minutes to dig a distance of one metre. Calculate the minimal number of hours that it takes to tunnel to the trapped miners.

*Solution.* Let us denote by $x$ the length of the horizontal tunnel the rescue team has dug to the East before starting to dig directly towards the miners. $x$ must be in $[0,600]$. Once the rescue team has dug this horizontal tunnel, they are located $600-x$ metres West of the miners, and 200 metres above them. The distance between the team and the miners is thus $\sqrt{(600-x)^2+200^2}$. As the time it took the rescue team to dig the first stretch is $5x$ minutes (soft rock), and the time it took it to dig the second stretch is $13\sqrt{(600-x)^2+200^2}$ minutes (hard rock) we see that the total time in minutes the team must dig is

$$T(x)=5x+13\sqrt{(600-x)^2+200^2}.$$

We want to find the minimum of $T(x)$ over $[0,600]$. Differentiating $T(x)$ we see that

$$T'(x)=5+\frac{13}{2\sqrt{(600-x)^2+200^2}}\cdot(-2(600-x)).$$

The equation $T'(x)=0$ is equivalent to

$$\frac{13(600-x)}{\sqrt{(600-x)^2+200^2}}=5$$

or

$$13(600-x)=5\sqrt{\left(600-x^2\right)+200^2}.$$

Squaring the above we get that

$$169(600-x)^2=25(600-x)^2+25\cdot 200^2$$

or $144(600-x)^2=5^2\cdot 200^2$. Since $600-x\geq 0$ we conclude that

$$(600-x)=\sqrt{\frac{5^2\cdot 200^2}{12^2}}=\frac{5\cdot 200}{12}=\frac{250}{3}.$$

Our potential minimums are $x=0$, $x=600-\frac{250}{3}=\frac{1550}{3}$ and $x=600$ As

$$T(600)=5600,\quad T\left(\frac{1550}{3}\right)=5400,\quad T(0)=2600\cdot\sqrt{10}>5400,$$

we see that the minimal time the rescue team will dig is 5400 minutes, i.e. 90 hours. $\square$

## 3. Integration Part I - The Indefinite Integral (or Anti Derivative)

The indefinite integral is nothing more than the reverse of differentiation. Denoted by $\int f(x)dx$, the indefinite integral answers the question

*Which function has $f$ as its derivative?*

A function whose derivative is $f$ is called *a primitive function* of $f$. The indefinite integral of $f$, also known as the anti derivative of $f$, gives us the family of primitive functions of $f$. A natural question is: Is it well defined? The answer to this question lies with the next theorem

**Theorem 3.1.** *If $F$ and $G$ are primitive functions of $f$ then there exists a constant $C \in \mathbb{R}$ such that $F = G + C$.*

The above theorem gives us the tools to find the anti derivative of $f$:

- Find a primitive function to $f$, $F$.
- The anti derivative of $f$, $\int f(x)dx$, equals to the family $\{F(x) + C \mid C \in \mathbb{R}\}$. We will write
$$\int f(x)dx = F(x) + C$$
to represent this family.

The above brings into light one big issue with the anti derivative: It is not unique. A more substantial issue with it is its complexity. While differentiation is quite straight forward, we can't compute explicitly most anti derivatives.
Much like differentiation, however, and in fact stemming from the appropriate rules for differentiation, the anti derivative does enjoy some properties that will help us compute it:

**Theorem 3.2** (Rules of Integration)**.** *Let $f$ and $g$ be real valued functions. Then*

*(i) (Linearity) For any $\alpha, \beta \in \mathbb{R}$*
$$\int \left(\alpha f(x) + \beta g(x)\right) dx = \alpha \int f(x)dx + \beta \int g(x)dx.$$
*(ii) (Integration by Parts {reverse product rule})*
$$\int f'(x)g(x)dx = f(x)g(x) - \int f(x)g'(x)dx.$$
*(iii) ($u-$substitution {reverse chain rule})*
$$\int f\left(g(x)\right)g'(x)dx\Big|_{\text{at the point } x} = \int f(u)du\Big|_{\text{at the point } u=g(x)}.$$

We also have the following table to assist us

Table of Integration

| Function | Anti Derivative |
|---|---|
| $f(x) = 1$ | $\int f(x)dx = x + C$ |
| $f(x) = x^n$ | $\int f(x)dx = \frac{x^{n+1}}{n+1} + C$ |
| $f(x) = x^\alpha,\ \alpha \neq -1$ | $\int f(x)dx = \frac{x^{\alpha+1}}{\alpha+1} + C$ |
| $f(x) = \sin(x)$ | $\int f(x)dx = -\cos(x) + C$ |
| $f(x) = \cos(x)$ | $\int f(x)dx = \sin(x) + C$ |
| $f(x) = \frac{1}{\cos^2(x)}(x)$ | $\int f(x)dx = \tan(x) + C$ |
| $f(x) = -\frac{}{\sin^2(x)}$ | $\int f(x)dx = \cot(x) + C$ |
| $f(x) = e^x$ | $\int f(x)dx = e^x + C$ |
| $f(x) = a^x,\ a > 0,\ a \neq 1$ | $\int f(x)dx = \frac{a^x}{\ln(a)} + C$ |
| $f(x) = \frac{1}{x}$ | $\int f(x)dx = \ln\lvert x\rvert + C$ |

A few very useful formulas, stemming from the above and $u-$substitution, are

$$\int u(x)^\alpha u'(x)dx = \begin{cases} \frac{u(x)^{\alpha+1}}{\alpha+1} + C & \alpha \neq -1 \\ \ln|u(x)| + C & \alpha = -1 \end{cases}$$

$$\int \sin(u(x))\, u'(x)dx = -\cos(u(x)) + C$$

$$\int \cos(u(x))\, u'(x)dx = \sin(u(x)) + C$$

$$\int e^{u(x)} u'(x)dx = e^{u(x)} + C$$

$$\int a^{u(x)} u'(x)dx = \frac{a^{u(x)}}{\ln a} + C, \quad \text{if } a > 0 \text{ and } a \neq 1.$$

These expressions are all particular cases of the following formula:

$$\int f'(u(x))\, u'(x)dx = f(u(x)) + C.$$

**Example 6.** Compute the following integrals:

a) $\int \left(x^2 + 1\right) dx$.
b) $\int (x+1)\left(x^3 - 2\right) dx$.
c) $\int \left(1 + \sqrt{x}\right)^2 dx \quad$ (for $x > 0$).

*Solution.* a) Due to linearity we see that

$$\int \left(x^2 + 1\right) dx = \int x^2 dx + \int 1\, dx = \frac{x^3}{3} + x + C.$$

b) Using linearity again we find that

$$\int (x+1)\left(x^3 - 2\right) dx = \int \left(x^4 + x^3 - 2x - 2\right) dx = \frac{x^5}{5} + \frac{x^4}{4} - x^2 - 2x + C.$$

c) Again, the linearity of the anti derivative implies that

$$\int \left(1+\sqrt{x}\right)^2 dx = \int \left(1+2\sqrt{x}+x\right) dx = x + 2\frac{x^{\frac{3}{2}}}{\frac{3}{2}} + \frac{x^2}{2} + C$$

$$= x + \frac{4x^{\frac{3}{2}}}{3} + \frac{x^2}{2} + C.$$

$\square$

**How does one use integration by parts?**

- identify a function which we would like to differentiate to simplify our expression This function will play the role of $g(x)$. The remaining function will play the role of $f'(x)$.
- Find *any* choice for a primitive of $f'$, $f$.
- Use the integration by parts formula and hope we get something we can solve[2].

**Example 7.** Find the following integrals using integration by parts:

a) $\int x \sin(x)\, dx$.
b) $\int x e^x\, dx$.
c) $\int \ln(x)\, dx$.
d) (*) $\int x \ln(x)\, dx$.

*Solution.* a) As the derivative of $x$ is 1, we would like to define $g(x) = x$. Consequently $f'(x) = \sin(x)$. As a primitive to $f'(x)$ we choose $f(x) = -\cos(x)$, and naturally $g'(x) = 1$. Thus

$$\int x \sin(x)\, dx = x(-\cos(x)) - \int 1 \cdot (-\cos(x))\, dx$$

$$= -x\cos(x) + \int \cos(x) dx = -x\cos(x) + \sin(x) + C.$$

b) Similar to the previous problem, we choose $f'(x) = e^x$ and $g(x) = x$. Then, a choice of primitive to $f'(x)$ is $f(x) = e^x$ and since $g'(x) = 1$ we have that

$$\int x e^x dx = x e^x - \int 1 \cdot e^x dx = x e^x - \int e^x dx = x e^x - e^x + C.$$

c) This integration seems more complicated to compute with integration by parts as we only have one function in the integral! This can be rectified by noticing that

$$\ln(x) = 1 \cdot \ln(x) = (x)' \ln(x).$$

[2]There are cases, which we won't discuss in these notes, where using integration by parts results in a recursive formula.

This time our choice is motivated by the identification of the derivative. Defining $f'(x) = 1$ and $g(x) = \ln(x)$, we have that a possible primitive to $f'(x)$ is $f(x) = x$ and $g'(x) = \frac{1}{x}$. Thus

$$\int \ln(x)\,dx = x\ln(x) - \int x \cdot \frac{1}{x} dx = x\ln(x) - \int 1\,dx = x\ln(x) - x + C.$$

d) This time the function that we would like to differentiate is $g(x) = \ln(x)$ as $g'(x) = \frac{1}{x}$. Thus $f'(x) = x$ and a possible primitive to $f'(x)$ is $f(x) = \frac{x^2}{2}$.Thus

$$\int x\ln(x)\,dx = \frac{x^2}{2} \cdot \ln(x) - \int \frac{x^2}{2} \cdot \frac{1}{x} dx = \frac{x^2\ln(x)}{2} - \frac{1}{2}\int x dx.$$

$$= \frac{x^2\ln(x)}{2} - \frac{x^2}{4} + C.$$

□

**How does one uses $u-$substitution?**

- Identify a function that seems to "complicate" the integration, and name it $u(x)$.
- Find $u'(x)$ and using the symbolic notation $du = u'(x)dx$ get rid of the symbolic $dx$.
- Write the remaining functions with $u$ - "eliminating" all the $x-$s.
- Integrate in $u$, and insert the expression for $u(x)$ in the final result.
- If possible *always* choose $u(x)$ in such a way that you can recognise $u'(x)dx$ in the integrand.

**Example 8.** Find

a) $\int \frac{\sin x}{\sqrt{\cos^3 x}} dx$.

b) $\int \frac{x^5}{\sqrt{1-x^6}} dx$.

c) $\int \frac{\sin(\sqrt{x+1})}{\sqrt{x+1}} dx$.

*Solution.* a) Recognising $\sin(x)$ as the almost derivative of $\cos(x)$ we define $u(x) = \cos(x)$. We have that $u'(x) = -\sin(x)dx$ and with the symbolic notation

$$du = -\sin(x)dx$$

we see that

$$\int \frac{\sin x}{\sqrt{\cos^3 x}} dx = \int \frac{-du}{\sqrt{u^3}} = -\int u^{-\frac{3}{2}} du = -\frac{u^{-\frac{1}{2}}}{-\frac{1}{2}}$$

$$= \frac{2}{\sqrt{u}} + C \underset{\text{bring } x \text{ back}}{=} \frac{2}{\sqrt{\cos(x)}} + C.$$

b) Recognising $x^5$ as almost the derivative of $1-x^6$ we define $u(x) = 1-x^6$. When have that $u'(x) = -6x^5$ and with the symbolic notation

$$du = -6x^5 dx$$

we see that

$$\int \frac{x^5}{\sqrt{1-x^6}} dx = \int \frac{-\frac{du}{6}}{\sqrt{u}} = -\frac{1}{6}\int u^{-\frac{1}{2}} du = -\frac{1}{6}\frac{u^{\frac{1}{2}}}{\frac{1}{2}}$$

$$= -\frac{\sqrt{u}}{3} + C \underset{\text{bring } x \text{ back}}{=} -\frac{\sqrt{1-x^6}}{3} + C.$$

c) A derivative is not so easily recognisable here. Therefore, we will take the function that seems to "complicate" matters, $\sqrt{1+x}$, and call it $u(x)$. We have that

$$u'(x) = \frac{1}{2\sqrt{1+x}}$$

meaning that

$$du = \frac{1}{2\sqrt{1+x}} dx.$$

Noting that the right hand side appears in our integral, up to a factor of 2, we find that

$$\int \frac{\sin\left(\sqrt{x+1}\right)}{\sqrt{x+1}} dx = \int \sin(u) \cdot 2du = 2\int \sin(u) du = -2\cos(u) du + C$$

$$\underset{\text{bring } x \text{ back}}{=} -2\cos\left(\sqrt{1+x}\right) + C.$$

□

## 4. Integration Part II - The Definite Integral

The definite integral of a function $f$ on the interval $[a,b]$, denoted by $\int_a^b f(x)dx$, answers the following question:

*What is the area ,with allowed negative heights, "under" the graph of the function $f$ over the interval $[a,b]$?*

At first glance, the definite integral seems completely unrelated to the indefinite integral. A beautiful connection exists, however, and is given in the following theorem:

**Theorem 4.1.** *Let $F$ be a primitive function of $f$ over $[a,b]$. Then*

$$\int_a^b f(x)dx = F(b) - F(a).$$

It is worth to note, even though we won't use it in our module, that the above is a consequence of the following pivotal theorem:

**Theorem 4.2** (The Fundamental Theorem of Calculus)**.** *Let $f$ be an integral function on $[a,b]$. Then the function*

$$F(x)=\int_a^x f(t)dt$$

*is a primitive function for $f$ on $(a,b)$.*

Most of the anti derivative rules we've learned transfer seamlessly to the definite integral in the following way:

**Theorem 4.3.** *Let $f$ and $g$ be real valued functions. Then*

*(i) (Linearity) For any $\alpha,\beta\in\mathbb{R}$*

$$\int_a^b \left(\alpha f(x)+\beta g(x)\right)dx=\alpha\int_a^b f(x)dx+\beta\int_a^b g(x)dx.$$

*(ii) (Integration by Parts {reverse product rule})*

$$\int f'(x)g(x)dx=f(x)g(x)\Big|_a^b-\int_a^b f(x)g'(x)dx,$$

*where $F(x)\Big|_a^b=F(b)-F(a)$.*

*(iii) ($u-$substitution {reverse chain rule})*

$$\int_a^b f\left(g(x)\right)g'(x)dx=\int_{u(a)}^{u(b)} f(u)du.$$

Questions relating to definite integrals consist, mostly *but not always*, of finding a primitive function as we've done in the previous section.

**Example 9.** Compute the area beneath the function $f(x)=x^3-2x+4$ over the interval $[0,1]$.

*Solution.* The required area is $\int_0^1 f(x)dx$. We have that

$$\begin{aligned}\int_0^1 f(x)dx=\int_0^1\left(x^3-2x+4\right)dx&=\left(\frac{x^4}{4}-x^2+4x\right)\Big|_0^1\\&=\left(\frac{1}{4}-1+4\right)-\left(\frac{0}{4}-0+0\right)=\frac{17}{4}.\end{aligned}$$

□

**Example 10.** Compute the area beneath the function $f(x)=x\cos(x)$ over the interval $[0,\pi]$.

*Solution.* As the anti derivative is not immediate, we turn our attention to other methods. Since we have a multiplication of two functions, one of which we would be happy to differentiate (the function $x$), we consider using integration by parts. Setting $f'(x)=\cos(x)$ and $g(x)=x$, we have that $g'(x) = 1$ and a choice for a primitive function for $f'(x)$ is $f(x) = \sin(x)$. Thus

$$\int_0^{\pi} f(x)dx = \int_0^{\pi} x\cos(x)dx = x\sin(x)\Big|_0^{\pi} - \int_0^{\pi} 1\cdot\sin(x)dx$$

$$= \underbrace{(\pi\sin(\pi) - 0\cdot\sin(0))}_{=0} - \int_0^{\pi} \sin(x)dx = \cos(x)\Big|_0^{\pi} = \cos(\pi) - \cos(0) = -2.$$

$\square$

**Example 11.** Compute the area beneath the function $f(x) = x^2\sin\left(x^3\right)$ over the interval $\left[0, \sqrt[3]{\frac{\pi}{2}}\right]$.

*Solution.* Recognising $x^2$ as the almost derivative of $x^3$ we find ourselves tempted to use $u$−substitution with $u(x) = x^3$. In that case we have that $u'(x) = 3x^2$, meaning that

$$du = 3x^2dx.$$

Thus

$$\int_0^{\sqrt[3]{\frac{\pi}{2}}} f(x)dx = \int_0^{\sqrt[3]{\frac{\pi}{2}}} x^2\sin\left(x^3\right)dx = \int_{0^3}^{\left(\sqrt[3]{\frac{\pi}{2}}\right)^3} \sin(u)\frac{du}{3}$$

$$= \frac{1}{3}\int_0^{\frac{\pi}{2}} \sin(u)du = \frac{-\cos(u)}{3}\Big|_0^{\frac{\pi}{2}} = \frac{-\cos\left(\frac{\pi}{2}\right) - (-\cos(0))}{3} = \frac{1}{3}.$$

$\square$

# Condensed notes on linear algebra for IMDS

Jeffrey Giansiracusa

October 15, 2021

## 1 Vectors

A vector is a list of numbers (usually these will be real number, but you might encounter complex numbers or more exotic things out in the world). We will usually write vectors as columns, e.g.:

$$\mathbf{v} = \begin{pmatrix} 1 \\ 4 \\ 7 \\ \vdots \\ 9 \end{pmatrix}$$

A component of a vector is just one of the entries in the list.

We write $\mathbb{R}^n$ for the set of all vectors with $n$-components, where $n$ is some number. E.g., $\mathbb{R}^2$ is the set of all vectors $\mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$.

We think of $\mathbb{R}^2$ as a plane (2-dimensional space) and $\mathbb{R}^3$ as 3-space, so we can draw pictures of vectors in 2 and 3 dimensions.

### 1.1 Vector addition

We can add vectors by adding them component by component.

**Example 1.**

$$\begin{pmatrix} 1 \\ 2 \\ 5 \end{pmatrix} + \begin{pmatrix} 0.1 \\ 0.9 \\ 2.5 \end{pmatrix} = \begin{pmatrix} 1.1 \\ 2.9 \\ 7.5 \end{pmatrix}$$

Note that you are only allowed to add vectors if they have the same length.

### 1.2 Scalar multiplication

When talking about vectors, we often use the word scalar as a fancy-sounding word for number. The idea is that a scalar can be used to scale (stretch) vectors (make them shorter or longer without changing their direction).

If $\lambda \in \mathbb{R}$ is a scalar and $\begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} \in \mathbb{R}^n$ is a vector, then $\lambda v \begin{pmatrix} \lambda v_1 \\ \vdots \\ \lambda v_n \end{pmatrix}$.

**Example 2.**

$$2 \begin{pmatrix} 4 \\ 7 \end{pmatrix} = \begin{pmatrix} 8 \\ 14 \end{pmatrix}.$$

## 2 Linear combinations and span

A linear combination of a bunch of vectors $\mathbf{v}_1, \ldots \mathbf{v}_k$ is simply a vector we make from these using scalar multiplication and addition:

$$\lambda_1 \mathbf{v}_1 + \cdots + \lambda_n \mathbf{v}_n,$$

where the $\lambda_i$ are scalars.

The span of a set of vectors is the set of all vectors we can make by taking linear combinations.

**Example 3.** The span of the vectors

$$\mathbf{b}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \mathbf{b}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \mathbf{b}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

is all of $\mathbb{R}^3$ since we can make an arbitrary vector $\mathbf{v} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$ as a linear combination

$$x\mathbf{b}_1 + y\mathbf{b}_2 + z\mathbf{b}_3.$$

The span of $\mathbf{b}_1$ and $\mathbf{b}_2$ is just the $xy$-plane (where $z = 0$).

### 2.1 Linear dependence and independence

A set of vectors $\mathbf{v}_1, \ldots, \mathbf{v_n}$ is linearly dependent if it is possible to make one of them as a linear combination of the others.

**Example 4.** The list of vectors

$$\mathbf{v}_1 = \begin{pmatrix} 1 \\ 4 \end{pmatrix}, \mathbf{v}_2 = \begin{pmatrix} 3 \\ -1 \end{pmatrix}, \mathbf{v}_3 = \begin{pmatrix} 5 \\ 7 \end{pmatrix}$$

is a linearly dependent set of vectors because we can make $v_3$ as a linear combination $2\mathbf{v}_1 + \mathbf{v}_2$. In fact, in this example we can make any one of the vectors from the other two. We can plug this expression for $\mathbf{v}_3$ into any linear combination $\lambda_1\mathbf{v}_1 + \lambda\mathbf{v}_2 + \lambda\mathbf{v}_3$ to turn it into a linear combination of just $\mathbf{v}_1$ and $\mathbf{v}_2$,

$$\lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2 + \lambda_3\mathbf{v}_3 = \lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2 + \lambda_3(2\mathbf{v}_1 + \mathbf{v}_2) = (\lambda_1 + 2\lambda_3)\mathbf{v}_1 + (\lambda_2 + \lambda_3)\mathbf{v}_2.$$

Thus the span of all three is equal to the span of just the first two vectors in this example.

**Example 5.** For another example, consider

$$\mathbf{v}_1 = \begin{pmatrix} 2 \\ 0 \\ 0 \end{pmatrix}, \mathbf{v}_2 = \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}, \mathbf{v}_3 = \begin{pmatrix} 0 \\ 3 \\ 1 \end{pmatrix}, \mathbf{v}_4 = \begin{pmatrix} 0 \\ 4 \\ 0 \end{pmatrix}.$$

In this example, we can make $\mathbf{v}_4$ as a linear combination $\mathbf{v}_4 = \mathbf{v}_2 + \mathbf{v}_3$, and we can rearrange the equation to make either $\mathbf{v}_3$ or $\mathbf{v}_2$ out of the others. Note however that we cannot make $\mathbf{v}_1$ as a linear combination of the others.

As a special case, two vectors are linearly dependent if and only if one is scalar multiple of the other.

A set of vectors is <u>linearly independent</u> if we cannot make any one as a linear combination of the others. For example, if we delete $\mathbf{v}_4$ from the above set of vectors then we would have an independent set of vectors.

Here are three useful equivalent ways to think about dependence and independence:

1. A set of vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ is independent if removing any one vector from the list makes the span smaller.

2. A set of vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ is independent if the span gets bigger each time we add one. I.e., the span of $\{\mathbf{v}_1, \mathbf{v}_2\}$ is bigger than the span of $\mathbf{v}_1$, and the span of the first 3 is bigger than the span of the first two, and so on.

3. A set of vectors is independent if there each point in the span can be made as a linear combination in only one unique way.

**Remark 1.** To see why the third condition is equivalent: if there is some vector $\mathbf{u}$ that we can make as a linear combination in two different ways,

$$u = \lambda_1\mathbf{v}_1 + \cdots + \lambda_n\mathbf{v}_n = \mu_1\mathbf{v}_1 + \cdots + \mu_n\mathbf{v}_n,$$

then the difference

$$(\lambda_1 - \mu_1)\mathbf{v}_1 + \cdots + (\lambda_n - \mu_n)\mathbf{v}_n = 0$$

gives us a nontrivial recipe for 0, and then we can use this to write one of the vectors $\mathbf{v}_i$ as a linear combination of the others as long as the coefficient $(\lambda_i - \mu_i)$ is not zero.)

Given a list of vectors that is dependent, we can always delete vectors from the list one at a time if they can be made as a linear combination of the others, until we arrive at a list of linearly independent vectors.

## 2.2 Determining if something is in the span

Suppose we have a vector $\mathbf{u}$ and we want to know if it is in the span of a list of vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$. I.e., we want to find scalars $\lambda_1, \ldots, \lambda_n$ such that

$$\lambda_1 \mathbf{v}_1 + \cdots + \lambda_n \mathbf{v}_n = \mathbf{u}.$$

When we write this out component by component, we get a system of linear equations in the variables $\lambda_1, \ldots, \lambda_n$. Either this system of equations has a solution, in which case $\mathbf{u}$ is in the span, or it does not have a solution and $\mathbf{u}$ is outside the span.

**Example 6.** Consider the vectors:

$$\mathbf{v}_1 = \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}, \mathbf{v}_2 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \mathbf{u} = \begin{pmatrix} 3 \\ 4 \\ 1 \end{pmatrix}.$$

Is the $\mathbf{u}$ in the span of $\mathbf{v}_1$ and $\mathbf{v}_2$?

We write out the system of equations:

$$\lambda_1 \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix} + \lambda_2 \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 3 \\ 4 \\ 1 \end{pmatrix}.$$

and then try to solve for $\lambda_1$ and $\lambda_2$. Looking at the middle component, we see that $\lambda_2 = 4$. Now looking at the first component and plugging in $\lambda_2 = 4$, we have $\lambda_1 + 4 = 3$, so $\lambda_1$ must be -1. But then we find that these values don't work for the equation of the third component equation, $2\lambda_1 + 0\lambda_2 = 1$. Thus there is no solution, so $\mathbf{u}$ is not in the span.

## 2.3 Linear spaces: lines, planes, etc

Here is a somewhat abstract definition to start with: A <u>linear space</u> in $\mathbb{R}^n$ is something that is the span of a set of vectors, and its <u>dimension</u> is the size of the largest set of independent vectors you can find in it.

More concretely, a line in $\mathbb{R}^n$ is the span of a single (nonzero) vector.

If we take the span of two vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ then there are two possibilities: either they are dependent (so $\mathbf{v}_2$ is a scalar multiple of $\mathbf{v}_1$ and the span is still a line), or they are independent and the span is something bigger.

A plane in $\mathbb{R}^n$ is what we get when we take the span of two independent vectors.

| |
|---|
| **Important fact:** In $\mathbb{R}^n$, you can find at most $n$ independent vectors. |

Thus you will never find a 4-dimensional space inside $\mathbb{R}^3$, etc.

### 2.4 Basis

A set of vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k$ in $\mathbb{R}^n$ is a basis if they are linearly independent and they span all of $\mathbb{R}^n$. It turns out that this can only happen when the number of vectors is equal to the dimension $n$.

The most common example of a basis is the *standard basis*:

$$\mathbf{b}_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \mathbf{b}_2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \quad \ldots, \quad \mathbf{b}_n = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}$$

(so $\mathbf{b}_i$ has a 1 in the $i^{th}$ positions and zeros everywhere else).

This is certainly not the only example of a basis, and many tasks in geometry and data science involve finding a basis that is particularly nice for the problem you are looking at.

If $\mathbf{v}_1, \ldots, \mathbf{v}_n$ is a basis, then any vector can be written in a unique way as a linear combination of these basis vectors:

$$\mathbf{u} = \lambda_1 \mathbf{v}_1 + \cdots + \lambda_n \mathbf{v}_n.$$

We can think of the list of numbers $\lambda_i$ as representing $\mathbf{u}$ relative to this basis.

## 3 Lengths, angles, and the dot product

### 3.1 The length of a vector

Pythagoras tells us that if we have a right triangle where the legs have length $a$ and $b$ and hypotenuse has length $c$, then

$$a^2 + b^2 = c^2.$$

We can use this to measure the lengths of vectors in the plane. Given a vector $\mathbf{v} = \begin{pmatrix} a \\ b \end{pmatrix}$, we have a right triangle by moving along the $x$-axis a distance $a$ and the parallel to the $y$-axis a distance $b$. The hypotenuse is precisely our vector, and so Pythagoras says that the length or the vector, written $||\mathbf{v}||$, is $\sqrt{a^2 + b^2}$.

It follows that if we want to measure the length of a vector $= \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}$ in $\mathbb{R}^n$, then we should use the formula

$$||\mathbf{v}|| = \sqrt{v_1^2 + \cdots + v_n^2}.$$

## 3.2 Angles

Suppose we have two vectors $\mathbf{u}$ and $\mathbf{v}$. If they are linearly dependent then we say that the angle between them is either 0 or 180 degrees (depending on whether they point in the same or opposite directions). If they are independent then they span a plane and in that plane we can measure the angle between them. (Note that we can't really tell the difference between an angle $\theta$ and $-\theta$ or $360 - \theta$ since we haven't decided which side of the plane we should look at.)

## 3.3 The dot product

The dot product of two vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$ is the number we obtain by multiplying corresponding components of the vectors and summing these up. We write $\mathbf{u} \cdot \mathbf{v}$.

**Example 7.**

$$\begin{pmatrix} 5 \\ 2 \\ 9 \end{pmatrix} \cdot \begin{pmatrix} 7 \\ -1 \\ 6 \end{pmatrix} = (5)(7) + (2)(-1) + (9)(6) = 35 - 2 + 54 = 87.$$

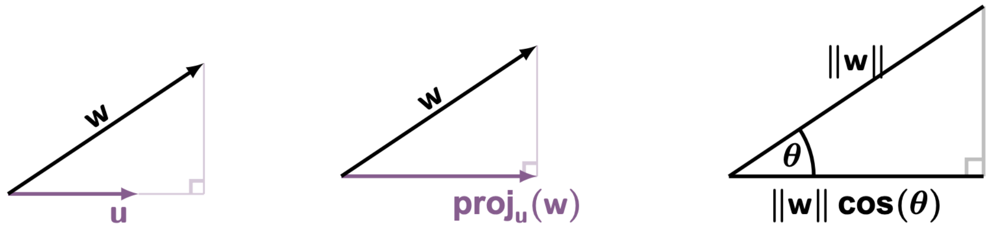
The dot product seems a bit weird, but it has a lovely geometric interpretation in terms of lengths and angles:

**Fact:** $\mathbf{u} \cdot \mathbf{v} = ||\mathbf{u}|| \, ||\mathbf{v}|| \cos\theta$, where $\theta$ is the angle between the two vectors.

Since the cosine of an angle is 0 if and only if the angle is a right angle (90 or -90 or 270), the dot product of two vectors is 0 if and only if they make a right angle.

### 3.4 Projection

Let $\mathbf{u}$ be a nonzero vector. There is a useful way to write any vector $\mathbf{w}$ as the sum of two pieces: one that points along the direction of $\mathbf{u}$, and one that is perpendicular to $\mathbf{u}$.




The piece of $\mathbf{w}$ that points along the direction of $\mathbf{u}$ is called the *projection* of $\mathbf{w}$ onto $\mathbf{u}$, and we write $\text{proj}_{\mathbf{u}}(\mathbf{w})$. This is a vector proportional to $\mathbf{u}$, and the length is $||\mathbf{w}|| \cos\theta = \frac{1}{||\mathbf{u}||}(\mathbf{w}\cdot\mathbf{u})$. We thus make $\mathbf{u}$ into the length we want by diving by the length of $\mathbf{u}$ and then multiplying by $\frac{1}{||\mathbf{u}||}(\mathbf{w}\cdot\mathbf{u})$. Thus

$$\text{proj}_{\mathbf{u}}(\mathbf{w}) = \left(\frac{\mathbf{u}\cdot\mathbf{w}}{\mathbf{u}\cdot\mathbf{u}}\right)\mathbf{u}.$$

The piece of $\mathbf{w}$ that is perpendicular to $\mathbf{u}$ is then what is left over: $\mathbf{w} - \text{proj}_{\mathbf{u}}(\mathbf{w})$.

# 4 Matrices

A matrix is a rectangular array of numbers (or variables). We say that a matrix $A$ has shape $n \times k$ if it has $n$ rows and $k$ columns. For example

$$\begin{pmatrix} 2 & 4 & 5 \\ 1 & 0 & 9 \end{pmatrix}$$

has shape $2 \times 3$. Note that we can think of the vectors we've been working with as $n \times 1$ matrices. Given a matrix $A$, we will often write $A_{ij}$ for the component in row $i$ and column $j$. In the matrix above, we have $A_{12} = 4$ and $A_{21} = 1$.

There are 4 operations we often do with matrices:

- We can add two matrices if they have exactly the same shape. As with vectors, we simply add them component by component. We can express this as: $(A + B)_{ij} = A_{ij} + B_{ij}$.
- We can multiply a matrix $A$ by a scalar $\lambda$ in essentially same way as with vectors. Just multiply each component of $A$ by $\lambda$. In terms of symbols, this is $(\lambda A)_{ij} = \lambda(A_{ij})$.

- You may encounter the transpose of a matrix, denoted $A^T$. This is simply the matrix whose rows are the columns of $A$. I.e., $(A^T)_{ij} = A_{ji}$.
- Matrix multiplication: this one is a little more complicated, so we'll devote the next section to it. To multiply $AB$ we will require that the number of columns of $A$ is equal to the number of rows of $B$.

## 4.1 Multiplying matrices

Suppose $A$ has shape $n \times k$ and $B$ has shape $k \times m$. Then we define the product $AB$ by the rule: $(AB)_{ij}$ is the dot product of the $i^{th}$ row of $A$ with the $j^{th}$ column of $B$.

**Example 8.** Consider the matrices

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 4 & 1 \\ 10 & 6 & 1 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 2 & 4 \\ 4 & 1 & 5 \\ 3 & 1 & 1 \end{pmatrix}.$$

We have

$$AB = \begin{pmatrix} 18 & 7 & 17 \\ 21 & 9 & 29 \\ 37 & 27 & 71 \end{pmatrix}, \quad BA = \begin{pmatrix} 45 & 34 & 9 \\ 56 & 42 & 18 \\ 15 & 16 & 11 \end{pmatrix}.$$

**Example 9.** Consider the matrices

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 2 & 0 \\ 4 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}.$$

Which multiplication is allowed: $AB$, $BA$, neither or both? Since $A$ is $2 \times 3$ matrix and $B$ is a $3 \times 3$ matrix, we are only allowed to compute $AB$. We find that

$$AB = \begin{pmatrix} 12 & 5 & 2 \\ 15 & 4 & 3 \end{pmatrix}.$$

## 4.2 Matrix times vector

A very important special case of matrix multiplication is when we multiply an $n \times m$ matrix $A$ by a vector $\mathbf{v} = \begin{pmatrix} v_1 \\ \vdots \\ v_m \end{pmatrix}$ of length $m$ (which is the same as a matrix of shape $m \times 1$).

| |
|---|
| **Important fact:** $A\mathbf{v}$ is a linear combination of the columns of $A$. Take $v_j$ times the $j^{th}$ column of $A$ and sum these up. |

Explanation: If we let $A_{*j}$ denote the $j^{th}$ column of $A$, then the we can use the components of $\mathbf{v}$ to make a linear combination of the columns of $A$ by

$$v_1 A_{*1} + \cdots + v_m A_{*m}$$

and the $i^{th}$ component of this vector is $\sum_{j=1}^{m} v_j A_{ij}$, which is exactly the $i^{th}$ component of $A\mathbf{v}$.

**Example 10.**

$$\begin{pmatrix} 2 & 3 & 5 \\ 1 & 9 & -2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = x \begin{pmatrix} 2 \\ 1 \end{pmatrix} + y \begin{pmatrix} 3 \\ 9 \end{pmatrix} + z \begin{pmatrix} 5 \\ -2 \end{pmatrix}$$

# 5 Solutions to systems of linear equations

Now we will look at the connection between matrices, linear (in)dependence, and finding solutions to systems of linear equations. Suppose we have a list of unknowns $x_1, \ldots, x_n$. A linear equation is an equation that says a certain linear combination of the unknowns is equal to a certain number. E.g., $3x_1 + 2x_2 + 9x_3 = 1$. A system of linear equations is a list of linear equations:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m \end{aligned}$$

We can collect the unknowns $x_i$ together into a vector $\mathbf{x}$. We can also collect the right hand side numbers $b_i$ into a vector $\mathbf{b}$, and we collect the coefficients $a_{ij}$ into a matrix $A$:

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}, \quad A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

The we can write the system of linear equations in the very compact form

$$A\mathbf{x} = \mathbf{b}$$

**Example 11.** Consider

$$\begin{aligned} 2x_1 + 3x_2 &= 4 \\ 9x_1 - x_2 &= 10 \end{aligned}$$

In terms of matrices, we can rewrite this as

$$\begin{pmatrix} 2 & 3 \\ 9 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 4 \\ 10 \end{pmatrix}$$

Recalling from Section 4.2 that a matrix times a vector gives us a linear combination of the columns of the matrix, trying to find a solution to an equation of the form

$$A\mathbf{x} = \mathbf{b}$$

is the same as trying to find a way to make $\mathbf{b}$ as a linear combination of the columns of $A$. There are 3 possibilities:

1. $\mathbf{b}$ is not in the span of the columns, so we can't make $\mathbf{b}$ as a linear combination of the columns and hence there is no solution.
2. $\mathbf{b}$ is in the span of the columns, and the columns are linearly independent, so there is a unique recipe to make $\mathbf{b}$ as a linear combination, and hence there is exactly one solution.
3. $\mathbf{b}$ is in the span of the columns, and the columns are linearly dependent. In this case, there are infinitely many solutions because there are infinitely many ways to make $\mathbf{b}$ as a linear combination of the columns.

# 6 Linear transformations

**Definition 1.** A function (we often say 'map' or 'mapping') $T : \mathbb{R}^n \to \mathbb{R}^m$ is called a linear transformation if

- For any $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ we have that

$$T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v}) \quad \textit{addition condition.}$$

- For any $\mathbf{u} \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$ we have that

$$T(\lambda\mathbf{u}) = \lambda T(\boldsymbol{u}) \quad \textit{scalar multiplication condition.}$$

A few notations we should remember when considering linear transformations:

- The *domain* of $T$ is the set of allowed inputs for $T$. In the case above, this it $\mathbb{R}^n$.
- The *codomain* of $T$ is the set in which the outputs of $T$ live. In the case above, this is $\mathbb{R}^m$. It might be the case that there are vectors in the codomain that are not hit by $T$ of anything.
- The *range* of $T$ is the set of outputs that can actually be produced by $T$:

$$\text{Range}(T) = \{\mathbf{w} \in \mathbb{R}^m \mid \text{there exists} \mathbf{v} \in \mathbb{R}^n \text{ such that } T(\mathbf{v}) = \mathbf{w}\}.$$

The range of a transformation is contained in its codomain.

**Example 12.** Find the domain, codomain of the following maps, and determine weather or not they are linear. In the case that they are, find Range $(T)$

a) $T\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} y \\ x \end{pmatrix}$.

b) $T\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} y \\ 1 \end{pmatrix}$.

c) $T(x) = \begin{pmatrix} x \\ x^2 \\ 2x \end{pmatrix}$.

*Solution.* a) $T$ acts on vectors in $\mathbb{R}^2$ and returns vectors in $\mathbb{R}^2$. Thus, Dom $(T) = \mathbb{R}^2$ and Codom $(T) = \mathbb{R}^2$. To show linearity we check the addition and scalar multiplication conditions:

addition condition we have that

$$T\left(\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} + \begin{pmatrix} x_2 \\ y_2 \end{pmatrix}\right) = T\begin{pmatrix} x_1 + x_2 \\ y_1 + y_2 \end{pmatrix} = \begin{pmatrix} y_1 + y_2 \\ x_1 + x_2 \end{pmatrix}$$

$$\underset{\text{separate the vectors}}{=} \begin{pmatrix} y_1 \\ x_1 \end{pmatrix} + \begin{pmatrix} y_2 \\ x_2 \end{pmatrix} = T\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} + T\begin{pmatrix} x_2 \\ y_2 \end{pmatrix}.$$

multiplication condition we see that

$$T\left(\lambda\begin{pmatrix} x \\ y \end{pmatrix}\right) = T\begin{pmatrix} \lambda x \\ \lambda y \end{pmatrix} = \begin{pmatrix} \lambda y \\ \lambda x \end{pmatrix} = \lambda\begin{pmatrix} y \\ x \end{pmatrix} = \lambda T\begin{pmatrix} x \\ y \end{pmatrix}.$$

Both conditions are valid, and as such $T$ is a linear transformation. Let us find the range:

Range $(T)$: We'd like to find for which $u, v \in \mathbb{R}$ we can find $x, y \in \mathbb{R}$ such that

$$T\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} u \\ v \end{pmatrix}.$$

As

$$T\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} y \\ x \end{pmatrix}$$

we see that $u = y$ and $v = x$. This means that we can get *any vector*. Indeed

$$T\begin{pmatrix} v \\ u \end{pmatrix} = \begin{pmatrix} u \\ v \end{pmatrix},$$

showing that Range $(T) = \mathbb{R}^2$.

There is a simpler way to find Range $(T)$, which we will discuss shortly.

b) $T$ acts on vectors in $\mathbb{R}^3$ and returns vectors in $\mathbb{R}^2$. Thus, $\mathrm{Dom}(T) = \mathbb{R}^3$ and $\mathrm{Codom}(T) = \mathbb{R}^2$. To show linearity we check the addition and scalar multiplication conditions:
<u>addition condition</u> we have that

$$T\left(\begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix} + \begin{pmatrix} x_2 \\ y_2 \\ z_2 \end{pmatrix}\right) = T\begin{pmatrix} x_1 + x_2 \\ y_1 + y_2 \\ z_1 + z_2 \end{pmatrix} = \begin{pmatrix} y_1 + y_2 \\ 1 \end{pmatrix}.$$

On the other hand

$$T\begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix} + T\begin{pmatrix} x_2 \\ y_2 \\ z_2 \end{pmatrix} = \begin{pmatrix} y_1 \\ 1 \end{pmatrix} + \begin{pmatrix} y_2 \\ 1 \end{pmatrix} = \begin{pmatrix} y_1 + y_2 \\ 2 \end{pmatrix}.$$

The addition condition is not satisfied, so the map is not a linear transformation. Regardless, let is check the second condition.
<u>multiplication condition</u> we see that

$$T\left(\lambda\begin{pmatrix} x \\ y \\ z \end{pmatrix}\right) = T\begin{pmatrix} \lambda x \\ \lambda y \\ \lambda z \end{pmatrix} = \begin{pmatrix} \lambda y \\ 1 \end{pmatrix}.$$

On the other hand

$$\lambda T\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \lambda\begin{pmatrix} y \\ 1 \end{pmatrix} = \begin{pmatrix} \lambda y \\ \lambda \end{pmatrix}.$$

In order for the scalar multiplication condition to be satisfied we must have that $\lambda = 1$. This means that the scalar multiplication is not valid for all $\lambda \in \mathbb{R}$ - showing us again that the map can't be a linear transformation.

c) $T$ acts on numbers in $\mathbb{R}$ and returns vectors in $\mathbb{R}^3$. Thus, $\mathrm{Dom}(T) = \mathbb{R}$ and $\mathrm{Codom}(T) = \mathbb{R}^3$. To show linearity we check the addition and scalar multiplication conditions:
<u>addition condition</u> we have that

$$T(x+y) = \begin{pmatrix} x+y \\ (x+y)^2 \\ 2(x+y) \end{pmatrix} = \begin{pmatrix} x+y \\ x^2+2xy+y^2 \\ 2(x+y) \end{pmatrix}.$$

On the other hand

$$T(x) + T(y) = \begin{pmatrix} x \\ x^2 \\ 2x \end{pmatrix} + \begin{pmatrix} y \\ y^2 \\ 2y \end{pmatrix} = \begin{pmatrix} x+y \\ x^2+y^2 \\ 2(x+y) \end{pmatrix}.$$

We see that the addition condition is satisfied if and only if $2xy = 0$, i.e. $x = 0$ or $y = 0$. This means that it is not satisfied for all $x$ and $y$ in $\mathbb{R}$, showing that the map is not a linear transformation. Like before, let us check the scalar multiplication as well.
multiplication condition we see that

$$T(\lambda x) = \begin{pmatrix} \lambda x \\ (\lambda x)^2 \\ 2\lambda x \end{pmatrix} = \begin{pmatrix} \lambda x \\ \lambda^2 x^2 \\ 2\lambda x \end{pmatrix}.$$

On the other hand

$$\lambda T(x) = \lambda \begin{pmatrix} x \\ x^2 \\ 2x \end{pmatrix} = \begin{pmatrix} \lambda x \\ \lambda x^2 \\ 2\lambda x \end{pmatrix}.$$

In order for the scalar multiplication condition to be satisfied we must have that $\lambda x^2 = \lambda^2 x^2$, i.e. $x = 0$ or $\lambda = 0$ or $\lambda = 1$. This means that it is not satisfied for all $x \in \mathbb{R}$ and $\lambda \in \mathbb{R}$, showing (yet again) that the map is not a linear transformation. □

While it may seem that linear transformation are detached from what we've learned so far, such maps are, in fact, intimately connected to matrices. This is expressed in the following theorem:

**Theorem 6.1.** *Any $n \times m$ matrix $A$ determines a linear transformation $T : \mathbb{R}^m \to \mathbb{R}^n$ by the simple formula $T(\mathbf{v}) = A\mathbf{v}$, and every linear transformation comes from a matrix in this way.*

*Proof.* You should check that the function defined by a matrix is a linear transformation. To show that every linear transformation comes from a matrix, we will start from a linear transformation $T$ and build a matrix $[T]$ from it, and then show that $T(\mathbf{v})$ is equal to the product of the matrix $[T]$ times the vector $\mathbf{v}$.

To build $[T]$, we will need the list of of standard basis vectors

$$\mathbf{b}_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \mathbf{b}_2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \quad \ldots, \quad \mathbf{b}_n = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}$$

(so $\mathbf{b}_i$ has a 1 in the $i^{th}$ positions and zeros everywhere else). We apply $T$ to these to get a new list of vectors $T(\mathbf{b}_1), \ldots, T(\mathbf{b}_n)$. Now we make a matrix by putting $T(\mathbf{b}_i)$ as the $i^{th}$ column.

Then we just need to check that the matrix multiplication $[T]\mathbf{v}$ gives the same result as $T(\mathbf{v})$ for any vector $\begin{pmatrix} v_1 \\ \vdots \\ v_m \end{pmatrix}$. The matrix multiplication $[T]\mathbf{v}$ is the linear combination of the columns of $[T]$ with coefficients $v_i$, so this is

$$\sum_{i=1}^{m} v_i T(\mathbf{b}_i).$$

On the other hand, we can write $\mathbf{v} = \sum_{i=1}^{m} v_i \mathbf{b}_i$, and then we use the fact that $T$ is a linear transformation to get

$$T(\mathbf{v}) = T\left(\sum_{i=1}^{m} v_i T(\mathbf{b}_i)\right) = \sum_{i=1}^{m} v_i T(\mathbf{b}_i). \qquad \square$$

**Example 13.** Consider the following map $T : \mathbb{R}^3 \to \mathbb{R}^2$:

$$T\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x+y \\ y+z \end{pmatrix}.$$

a) Show that the map is a linear transformation.

b) Find the matrix representation of $T$, $[T]$.

*Solution.* a) To show linearity we check the addition and scalar multiplication conditions:

<u>addition condition</u> we have that

$$T\left(\begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix} + \begin{pmatrix} x_2 \\ y_2 \\ z_2 \end{pmatrix}\right) = T\begin{pmatrix} x_1+x_2 \\ y_1+y_2 \\ z_1+z_2 \end{pmatrix} = \begin{pmatrix} (x_1+x_2)+(y_1+y_2) \\ (y_1+y_2)+(z_1+z_2) \end{pmatrix}$$

$$\underset{\text{separate the vectors}}{=} \begin{pmatrix} x_1+y_1 \\ y_1+z_1 \end{pmatrix} + \begin{pmatrix} x_2+y_2 \\ y_2+z_2 \end{pmatrix} = T\begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix} + T\begin{pmatrix} x_2 \\ y_2 \\ z_2 \end{pmatrix}.$$

<u>multiplication condition</u> we see that

$$T\left(\lambda\begin{pmatrix} x \\ y \\ z \end{pmatrix}\right) = T\begin{pmatrix} \lambda x \\ \lambda y \\ \lambda z \end{pmatrix} = \begin{pmatrix} \lambda x + \lambda y \\ \lambda y + \lambda z \end{pmatrix} = \begin{pmatrix} \lambda\,(x+y) \\ \lambda\,(y+z) \end{pmatrix} = \lambda\begin{pmatrix} x+y \\ y+z \end{pmatrix} = \lambda T\begin{pmatrix} x \\ y \\ z \end{pmatrix}.$$

Both conditions are valid, and as such $T$ is a linear transformation.

b) We have that

$$T\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad T\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad T\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

and as such

$$[T] = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}.$$

$\square$

# 7 Determinants and inverses

*coming soon... for now, see lecture slides.*

The inverse of a square matrix $A$ is a matrix $A^{-1}$ such that the products $AA^{-1}$ and $A^{-1}A$ are both equal to the identity matrix $\mathbb{1}$.

Finding the inverse of a matrix $A$ is useful because it gives us a very quick way to write down solution to systems of equations $A\mathbf{x} = \mathbf{b}$. Just multiply both sides by $A^{-1}$ and we get $\mathbf{x} = A^{-1}\mathbf{b}$. This shows us that, for a square system of equations (the number of equations equals the number of unknowns), there is a unique solution if and only if the matrix has an inverse.

**Important fact:** The inverse of $A$ exists if and only if the determinant of $A$ is nonzero.

# 8 Eigenvalues and eigenvectors

An eigenvector for a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ is a nonzero vector $\mathbf{v}$ such that $T(\mathbf{v})$ is a scalar multiple of $\mathbf{v}$. I.e.,

$$T(\mathbf{v}) = \lambda\mathbf{v}$$

for some scalar $\lambda$ which is called the corresponding eigenvalue. (We exclude the zero vector because it satisfies this trivially for any value of $\lambda$.) Notice that if $\mathbf{v}$ is an eigenvector, then any scalar multiple of $\mathbf{v}$ is also an eigenvector.

If $T$ is a shear transformation of $\mathbb{R}^2$, then a vector pointing along the shearing axis is an eigenvector with eigenvalue 1. Not every linear transformation has eigenvectors. For instance, a rotation of $\mathbb{R}^2$ has none (unless we decide to work with complex numbers, but that is a story for another day...)

Suppose we have a linear transformation given by a matrix $A$ and we want to find an eigenvector. We are looking for a solution to the equation

$$A\mathbf{v} = \lambda\mathbf{v}$$

where both the vector $\mathbf{v}$ and the scalar $\lambda$ are unknown. Subtracting $\lambda\mathbf{v}$ from both sides, we can rearrange this as

$$(A - \lambda\mathbb{1})\mathbf{v} = 0.$$

If we forget around the fact that we don't know what $\lambda$ is yet, this looks like a system of linear equations. It has one obvious solution: $\mathbf{v} = 0$. But that is not a very interesting solution because eigenvectors are supposed to be nonzero vectors. So, in order to find a nonzero $\mathbf{v}$, we need this system of equations to have infinitely many solutions. This happens if and only if $(A - \lambda\mathbb{1})$ does not have an inverse, which happens if and only if its determinant is 0.

Thus, the procedure to find an eigenvector is to first find the possible eigenvalues by solving the equation

$$\det(A - \lambda \mathbb{1}) = 0$$

for $\lambda$. Then, once you have a possible value of $\lambda$, you go back and look for a non-zero $\mathbf{v}$ that gives a solution to

$$(A - \lambda \mathbb{1})\mathbf{v} = 0.$$

Solving $\det(A - \lambda \mathbb{1}) = 0$ amounts to finding the roots of a polynomial $\lambda^n + \cdots = 0$. Unless the numbers happen to be very carefully selected, it is best to get a computer to do this. But in the $2 \times 2$ case we can easily do it by hand using the quadratic formula. Given a polynomial $a\lambda^2 + b\lambda + c$, the roots are given by the formula

$$\frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

The quantity $b^2 - 4ac$ has the fancy name *discriminant*.

1. If it is negative then we are taking the square root of a negative number and we are going to need to use complex numbers.
2. If it is zero then the choice of $\pm$ is irrelevant and there is only one root. In fact, the polynomial can be written as $a(\lambda + b/(2a))^2$.
3. If the discriminant is positive then there are two distinct roots, one given by taking the $\pm$ to be $+$, and the other given by choosing it to be $-$.

**Example 14.** Let us try to find the eigenvectors and eigenvalues of $A = \begin{pmatrix} 3 & 2 \\ 2 & 3 \end{pmatrix}$. First we compute

$$0 = \det(A - \lambda \mathbb{1}) = \det \begin{pmatrix} 3 - \lambda & 2 \\ 2 & 3 - \lambda \end{pmatrix} = (3 - \lambda)^2 - 4 = \lambda^2 - 6\lambda + 5.$$

This is a quadratic equation, so we use the quadratic formula:

$$\lambda = \frac{6 \pm \sqrt{36 - 4 \cdot 5}}{2} = 3 \pm 2 = 1 \text{ or } 5.$$

For each of these possible values for $\lambda$, we can now look for an eigenvector.

Let us first take $\lambda = 5$. We want to find a nonzero solution to the equation

$$(A - 5\mathbb{1})\mathbf{v} = 0.$$

Writing this out, we have

$$\begin{pmatrix} -2 & 2 \\ 2 & -2 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

This is the system of linear equations

$$\begin{aligned} -2v_1 + 2v_2 &= 0 \\ 2v_1 - 2v_2 &= 0 \end{aligned}$$

The second equation is just -1 times the first, so if we find a solution to the first then it is automatically a solution to the second equation as well. The first equation tell us that $v_2 = v_1$, and nothing more. So $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ is a solution, and any other solution is a scalar multiple of this. Thus we have found an eigenvector with eigenvalue 5.

If we instead take $\lambda = 1$, we arrive at the system of equations

$$\begin{aligned} 2v_1 + 2v_2 &= 0 \\ 2v_1 + 2v_2 &= 0 \end{aligned}$$

and the solutions to this system are all the vectors in the span of $\begin{pmatrix} 1 \\ -1 \end{pmatrix}$.