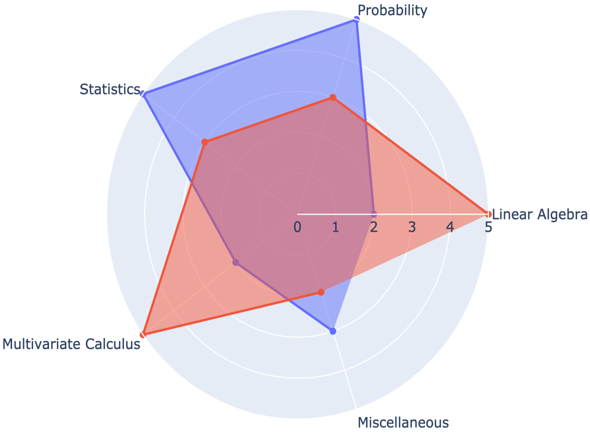
A GUIDED TOUR OF AI: FROM FOUNDATIONS TO LATEST APPLICATION

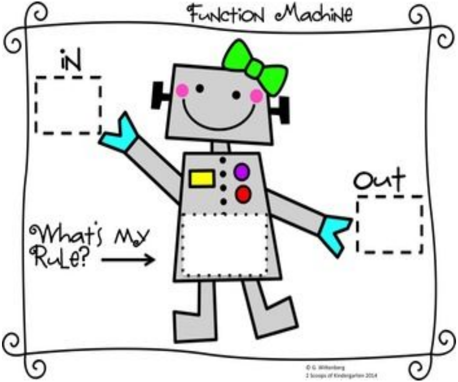
Introduction to functions

SCAI, Sorbonne University Abu Dhabi, Abu Dhabi, UAE

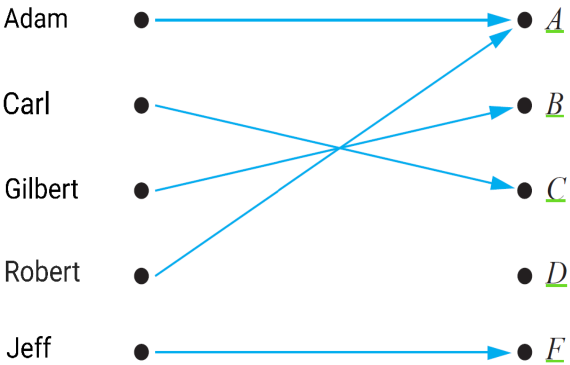
Roots of Machine Learning



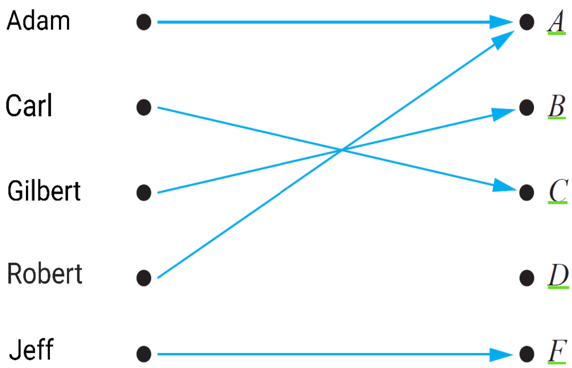
Importance of Functions in ML



Suppose that each student in a class is assigned a letter grade from the set $\{A, B, C, D, F\}$. And suppose that the grades are A for Adam, C for Carl, B for Gilbert, A for Robert, and F for Jeff.



This assignment is an example of a function.



- The concept of a function is extremely important in mathematics and computer science.
- For example, functions are used to represent how long it takes a computer to solve problems of a given size.
- Many computer programs and subroutines are designed to calculate values of functions.

Definition

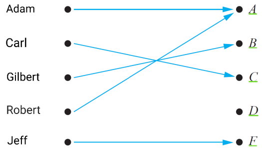
- Let A and B be collections of elements (that we call **sets**).
- A **function** f from A to B is an assignment of **exactly one element** of B to each element of A.
- We write $f(a) = b$ if b is the unique element of B assigned by the function f to the element a of A.
- If f is a function from A to B, we write

$$f : A \longrightarrow B$$

Remark: Functions are sometimes also called **mappings** or **transformations**.

Functions are specified in many different ways.

- Sometimes we explicitly state the assignments, as in



- Often we give a formula, such as $f(x) = x + 1$, to define a function.

Some terminologies

- If f is a function from A to B, we say that A is the **domain** of f and B is the **codomain** of f.
- If $f(a) = b$, we say that b is the **image** of a and a is a **preimage** of b.
- The **range**, or **image**, of f is the collection of all images of elements of A.
- Also, if f is a function from A to B, we say that f **maps** A to B.

Example

What are the **domain**, **codomain**, and **range** of the function that assigns grades to students described in the first slide?

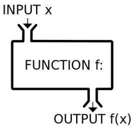
Example

What are the **domain**, **codomain**, and **range** of the function that assigns grades to students described in the first slide?

Solution:

- Let G be the function that assigns a grade to a student in the class.
- Note that $G(\text{Adam}) = A$, for instance.
- The **domain** of G is the set Adam, Carl , Gilbert, Robert, Jeff.
- The **codomain** is the set A,B,C,D, F.
- The **range** of G is the set A,B,C, F, because each grade except D is assigned to some student.

- To recap, a function is mainly a mathematical relationship from a set of inputs to a set of outputs.
  - The output value depends on (is a function of) the input value.
  - Each input produces exactly one output.



- Let's denote the input by the variable x. In this case, the output would be $f(x)$.
  - From now on, we consider functions defined by their algebraic formula and depending on a variable x for example

$$f(x) = 3x^2 + 6$$

Function composition

- Given two functions $f: X \to Y$ and $g: Y \to Z$ such that the domain of g is the codomain of f, their composition is the function $g \circ f: X \to Z$ defined by

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

- That is, the value of $g \circ f$ is obtained by first applying f to x to obtain $y = f(x)$ and then applying g to the result y to obtain $g(y) = g(f(x))$.
- In the notation the function that is applied first is always written on the right.

- Functions can be identified using **graphs**.
- Graphs display many input-output pairs in a small space. The visual information they provide often makes relationships easier to understand.
- We typically construct graphs with the input values along the horizontal axis and the output values along the vertical axis.

- In general, the domain of a given function can be given by different sets (intervals) where for each set, a different rule is assigned.
- But in the sequel, we shall consider only the functions that are ruled by one algebraic formula all over the domain, that assigns to each input value one and only one output value.
- We shall also deal only with "nice enough" function, or mathematically, **continuous functions**, where we can draw the graph from start to finish without ever once picking up our pencil.

- For computing the image of a number (being in the domain of the given function), we replace x by the number in the algebraic form, then we simply compute.

Example

- Given the function $f(x) = 3x^2 - 2$.
- What is the image of -2 by f?

- For computing the image of a number (being in the domain of the given function), we replace x by the number in the algebraic form, then we simply compute.

Example

- Given the function $f(x) = 3x^2 - 2$.
- what is the image of -2 by f?

Solution:

- For computing the image of -2 by f, we replace x by -2 in the expression $3x^2 - 2$.
- In other words, we compute $f(-2) = 3(-2)^2 - 2 = 3 \cdot 4 - 2 = 12 - 2 = 10$.
- Finally, the image of -2 by f is 10.

In the coordinates plane, the **graphic representation** of a function consists of all the points of coordinates $(x, f(x))$, where x belongs to the domain of f.

Example

Make a sketch of the graphic representation of the function $f(x) = 3x^2 - 2$.

Example

Make a sketch of the graphic representation of the function $f(x) = 3x^2 - 2$.

Solution:

Since the **graphic representation** of a function consists of all the points of coordinates $(x, f(x))$, then

- We can start by computing the images of different points by f, and then gathering them eventually in a table of values.

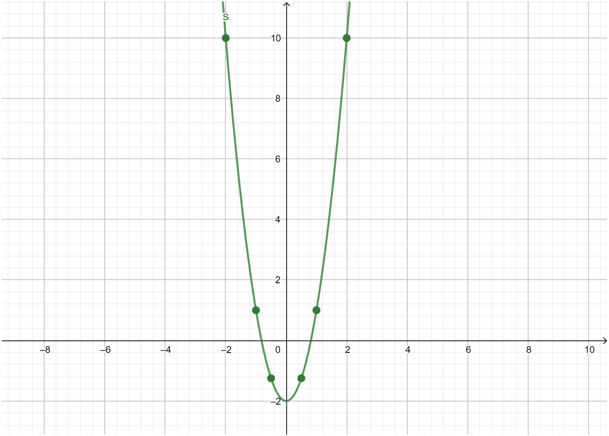
x	-2	-1	-0.5	0	0.5	1	2
f(x)	10	1	-1.25	-2	-1.25	1	10

- Then, we can place the obtained points $(x, f(x))$ on the coordinate plane and connect them by considering that between two points, the evolution is progressive.

Example

Make a sketch of the graphic representation of the function $f(x) = 3x^2 - 2$.

Solution:



- In general, functions can have many different algebraic expressions ($\ln(x)$, $\exp(x)$, $\cos(x)$, $\sin(x)$...)
- The most nice functions are the polynomial ones.Their **domain** is the line of all the real values denoted $\mathbb{R}$.
- Generally speaking, a polynomial is a sum and subtraction of terms of the form kx^n where k is any number and n is a positive integer. In this case, the maximal occuring power is called **degree** of the polynomial. For example, $3x^2 + 2x - 5$ is a polynomial of degree 2.
- Polynomial functions of degree 1 are the affine function. Their graphs are straight lines.
- Polynomial functions of degree 2 are called quadratic functions. Their graphs are parabolas $\left(\cup, \cap\right)$.

Example

Again, make a sketch of the graphic representation of the function

$$g(x) = 2x + 3$$

Example

Again, make a sketch of the graphic representation of the function

$$g(x) = 2x + 3$$

Solution:

- Let's make a table of values.

x	-3	-2	-1	0	1	2	3
f(x)	-3	-1 (=-3 + 2)	1 (=-1 +2)	3 (= 1 + 2)	5 (=3 +2)	7 (=5 +2)	9 (=7 +2)

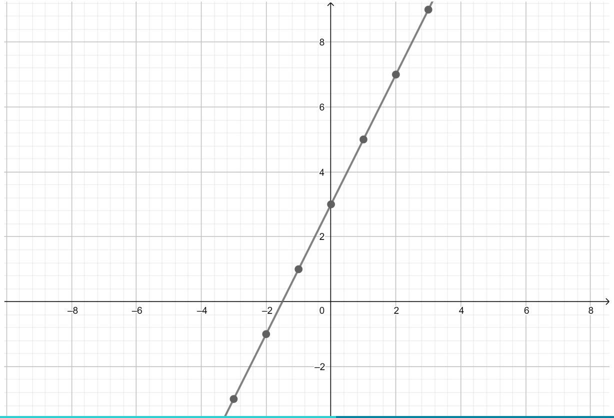
- Then, we obtain the graph

Example

Again, make a sketch of the graphic representation of the function

$$g(x) = 2x + 3$$

Solution:



What can we notice?

- As we can clearly see, the graph of the function $f(x) = 2x + 3$ is a straight line.
- Moreover, from the table of values, we can notice that for $x = 0$, $f(0) = 3$, that is the constant appearing in the function.
- From the graph, we see that the line crosses the y-axis exactly at the point of coordinates $(0, f(0)) = (0, 3)$
- Also from the table of values, we can see that by increasing the value of x by 1, the value of $f(x)$ is increasing by 2, that is the the value attached to x in the expression of $f(x)$.

In general

- The graph of any function of algebraic expression $f(x) = mx + b$ is a straight line.
- This algebraic expression can also be simply given by the linear equation $y = mx + b$ since any point on the graph (straight line) is of coordinates $(x, y) = (x, f(x))$.
- These functions are called **affine functions**.
- The number attached to x is called the slope.
- The number unattached to x is called the y-intercept.
- This specific form of an affine function $f(x) = mx + b$ is also called **slope-intercept** form.
- Cause simply by reading the expression, we can simply extract the slope and the y-intercept!

Example

- Which of the following expressions is an affine function?

$$f(x) = 2x - \frac{1}{5}, \quad g(x) = 3x, \quad h(x) = \frac{1}{x} + 7, \quad s(x) = 4,$$

$$r(x) = 2(x-2) - 5$$

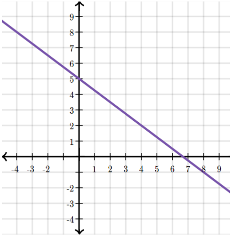
- When it applies, extract the slope and the y-intercept of each affine function.

Some properties about the slope

- Slope is a measure of the steepness of a line.
- It can be seen as

$$\text{Slope} = \frac{\text{Rise}}{\text{Run}} = \frac{\Delta y}{\Delta x}$$

- Example, consider that we're given the graph of a line and asked to find its slope.

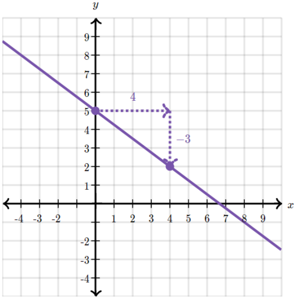


- The line appears to go through the points $(0, 5)$ and $(4, 2)$.

$$\text{Slope} = \frac{\Delta y}{\Delta x} = \frac{2-5}{4-0} = \frac{-3}{4}.$$

Some properties about the slope

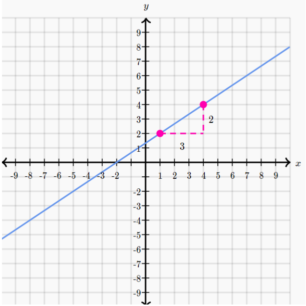
- Slope is a measure of the steepness of a line.
- Example, consider that we're given the graph of a line and asked to find its slope.



- The line appears to go through the points $(0, 5)$ and $(4, 2)$.

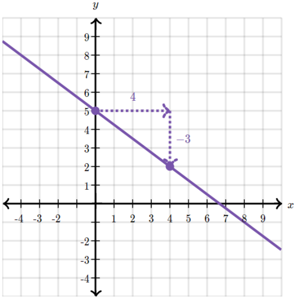
$$\text{Slope} = \frac{\Delta y}{\Delta x} = \frac{2-5}{4-0} = \frac{-3}{4}.$$

- We say that a function is **increasing** if "while running we are also rising!"
- In other words, when the slope is **positive**
- Take for example $f(x) = \frac{2}{3}x + \frac{4}{3}$.



- Here we see that running 3 steps along the x-axis yields to rising 2 steps along the y-axis.

- We say that a function is **decreasing** if "while running we are also dropping!"
- In other words, when the slope is **negative**
- Take for example $f(x) = \frac{-3}{4}x + 5$.

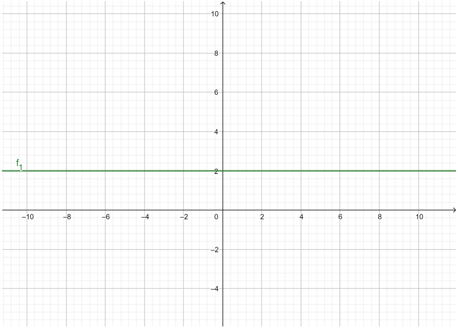


- Here we see that running 4 steps along the x-axis yields to dropping 3 steps along the y-axis.

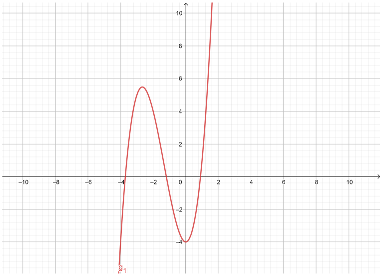
But what if the slope is 0?

But what if the slope is 0?

- In this case, running along the x-axis will always keep us on the same level regarding the y axis.
- Our trajectory is thus a horizontal line.

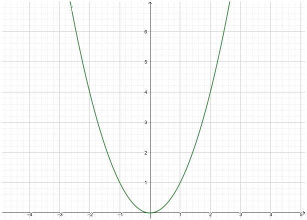


Can you tell me in this case what is the expression of the corresponding function?



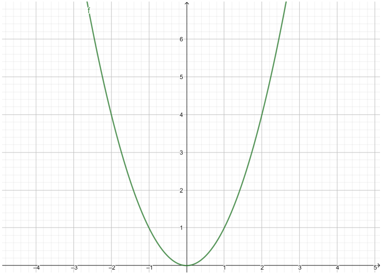
Now what if the function is not affine? Meaning when its graph is not a line.
What would be its variation?

- Take for example $f(x) = x^2$. Its graph is given by



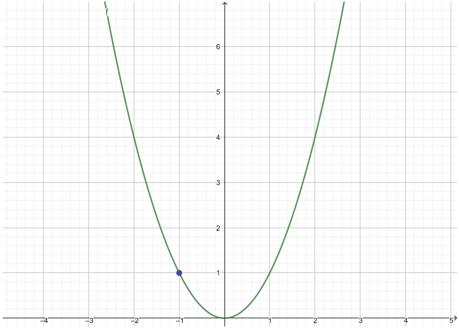
- As we can see, there no constant variation all over the plane (*and when it's not constant, maybe it's a function of* x. ) In fact
- when $x < 0$, i.e., on the left hand side of the plane, we "see" that the function is decreasing (while running we are dropping!),
- when $x > 0$, i.e., on the right hand side of the plane, we "see" that the function is increasing (while running we are rising!).

- Take for example $f(x) = x^2$. Its graph is given by



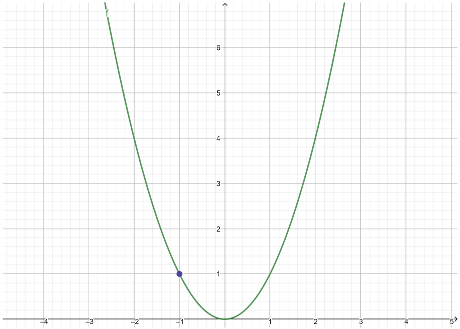
- Thus, since there's no constant variation all over the plane, let's start by studying **local** variation.
- For this, let's chose a point on the graph (of coordinates $(x, f(x)) = (x, x^2)$), take for example the point $(-1, 1)$.

- Take for example $f(x) = x^2$. Its graph is given by

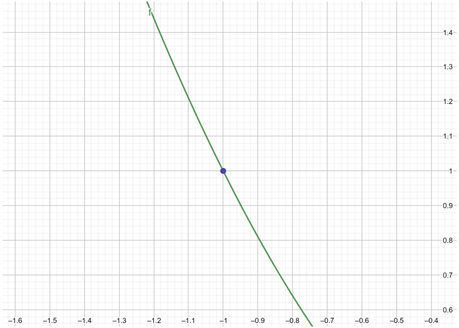


- Thus, since there's no constant variation all over the plane, let's start by studying **local** variation.
- For this, let's chose a point on the graph (of coordinates $(x, f(x)) = (x, x^2)$), take for example the point $(-1, 1)$,
- and let's study the variation of the graph at the moment it passes by this point.

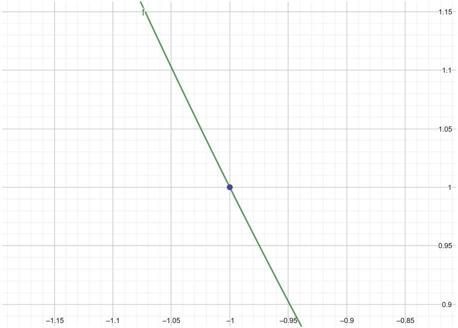
- Take for example $f(x) = x^2$. Its graph is given by



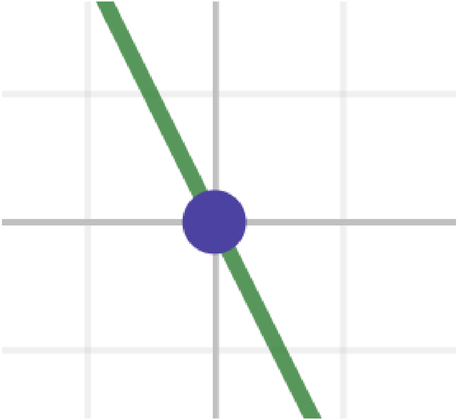
- Let's zoom in as much as we can to our point $(-1, 1)$



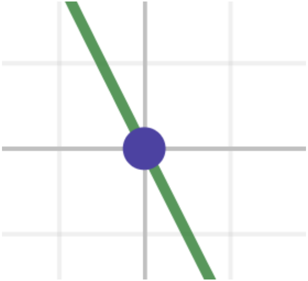
- Let's zoom in as much as we can to our point $(-1, 1)$



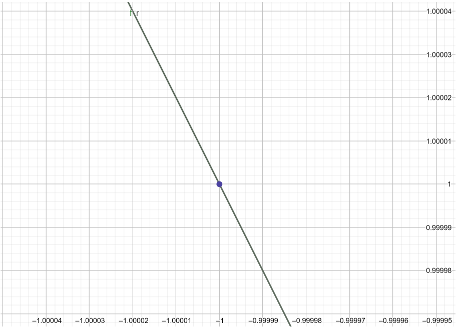
- Let's zoom in as much as we can to our point $(-1, 1)$



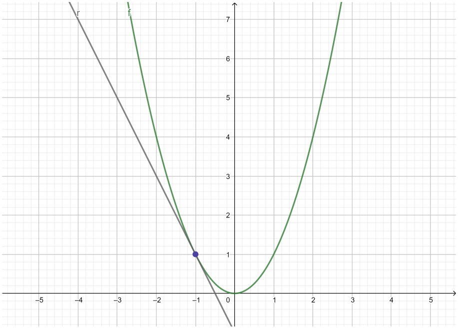
- Let's zoom in as much as we can to our point $(-1, 1)$



- As we can clearly see, around the point $(-1, 1)$, the graph resembles to a straight line! This is a good new since we already now how to find the slope of a line.
- We recall it however: given a line corresponding to a function $g(x)$ and two points on this line of coordinates $(a, g(a))$, $(b, g(b))$, then the slope is given by $\dfrac{g(b) - g(a)}{b - a}$.



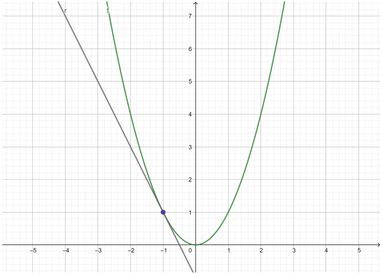
- From this zoomed in place, let's draw the line we see, then zoom out.



- From this zoomed in place, let's draw the line we see then zoom out.

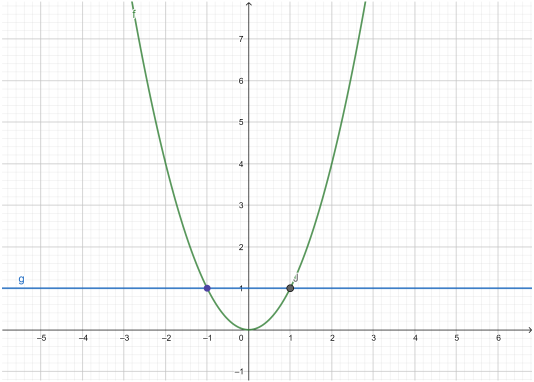
- The grey line that we've obtained is called the **tangent line** to the green curve at the point $(-1, 1)$.
- In general, the tangent line to a curve at a point is that straight line that best approximates (or "clings to") the curve near that point.
- The term slope is defined for a straight line and the variation of the line is given by its slope.
- But for a curve, at a given point, its variation is given locally by the slope of the tangent at that point.
- Take the point $A = (a, f(a))$ on the graph of f. Then the slope of the tangent at the point A is said to be **the point derivative of f at a and is denoted by $f^{'}(a)$**.
- For example for $f(x) = x^2$, when you are exactly at $x = -1$, i.e., on the point $(-1, 1)$, the slope of the tangent at that point is denoted by $f^{'}(-1)$ and it is the point derivative of f at $x = -1$.

- Back to the example $f(x) = x^2$. Let's compute $f'(-1)$.



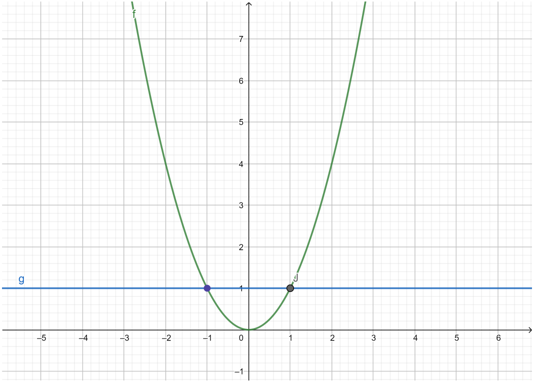
- Of course, we can compute the slope of the grey tangent at $(-1, 1)$ by reading the graph and considering two different points on the tangent. Can you tell me how and how much is it?
- But in general, We need the slope in order to be able to draw the tangent, that as we said, it best approximates the curve near that point.

- Back to the example $f(x) = x^2$. Let's compute $f'(-1)$.



- Let's start by drawing a line from the point $(-1, 1)$ to any other point <u>on the curve</u>, say $\left(-1+2, f((-1+2)\right) = (1, 1)$.
- But, for simplifying, let's write h instead of 2, i.e., let's write $\left(-1+h, f((-1+h)\right)$

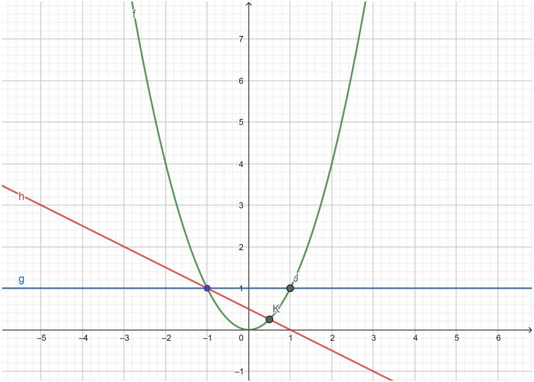
- Back to the example $f(x) = x^2$. Let's compute $f'(-1)$.



- The slope of the blue line passing by $(-1, 1)$ and $\big(-1+h, f(-1+h)\big)$ is then

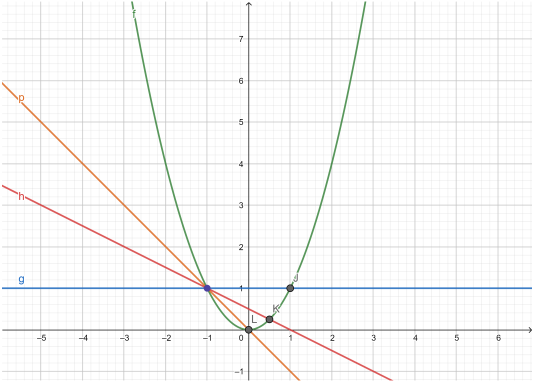
$$\frac{f(-1+h) - f(-1)}{(-1+h) - (-1)} = \boxed{\frac{f(-1+h) - f(-1)}{h}}$$

- Back to the example $f(x) = x^2$. Let's compute $f'(-1)$.



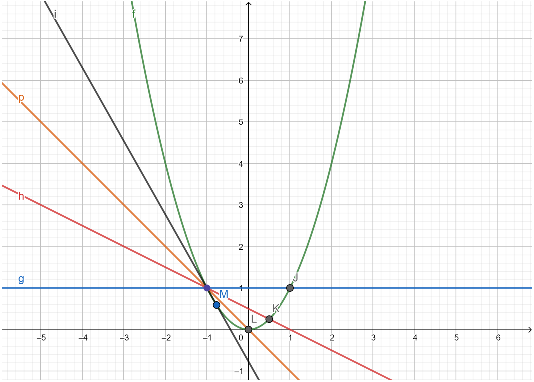
- Now, if we keep on decreasing the value of h, i.e., we take h from the value 2 to values closer and closer to zero, we get each time a line that is closer and closer to the tangent we aim at computing its slope.

- Back to the example $f(x) = x^2$. Let's compute $f'(-1)$.



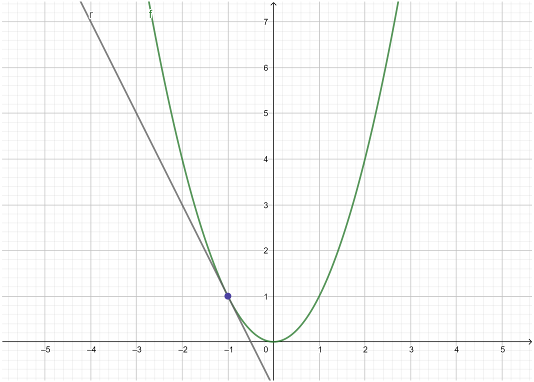
- Now, if we keep on decreasing the value of h, i.e., we take h from the value 2 to values closer and closer to zero, we get each time a line that is closer and closer to the tangent we aim at computing its slope.

- Back to the example $f(x) = x^2$. Let's compute $f'(-1)$.



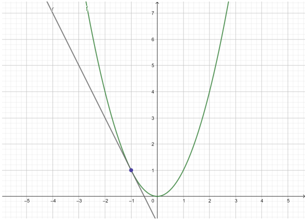
- Now, if we keep on decreasing the value of h, i.e., we take h from the value 2 to values closer and closer to zero, we get each time a line that is closer and closer to the tangent we aim at computing its slope.

- Back to the example $f(x) = x^2$. Let's compute $f'(-1)$.



- Now, if we keep on decreasing the value of h, i.e., we take h from the value 2 to values closer and closer to zero, we get each time a line that is closer and closer to the tangent we aim at computing its slope.

- Back to the example $f(x) = x^2$. Let's compute $f'(-1)$.



- That means that if we reduce h as much as possible to be extremely close to 0 (to the much we may think it's 0!) in the expression

$$\frac{f(-1+h) - f(-1)}{(-1+h) - (-1)} = \boxed{\frac{f(-1+h) - f(-1)}{h}}$$

then we get the slope of the tangent at $(-1, 1)$!

- That means that if we reduce h as much as possible to be extremely close to 0 (to the much we may think it's 0!) in the expression

$$\frac{f(-1+h)-f(-1)}{(-1+h)-(-1)} = \boxed{\frac{f(-1+h)-f(-1)}{h}}$$

then we get the slope of the tangent at $(-1, 1)$!

- But for this value to be defined (we can't have a zero in the denominator), let's simplify the expression

$$\boxed{\frac{f(-1+h)-f(-1)}{h}}$$

then simply replace h by 0:

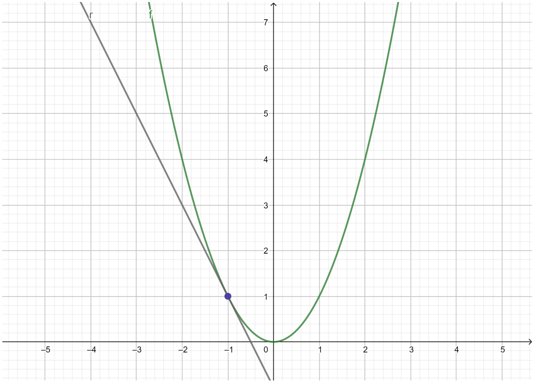
- But for this value to be defined (we can't have a zero in the denominator), let's simplify the expression

$$\boxed{\frac{f(-1+h)-f(-1)}{h}}$$

then simply replace h by 0:

$$\begin{aligned}\frac{f(-1+h)-f(-1)}{h} = \frac{(-1+h)^2-(-1)^2}{h} &= \frac{(1-2h+h^2)-1}{h}\\ &= \frac{-2h+h^2}{h}\\ &= \frac{h(-2+h)}{h}\\ &= -2+h \underset{h\longrightarrow 0}{\longrightarrow} \boxed{-2}\end{aligned}$$

- $f(x) = x^2$:



$$f'(-1) = -2$$

- For $f(x) = x^2$, $f^{'}(-1) = -2$
- And here we go! We've found the slope of the tangent at $(-1, 1)$, that is $f^{'}(-1)$ and we see it's negative, and the function f is decreasing when x is very close to the value -1.

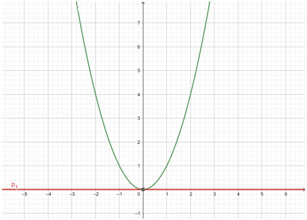
But what about all the other values on the x-axis?
Should we do this computation all over again each time we want to study the variation of a curve at a given point?
No!
We can simply define a new function and name it the derivative function denoted $f^{'}$, that for any input value a returns $f^{'}(a)$.

Function	Derivative
$f(x) = a$ (a constant value)	$f^{'}(x) = 0$
$f(x) = x^n$	$f^{'}(x) = nx^{n-1}$
$f(x) = \frac{1}{x}$	$f^{'}(x) = -\frac{1}{x^2}$
$f(x) = a \cdot g(x)$	$f^{'}(x) = a \cdot g^{'}(x)$
$f(x) = g(x) + h(x)$	$f^{'}(x) = g^{'}(x) + h^{'}(x)$
$f(x) = g(x) \cdot h(x)$	$f^{'}(x) = g^{'}(x)h(x) + h^{'}(x)g(x)$
$f(x) = (g \circ h)(x)$ (chain rule)	$f^{'}(x) = h^{'}(x)(g^{'}(h(x)))$
$f(x) = (g(x))^n$ (chain rule)	$f^{'}(x) = ng^{'}(x)(g(x))^{n-1}$

Example

Function	Derivative
$f(x) = 3$	$f^{'}(x) = 0$
$f(x) = x^3$	$f^{'}(x) = 3x^2$
$f(x) = 3 \cdot x^2$	$f^{'}(x) = 3 \cdot (2x) = 6x$
$f(x) = 3x^2 + x^3$	$f^{'}(x) = 6x + 3x^2$
$f(x) = (3x^2)(x+1)$	$f^{'}(x) = 6x(x+1) + 3x^2$
$f(x) = (x^2+2)^3$	$f^{'}(x) = 3(2x)(x^2+2)^2$

- Now back to our example $f(x) = x^2$.
- We have $f'(x) = 2x$.
- For example, $f'(-1) = -2$, $f'(2) = 2(2) = 4$, $f'(0) = 0$.
- In this case, we see that the slope of tangent at the point $(0, 0)$ is equal to 0, i.e., the tangent at the point $(0, 0)$ is a horizontal line!



- Moreover, for $x < 0$, $f'(x) = 2x < 0$, i.e., for $x < 0$, f is decreasing.
- For $x > 0$, $f'(x) = 2x > 0$, i.e., for $x > 0$, f is increasing.

- Now back to our example $f(x) = x^2$, $f^{'}(x) = 2x$.
- $f^{'}(0) = 0$.
- Moreover, for $x < 0$, $f^{'}(x) = 2x < 0$, i.e., for $x < 0$, f is decreasing.
- For $x > 0$, $f^{'}(x) = 2x > 0$, i.e., for $x > 0$, f is increasing.
- Then we can draw a table that better illustrates the explained variation:

x	$-\infty$		0		$+\infty$
$f'(x)$		*negative*	0	*positive*	
$f(x)$		$\searrow$	0	$\nearrow$	

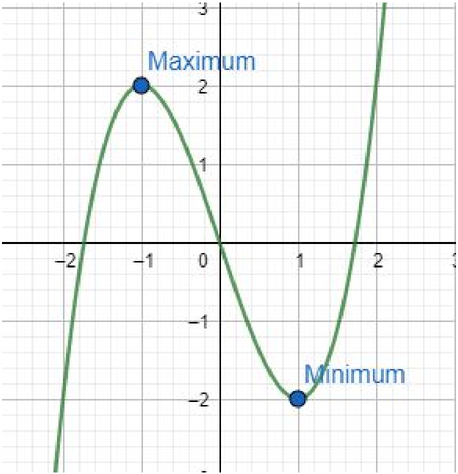
- Now back to our example $f(x) = x^2$, $f'(x) = 2x$.
- $f'(0) = 0$.
- Moreover, for $x < 0$, $f'(x) = 2x < 0$, i.e., for $x < 0$, f is decreasing.
- For $x > 0$, $f'(x) = 2x > 0$, i.e., for $x > 0$, f is increasing.
- Then we can draw a table that better illustrates the explained variation:

x	$-\infty$		0		$+\infty$
$f'(x)$		-	0	$+$	
$f(x)$		$\searrow$	0	$\nearrow$	

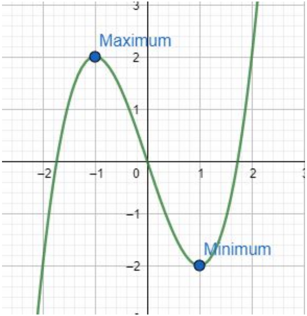
x	$-\infty$		0		$+\infty$
$f'(x)$		-	0	$+$	
$f(x)$		$\searrow$	0	$\nearrow$	

- The final point we see about the function $f(x) = x^2$ is that when $f^{'}(x) = 0$, i.e., at $x = 0$, the function reaches its minimal y-value that is $f(0) = 0$.

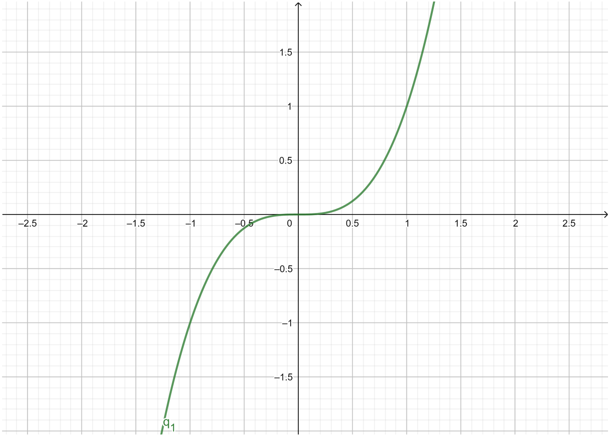
- In general the local maximum value of a function is the place where a function reaches its **highest point**, or **vertex**, on a graph, at a given interval. In other words, it is the maximal y-value that the function can reach, at a given interval.
- The local minimum value of a function is the place where the graph has a vertex at its lowest point, at a given interval. In other words, it is the minimal y-value that the function can reach.
- A global maximum point refers to the point with the largest value on the graph of a function, when a largest value exists. A global minimum point refers to the point with the smallest value. Together these two values are referred to as global extrema.



In this graph, are the maximum and the minimum local or global?



- In general, both the maximum and the minimum (either local or global) are reached at $x = a$ if $f'(a) = 0$ and the signs of f' (positive or negative) are different for $x < a$ and for $x > a$.
- In this case, the point $f(a)$ is either a maximum or a minimum at $x = a$.



In this graph (of the function x^3), do we have any maxima or minima?

- Take a quadratic function $f(x) = ax^2 + bx + c$ $(a \neq 0)$.
- We recall that its graph is a parabola $\big(\cup$ or $\cap\big)$.
- Thus, it is clear that parabolas have either a global maximum or a global maximum.
- Let's draw its table of variation and find its maximum or minimum.
- For this, let's compute $f'(x)$.
- We have $f'(x) = 2ax + b$
- We recall that the maximum or the minimum occurs at the value x where $f'(x) = 0$.

$$f'(x) = 2ax + b = 0 \quad \Longleftrightarrow \quad 2ax = -b \quad \Longleftrightarrow \quad \boxed{x = -\frac{b}{2a}}$$

- Thus, $f\left(-\frac{b}{2a}\right)$ is either the global minimum or the global maximum.

Two cases arise:

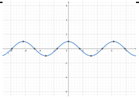
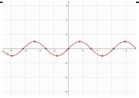
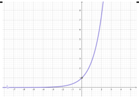
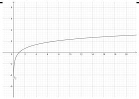
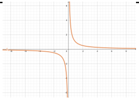
- If $a > 0$ (positive), then the parabola is smiling, like ☺
- In this case, its table of variation is as follows:

x	$-\infty$		$-\frac{b}{2a}$		$+\infty$
$f'(x)$		$-$	0	$+$	
$f(x)$		$\searrow$	$f\left(-\frac{b}{2a}\right)$	$\nearrow$	

Two cases arise:

- If $a < 0$ (negative), then the parabola is frowning, like ☹
- In this case, its table of variation is as follows:

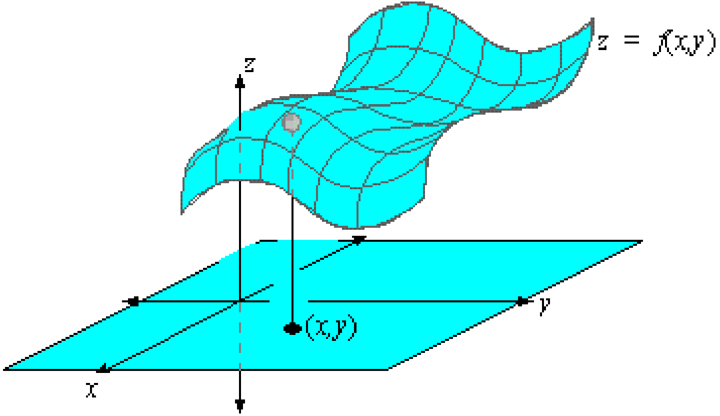
x	$-\infty$		$-\frac{b}{2a}$		$+\infty$
$f'(x)$		$+$	0	$-$	
$f(x)$		$\nearrow$	$f\left(-\frac{b}{2a}\right)$	$\searrow$	

Function	Domain	Graph
cosine: $\cos(x)$	$\mathbb{R}$	
sin: $\sin(x)$	$\mathbb{R}$	
Exponential:$\exp(x)$	$\mathbb{R}$	
Logarithmic: $\ln(x)$	$]0, +\infty]$	
Hyperbolic $\dfrac{1}{x}$	$\mathbb{R} - \{0\}$	

- A function of 2 variables is a function whose inputs are points (x, y) in the xy-plane and whose outputs are real numbers.
- We often denote functions of 2 variables by $f(x, y)$, which means "the output from an input of (x, y)".
- We often define these functions in the form

$$f(x,y) = \text{"expression in } x \text{ and } y\text{"}$$

- Equivalently, we can consider $f(x, y)$ to be the assignment of a real number to a point (x, y) in the xy-plane.
- The graph of $f(x, y)$ is the set of points in $\mathbb{R}^3$, i.e., point of the form (x, y, z), that satisfy $z = f(x, y)$.
- The graph of $f(x, y)$ is the **surface** $z = f(x, y)$ and the output z is the height of the surface at the point (x, y).


z = f(x,y)
z
y
(x,y)
x

Example

- Evaluate $f(1,2)$ and $f(2,5)$ if $f(x,y) = x^2 + 2xy$.

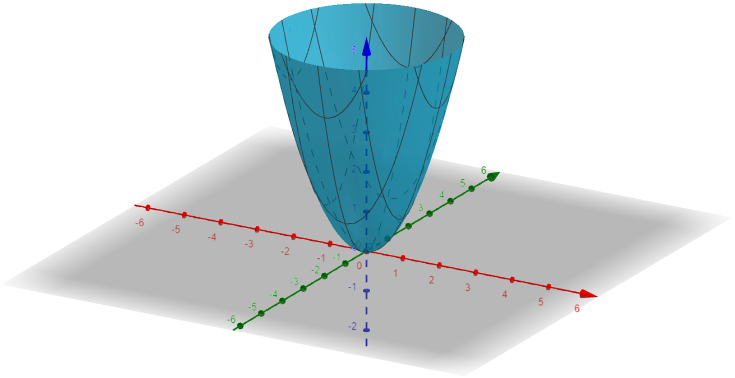
Example

- Evaluate $f(1,2)$ and $f(2,5)$ if $f(x,y) = x^2 + 2xy + x$.

Solution:

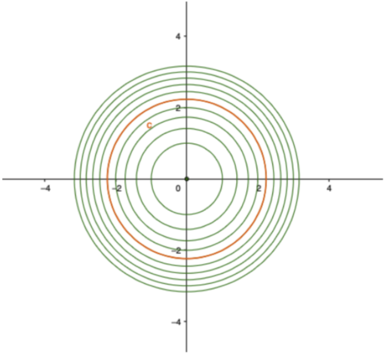
- To begin with, $f(1,2) = 1^2 + 2 \cdot 1 \cdot 2 + 1 = 6$, which is to say that $f(x,y) = x^2 + 2xy + x$ maps the point $(1,2)$ to the number 6.
- Likewise, $f(2,5) = 2^2 + 2 \cdot 2 \cdot 5 + 2 = 26$.

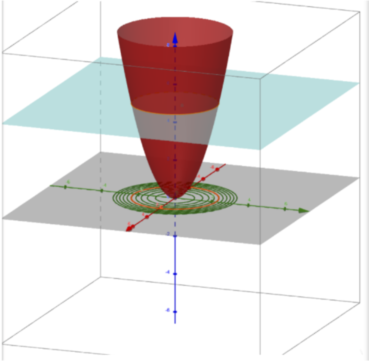
- A graphing calculator (Geogebra) or computer algebra system is often used to produce an approximation of the graph of a function.
- $f(x,y) = x^2 + y^2$:



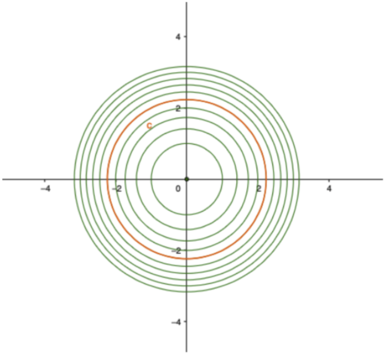
- Contour plots are a way to visualize bivariate functions on a 2D plane. They show curves, called contours, that connect points with the **same output value**.
- These curves represent the "**level sets**" of the function, where the function's value remains constant along each contour.

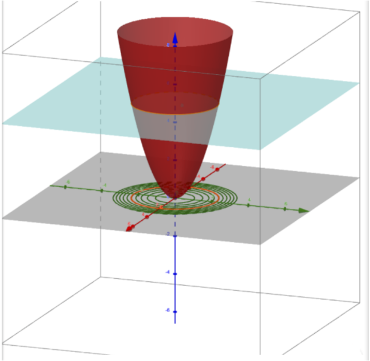
- Consider the example of $f(x, y) = x^2 + y^2$, represented by the equation $z = x^2 + y^2$.
- Here's its contour plot.





- As we mentioned, in the contour plot, we draw curves to represent points with the **same height**.
- The orange curve we see in the contour plot corresponds to the height $z = 4$, and it's simply the circle $x^2 + y^2 = 4$.
- Simply put, a contour plot is like cutting the surface with a paper at a specific height (in this case, 4) and tracing the shape of the cut on the surface.





- Now, let's discuss partial derivatives, extrema, and critical points.
- Partial derivatives help us understand how a bivariate function changes **in the direction of one variable** while keeping the other constant.
- To denote the partial derivative with respect to x, we use

$$\frac{\partial f}{\partial x}$$

  and it tells us how the function changes concerning x while y remains fixed.
- Similarly, to represent the partial derivative with respect to y, we use

$$\frac{\partial f}{\partial y}$$

  which indicates how the function changes concerning y while x stays constant.

Let's take the function $f(x, y) = x^2 + y^2$ as an example.

- First, let's find $\dfrac{\partial f}{\partial x}$:

  $$\boxed{\frac{\partial f}{\partial x} = \frac{d}{dx}\left(x^2 + y^2\right) = 2x + 0 = 2x}$$

  This result means that when we change the value of x while keeping y constant, the rate of change in the function is simply 2 times the value of x.

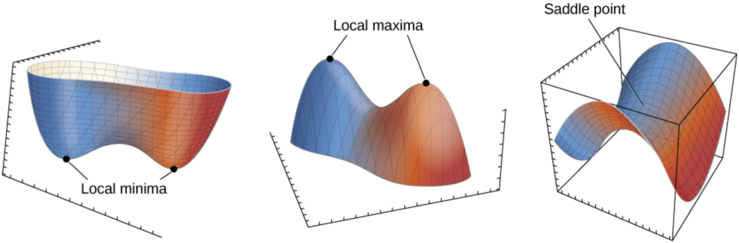
- Next, let's find $\dfrac{\partial f}{\partial y}$:

  $$\boxed{\frac{\partial f}{\partial y} = \frac{d}{dy}\left(x^2 + y^2\right) = 0 + 2y = 2y}$$

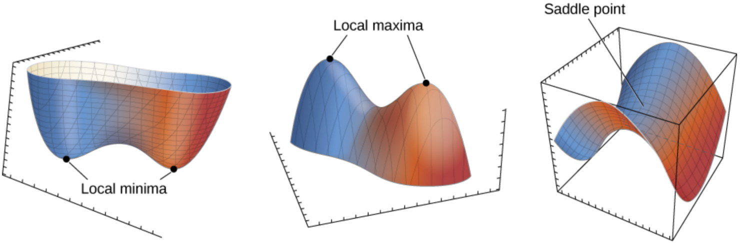
  Similarly, this means that when we change the value of y while keeping x constant, the rate of change in the function is 2 times the value of y.

By calculating these partial derivatives, we gain insights into how $f(x, y)$ changes concerning each variable independently, making it easier to analyze the behavior of the function in different directions.

Let's talk now about the extrema, the saddle points, and the critical points of a bivariate function $f(x, y)$.



Just like when we climb a hill in the real world, we want to find the **highest point** (maximum) and the **lowest point** (minimum) on this hill. We call them extrema.



- When we are situated at the highest point (maximum), moving in any direction will result in descending altitude.
- Conversely, when we are positioned at the lowest point (minimum), moving in any direction will lead to an ascending altitude.
- However, a saddle point is the point where moving in one direction can cause an increase in altitude, while moving in another direction can cause a decrease.

- For a univariate function, we can look at where the slope of the curve is zero and changes direction to find the maximum or minimum.
- Now, for a bivariate function, to find the maximum and minimum points, we need to look for places where the slopes are zero or change direction in both the x and y directions.
- These points are potential candidates for being the highest or lowest points on the hill-like surface. We call them **Critical points**

To find the critical points of a bivariate function:

1. First, we find the partial derivatives with respect to both x and y.
   - For $f(x,y) = x^2 + y^2$, we compute

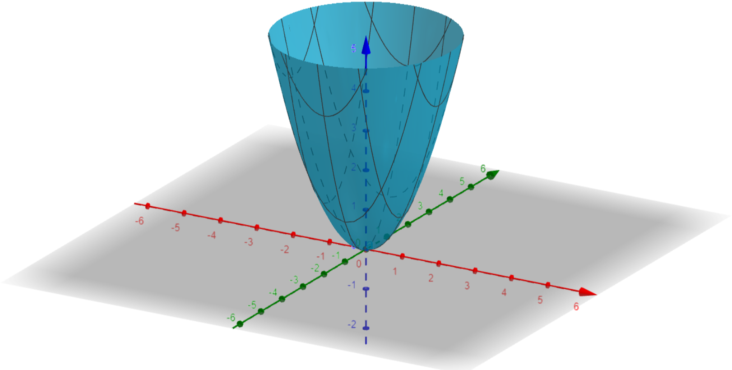
$$\begin{cases} \dfrac{\partial f}{\partial x} = 2x \\ \dfrac{\partial f}{\partial y} = 2y \end{cases}$$

2. Next, we set both of these partial derivatives equal to zero and solve for x and y.
   -

$$\begin{cases} \dfrac{\partial f}{\partial x} = 2x = 0 \\ \dfrac{\partial f}{\partial y} = 2y = 0 \end{cases} \quad \text{then} \quad x = 0 \text{ and } y = 0$$

3. The points (x,y) that we find are the **critical points** we are searching for
   - In our example we have only one critical point $(x,y) = (0,0)$.

- After identifying the critical points, we can analyze the function further to determine if these points correspond to a maximum, minimum, or neither (saddle point).
- For example, the critical point of $f(x, y) = x^2 + y^2$ is $(x, y) = (0, 0)$. It corresponds to the altitude $f(0, 0) = 0$. Is 0 the highest point, the lowest point or neither?



Thank you!