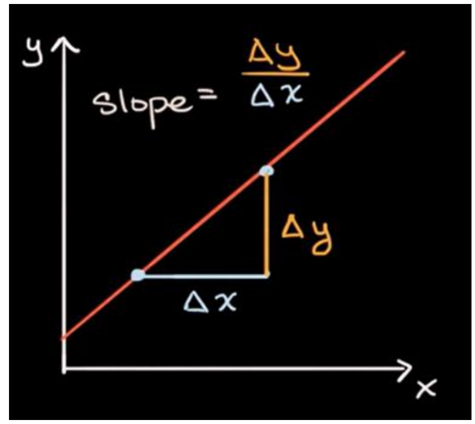
Revision:

The slop:

For any line which has a constant rate of change

F prime = m = y2-y1/x2-x1



$$= \frac{\Delta y}{\Delta x} = \frac{y_2 - y_1}{x_2 - x_1}$$

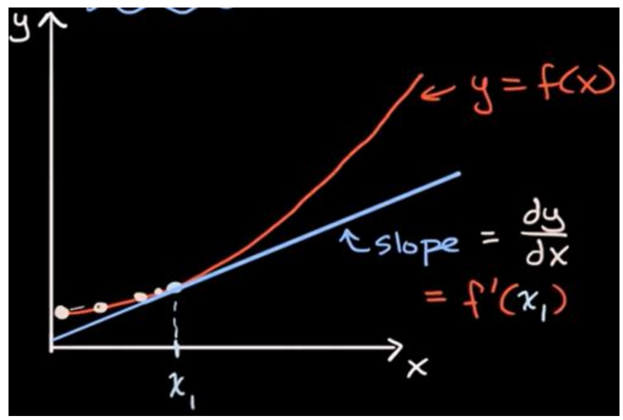
Instantaneous rate of change at specific point for a curve which changing:

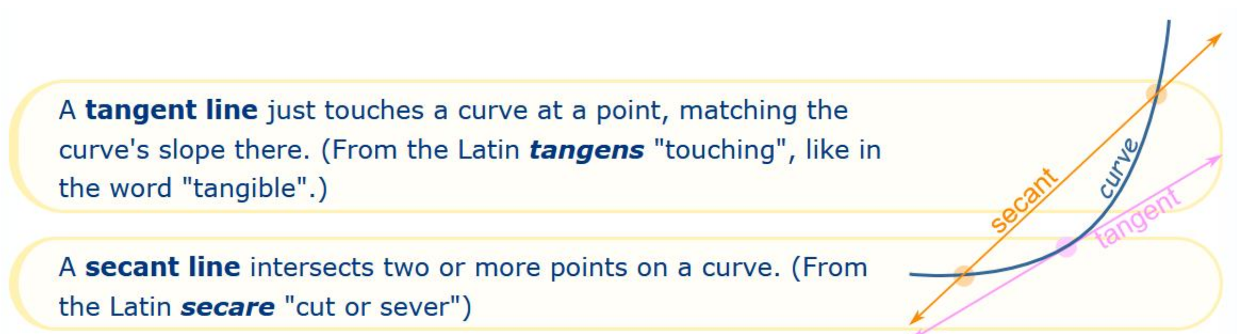
dy/dx means changes x and y become close to zero

super small change of y / super small change of x

when the changes approach to 0

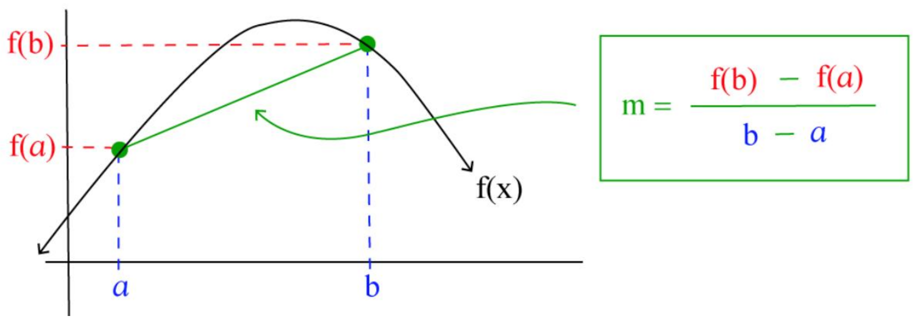
f prime is the slop of the tangent line at that point.





The slope of the secant line passing between two points.

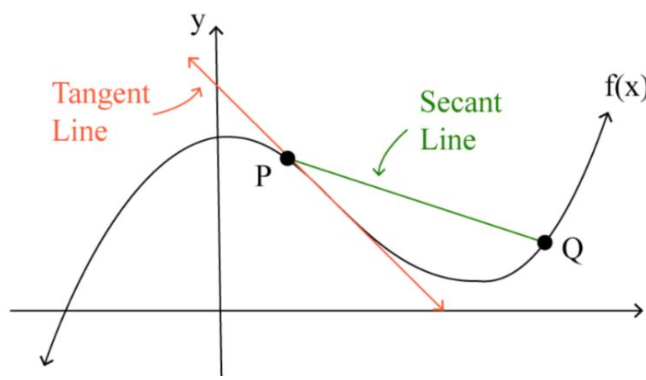
$$\textit{Average Rate of Change} = \frac{\Delta y}{\Delta x} = \frac{y_2 - y_1}{x_2 - x_1} = \frac{f(b) - f(a)}{b - a}$$



Average Vs Instantaneous Rate of Change
While both used to find the slope,

The average rate of change:
- Calculates the slope of the secant line using the **slope formula** from algebra.
- finds the **slope over an interval**

The instantaneous rate of change:
- Calculates the **slope of the tangent line** using derivatives.
- finds the **slope at a particular point**



Given $f(x) = 1 - 5x - x^3$ find the following

a. **Average rate of change** over $[1,3]$

$$Avg = \frac{f(3) - f(1)}{3-1} = \frac{\left[1-5(3)-(3)^3\right] - \left[1-5(1)-(1)^3\right]}{3-1} = \frac{(-41)-(-5)}{2} = -18$$

b. **Instantaneous rate of change** at $x = 2$

$$f'(x) = -5 - 3x^2$$

$$f'(2) = -5 - 3(2)^2 = -17$$

Formal definition of the derivative as a limit:

To find the slop of the line at specific point, the **derivative** of function f at x=c is the **limit** of the slope of the secant line from x=c to x=c+h as h approaches 0. Symbolically, this is the **limit** of [f(c)-f(c+h)]/h as h→0.

With a small change of notation, this limit written as:

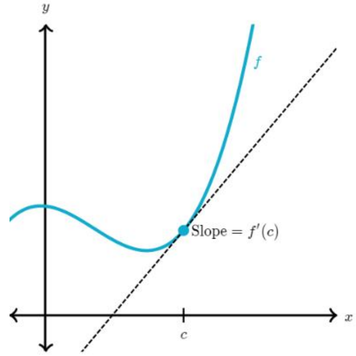
$$\lim_{h \to 0} \frac{f(a+h) - f(a)}{h}$$

The **derivative of** f(x) **with respect to *x*** is the function f′(x) and is defined as: (the slop of the tangent line)

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

We can calculate the slope of a tangent line using the definition of the derivative of a function f at $x = c$, (provided that limit exists):

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



Once we've got the slope, we can find the equation of the line.

Example: Finding the equation of the line tangent to the graph of

$f(x) = x^2$ **at** $x = 3$

The expression for the derivative of f(x)

$$f'(3) = \lim_{h \to 0} \frac{(3+h)^2 - 3^2}{h}$$

It gives us the slope of the tangent line.

$f'(3) =$ 6

To find the complete equation, we need a point the line goes through.
We are looking for the equation of the line whose slope is 6, and that goes through the point (3,9). To do that, we can use the definition of slope:

$$6 = \frac{y-9}{x-3}$$

Now we can isolate y:

$$6 = \frac{y-9}{x-3}$$

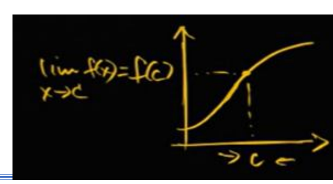
$$6(x-3) = y - 9$$

$$6(x-3) + 9 = y$$

Therefore, the equation of the line is $y = 6(x-3) + 9$**.**

Continuity

A function is said to be continuous at a point *x* = *a*, if

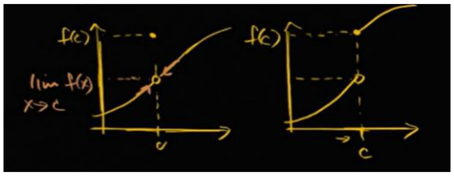


$\lim_{x \to a}$ f(x) Exists, and

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

It implies that if the left-hand limit (L.H.L), right hand limit (R.H.L) and the value of the function at $x=a$ exists and these parameters are equal to each other, then the function f is said to be continuous at $x=a$.

Discontinuous



If the function is undefined or does not exist, then we say that the function is discontinuous.

Continuity in open interval (a, b)

$f(x)$ will be continuous in the open interval (a,b) if at any point in the given interval the function is continuous.

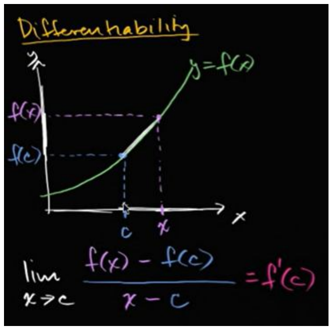
Continuity in closed interval [a, b]

A function $f(x)$ is said to be continuous in the closed interval $[a,b]$ if it satisfies the following three conditions.

1) $f(x)$ is be continuous in the open interval (a, b)

2) $f(x)$ is continuous at the point a from right i.e. $\lim_{x\to a} f(x)=f(a)$
3) $f(x)$ is continuous at the point b from left i.e. $\lim_{x\to b} f(x)=f(b)$

Differentiability



$f(x)$ is said to be differentiable at the point $x = a$ if the derivative f '(a) exists at every point in its domain. It is given by

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

For a function to be differentiable at any point $x=a$ in its domain, it must be continuous at that particular point but vice-versa is not always true.

Example: Consider the function $f(x) = (2x-3)^{\frac{1}{5}}$.Discuss its continuity and differentiability at $x = \frac{3}{2}$.

Solution: For checking the continuity, we need to check the left hand and right-hand **limits** and the value of the function at a point x=a.

L.H.L. = $\lim\limits_{x \to a^-} f(x) = \lim_{x \to \frac{3}{2}} (2x-3)^{\frac{1}{5}}$

$= \left(2 \times \frac{3}{2} - 3\right)^{\frac{1}{5}}$

= 0

R.H.L. = $\lim\limits_{x \to a^+} f(x) = \lim_{x \to \frac{3}{2}} (2x-3)^{\frac{1}{5}}$

$= \left(2 \times \frac{3}{2} - 3\right)^{\frac{1}{5}}$

= 0

L.H.L = R.H.L = f(a) = 0.

Thus the function is continuous at about the point $x = \frac{3}{2}$.

Now to check differentiability at the given point, we know

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

= $\lim\limits_{h \to 0} \frac{f(\frac{3}{2}+h)-f(\frac{3}{2})}{h}$

= $\lim\limits_{h \to 0} \frac{\left([2(\frac{3}{2})+h]-3\right)^{\frac{1}{5}} - \left(2(\frac{3}{2})-3\right)^{\frac{1}{5}}}{h}$

= $\lim\limits_{h \to 0} \frac{(3+2h-3)^{\frac{1}{5}} - (3-3)^{\frac{1}{5}}}{h}$

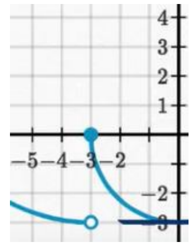
= $\lim\limits_{h \to 0} \frac{(2h)^{\frac{1}{5}} - 0}{h}$

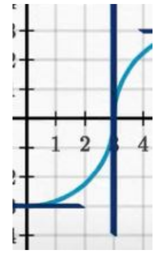
= $\lim\limits_{h \to 0} \frac{2^{\frac{1}{5}}}{h^{\frac{4}{5}}} = \infty$

Thus f is not differentiable at $x = \frac{3}{2}$.

A function f is not differentiable in the below scenario:

1- Vertical tangent.
2- Not continuous function
3- Sharp turn function





Example:

Is the function given below continuous/differentiable at $x = 3$?

$$f(x) = \begin{cases} x^2 & , x < 3 \\ 6x - 9 & , x \geq 3 \end{cases}$$

function f to be continues:
Limit from left side should equal limit from right side by applying the values directly in the function.

Check the function f(3) if equals to the limit for both sides:

$$\lim_{x \to 3^-} x^2 = 9$$

$$\lim_{x \to 3^+} 6x - 9 = 9$$

9 Cont.

$$f(\cancel{3}) = \lim_{x \to 3} f(x)$$

function f to be differentiable:
Limit from left side should equal limit from right side.

To define the limit:

differentiable

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

Test both sides if equals:

$$\lim_{x \to 3^-} \frac{(x+3)\cancel{(x-3)}}{\cancel{x-3}} = 6$$

$$\lim_{x \to 3^+} \frac{\cancel{6x-18}^{\,6(\cancel{x-3})}}{\cancel{x-3}} = 6$$

If a function is differentiable then it's also continuous. This property is very useful when working with functions, because if we know that a function is differentiable, we immediately know that it's also continuous.

The power rule:
to find the derivative of any expression in the form x^n:

$$\frac{d}{dx}[x^n] = n \cdot x^{n-1}$$

*Power Rule

$$\lim_{\Delta x \to 0} \frac{f(x+\Delta x) - f(x)}{\Delta x}$$

$$f(x) = x^n, \quad n \neq 0$$

$$f'(x) = n x^{n-1}$$

$$\frac{d}{dx}[A f(x)] = A \underbrace{\frac{d}{dx}[f(x)]} = A f'(x)$$

$$\frac{d}{dx}[f(x) + g(x)] = \frac{d}{dx}[f(x)] + \frac{d}{dx}[g(x)] = f'(x) + g'(x)$$

the basic derivative rules

Constant rule	$\frac{d}{dx}k = 0$
Constant multiple rule	$\frac{d}{dx}[k \cdot f(x)] = k \cdot \frac{d}{dx}f(x)$
Sum rule	$\frac{d}{dx}[f(x) + g(x)] = \frac{d}{dx}f(x) + \frac{d}{dx}g(x)$
Difference rule	$\frac{d}{dx}[f(x) - g(x)] = \frac{d}{dx}f(x) - \frac{d}{dx}g(x)$

The derivatives of the basic trigonometric functions are;

1. $\frac{d}{dx}(\sin x) = \cos x$
2. $\frac{d}{dx}(\cos x) = -\sin x$
3. $\frac{d}{dx}(\tan x) = \sec^2 x$
4. $\frac{d}{dx}(\cot x) = -\operatorname{cosec}^2 x$
5. $\frac{d}{dx}(\sec x) = \sec x \tan x$
6. $\frac{d}{dx}(\operatorname{cosec} x) = -\operatorname{cosec} x \cot x$

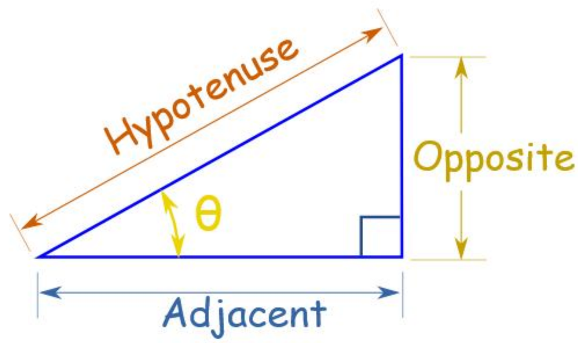
$$\cot x = \frac{\cos x}{\sin x}, \qquad \frac{d}{dx}(\cot x) = -cosec^2 x$$

$$sec(x) = \frac{1}{\cos(x)}, \qquad \frac{d}{dx}(\sec x) = \frac{\sin x}{cos^2 x} = \sec x \tan x$$

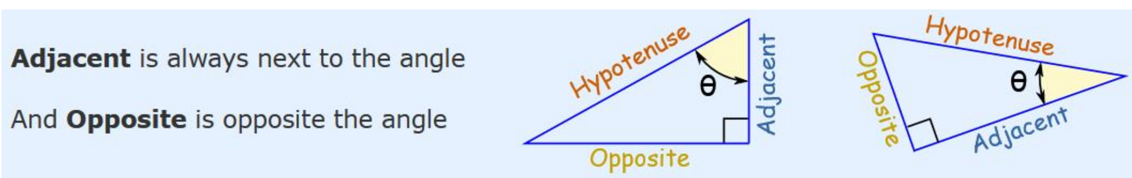
$$cosec(x) = \frac{1}{\sin(x)}, \qquad \frac{d}{dx}(\operatorname{cosec} x) = \frac{\cos x}{sin^2 x} = \operatorname{cosec} x \cot x$$

Sine, **Cosine** and **Tangent** are the main functions used in Trigonometry and are based on a Right-Angled Triangle.

Before getting stuck into the functions, it helps to give a **name** to each side of a right triangle:

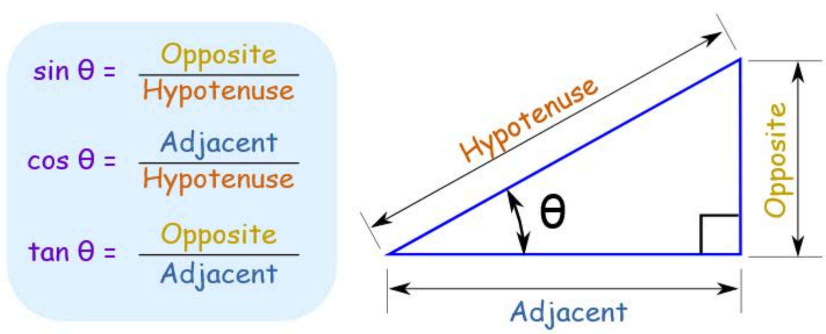


- "Opposite" is opposite to the angle θ
- "Adjacent" is adjacent (next to) to the angle θ
- "Hypotenuse" is the long one



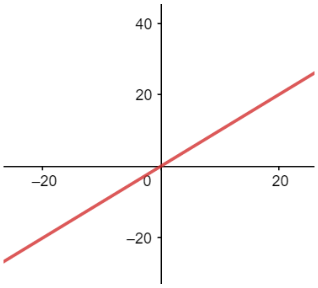
Sine, Cosine and Tangent

Sine, **Cosine** and **Tangent** (often shortened to **sin**, **cos** and **tan**) are each a **ratio of sides** of a right-angled triangle:

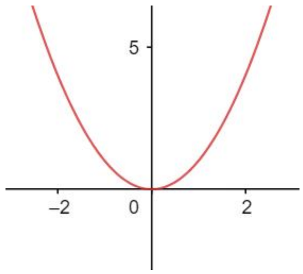


For a given angle ***θ*** each ratio stays the same
no matter how big or small the triangle is

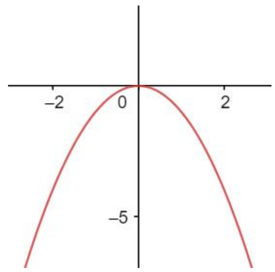
Parent Functions



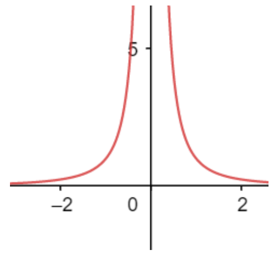
$f(x) = x$
Linear



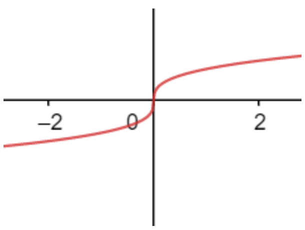
$f(x) = x^2$
quadratic



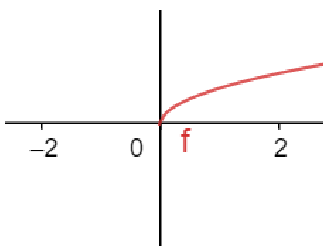
$f(x) = -x^2$



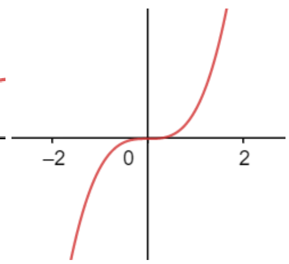
$f(x) = \dfrac{1}{x^2}$



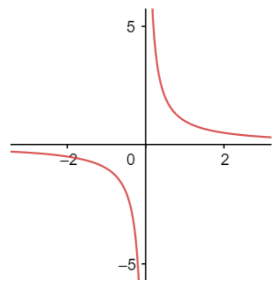
$f(x) = \sqrt[3]{x}$
cube root



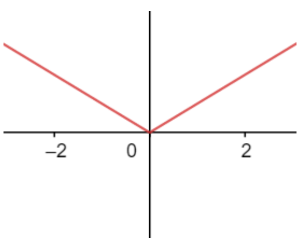
$f(x) = \sqrt{x}$
square root



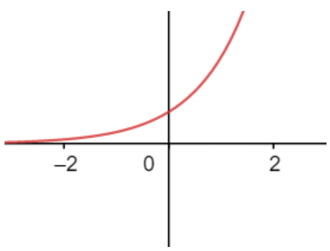
$f(x) = x^3$
Cubic



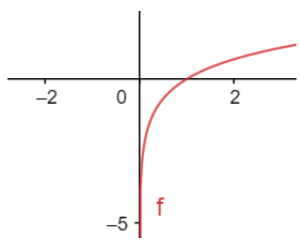
$f(x) = \dfrac{1}{x}$



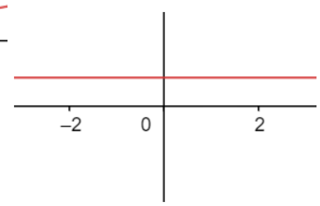
$f(x) = |x|$
Absolute Value



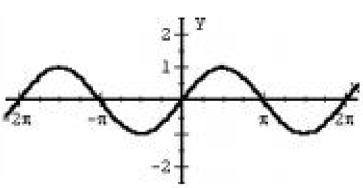
$f(x) = e^x$
exponential



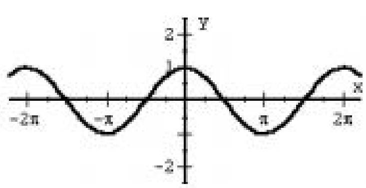
$f(x) = \ln(x)$
natural logarithmic



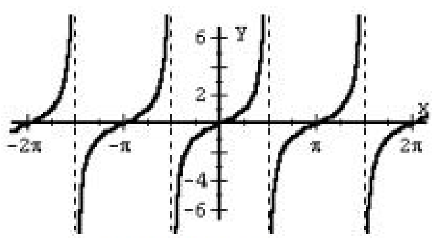
$f(x) = 1$
constant



f(x) = sin x

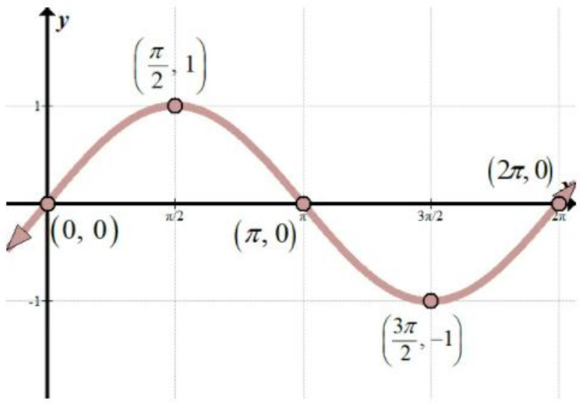


f(x) = cos x

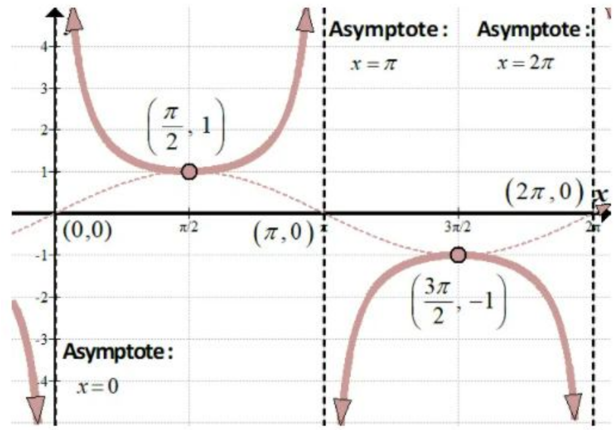


f(x) = tan x

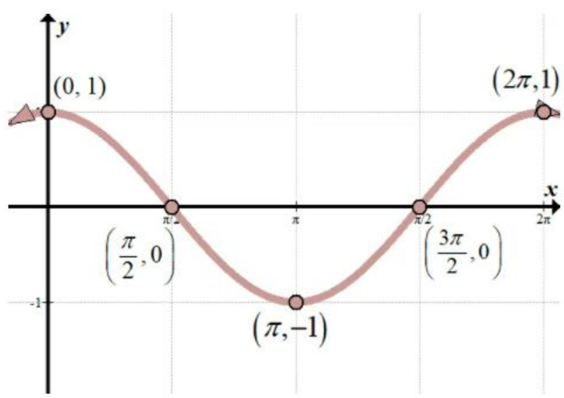
Trigonometric Functions



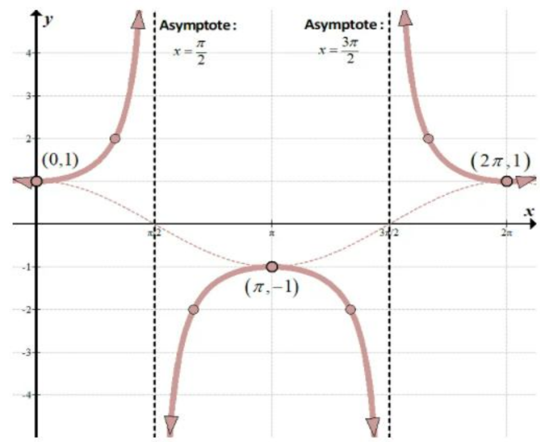
$f(x) = \sin(x)$



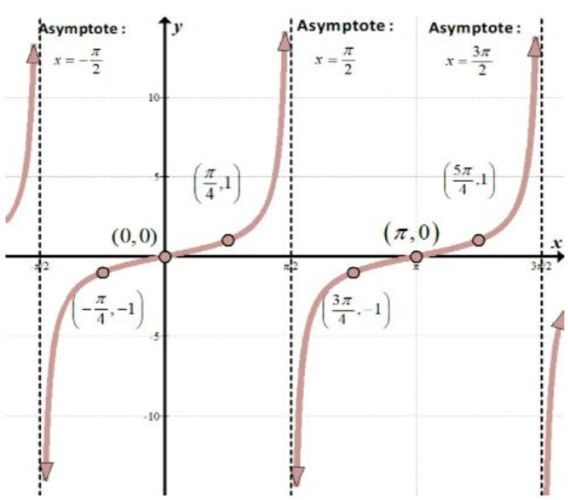
$f(x) = \csc(x) = \dfrac{1}{\sin(x)}$



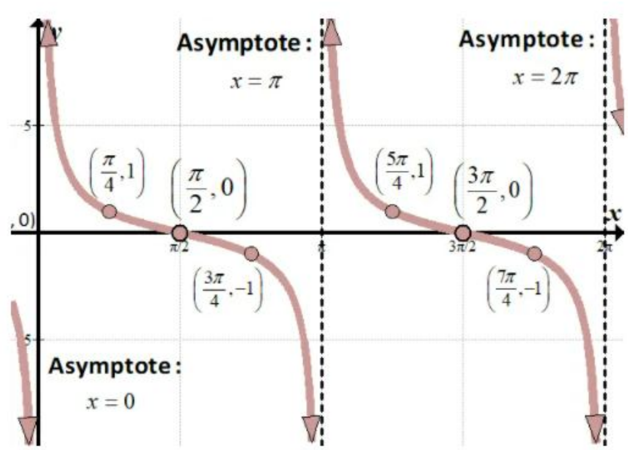
$f(x) = \cos(x)$



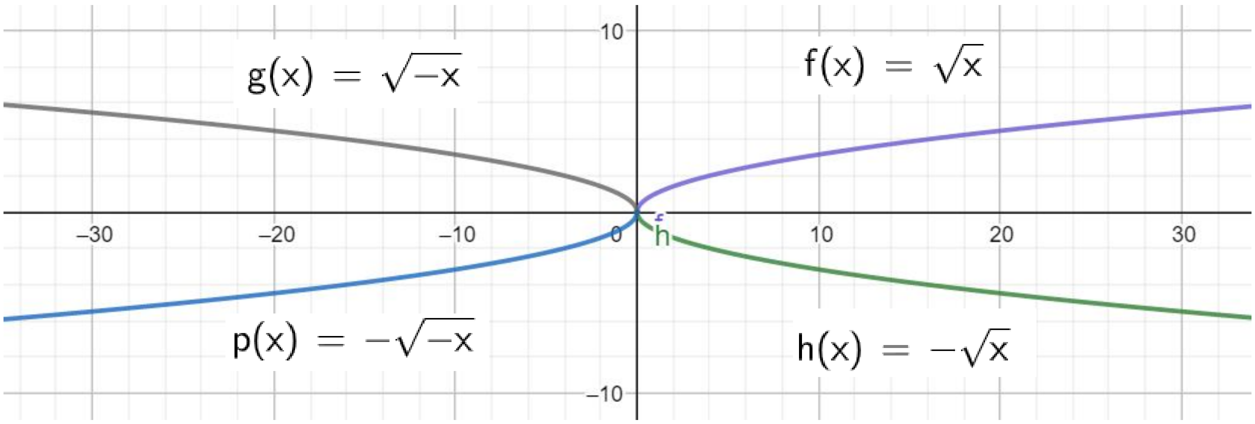
$f(x) = \sec(x) = \dfrac{1}{\cos(x)}$



$f(x) = \tan(x)$



$f(x) = \cot(x)$



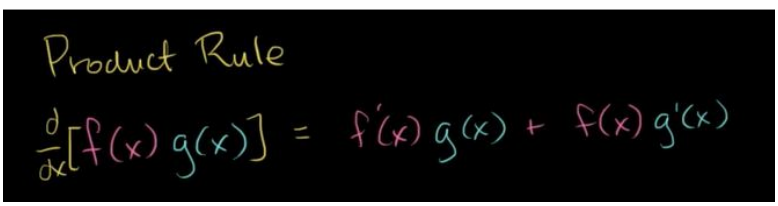
Derivative of e^x: is the only function that is the derivative of itself!

$$f(x) = e^x, \qquad \frac{d}{dx}[e^x] = e^x$$

Derivative of ln(x):

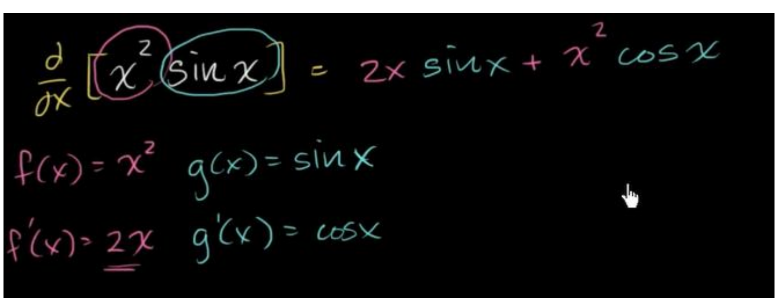
$$f(x) = \ln(x), \qquad \frac{d}{dx}[\ln(x)] = \frac{1}{x}$$

Product rule:



$$\frac{d}{dx}[f(x) \cdot g(x)] = \frac{d}{dx}[f(x)] \cdot g(x) + f(x) \cdot \frac{d}{dx}[g(x)]$$

Example:



Quotient rule:

Quotient Rule

$$f(x) = \frac{u(x)}{v(x)} \qquad f'(x) = \frac{u'(x)\,v(x) - u(x)\,v'(x)}{[v(x)]^2}$$

$$\frac{d}{dx}\left[\frac{f(x)}{g(x)}\right] = \frac{\frac{d}{dx}[f(x)]\cdot g(x) - f(x)\cdot \frac{d}{dx}[g(x)]}{[g(x)]^2}$$

Example:

$$f(x) = \frac{\boxed{x^2}}{\boxed{\cos x}} \quad \begin{array}{l} \leftarrow u(x) \\ u'(x) = 2x \\ \leftarrow v(x) \\ v'(x) = -\sin x \end{array}$$

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

Chain rule

✱ Chain Rule

$$\frac{d}{dx}[f(g(x))] = f'(g(x))\cdot g'(x)$$

$$\frac{d}{dx}\left[f\Big(g(x)\Big)\right] = f'\Big(g(x)\Big)g'(x)$$

Example

$$\frac{d}{dx}[f(g(x))] = \underbrace{f'(g(x))}\cdot g'(x)$$

$$\frac{d}{dx}\Big[\underbrace{\ln(\overbrace{\sin(x)}^{g(x)})}_{f(g(x))}\Big] = \frac{1}{\sin(x)}\cdot\cos(x)$$

$$\frac{d}{dx}\Big[\overbrace{\ln(x)}^{f(x)}\,\overbrace{\sin(x)}^{g(x)}\Big]$$

Product Rule

$$f'(g(x)) = \frac{1}{\sin(x)} \qquad g'(x) = \cos(x)$$

Composite Functions

$f(x) = 1 + x \qquad g(x) = \cos(x)$

$f(g(x)) = 1 + \cos(x)$

$x \to \boxed{g} \xrightarrow{g(x)} \boxed{f} \xrightarrow{f(g(x))}$

A function is *composite* if you can write it as $f(g(x))$. In other words, it is a function within a function, or a function of a function.

For example, $\cos(x^2)$ is composite, because if we let $f(x) = \cos(x)$ and $g(x) = x^2$, then $\cos(x^2) = f(g(x))$.

g is the function within f, so we call g the "inner" function and f the "outer" function.

$$\underbrace{\cos(\overbrace{x^2}^{\text{inner}})}_{\text{outer}}$$

Example:

$$h(x) = \underbrace{(\overbrace{5-6x}^{\text{inner}})^5}_{\text{outer}}$$

$$g(x) = 5 - 6x \qquad \text{inner function}$$

$$f(x) = x^5 \qquad \text{outer function}$$

Because h is composite, we can differentiate it using the chain rule:

$$\frac{d}{dx}\left[f\Big(g(x)\Big)\right] = f'\Big(g(x)\Big) \cdot g'(x)$$

Before applying the rule, let's find the derivatives of the inner and outer functions:

$$g'(x) = -6$$

$$f'(x) = 5x^4$$

Now let's apply the chain rule:

$$\frac{d}{dx}\left[f\Big(g(x)\Big)\right]$$

$$=f'\Big(g(x)\Big)\cdot g'(x)$$

$$=5(5-6x)^4\cdot -6$$

$$=-30(5-6x)^4$$

Derivative of a^x (for any positive base a)

$$\frac{d}{dx}[e^x]=e^x \qquad a=e^{\boxed{\ln a}} \qquad \ln a$$

$$\frac{d}{dx}[a^x]=\frac{d}{dx}\left[\left(e^{\ln a}\right)^x\right]=\frac{d}{dx}\left[e^{\boxed{(\ln a)x}}\right]=e^{(\ln a)x}\cdot\ln a$$

$$=(\ln a)\big(\underbrace{e^{\ln a}}_{a}\big)^x=(\ln a)\,a^x$$

example:

$$\frac{d}{dx}[8\cdot 3^x]=8\cdot(\ln 3)\cdot 3^x=(8\ln 3)\cdot 3^x$$

Derivative of $\log_a x$ (for any positive base $a\neq 1$)

$$\frac{d}{dx}[\ln x]=\frac{1}{x} \qquad \boxed{\log_a b=\frac{\log_c b}{\log_c a}}$$

$$\frac{d}{dx}[\log_a x]=\frac{d}{dx}\left[\frac{1}{\ln a}\cdot\ln x\right] \qquad \log_{\square}100=2$$

$$\log 100$$

$$\frac{1}{\ln a}\cdot\cancel{\frac{d}{dx}[\ln x]}^{\frac{1}{x}}=\frac{1}{(\ln a)x} \qquad \ln x=\log_e x$$

$$\log_3 8=\frac{\log_{10}8}{\log_{10}3}=\frac{\ln 8}{\ln 3}$$

Example:

$$f(x)=\log_7 x$$

$$f'(x)=\frac{1}{(\ln 7)x}$$

Derivative rules review

Basic Properties and Formulas

If $f(x)$ and $g(x)$ are differentiable functions (the derivative exists), c and n are any real numbers,

1. $(cf)' = cf'(x)$
2. $(f \pm g)' = f'(x) \pm g'(x)$
3. $(fg)' = f'g + fg'$ – **Product Rule**
4. $\left(\frac{f}{g}\right)' = \frac{f'g - fg'}{g^2}$ – **Quotient Rule**
5. $\frac{d}{dx}(c) = 0$
6. $\frac{d}{dx}(x^n) = nx^{n-1}$ – **Power Rule**
7. $\frac{d}{dx}(f(g(x))) = f'(g(x))g'(x)$

   This is the **Chain Rule**

Common Derivatives

$\frac{d}{dx}(x) = 1$

$\frac{d}{dx}(\sin x) = \cos x$

$\frac{d}{dx}(\cos x) = -\sin x$

$\frac{d}{dx}(\tan x) = \sec^2 x$

$\frac{d}{dx}(\sec x) = \sec x \tan x$

$\frac{d}{dx}(\csc x) = -\csc x \cot x$

$\frac{d}{dx}(\cot x) = -\csc^2 x$

$\frac{d}{dx}(\sin^{-1} x) = \frac{1}{\sqrt{1-x^2}}$

$\frac{d}{dx}(\cos^{-1} x) = -\frac{1}{\sqrt{1-x^2}}$

$\frac{d}{dx}(\tan^{-1} x) = \frac{1}{1+x^2}$

$\frac{d}{dx}(a^x) = a^x \ln(a)$

$\frac{d}{dx}(\mathbf{e}^x) = \mathbf{e}^x$

$\frac{d}{dx}(\ln(x)) = \frac{1}{x}, \ x > 0$

$\frac{d}{dx}(\ln|x|) = \frac{1}{x}, \ x \neq 0$

$\frac{d}{dx}(\log_a(x)) = \frac{1}{x \ln a}, \ x > 0$

Derivative of exponential functions

$$\frac{d}{dx}(e^x) = e^x \ln e = e^x(1) = e^x \qquad \frac{d}{dx}(a^x) = a^x \ln a$$

Derivatives of inverse trig functions

$$\frac{d}{dx}\left[a \sin^{-1}(f(x))\right] = \frac{a \cdot f'(x)}{\sqrt{1 - [f(x)]^2}} \qquad \frac{d}{dx}\left[a \csc^{-1}(f(x))\right] = -\frac{a \cdot f'(x)}{|f(x)|\sqrt{[f(x)]^2 - 1}}$$

$\tan^2 x + 1 = \sec^2 x$, $\cos^2 x + \sin^2 x = 1$, $\sec x = 1/\cos x$

Taylor series

Application: Taylor series expansion represents a function as a series, an infinite sum, of terms that are calculated using the function's derivatives at a point:

$$f(x) = \frac{f(a)}{0!}(x-a)^0 + \frac{f'(a)}{1!}(x-a)^1 + \frac{f''(a)}{2!}(x-a)^2 + \frac{f'''(a)}{3!}(x-a)^3 + \ldots$$

Note that the first term is just the function calculated at the point a, it is the 0th order derivative of the function and is called the 0th order term (all terms are referred to by the order of the derivative).

The Taylor series, in the LaGrange remainder formula, for $f(x)$ in a neighborhood of the value a is

$$f(x) = \sum_{k=0}^{n-1}\left[\frac{f^{(k)}(a)(x-a)^k}{k!}\right] + R_n \quad R_n = \frac{f^{(n)}(c)(x-a)^n}{n!}$$

A good example of the use of the Taylor series expansion is that for $f(x) = e^x$ around the point $a = 0$ (a Taylor series expansion around $a = 0$ is known as a Maclaurin series). Recall $f'(e^x) = e^x$, $e^0 = 1$ then

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \ldots + \frac{x^{n-1}}{(n-1)!} + R_n$$

$R_n = e^c x^n/n!$ and $c \in (0, x)$. The remainder term can be made as small as desired by taking larger n so that the approximation of e^x can be fixed as precise as desired.

Taylor series in two variables

The Taylor series expansion can be extended to 2 or more dimensions, we consider only $\mathbb{R}^2$ here. Consider the case of a function $f(x, y)$. The Taylor series around a point (a, b) gives an approximation of the function in the neighborhood of (a, b) as follows (using Leibniz' notation for clarity)

$$\begin{aligned} f(x,y) = f(a,b) &+ \frac{\partial f(a,b)}{\partial x}(x-a) + \frac{\partial f(a,b)}{\partial y}(y-b) + \\ &+ \frac{1}{2}\left[\frac{\partial^2 f(a,b)}{\partial x^2}(x-a)^2 + \frac{\partial^2 f(a,b)}{\partial x \partial y}(x-a)(y-b) + \frac{\partial^2 f(a,b)}{\partial y^2}(y-b)^2\right] + \ldots \end{aligned}$$

The general expression for the Taylor series in 2 variables can be written as

$$f(x,y) = \sum_{k=0}^{\infty}\left(\frac{1}{n!}\sum_{k=0}^{n}\binom{n}{k}\frac{\partial^n f(a,b)}{\partial x^{n-k}\partial y^k}(x-a)^{n-k}(y-b)^k\right) \qquad \binom{n}{k} = \frac{n!}{(n-k)!k!}$$

An example. Let

$$f(x, y) = x^3 + 3y - y^3 - 3x.$$

We find the second-degree Taylor polynomial as

$$\begin{aligned} f(x,y) \approx\ & f(a,b) + f_x(a,b)(x-a) + f_y(a,b)(y-b) + \\ & + \frac{1}{2}\left(f_{xx}(a,b)(x-a)^2 + 2f_{xy}(a,b)(x-a)(y-b) + f_{yy}(a,b)(y-b)^2\right) \end{aligned}$$

Using the point (2,1) for the approximation we have

$$\begin{aligned} & f(2,1) = 8 + 3 - 1 - 6 = 4 \\ & f_x = 3x^2 - 3 = 9 \qquad f_{xx} = 6x = 12 \\ & f_y = 3 - 3y^2 = 0 \qquad f_{yy} = -6y = -6 \\ & \qquad\qquad\qquad\qquad\quad f_{xy} = 0 \end{aligned}$$

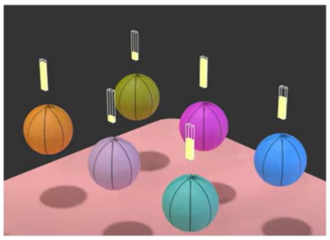
and

$$\begin{aligned} f(x,y) &\approx 4 + 9(x-2) + 0(y-1) + \frac{1}{2}\left(12(x-2)^2 + 2 \cdot 0 - 6(y-1)^2\right) \\ &\approx 4 + 9(x-2) + 6(x-2)^2 - 3(y-1)^2 \end{aligned}$$

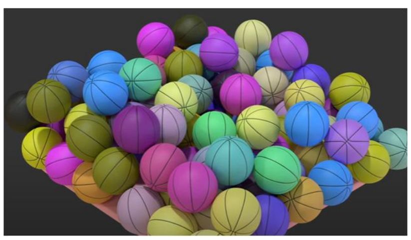
What is Optimization:

https://www.youtube.com/watch?v=d0CF3d5aEGc&ab_channel=VisuallyExplained

Whenever you have many options that you can choose from, while each option had a cost, your job to make the optimal chois (you need to pick the option with the smallest possible cost) sometimes we want to choose the option which maximizes like profit for example, and if we choose the maximum and we want to convert to minimum we multiply by (x-1) that convert the maximization problem to minimization problem.

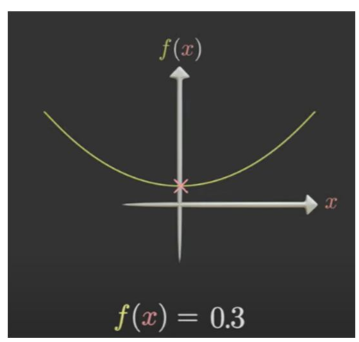


The number of options could be infinite.



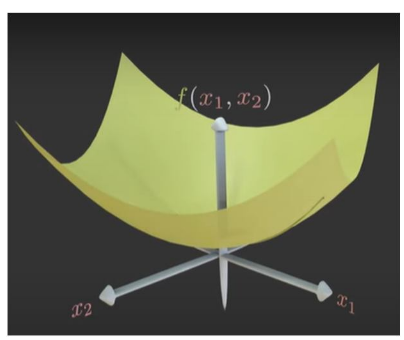
Try to pick a real number x that minimizes some given function f(x), try to pick a scaler x that refers as a decision variable such that the f(x) is as small as possible.

This in one dimensional scale



But it could also be multi-dimensional.

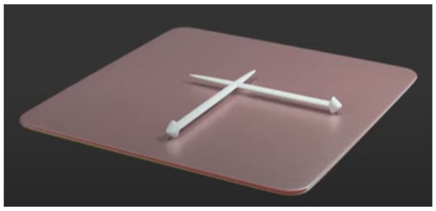
It could be 2 Dimensional x1, x2.



And 3 dimensional x1, x2, x3 and more...

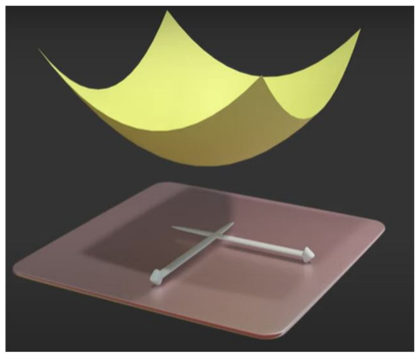
The optimization problem is given by 3 ingredients.

1- **The set** where our decisions lives **Decision Variable** $\boldsymbol{x} \in \boldsymbol{R^n}$

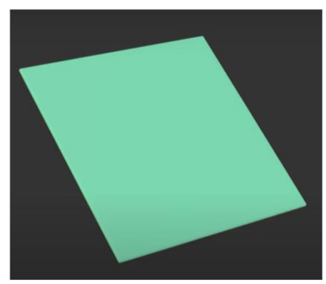


2- **Cost function f** that we want to minimize $f: R^n \rightarrow R$, this function de-set where our decision variable lives to R,

This function is known sometimes as the objective function.



3- **Constrains** of the decision variable

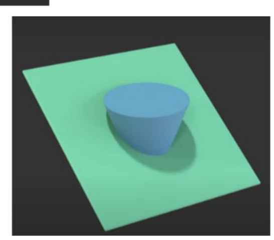


a. **Equality Constraints** $h_i(x) = 0 \quad i = 1, \ldots$

$$\text{Example: } x_1 + x_2 + x_3 = 0$$

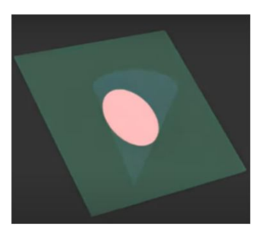
b. **Inequality constraints** $g_j(x) \leq 0 \quad j = 1, \ldots$

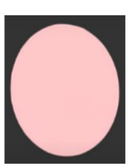
$$\text{Example: } x_1^2 + x_2^2 \leq x_3^2$$



The equality and inequality constraints together define which known as:

the visible set or the set of possibilities where you can pick the decision variable x1





optimization problem are always written in this form: $\min_{x \in R^n} f(x)$

$$\min_{x \in \mathbb{R}^n} f(x) \qquad \begin{aligned} h_i(x) &= 0 \quad i = 1, \dots \\ g_j(x) &\leq 0 \quad j = 1, \dots \end{aligned}$$

Examples: When the cost function f is a linear function

$$\min_{x \in \mathbb{R}^n} f(x) \longrightarrow \text{Linear}$$

And the functions $h_i(x)$ and $g_j(x)$ as linear function as well

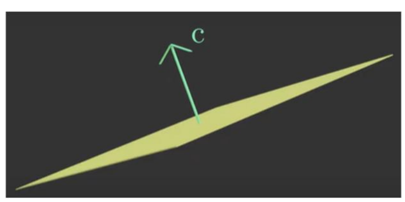
$$\begin{aligned} \text{Linear} &\longleftarrow h_i(x) = 0 \quad i = 1, \dots \\ \text{Linear} &\longleftarrow g_j(x) \leq 0 \quad j = 1, \dots \end{aligned}$$

Then we have a linear program.

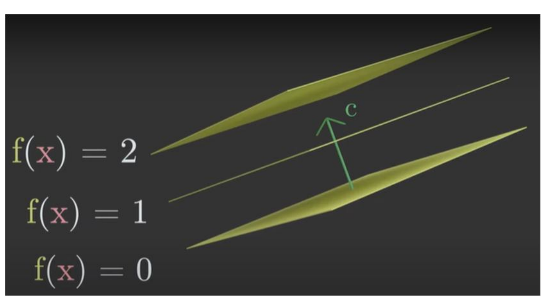
Visualize Linear functions like this one:

$$f(x) = \underset{\in \mathbb{R}^n}{\underset{\uparrow}{c}}{}^T x + \underset{\in \mathbb{R}}{\underset{\uparrow}{d}}$$

Represented hyperplane with equation **f(x)=0** or with the vector c that is orthogonal to this hyperplane

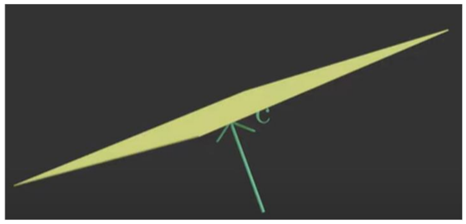


The hyperplane representation is useful for visualizing where the linear function f takes some given value like f(x) takes 0 or 1 or 2



In the other hand the vector representation is useful for understanding the direction that increases or decreases this linear function.

f(x) = -1.9

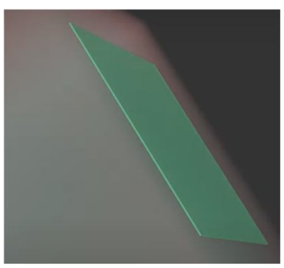


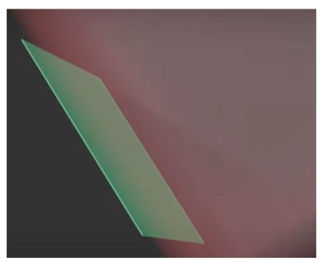
For linear cost functions we prefer the vector representation, the reason is that in the absence of constraints and in order to minimize the cost function **f** what you need to do is simply move along the vector -c as much as possible.

Visualize Linear constraints.

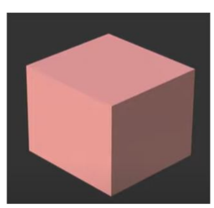
For the linear functions that define linear constraints we prefer the hyperplane representation $x_1 + x_2 + x_3 = 0$ divides the plans in 3 regions:

$x_1 + x_2 + x_3 \geq 0$ Positive, $x_1 + x_2 + x_3 \leq 0$ negative, region itself = 0

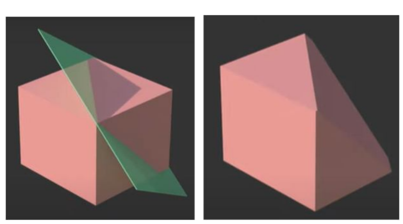




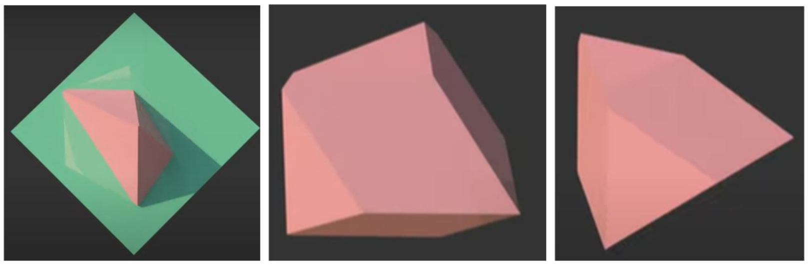
Now if you think of the whole space where your decision variable lives as this cube:



Then a linear constraint for example an inequality constraint will cut out a part of the cube $x_1 + x_2 + x_3 \geq 0$



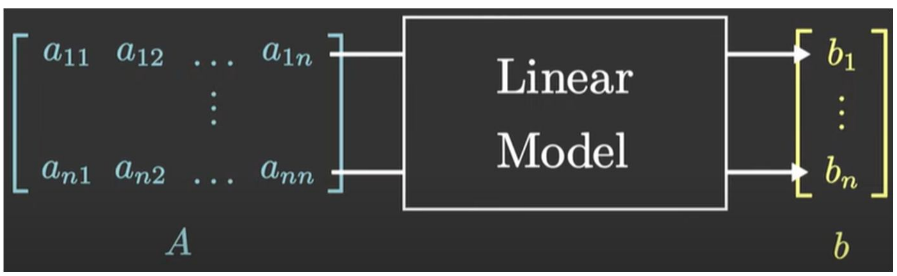
And as you add mor and more linear constraint. More and more parts of the cube are gone be cut off.



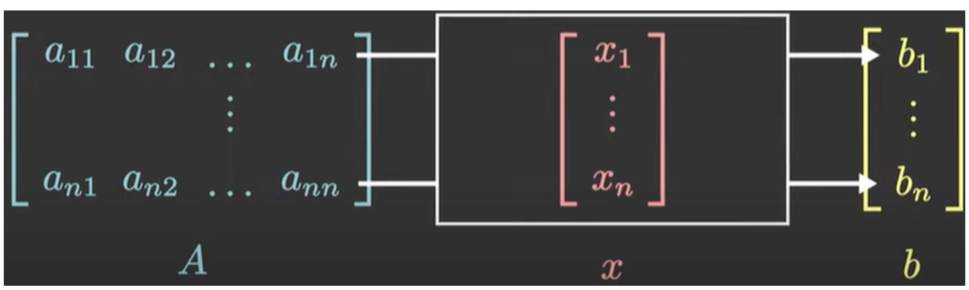
And the region of the space we are left with is our visible region.

Second example of optimization problem: **Linear Regression**

Here we try to model the relation between **inputs A** and **outputs b,** we try to fit a Linear model



So our decision variable is a vector of weight



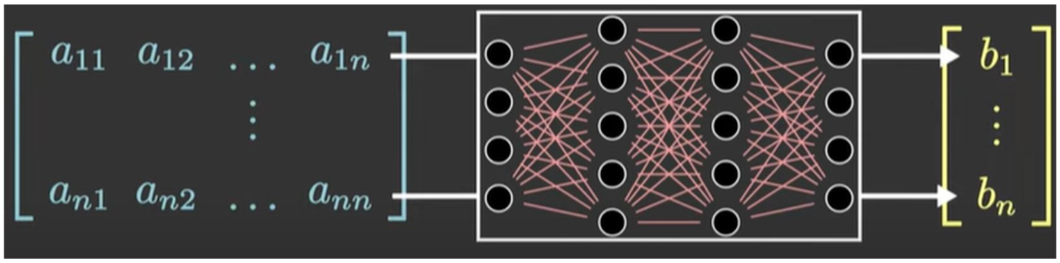
Our objective function measures the **error** in our linear model.

$$\min || A x - b ||^2$$

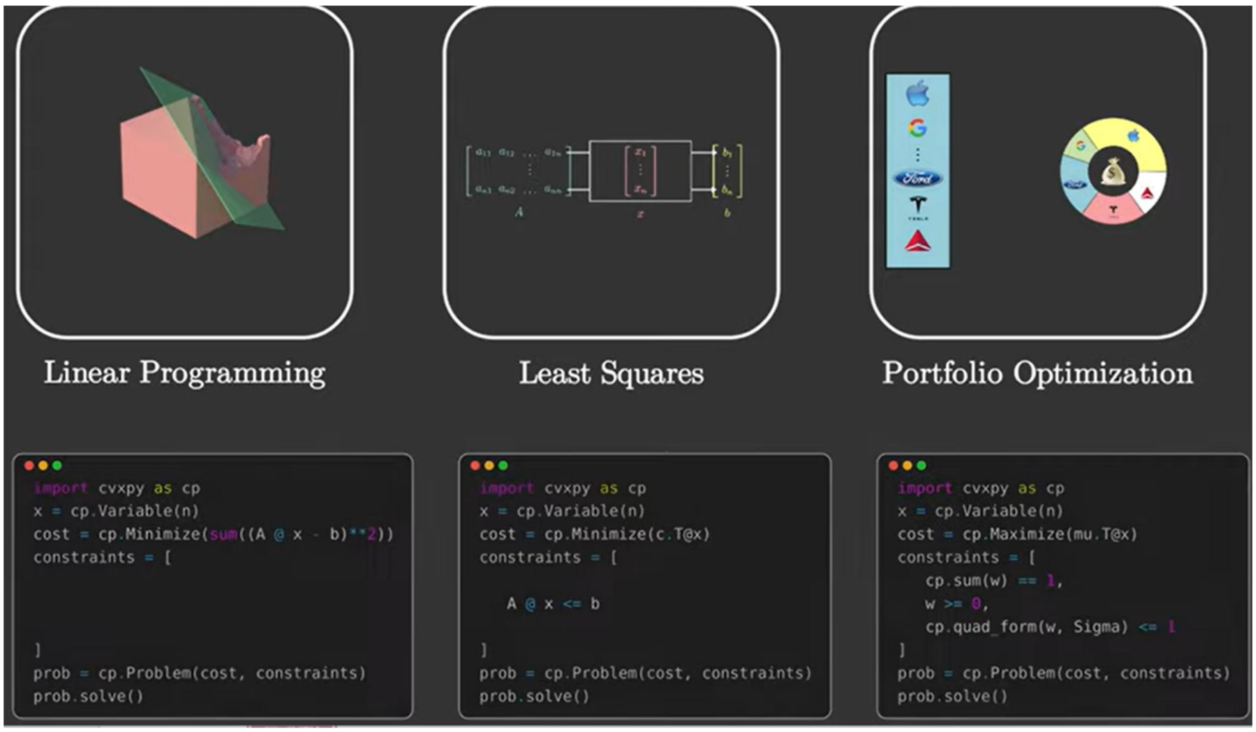
And note that the objective function is not a linear function but a quadratic function.

T

here exist quite few variations of this problem, for example we could replace the linear model with more complicated function like neural networks.



In that case the regression problem becomes considerably a hardware problem

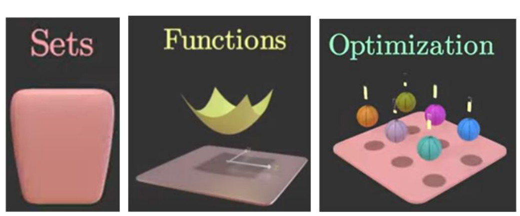


The common property that's allows us to solve the optimization problems in a unified fashion.

Convex Optimization

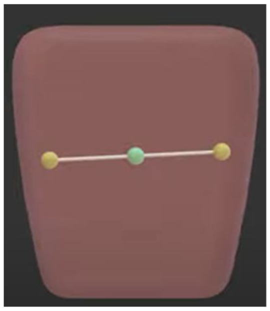
What is convexity?

In 3 different contexts:

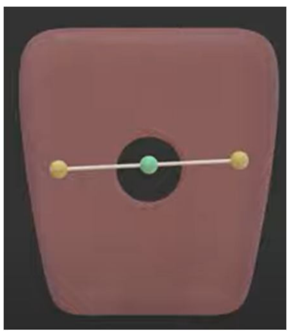
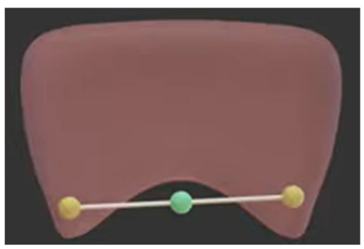


1- **Definition relates to Sets**:

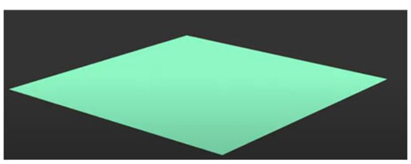
Set is convex for any pair of points inside the set if the segment between the two points falls entirely inside the set.

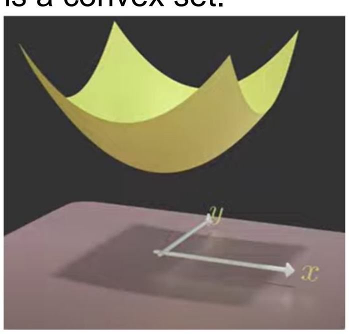
For example this set with a hole inside the set fails to be convex because the middle point falls outside the set.

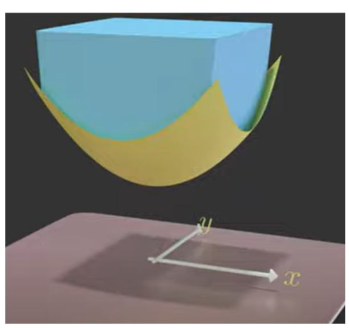



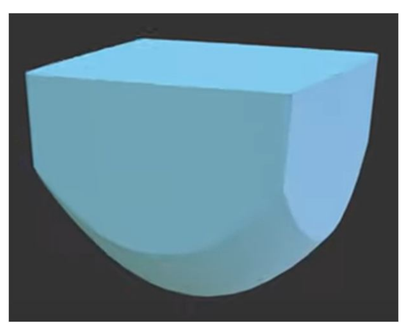
Exact convex set include plains, balls, potato shape,



2- **Definition relates to functions.**
A function f is convex if it's epigraph that is the region above of the graph of the function is a convex set.

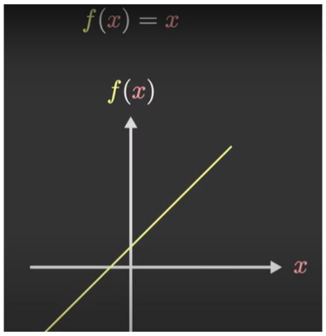


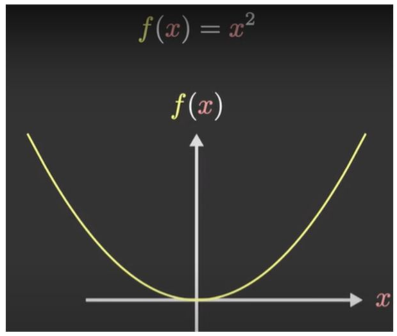


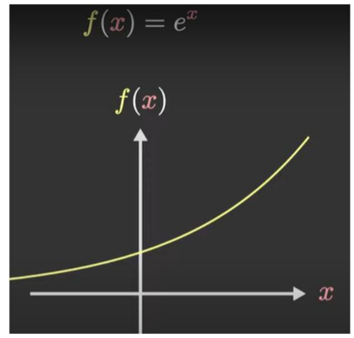


Intuitively that means the function is ball shaped or curved upward.

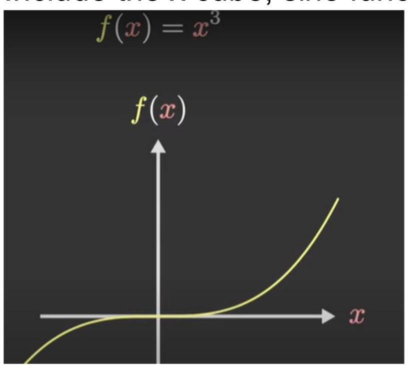
Convex functions include the function x squared, exponential function etc

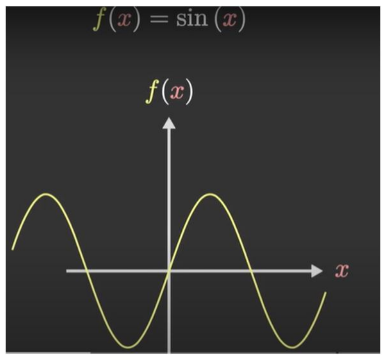






Non-Convex functions
Include the x cube, sine function and etc.





A nice way to construct new complex functions from existing ones, is to take a function that is known to be convex (x^2) and scale it with the positive constant (3) $3\,x^2$ or take two convex functions and add them together.

$$3\,x^2 + e^x$$

3- **Definition relates to Optimization.**
Where one encounters convexity is that of optimization problems.

Optimization problem is said to be convex:

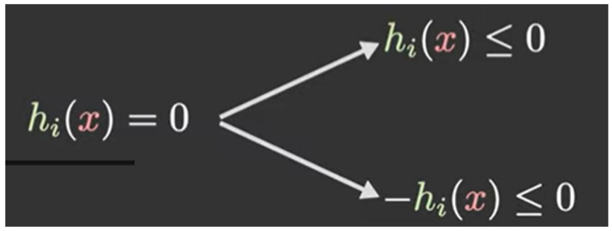
- If the objective function f is convex $\min_{x \in \mathbb{R}^n} f(x) \longrightarrow \text{Convex}$.
- The function gi that give the inequality constraints are convex

$$\text{Convex} \longleftarrow g_j(x) \leq 0 \quad j = 1, \ldots$$

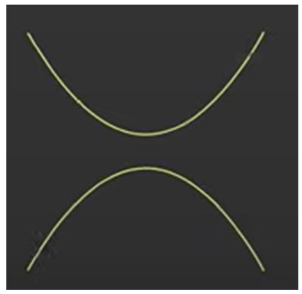
- And the function hi are linear and not just convex.

$$\text{Linear} \longleftarrow h_i(x) = 0 \quad i = 1, \ldots$$

the reason that we said linear and not just convex is that we can replace an equality h equals zero with two inequalities is h equal or smaller zero and -h equal or smaller than 0



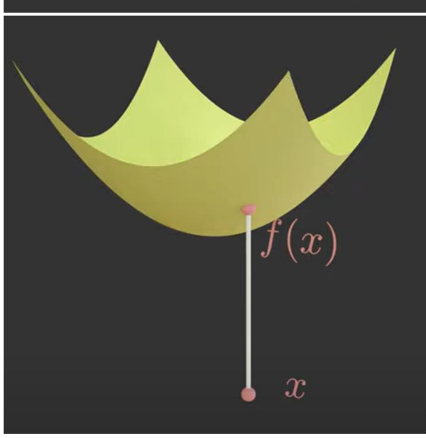
The only function for which h and – h are both convex are the linear ones.

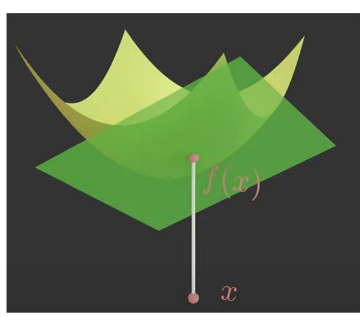


Convexity generalizes the notion of just linearity

If we want to talk about local behavior of a function we have to mention taylor's approximation theorem.
Taylor's in simplest form: if you zoom in on a function around the point x, then it will look a lot like a linear function, or in other words the line or in higher dimension the hyperplane given by this equation is a good approximation of the graph of f around the point x

$$f(y) \approx f(x) + \nabla f(x)^T (y - x)$$



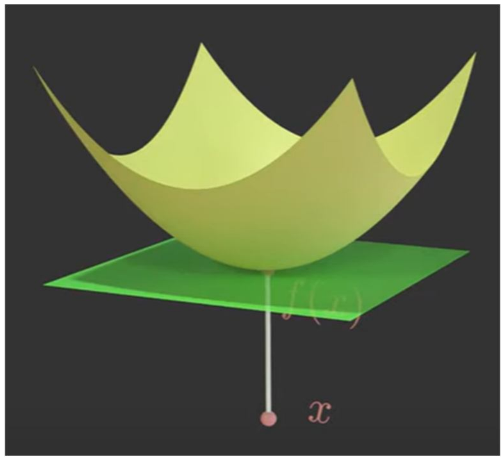


This hyperplane is called tangent hyperplanes.
The dual definition of the convex function is as follows:

- f is convex if and only if its graph is always above its tangent hyperplanes.

$$f(y) \geq f(x) + \nabla f(x)^T (y - x)$$

Let say you find a point f where the gradient of f is equal to zero, then the tangent hyperplane at that point is horizontal



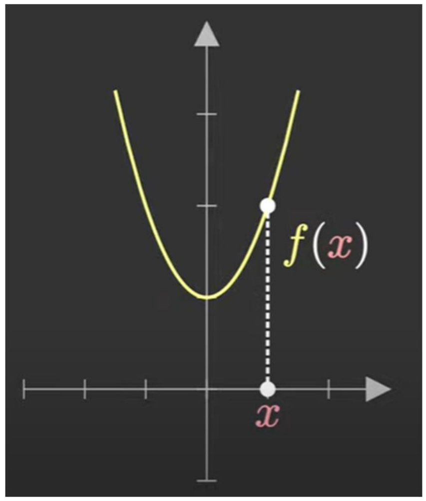
And since convex function will always be above their tangent planes, this mean that the point x is actually a global minimizer of f

And this already gives us a way to solve:

Unconstrained minimization problems.

Take the function x squared plus one as an example

$$f(x) = x^2 + 1 \quad , \quad f'(x) = 2x$$



- this function is convex
- and its gradient or derivative is 2 times x, so when x =0 the derivative equal to 0
- so x=0 is where this function attains its minimum

another example of unconstrained optimization problem is the last problem we saw earlier

$$\min \| A x - b \|^2$$

To find its minimum we simply set the gradient to 0

$$\nabla \| A x - b \|^2 = 0$$

And then we solve for x

$$2A^T (A x - b) = 0$$

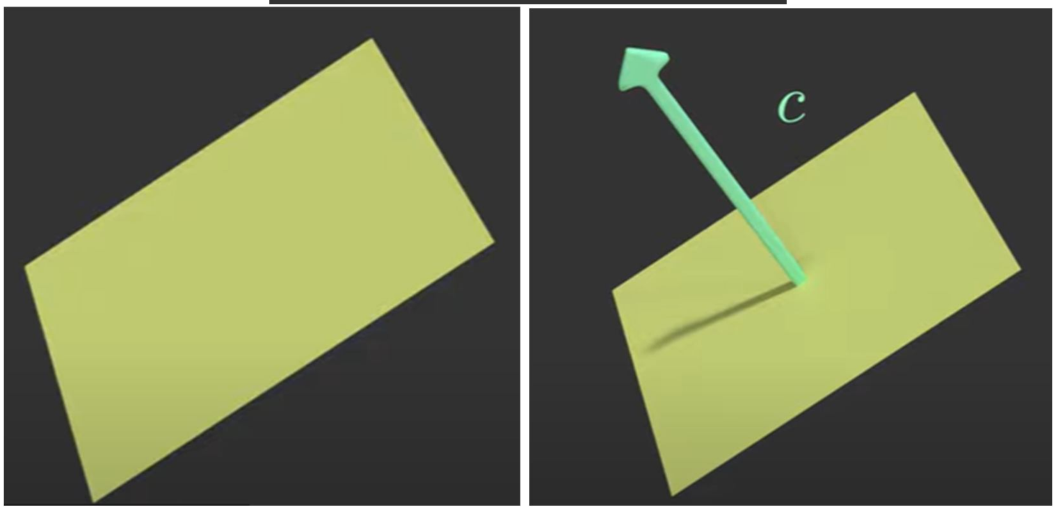
$$A^T A x - A^T b = 0$$

$$A^T A x = - A^T b$$

$$x = - (A^T A)^{-1} A^T b$$

The third example non-constant linear function

$$f(x) = c^T x + d$$



The gradient of that function is a constant given by a non-zero constant vector, and since this gradient can never vanish we can conclude that this linear function doesn't have a minimizer, and indeed it is not hard to see that non-constant linear functions are unbounded

$$\nabla f(x) = c$$

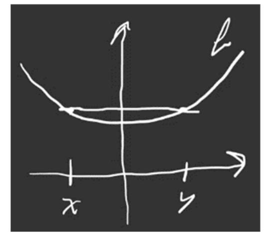
Convex function >> second order derivative is positive (non-negative), that mean it's gradient always increasing.

Convex but not strictly convex, if picked 2 points in the constant line will be equal not strictly below.

0th order:

Convex

The graph under any line between 2 points in the function



$f(\theta x + (1-\theta)y) \leq \ \theta f(x) + (1-\theta) f(x)$

Strictly convex

Same like convex but sharp inequality

$(\theta x + (1-\theta)y) < \ \theta f(x) + (1-\theta) f(x)$

Alpha α-convex

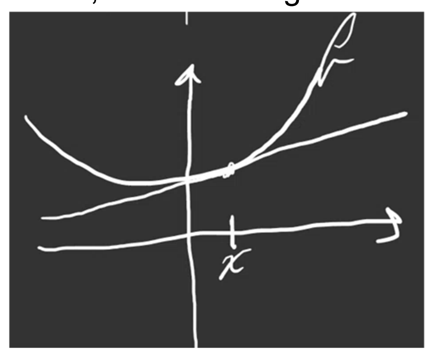
Not only below the line but below the line minus some distance between y and x (the graph is even more below the line)

$(\theta x + (1-\theta)y) < \ \theta f(x) + (1-\theta) f(x) - \frac{\alpha}{2}||y-x||^2$

1st order

Convex

the tangent is below the graph (so take any points in the function, draw the tangent it will be below the graph)



$f(y) \geq \ f(x) + f^{`}(x,y)$

Strictly convex

$f(y) > \ f(x) + f^{`}(x,y)$

Alpha α-convex

$f(y) > \ f(x) + f^{`}(x,y) + \frac{\alpha}{2}||y-x||^2$

2nd order (U-Shaped)

Convex

$\forall x \in k \qquad f^{``}(x,y,y) \geq 0$

Strictly convex

$\forall x \in k \qquad f^{``}(x,y,y) > 0$

Alpha α-convex

$\forall x \in k \qquad f^{``}(x,y,y) > 0 + \ \alpha||y||^2$

Gateaux derivative (directional derivative)

$$f^{`}(x,y) \coloneqq \lim_{\varepsilon \to 0} \frac{f(x+\varepsilon y) - f(x)}{\varepsilon}$$

If $f^{`}(x,y)$ exist for all ∀ y, and derivative is linear in y.

We can apply Riese Representation Theorem which say: $f^{`}(x,y) = < \nabla f(x), y >$

When we want to calculate the derivative normally no need to do the limit, just calculate the gradient and take inner product with y

Minima:

$$\min_{x \in k} f(x)$$

Existence and Uniqueness of solutions of minimization problems f(x), x∈k:

Existence we should have:

- The set we want to minimize k needs to be closed, non-empty.
- f needs to be continuous, infinite at infinity in any direction.
    - Inf at inf: $f(x) \to +\infty \; as \; ||x|| \to +\infty$
    - As the norm ||x|| goes to infinity the function goes to infinity
    - example: $f(x) = e^x + y^2, \gg \propto \begin{pmatrix} -1 \\ 0 \end{pmatrix} \gg f\left(\propto \begin{pmatrix} -1 \\ 0 \end{pmatrix}\right) \gg 0$ not infinite at infinity

Uniqueness:

- k convex.
- f strictly convex.

  Convex x, y ∈ k, θ ∈ [0,1], Then αX+(1-α)Y ∈ k

Charactarise minima (if we have a minimum already)

- k is open (you can travel in every direction and you will still inside)
- k is gateaux differentiable

  then for X* is minimum it means gradient $\nabla f(x^*) = 0 \; (euler \; equality) \; or$

  $$< \nabla f(x^*), (y - x^*) > = 0$$

  That means if it is a minimum, no matter in which direction you go, you are going to increase the function.

  This call Euler's equality

  On the other hand, Euler's inequality

- k is convex and
- f is Gateaux differentiable

  then for X* is minimum it means $\forall y \in k \; < \nabla f(x^*), (y - x^*) > \geq 0$ positive.

2nd order

- k is open and
- f is 2nd order differentiable

  then for X* is minimum then (the gradient is 0, and hessian is positive)

  $\nabla f(x^*) = 0 \, , \nabla^2 f(x^*, y) \; is \; positive$ means eigen values $\lambda_i \geq 0 \; positive$

Constraint optimization:

Inequality constraints

$$\min_{g_i(x)=0} f(x)$$

We make it large in a way then it will be easy to solve we use Lagrange. We use the original function adding the logrunner multiplier multiplied by the constraint g

$$\mathcal{L}(x,\lambda) = f(x) + \sum_{i=1} \lambda_i g_i(x)$$

When we solve that it will solve the original problem

Lagrange $\mathcal{L}(\boldsymbol{x},\boldsymbol{\lambda})$

- f is differentiable, g_i are regular constraint (the gradient of ∇g_i linearly independent)
  that means for constant a , $a_i \in \mathbb{R}$ $\sum_i a_i \nabla g_i(x) = 0 \iff a_i = 0\ \forall i$
  if these are met (f is diff, gi are regular) then:
  then (x,λ) is critical point of the Lagrange. $\iff$
  $$g_i(x) = 0$$
  $$\nabla f(x) + \sum_i \lambda_i \nabla g_i(x) = 0$$
  Same as (x, λ) $is\ saddle\ point\ of\ lagrange$ So, we can solve this system of equations, it will give us candidates, we test individually to find the minimum of the original problem

Inequality constraints:

$$\min_{g_i(x)\le 0} f(x)$$

- f and g are differentiable, convex.
- K closed, and convex
- $there\ exist\ \exists x : s.t.\ \ g_i(\bar{x}) < 0\ (or\ g_i(\bar{x}) \le 0\ and\ g_i\ is\ affine$
  then $(x,\lambda) \in \mathbb{R}^n X\, \mathbb{R}_+^m , \lambda_i \ge 0$ (positive) is a saddle point of the Lagrange.
  $\iff$ KT relations, it says.
  $\exists\ \lambda \in \mathbb{R}_+^m, g_i(x) \le 0 , \lambda_i \ge 0$
  $\lambda_i g_i(x) = 0\ \forall i$
  $$\nabla f(x) + \sum_i \lambda_i \nabla g_i(x) = 0$$
  It gives critical points of λ and among them the point that solve the minimization problem.

Gradient algorithms

We want the iterative method so that every time we do an update adding some vector y which gives us an update value. y is the direction in which we are going to search

The idea comes from Taylor expansion using algorithms to find a sequence of points that will lead us to the minimum.

$f(x_{k+1}) = f(x_k + y) \approx f(x_k) + f^{`}(x_k + y) + \cdots$ (Taylor expansion)

We want to make it small as possible.

$$f(x_k) + < \nabla f(x_k), y >$$

We use cushy swats inequality $|< \nabla f(x_k), y >| \leq ||\nabla f(x_k)||.||y||$ (so max when y=$\nabla f(x_k)$ the biggest increase)

Y=$-\nabla f(x_k)$ (the largest size but with minus -)

$x_{k+1} = x_k - \rho_x \nabla f(x_k)$

The algorithm the new point equals the point where we be before – a step size ρ_x times the gradient at this point.

To choose the step size we have 2 ways:

Fixed step $\rho_k = \rho$

f $\alpha - convex, and\ f^{`}\ in\ \mu - lipshitz$, then $x_k \to x^* \ (the\ minimum)$ if $0 < \rho < \frac{2\alpha}{\mu^2} \Rightarrow$ (we sure that is converges and we sure if we go long enough we will reach to the minimum)

Lipschitz continuity:

the gradient in x – the gradient in y <= μ times norm of x-y

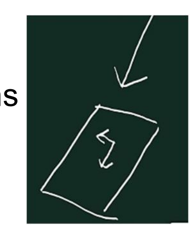
$\left||\nabla f(x) - \nabla f(y)|\right| \leq \mu\left||x - y|\right|$

Where $\mu = max\ \lambda_i, \lambda\ eigenvalues\ of\ hissean\ \nabla^2 f(x)$

Optimal step ρ_k at each iteration we find ρ_k by solving minimization problem.

$$f(x_k - \rho_k \nabla f(x_k)) = \inf_{\rho > 0} f(x_k - \rho \nabla f(x_k))$$

Conjugate gradient (optimal step) Ax=b



You take the best combination of gradients. To solve linear system of equations

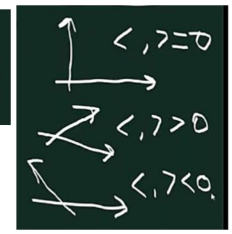
$x_{k+1} = x_k - \sum_i \rho_{ik} \nabla f(x_k)$

Every time we add a gradient to our collection, and we go to the optimal length in this direction, the next step we add another gradient, we make sure it's orthogonal to that direction. Which will go to different dimension in the space up to n

Instead of looking for our line, we will look to a plan. X $\in \mathbb{R}^n$

It converges in n steps (the minimum living somewhere in $\mathbb{R}^n$, every time we are searching in one direction (dimension) more until finding it at the end)

if



- k convex and closed, gradient step with projection
- (the closest point in the set to the point we are projecting) when satisfies

$the\ projection\ \Pi_k(x): < x - \Pi_k(x), \Pi_k(x) - z > \ \leq 0$ (negative)

If you go from that point to anything else in the set, it will have more than 90 degrees angle. $\Pi_k(x)$: $\mathbb{R}^n \rightarrow k$ $(function\ going\ from\ Rn\ to\ k\ it\ returns\ vector\ in\ k)$

What do we use it for? if we have a constraint problem, at every step we can project back into our set.

$x_{k+1} = \ \Pi_k(x_k - \rho_k \nabla f(x_k))$ (if the point inside k then projection is the identity will not do anything)

Practice Exam Exercise:

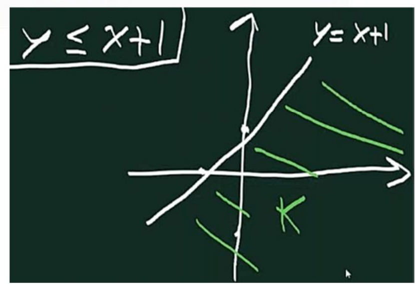
Exercise 1.

We are dealing with the following problem on $\mathbb{R}^2$:

$$(\mathcal{P}) \qquad \min_{x-y\geq -1} x^2+2x+xy+y^2-4y+3.$$

1. Is the set $K=\{(x,y)/x-y\geq -1\}$ closed in $\mathbb{R}^2$? open? convex? (explain in two words for each question), how would you call this set?
2. Is the function $f(x,y)=x^2+2x+xy+y^2-4y+3$ convex? strictly convex? α-convex? (if yes, give the value of α)
3. Justify the existence and uniqueness of the solution of $(\mathcal{P})$.
4. Apply the Kuhn-Tucker theorem : are the hypotheses verified? write the Kuhn-Tucker relations.
5. Solve $(\mathcal{P})$ and give the point where the minimum is reached as well as the minimal value of the function f on K.

1. Is the set $K=\{(x,y)/x-y\geq -1\}$ closed in $\mathbb{R}^2$? open? convex? (explain in two words for each question), how would you call this set?



We can rewrite k to be:

$x-y\geq -1 \Rightarrow x+1\geq y$

Large inequality then k is closed, k is not open continuous at g

Less than or equal that mean the line is included in the set

Closed mean the boundary of k is inside k (because it is soft inequality <=)

K is convex (if we draw a line between any 2 points it will be inside the set)

$x_1, x_2 \in k, \theta\in[0,1], \qquad x_1=\begin{pmatrix}x_1\\y_1\end{pmatrix}, x_2=\begin{pmatrix}x_2\\y_2\end{pmatrix}$

$\theta x_1+(1-\theta)x_2=\theta\begin{pmatrix}x_1\\y_1\end{pmatrix}+(1-\theta)\begin{pmatrix}x_2\\y_2\end{pmatrix}=\begin{pmatrix}\theta x_1+(1-\theta)x_2\\\theta y_1+(1-\theta)y_2\end{pmatrix}$

$\gg$ *we want to know if it's in k* ? Is x-y >= -1?>

$(\theta x_1+(1-\theta)x_2)-(\theta y_1+(1-\theta)y_2)=$

Now taking the first minus the second component

$=\theta(x_1-y_1)+(1-\theta)(x_2-y_2)$

And because $x_1-y_1\geq -1, x_2-y_2\geq -1$ we can replace them by -1

$\geq -\theta-(1-\theta)$ >> $\geq -\theta-1+\theta\geq\ -1$

That mean $\theta x_1+(1-\theta)x_2\ \in k$ *then k is convex*

2. Is the function $f(x, y) = x^2 + 2x + xy + y^2 - 4y + 3$ convex ? strictly convex ? α-convex ? (if yes, give the value of α)

To solve it we should calculate the hessian.

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

$$\nabla^2 f(x,y) = \begin{pmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \\ \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$$

Characteristic equation:

det $(A - \lambda \mathrm{I}) = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} - \lambda \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{vmatrix} 2-\lambda & 1 \\ 1 & 2-\lambda \end{vmatrix} = (2-\lambda)^2 - 1$

we can make it like $(2-\lambda)^2 - 1^2\ to\ use\ a^2 - b^2 = (a+b)(a-b)$

$$(2-\lambda-1)(2-\lambda+1) = (1-\lambda)(3-\lambda) = 0$$

$$1 - \lambda = 0 \gg \lambda_1 = 1$$

$$3 - \lambda \gg \lambda_2 = 3$$

Eigen values both are positive.

f is $\alpha\ convex\ and\ \alpha = 1\ which\ is\ the\ minimum\ of\ lamdas\ \lambda_i$

$which\ implies\ to\ be \gg strictly\ convex \gg convex$

3. Justify the existence and uniqueness of the solution of $(\mathcal{P})$.

Existence :

- k is closed, non-empty,
- f function is continuous, the function is infinite at infinity.
- f is $\alpha - convex \gg f\ infinite\ at\ infinity$
    - $f(y) \geq f(x) + f`(x, y-x) + \frac{\alpha}{2}||y-x||^2$
    - $linear\ in\ y, so\ if\ \left||y|\right| \to\ +\infty\ , then\ f(y) \to\ +\infty$

Uniqueness:

- k is convex
- f is $\alpha - convex$ >> strictly convex

4. Apply the Kuhn-Tucker theorem : are the hypotheses verified ? write the Kuhn-Tucker relations.

KT hypotheses:

- f and g are differentiable, convex
  $there\ exist\ \exists x: g(x) < 0\ (\leq 0\ if\ \ g\ affine)$

  Already done.
  f is twice differentiable
  from the question definition the constraint: x-y+1≥0

  g(x,y) = y-x-1≤ 0 >> g(0,0)=-1 <0
  KT: g(x,y) ≤ 0 , $\lambda \geq 0, \to \nabla^2 g = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \to convex$

  g(x,y) ≤ 0
  $\nabla f(x,y) + \lambda \nabla g(x,y) = 0$ the gradient of f + lambda gradient of g = 0
  $\lambda g(x,y) = 0$

5. Solve $(\mathcal{P})$ and give the point where the minimum is reached as well as the minimal value of the function f on K.

$\nabla f(x,y) + \lambda \nabla g(x,y) = 0$

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

$g(x,y) = y - x - 1$

$$\nabla g(x,y) = \begin{pmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

$\nabla f(x,y) + \lambda \nabla g(x,y) = 0$

$$\begin{pmatrix} 2x+2+y \\ x+2y-4 \end{pmatrix} + \lambda \begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$2x+2+y-\lambda = 0$, $x+2y-4+\lambda = 0$

$\lambda g(x,y) = 0$, $g(x,y) = y-x-1$

$\lambda(y-x-1) = 0$

Case $\lambda \neq 0$ >> y = x+1

$2x+2+y-\lambda = 0$, >> $2x+x+1+2-\lambda = 0 \gg \lambda = 3x+3$

$x+2y-4+\lambda = 0$, >>

$2(x+1)+x-4+3x+3 = 0 \gg 6x+1 = 0 \gg x = -1/6$ >> $(x, y, \lambda) = (-\frac{1}{6}, \frac{5}{6}, \frac{5}{2})$

Case $\lambda = 0$

$2x+y+2 = 0 \gg y = -2x-2$

$2(-2x-2)+x-4 = 0 \gg -3x-8 = 0 \gg x = -\frac{8}{3} \gg y = \frac{16}{3} - \frac{6}{3} = 10/3$

$\left(-\frac{8}{3}, \frac{10}{3}, 0\right), \left(-\frac{1}{6}, \frac{5}{6}, \frac{5}{2}\right)$

$g\left(-\frac{8}{3}, \frac{10}{3}\right) = \frac{10}{3} + \frac{8}{3} - 1 = \frac{15}{3} = 5 > 0$ not in K

$g\left(-\frac{1}{6}, \frac{5}{6}\right) = \frac{5}{6} + \frac{1}{6} - 1 = 0 \leq 0$ satisfy g(x,y) $\leq$ 0

Then the minimum $(-\frac{1}{6}, \frac{5}{6}, \frac{5}{2})$, $f\left(-\frac{1}{6}, \frac{5}{6}\right) = -\frac{1}{12}$

Exercise 2.

One considers on $\mathbb{R}^3$ the following minimisation problem : $min\ f(x,y,z)$ under the constraints $g_1(x,y,z)=0$ and $g_2(x,y,z)=0$, with $f(x,y,z)=\frac{1}{2}(x^2+y^2+z^2)$, $g_1(x,y,z)=x+y-1$ and $g_2(x,y,z)=x+z-1$.

1. Study the regularity of the constraints.
2. Write the equations for the optimum point and the Lagrange multipliers.
3. Find the solutions $(x,y,z,\lambda_1,\lambda_2)$ and determine the candidate points for being a minimum and deduce the solution of the problem by giving the point(s) where the minimum is reached and the minimal value of the cost-function.
4. Give a geometric interpretation of this minimisation problem.

1. Study the regularity of the constraints.

Constraints are regular if $\nabla g_i(x) is\ linearly\ independent$

$g_1(x,y,z)=x+y-1 \quad , \ g_2(x,y,z)=x+z-1$

$\nabla g_i(x) is\ linearly\ independent$

$$\nabla g_1(x,y,z)=\begin{pmatrix}\frac{\partial f}{\partial x}\\ \frac{\partial f}{\partial y}\\ \frac{\partial f}{\partial z}\end{pmatrix}=\begin{pmatrix}1\\1\\0\end{pmatrix}, \qquad \nabla g_2(x,y,z)=\begin{pmatrix}\frac{\partial f}{\partial x}\\ \frac{\partial f}{\partial y}\\ \frac{\partial f}{\partial z}\end{pmatrix}=\begin{pmatrix}1\\0\\1\end{pmatrix}$$

Check linearity dependency.

Vi are linearly independent if any linear combination gives zero >> $\sum a_i v_i = 0, only\ if\ all\ of\ the\ scalers\ are\ zero\ , a_i=0\ \forall i$

Example $v_1\begin{pmatrix}1\\2\end{pmatrix}$, $v_2\begin{pmatrix}2\\4\end{pmatrix}$ $are\ linearly\ dependent\ because$

$2v_1 - v_2 = 0$

$a_2 = 2, \quad a_2 = -1$

$$\lambda_1\nabla g_1+\lambda_2\nabla g_2 = \lambda_1\begin{pmatrix}1\\1\\0\end{pmatrix}+\lambda_2\begin{pmatrix}1\\0\\1\end{pmatrix}=\begin{pmatrix}0\\0\\0\end{pmatrix}$$

$\lambda_1+\lambda_2=0$, $\lambda_1=0$, $\lambda_2=0$, that means $\nabla g_1, \nabla g_2\ are\ linearly\ independent.$

2. Write the equations for the optimum point and the Lagrange multipliers.

$g_1(x,y,z) = x + y - 1 = 0 \quad ,$
$g_2(x,y,z) = x + z - 1 = 0$
$f(x,y,z) = \frac{1}{2}(x^2 + y^2 + z^2) = 0$

$$\nabla f(x,y,z) + \lambda_1 \nabla g_1(x,y,z) + \lambda_2 \nabla g_2(x,y,z) = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$$\nabla f + \lambda_1 \nabla g_1 + \lambda_2 \nabla g_2 = \begin{pmatrix} x \\ y \\ z \end{pmatrix} + \lambda_1 \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + \lambda_2 \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

3. Find the solutions $(x, y, z, \lambda_1, \lambda_2)$ and determine the candidate points for being a minimum and deduce the solution of the problem by giving the point(s) where the minimum is reached and the minimal value of the cost-function.

$g_1(x,y,z) = x + y - 1 = 0 \quad , x = 1 - y$
$g_2(x,y,z) = x + z - 1 = 0 \quad , x = 1 - z$
$f(x,y,z) = \frac{1}{2}(x^2 + y^2 + z^2) = 0$

$$\nabla g_1(x,y,z) = \begin{pmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \\ \frac{\partial f}{\partial z} \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \nabla g_2(x,y,z) = \begin{pmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \\ \frac{\partial f}{\partial z} \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \ \mathrm{f}(x,y,z) = \begin{pmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \\ \frac{\partial f}{\partial z} \end{pmatrix} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

$$\nabla f + \lambda_1 \nabla g_1 + \lambda_2 \nabla g_2 = \begin{pmatrix} x \\ y \\ z \end{pmatrix} + \lambda_1 \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + \lambda_2 \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

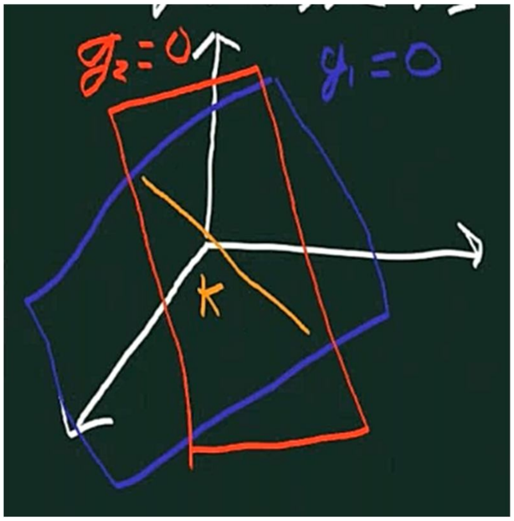
$\mathrm{x} + \lambda_1 + \lambda_2 = 0$
$\mathrm{y} + \lambda_1 = 0 \gg y = -\lambda_1 \gg \lambda_1 = -y = x - 1 \gg y = 1 - x$
$\mathrm{z} + \lambda_2 = 0 \gg z = -\lambda_2 \gg \lambda_2 = -z = x - 1 \gg z = 1 - x$
earlier : $x = 1 - y = 1 - z \gg y = z$
$y = z = -\lambda_1 = -\lambda_2$

$\mathrm{x} + \lambda_1 + \lambda_2 = 0 \gg x + (x-1) + (x-1) = 0 \gg 3x - 2 = 0 \gg x = \frac{2}{3}$
$z = y = -x + 1 = -\frac{2}{3} + 1 = \frac{1}{3} \quad , \lambda_1 = \lambda_2 = -\frac{1}{3}$
$\textit{the minimum point}: (x, y, z, \lambda_1, \lambda_2) = \left(\frac{2}{3}, \frac{1}{3}, \frac{1}{3}, -\frac{1}{3}, -\frac{1}{3}\right),$
$\mathcal{L}(x,y,z) = \frac{1}{2}(x^2 + y^2 + z^2)$
$\mathcal{L}\left(\frac{2}{3}, \frac{1}{3}, \frac{1}{3}\right) = \frac{1}{2}\left(\frac{4}{9} + \frac{1}{9} + \frac{1}{9}\right) = \frac{1}{2}\left(\frac{6}{9}\right) = \frac{1}{2}\left(\frac{2}{3}\right) = \frac{1}{3}$

4. Give a geometric interpretation of this minimisation problem.

$f(x,y,z) = \frac{1}{2}(x^2 + y^2 + z^2) = \frac{1}{2}\left\|\begin{pmatrix} x \\ y \\ z \end{pmatrix}\right\|^2$ (half of the norm gives the distance)

We are in R 3



The intersection between the plan g1, g2 is a line , k is the points where both of the constraints are satisfied, mean intersection.
How can we see the minimization problem given the set k, and the function f?
Which point on k are we looking for that minimizes f?

There is geometric interpretation for this.

$f(x,y,z) = d\left(\begin{pmatrix} x \\ y \\ z \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}\right)$ distance from the point (x,y,z) to (0,0,0)

$X = the\ vector \begin{pmatrix} x \\ y \\ z \end{pmatrix}$

$f(x,y,z) = \frac{1}{2}\left|\left|\begin{pmatrix} x \\ y \\ z \end{pmatrix} - \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}\right|\right|^2$

$g_1, g_2\ are\ places\ in\ \mathbb{R}^3$

The minimization problem is finding the minimal distance from $g_1\ and\ g_2\ or\ k$ to $\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$

Exercise 3.
One considers the minimisation on $\mathbb{R}^3_+$ of the function :

$$g(x, y, z) = x^2 + 2x + y^2 - 4y + z^2.$$

The set of constraints is $K = \mathbb{R}^3_+ = \{(x, y, z)/x \geq 0,\ y \geq 0,\ z \geq 0\}$.

1. Prove that K is convex.
2. Prove that g is α-convex.
3. Prove that the problem has (existence) a unique solution.
4. Characterize the minimum $X^* = (x^*, y^*, z^*)$ of g by an inequality (one will justify the characterization of the minimum with the help of propositions of the course).
5. Prove that the inequality in the previous question should be satisfied for each component independently, and deduce the values of x^*, y^* and z^*.
6. Apply the fixed step gradient algorithm with projection : first give the values of α (such that g is α-convex) and M (such that ∇g is M-Lipschitz). Deduce the interval in which the step ρ should be chosen.
7. We set $\rho = \frac{1}{2}$, and the first guess $X_0 = (0, 0, 0)$. Calculate the iterate X_1 of the gradient algorithm with projection on K.
8. Calculate the iterates X_2 and X_3. Conclude.

1. Prove that K is convex.

K= $\mathbb{R}^3_+$

$$x_1, x_2 \in K = \mathbb{R}^3_+, \theta \in [0,1],\ x_1 = \begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix},\ x_2 = \begin{pmatrix} x_2 \\ y_2 \\ z_2 \end{pmatrix},\ x_1, y_1, z_1,\ x_2,\ y_2,\ z_2 \geq 0$$

$$\theta x_1 + (1-\theta)x_2 = \begin{pmatrix} \theta x_1 + (1-\theta)x_2 \\ \theta y_1 + (1-\theta)y_2 \\ \theta z_1 + (1-\theta)z_2 \end{pmatrix}$$

$since\ \ x_1 \geq 0,\ \ x_2 \geq 0$

$\theta x_1 + (1-\theta)x_2 \geq 0\ , same\ for\ y, z \gg\ \theta x_1 + (1-\theta)x_2\ \in k$, so k is convex

2. Prove that g is α-convex.

$g(x,y,z) = x^2 + 2x + y^2 - 4y + z^2$

Calculate the hessian.

$$\nabla g(x,y,z) = \begin{pmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \\ \frac{\partial f}{\partial z} \end{pmatrix} = \begin{pmatrix} 2x+2 \\ 2y-4 \\ 2z \end{pmatrix},$$

$$\nabla^2 g(x,y,z) = \begin{pmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} & \frac{\partial f}{\partial z} \\ \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} & \frac{\partial f}{\partial z} \\ \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} & \frac{\partial f}{\partial z} \end{pmatrix} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$, the diagonal is eigen values.

$\det(\nabla^2 g(x,y,z) - \lambda \mathrm{I}) = 0$, *but can diagonaly read* $\nabla^2 g$ is diagonal

$\lambda_1 = \lambda_2 = \lambda_3 = 2$

Characteristic equation >> det : $(2-\lambda)^3 = 0$

$\alpha = \min \lambda_i , \lambda_i > 0 , \alpha = \min \lambda_i = 2$, convex

3. Prove that the problem has (existence) a unique solution.

Exist:

K is closed, non-empty (the inequality constraints are soft <=)

F is continuous, infinite at infinity g is alpha convex (<= $\alpha - convex$)

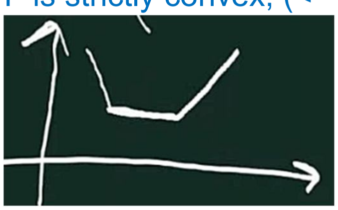
$$g(y) \geq g(x) + g`(x, y-x) + \frac{\alpha}{2}||y-x||^2$$

$so\ when\ ||y|| \rightarrow +\infty, RHS \rightarrow +\infty => g(y) \rightarrow +\infty$

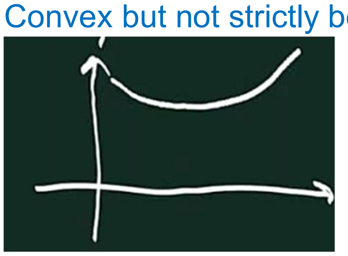
Uniqueness: (must be strictly convex like only one minimum)

K is convex

F is strictly convex, (<= $\alpha - convex$)



Convex but not strictly because there are many minimums.



Strictly convex

Convexity of the set we did earlier

4. Characterize the minimum $X^* = (x^*, y^*, z^*)$ of g by an inequality (one will justify the characterization of the minimum with the help of propositions of the course).

X^* *is* min => *the drivative in any direction y will be positive* $g^{`}(X^*, Y{-}X^*) \geq 0, \forall Y$

The inner product of these 2 vectors should be positive.

$< \nabla g(X^*), Y{-}X^* > \geq 0$

$$\nabla g(x,y,z) = \begin{pmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \\ \frac{\partial f}{\partial z} \end{pmatrix} = \begin{pmatrix} 2x+2 \\ 2y-4 \\ 2z \end{pmatrix} \quad , Y{-}X^* = \begin{pmatrix} x-x^* \\ y-y^* \\ z-z^* \end{pmatrix}$$

5. Prove that the inequality in the previous question should be satisfied for each component independently, and deduce the values of x^*, y^* and z^*.

$$< \nabla g(X^*), Y{-}X^* > = < \begin{pmatrix} 2x^*+2 \\ 2y^*-4 \\ 2z^* \end{pmatrix}, \begin{pmatrix} x-x^* \\ y-y^* \\ z-z^* \end{pmatrix} > =$$

$(2x^*+2)(x-x^*) + (2y^*-4)(y-y^*) + 2z^*(z-z^*) \geq 0, \forall Y$

Inequality should be satisfied for each component. So we will choose a point

$Choose \begin{pmatrix} x \\ y^* \\ z^* \end{pmatrix} => (2x^*+2)(x-x^*) + (2y^*-4)(y^*-y^*) + 2z^*(z^*-z^*) \geq 0$

=> $(2x^*+2)(x-x^*) \geq 0$

Same for others

$Choose \begin{pmatrix} x^* \\ y \\ z^* \end{pmatrix} => (2y^*-4)(y-y^*) \geq 0$

$Choose \begin{pmatrix} x^* \\ y^* \\ z \end{pmatrix} => 2z^*(z-z^*) \geq 0$

Deduce the values of x^*

$\forall x \gg (2x^*+2)(x-x^*) \geq 0$

$x^* = 0 \gg we\ got\ 2X0 + 2 = 2,$

$thta\ mean\ whatever\ positive\ value\ we\ put\ for\ x\ it\ will\ be\ positive$

$\forall y \gg (2y^*-4)(y-y^*) \geq 0$

$y^* = 2 \gg we\ got\ 2X2 - 4 = 4 - 4 = 0,$

$that\ mean\ whatever\ we\ put\ for\ y\ it\ will\ be\ 0$

$\forall z \gg 2z^*(z-z^*) \geq 0$

$z^* = 0 \gg same\ concept$

$$X^* = \begin{pmatrix} 0 \\ 2 \\ 0 \end{pmatrix}$$

6. Apply the fixed step gradient algorithm with projection : first give the values of α (such that g is α-convex) and M (such that ∇g is M-Lipschitz). Deduce the interval in which the step ρ should be chosen.

From first question

$g(x, y, z) = x^2 + 2x + y^2 - 4y + z^2$

Calculate the hessian.

$$\nabla g(x, y, z) = \begin{pmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \\ \frac{\partial f}{\partial z} \end{pmatrix} = \begin{pmatrix} 2x + 2 \\ 2y - 4 \\ 2z \end{pmatrix},$$

$$\nabla^2 g(x, y, z) = \begin{pmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} & \frac{\partial f}{\partial z} \\ \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} & \frac{\partial f}{\partial z} \\ \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} & \frac{\partial f}{\partial z} \end{pmatrix} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$, the diagonal is eigen values

$\lambda_1 = \lambda_2 = \lambda_3 = 2$

Characteristic equation >> det : $(2 - \lambda)^3 = 0$

$\alpha = \min \lambda_i , \lambda_i > 0 , \alpha = 2$

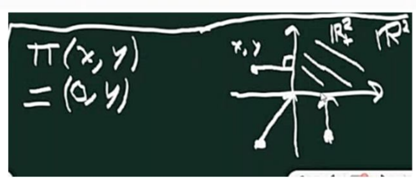
$\alpha = 2 = \min \lambda_i \ , \mu = 2 = \max \lambda_i$

$$0 < \rho < \frac{2\alpha}{\mu^2} = 1$$

7. We set $\rho = \frac{1}{2}$, and the first guess $X_0 = (0, 0, 0)$. Calculate the iterate X_1 of the gradient algorithm with projection on K.

K= $\mathbb{R}_+^3$

$$\text{projection } \Pi_x(x) = \begin{pmatrix} \max(0, x) \\ \max(0, y) \\ \max(0, z) \end{pmatrix}$$



In projection, any negative replaces with 0

The algorithm

$X_{k+1} = X_k - \rho \nabla g(X_k)$

apply the projection (inject in the projection)

$X_{k+1} = \Pi_k(X_k - \rho \nabla g(X_k))$

$$X_k = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \nabla g = \begin{pmatrix} 2x + 2 \\ 2y - 4 \\ 2z \end{pmatrix}$$

$$X_1 = \Pi_k\left(\begin{pmatrix}0\\0\\0\end{pmatrix} - \frac{1}{2}\begin{pmatrix}2\\-1\\0\end{pmatrix}\right) = \Pi_k\begin{pmatrix}-1\\2\\0\end{pmatrix} = \begin{pmatrix}0\\2\\0\end{pmatrix}$$

8. Calculate the iterates X_2 and X_3. Conclude.

Will use the previous point X1

$$X_1 = \begin{pmatrix}0\\2\\0\end{pmatrix} \gg , \nabla g = \begin{pmatrix}2x+2\\2y-4\\2z\end{pmatrix} = \begin{pmatrix}2\\0\\0\end{pmatrix}$$

$$X_2 = \Pi_k\left(\begin{pmatrix}0\\2\\0\end{pmatrix} - \frac{1}{2}\begin{pmatrix}2\\0\\0\end{pmatrix}\right) = \Pi_k\begin{pmatrix}-1\\2\\0\end{pmatrix} = \begin{pmatrix}0\\2\\0\end{pmatrix}$$

$$X_3 = \Pi_k\left(\begin{pmatrix}0\\2\\0\end{pmatrix} - \frac{1}{2}\begin{pmatrix}2\\0\\0\end{pmatrix}\right) = \Pi_k\begin{pmatrix}-1\\2\\0\end{pmatrix} = \begin{pmatrix}0\\2\\0\end{pmatrix} = X^* \textit{ the minimum}$$

In fact, we are at the minimum, X_k *has converges to* X^*

Is ρ *optimal point?*

we calculate $\gg$ $f(X_k - \rho\nabla g(X_k))$

Then differentiate $\frac{d}{d\rho}(\quad) = 0$

Exercise 1.

We are dealing with the following problem on $\mathbb{R}^3$:

$$(\mathcal{P}) \qquad \min_{x+y+z\leq 3} x^2 + 2y^2 + 3z^2 - 4x - 4y.$$

1. Is the set $K = \{(x,y,z)/x+y+z \leq 3\}$ closed in $\mathbb{R}^3$? open ? convex ? (explain in two words for each question), how would you call this set ?
2. Is the function $f(x,y,z) = x^2 + 2y^2 + 3z^2 - 4x - 4y$ convex ? strictly convex ? α-convex ? (if yes, give the value of α)
3. Justify the existence and uniqueness of the solution of $(\mathcal{P})$.
4. Apply the Kuhn-Tucker theorem : are the hypotheses verified ? write the Kuhn-Tucker relations.
5. Solve $(\mathcal{P})$ and give the point where the minimum is reached as well as the minimal value of the function f on K.

1. Is the set $K = \{(x, y, z)/x + y + z \leq 3\}$ closed in $\mathbb{R}^3$? open? convex? (explain in two words for each question), how would you call this set?

inequality then k is closed, k is not open continuous at g

Less than or equal that mean the line is included in the set

Closed mean the boundary of k is inside k (because it is soft inequality <=)

K is convex (if we draw a line between any 2 points it will be inside the set)

$x_1, x_2 \in k, \theta \in [0,1], \quad x_1 = \begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix}, x_2 = \begin{pmatrix} x_2 \\ y_2 \\ z_2 \end{pmatrix}$

$\theta x_1 + (1-\theta)x_2 = \theta \begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix} + (1-\theta) \begin{pmatrix} x_2 \\ y_2 \\ z_2 \end{pmatrix}$

$\gg$ *we want to know if it's in k*? Is x+y+z <= 3?>

$(x_1, y_1, z_1), (x_2, y_2, z_2) \in k$ then

$(\theta(x_1, y_1, z_1) + (1-\theta)(x_2, y_2, z_2) \leq 3\theta + 3(1-\theta) = 3$

That mean $(\theta(x_1, y_1, z_1) + (1-\theta)(x_2, y_2, z_2) \in k$ *then k is convex*

2. Is the function $f(x, y, z) = x^2 + 2y^2 + 3z^2 - 4x - 4y$ convex? strictly convex? α-convex? (if yes, give the value of α)

To solve it we should calculate the hessian.

$$\nabla f(x, y, z) = \begin{pmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \\ \frac{\partial f}{\partial z} \end{pmatrix} = \begin{pmatrix} 2x - 4 \\ 4y - 4 \\ 6z \end{pmatrix}$$

$$\nabla^2 f(x, y) = \begin{pmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} & \frac{\partial f}{\partial z} \\ \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} & \frac{\partial f}{\partial z} \\ \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} & \frac{\partial f}{\partial z} \end{pmatrix} = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 6 \end{pmatrix}$$

$\lambda_1 = 2, \quad \lambda_2 = 4, \quad \lambda_3 = 6$

f is α *convex and* $\alpha = 2 = \min \lambda_i$, *which is the minimum of lamdas* λ_i

Eigen values are positive.

f is α $convex\ and\ \alpha = 2\ which\ is\ the\ minimum\ of\ lamdas\ \lambda_i$

$which\ implies\ to\ be \gg strictly\ convex \gg convex$

3. **Justify the existence and uniqueness of the solution of $(\mathcal{P})$.**

Existence :

- k is closed, non-empty,
- f function is continuous, the function is infinite at infinity.
- f is $\alpha - convex \gg f\ infinite\ at\ infinity$
    - $f(y) \geq f(x) + f`(x, y - x) + \frac{\alpha}{2}||y - x||^2$
    - $linear\ in\ y, so\ if\ \left||y|\right| \to\ +\infty\ , then\ f(y) \to\ +\infty$

Uniqueness:

- k is convex
- f is $\alpha - convex$ >> strictly convex

Exercise 2: Convexity

Is the function $f(x) = ||x||^2$ defined on $\mathbb{R}^n$ convex? Is it strictly convex? Is it α-convex?

Exercise 2: $f(x) = ||x||^2 = \langle x, x \rangle$

Let $y \in \mathbb{R}^n$, $\varepsilon > 0$, $f(x+\varepsilon y) = \langle x+\varepsilon y, x+\varepsilon y \rangle = \langle x, x \rangle + \langle \varepsilon y, x \rangle + \langle x, \varepsilon y \rangle + \langle \varepsilon y, \varepsilon y \rangle$

$= f(x) + 2\varepsilon \langle x, y \rangle + \varepsilon^2 \langle y, y \rangle$

Then $\frac{f(x+\varepsilon y) - f(x)}{\varepsilon} = 2\langle x, y \rangle + \varepsilon \langle y, y \rangle \xrightarrow[\varepsilon \to 0]{} 2\langle x, y \rangle = (f'(x), y)$.

Let $z \in \mathbb{R}^n$, $\varepsilon > 0$, $\frac{(f'(x+\varepsilon z), y) - (f'(x), y)}{\varepsilon} = \frac{2\langle x+\varepsilon z, y \rangle - 2\langle x, y \rangle}{\varepsilon} = 2\langle z, y \rangle$

Then $(f''(x); y, z) = 2\langle y, z \rangle$.

As $(f''(x); y, y) = 2\langle y, y \rangle = 2||y||^2 \geq 2||y||^2$, we deduce that f is 2-convex. Then f is strictly convex, and convex.

Other solution: compute $\nabla f(x) = 2x$, $\nabla^2 f(x) = 2 I_n$ with all eigenvalues $= 2$.

Computation of Gateaux-derivatives and convexity

One considers $x \in \mathbb{R}^n \to f(x) = ||Ax - b||^2$, with $A \in \mathcal{M}_{n,n}(\mathbb{R})$ and $b \in \mathbb{R}^n$.

3.1. Compute the Gâteaux-derivative of f and deduce the gradient $\nabla f(x)$.

3.2. Compute the second-order Gâteaux-derivative of f and deduce the hessian $\nabla^2 f(x)$.

3.3. For which type of matrices $A \in \mathcal{M}_{n,n}(\mathbb{R})$, is the function f convex? strictly convex? α-convex?

Exercise 3: $f(x) = \|Ax-b\|^2 = \langle Ax-b, Ax-b \rangle$

1) Let $y \in \mathbb{R}^n$, $\varepsilon > 0$. $f(x+\varepsilon y) = \langle A(x+\varepsilon y)-b, A(x+\varepsilon y)-b \rangle$

$= \langle Ax+\varepsilon Ay-b, Ax+\varepsilon Ay-b \rangle = \langle Ax-b+\varepsilon Ay, Ax-b+\varepsilon Ay \rangle$

$= \langle Ax-b, Ax-b \rangle + \langle Ax-b, \varepsilon Ay \rangle + \langle \varepsilon Ay, Ax-b \rangle + \langle \varepsilon Ay, \varepsilon Ay \rangle$

$= f(x) + 2\varepsilon \langle Ax-b, Ay \rangle + \varepsilon^2 \langle Ay, Ay \rangle$

Then $\dfrac{f(x+\varepsilon y)-f(x)}{\varepsilon} = 2\langle Ax-b, Ay \rangle + \varepsilon \langle Ay, Ay \rangle \xrightarrow[\varepsilon \to 0]{} 2\langle Ax-b, Ay \rangle$

$\Rightarrow \boxed{(f'(x), y) = 2\langle Ax-b, Ay \rangle}$

$= 2\langle A^T(Ax-b), y \rangle = \langle \nabla f(x), y \rangle$ and then $\boxed{\nabla f(x) = 2A^T(Ax-b)}$

2) $\dfrac{(f'(x+\varepsilon z), y) - (f'(x), y)}{\varepsilon} = \dfrac{\langle 2(A(x+\varepsilon z)-b), Ay \rangle - \langle 2(Ax-b), Ay \rangle}{\varepsilon} = 2\langle Az, Ay \rangle$

$\Rightarrow \boxed{(f''(x), y, z) = 2\langle Ay, Az \rangle} = 2\langle A^T Ay, z \rangle = \langle \nabla^2 f(x)\, y, z \rangle$

$\Rightarrow \boxed{\nabla^2 f(x) = 2A^T A}$

3) f convex $\iff \forall x, \forall y, \ (f''(x), y, y) \geq 0$

$\iff \forall x, \forall y, \ 2\langle Ay, Ay \rangle \geq 0$

$\iff \forall x, \forall y, \ 2\|Ay\|^2 \geq 0$ which is true.

Then for any A, f is convex.

For strict convexity and α-convexity, the question is: $2\|Ay\|^2 > 0 \quad \forall y \neq 0$?

$\iff Ay \neq 0, \forall y \neq 0$ because only 0 has a zero norm.

We can conclude that for any matrix A invertible, f is strictly convex and α-convex