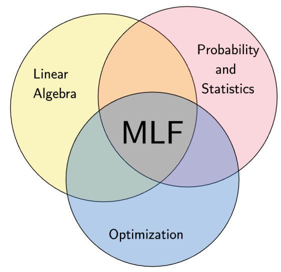
Week-2, Session-1



Karthik Thiagarajan

Disclaimer: These notes were generated as a part of this live session. It can't be used as a standalone resource or a set of notes for this week.

1. Concepts

- [ ] Sets
- [ ] Functions of one variable
    - [ ] Graphs
    - [ ] Limits
    - [ ] Continuity
    - [ ] Differentiability
    - [ ] Product, quotient and chain rules
    - [ ] Taylor series expansion
    - [ ] Linear approximation
    - [ ] Critical points, maxima and minima
    - [ ] Higher order approximation
- [ ] 3D Geometry
    - [ ] Points
    - [ ] Lines
    - [ ] Planes and hyperplanes
- [ ] Functions of two variables
    - [ ] Contours
    - [ ] Limits
    - [ ] Continuity
    - [ ] Partial derivatives, gradient
    - [ ] Directional derivative
    - [ ] Taylor series expansion
    - [ ] Linear approximation
    - [ ] Gradient $\rightarrow$ steepest ascent, orthogonality to contours
    - [ ] Critical points, maxima and minima

☐ Higher order approximation

Functions

$$f : X \to \mathbb{R}$$

$$x \to f(x)$$

Limits

$$\lim_{x \to c} f(x)$$

$$\{x_n\} = x_1, x_2, x_3, \ \ldots$$

$$f(x) = x^2 + 2x - 1$$

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

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

$$\lim_{x \to \infty} f(x) = 0$$

Continuity

A function f is said to be continuous at a point $x = a$ if:

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

As an example:

$$f(x) = \begin{cases} x^2 + 5, & x > 2 \\ m(x+1) + k, & -1 < x \leqslant 2 \\ 2x^3 + x + 7, & x \leqslant -1 \end{cases}$$

Is this function continuous?

It is continuous if $k = 4, m = \dfrac{5}{3}$.

$$\begin{aligned} 9 &= 3m + k \\ 4 &= k \end{aligned} \implies k = 4, m = \frac{5}{3}$$

Can you give me an example of a function for which limits exists at $x = c$ but the value of the function is different from the limit?

Consider $f\colon \mathbb{R} \to \mathbb{R}$.

$$f(x) = \begin{cases} x, & x > 0 \\ -x, & x < 0 \\ a, & x = 0 \end{cases}$$

If $a \neq 0$, then f is not continuous at $x = 0$.

Functions such as polynomials, trigonometric, exponential, log are continuous.

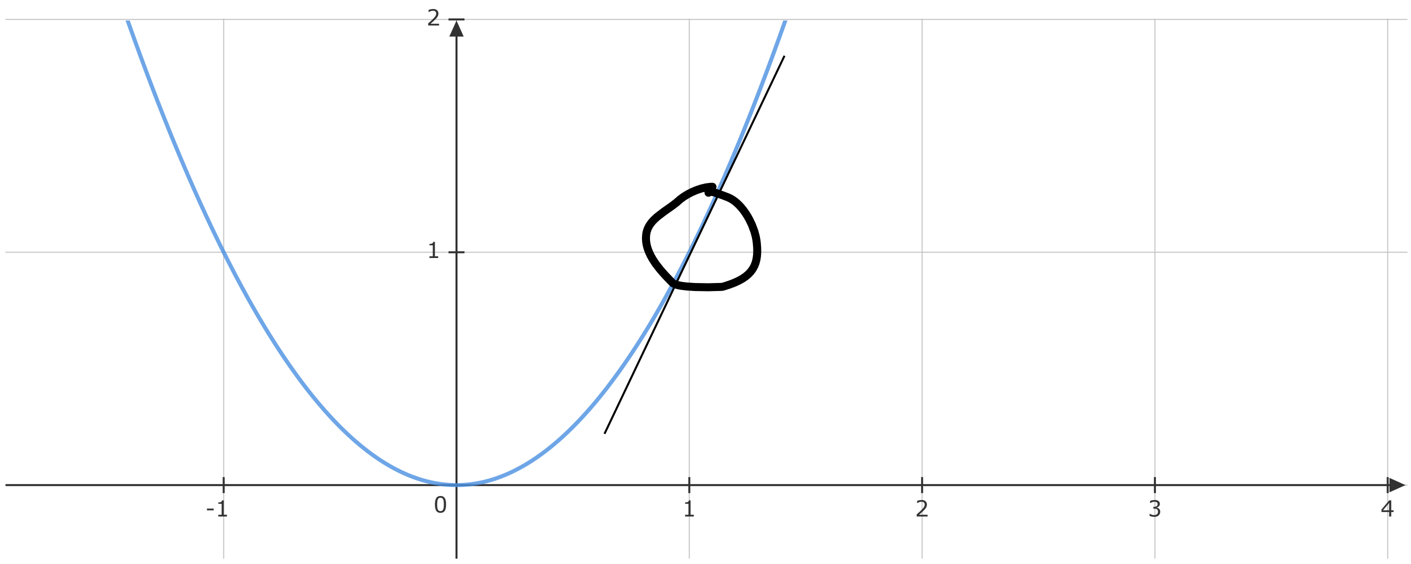
$$f\colon (0, \infty) \to \mathbb{R}$$

$$f(x) = \log(x)$$

Differentiability

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

$f'(a)$ is the **slope of the tangent** to the curve $(x, f(x))$ at $x = a$. The limiting process is the secant becoming the tangent at the point $x = a$.



Linear approximations

The idea is the following: you can approximate the behavior of a function at a point with the help of its derivatives.

If you want to know how the function behaves, in an approximate sense, around $x = a$

$$f(x) \boxed{\approx} f(a) + f'(a) \cdot (x - a)$$

The linear approximation to f at the point $x = a$:

$$L_a[f](x) = f(a) + f'(a) \cdot (x - a)$$

- Linear approximation to f at a point is a function.
- It gives an approximation to the function in and around that point.

If f is already linear, then the linear approximation is the function itself.

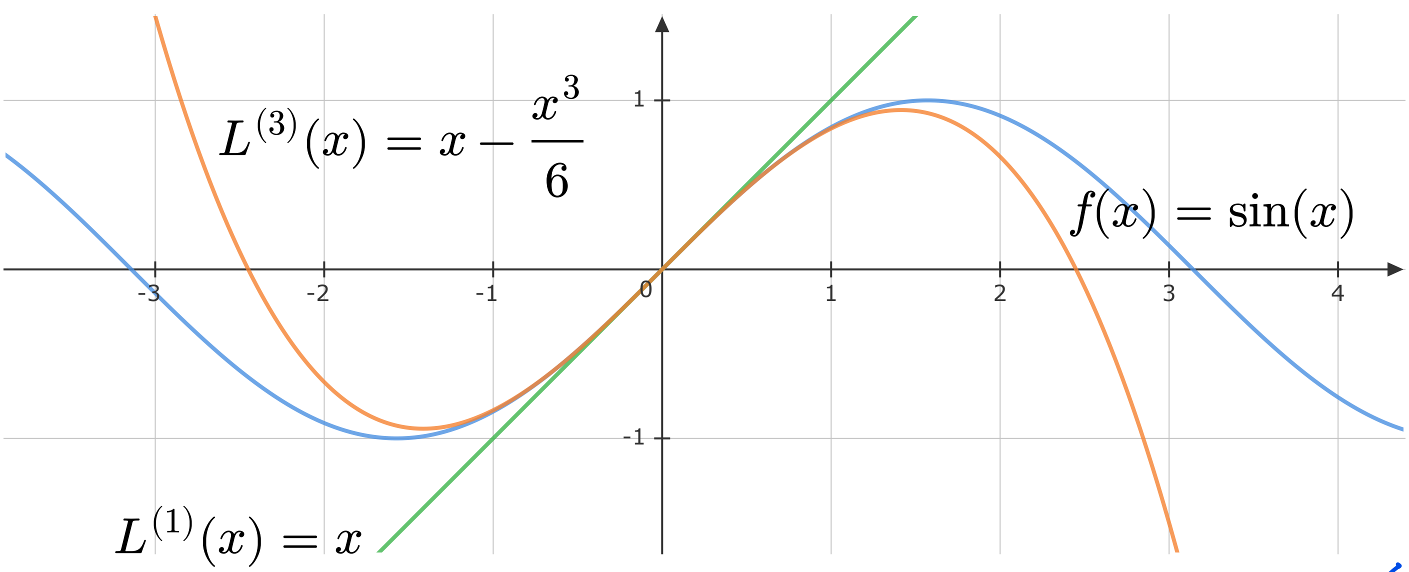
$$f(x) = 3x + 4$$

$$L_0[f](x) = f(0) + f'(0).\,(x - 0) = 4 + 3x = f(x)$$

Linear approximation to $\sin(x)$ at $x = \frac{\pi}{2}$.

$$L_{\pi/2} = \sin\left(\frac{\pi}{2}\right) + \cos\left(\frac{\pi}{2}\right) \cdot \left(x - \frac{\pi}{2}\right) = 1$$

Note that this is a line parallel to the x-axis, namely, $y = 1$.



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

$$L_0(x) = \sin 0 + \cos 0 \cdot (x - 0) = x$$

Second order approximation to $f(x)$

$$f(x) \approx f(a) + f'(a) \cdot (x - a) + \frac{f''(a)}{2!} \cdot (x - a)^2$$

Taylor polynomial.

$$\sin(0)+\cos(0)\cdot(x-0)-\frac{\sin(0)}{2!}\cdot(x-0)^2$$

Third order approximation to $f(x)$

$$f(x)\approx f(a)+f'(a)\cdot(x-a)+\frac{f''(a)}{2!}\cdot(x-a)^2+\frac{f'''(a)}{3!}(x-a)^3$$

$$\sin(0)+\cos(0)\cdot(x-0)-\frac{\sin(0)}{2!}\cdot(x-0)^2-\frac{\cos(0)}{3!}(x-0)^3$$

The third order approximation is the following Taylor polynomial:

$$x-\frac{x^3}{6}$$

In general, one can express $\sin(x)$ as:

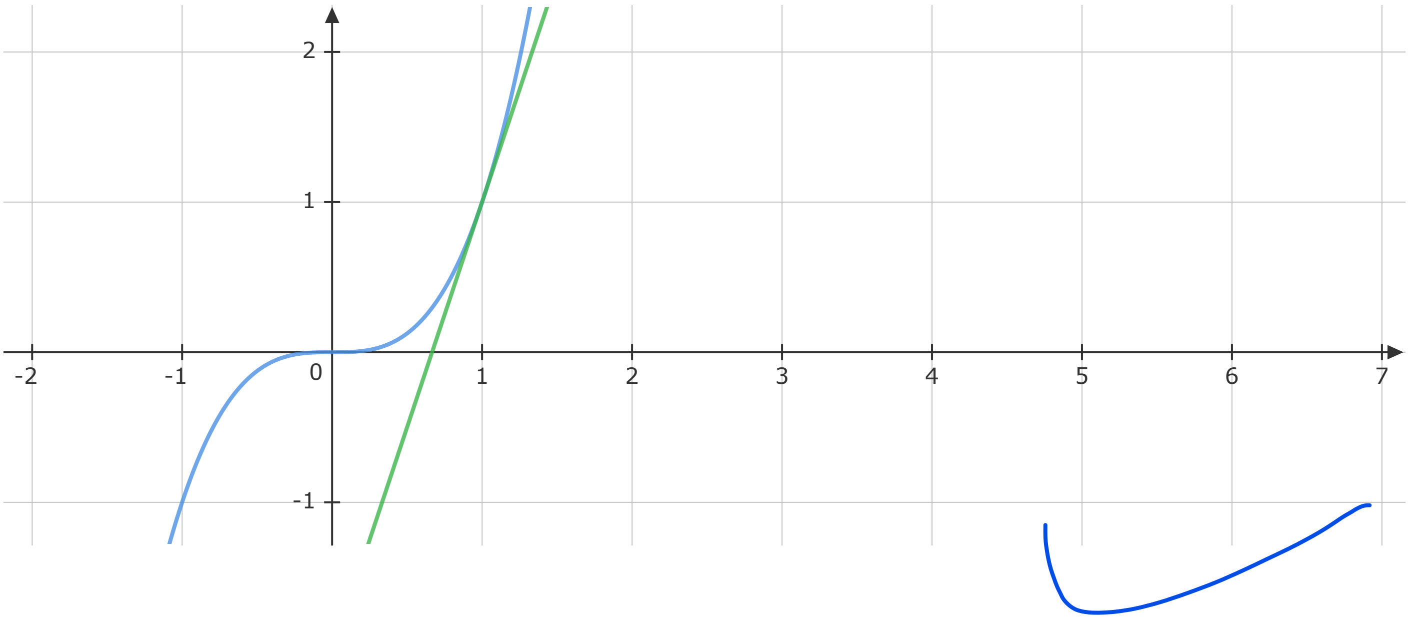
$$\boxed{\sin(x)=x-\frac{x^3}{3!}+\frac{x^5}{5!}-\cdots}$$

A linear approximation for $f(x)=x^3$ at $x=1$ is:

$$L_1[f](x)=f(1)+f'(1)\cdot(x-1)$$

$$L_1[f](x)=1+3(x-1)=3x-2$$

Visually, this is nothing but the tangent to f at $x=1$.



Maxima and Minima

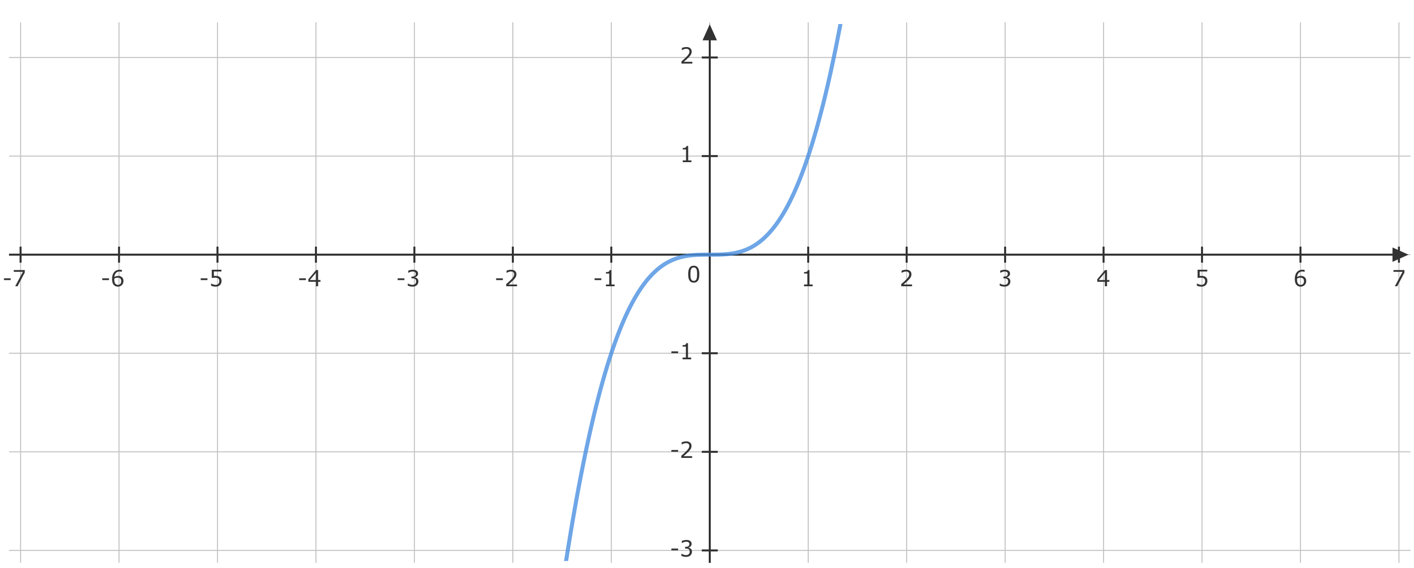
- If the function f is differentiable and if it has a local maximum/minimum at $x = a$, then $f'(a) = 0$
- **If the function f is differentiable and if $f'(a) = 0$, we will call $x = a$ a critical point.**

Note that critical points also include points where the function is not differentiable. For example, $x = 0$ is a critical point of $f(x) = |x|$. But to keep things simple, we shall only look at differentiable functions.

A classic example for a critical point that is not an extremum:

$$f(x) = x^3$$

$x = 0$



2D-3D Geometry

- [ ] Points
- [ ] Lines
- [ ] Planes and hyperplanes

$\mathbb{R}^2$

Points: (x, y)

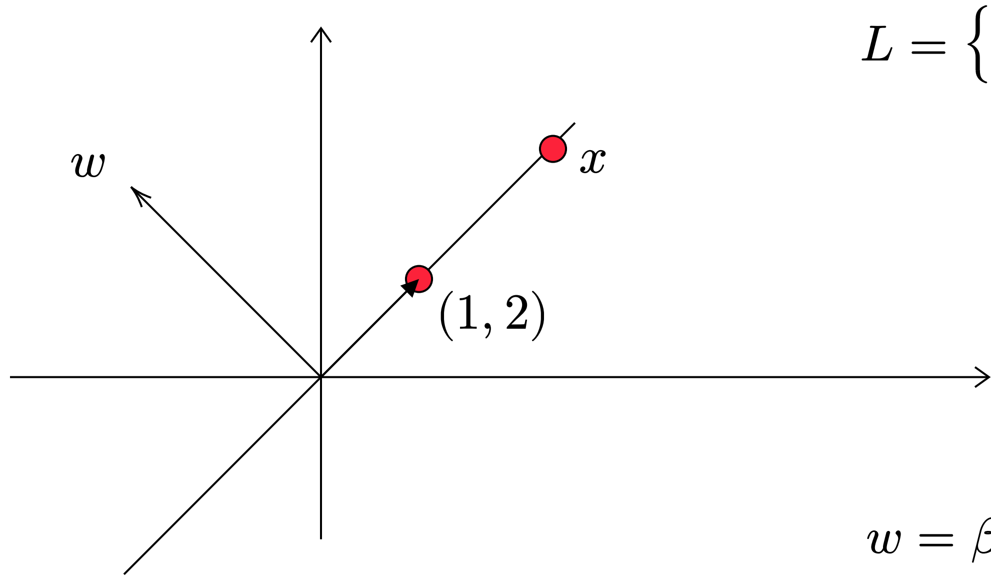
Points and vectors are one and the same thing:

$$(1, 2) \in \mathbb{R}^2$$

L is a line passing through the origin.

$$L = \{\alpha \cdot (1, 2) : \alpha \in \mathbb{R}\}$$

$$L = \left\{x : w^T x = 0\right\}$$



$$w = \beta \cdot (-2, 1), \ \ \beta \neq 0$$

Getting back to the line

$$L = \left\{x : w^T x = 0, w = \beta \cdot (-2, 1), \beta \neq 0\right\}$$

Vectors are represented as column vectors.

$$(1,2) \rightarrow \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

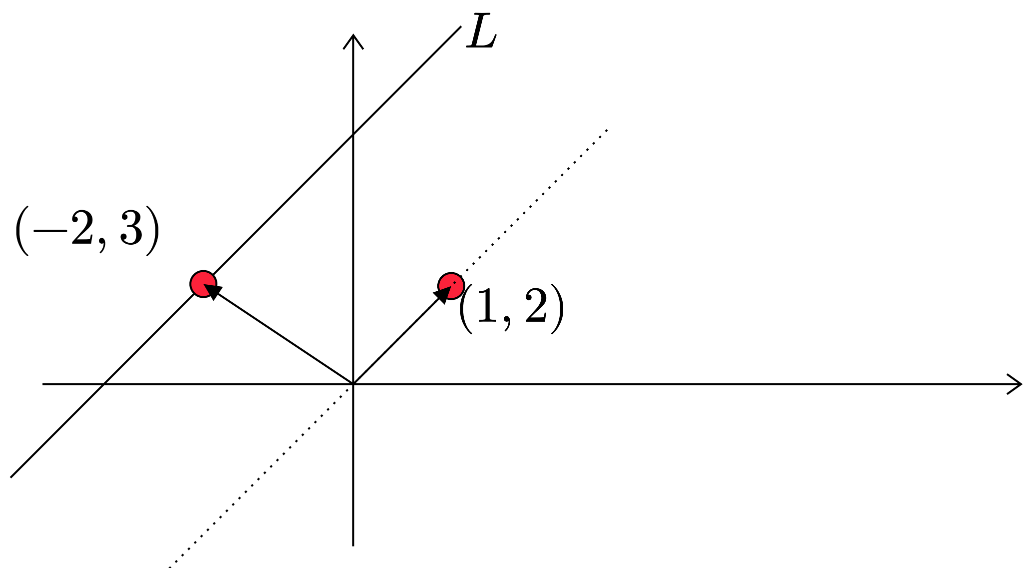
$$w = \begin{bmatrix} -2 \\ 1 \end{bmatrix}, w^T = \begin{bmatrix} -2 & 1 \end{bmatrix}$$

The dot product is often represented as $w \cdot x = w^T x$. Here, w^T is a row-vector and x is a column vector.

$$w^T x = w \cdot x$$

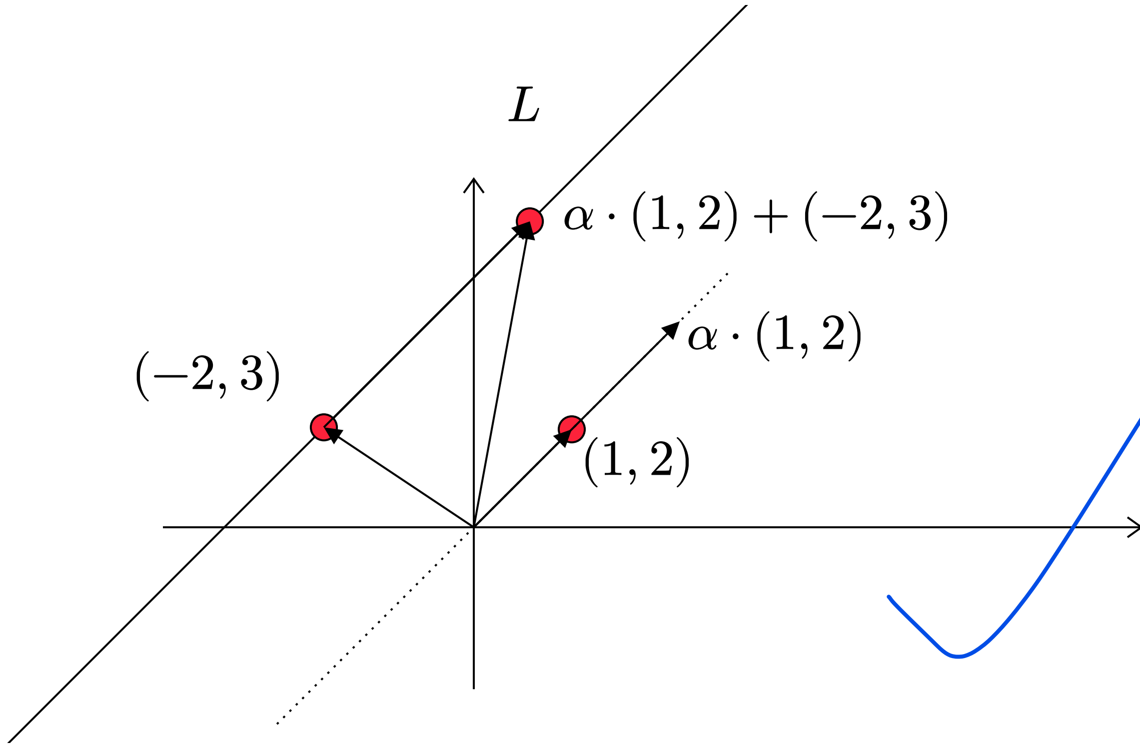
Geometrically, a line is perfectly determined if we have either of these things with us:

- A line parallel to it and some point on it.
- Two points on the line.



$$\alpha \cdot (1,2) + (-2,3)$$

$$L = \{\alpha \cdot (1,2) + (-2,3) : \alpha \in \mathbb{R}\}$$



w is going to be perpendicular to L also.

If you have two points x_1, x_2 on a line, then $x_1 - x_2$ is a vector parallel to the line.

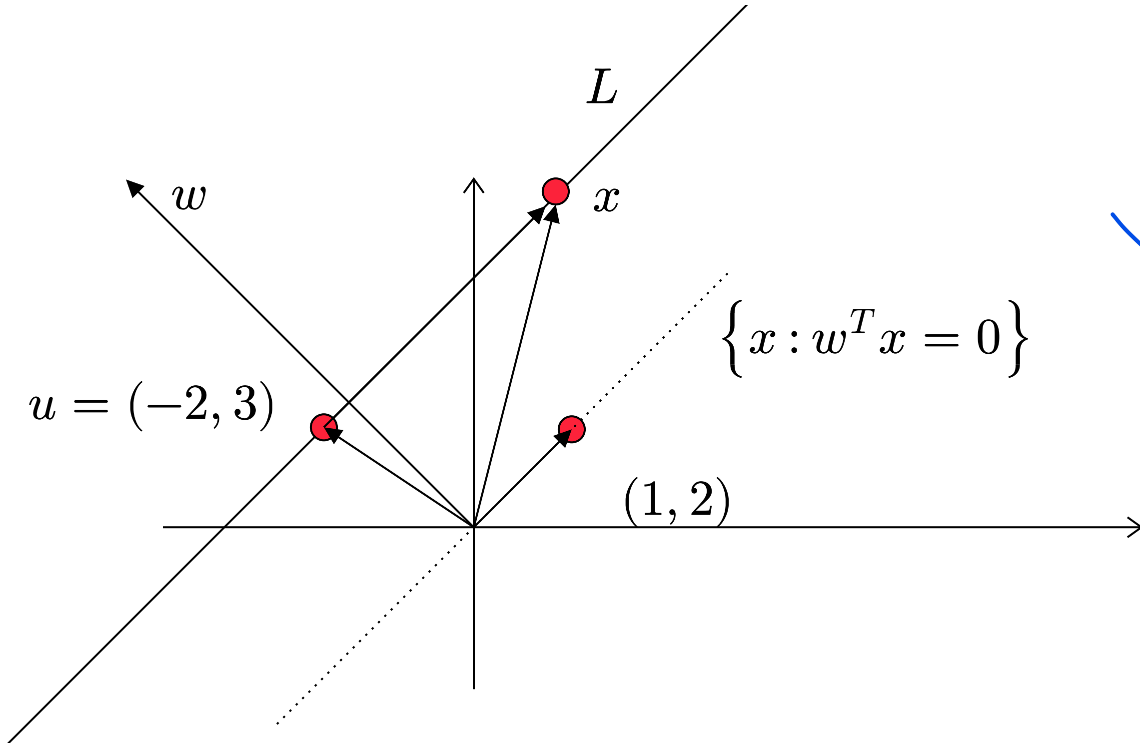
$$w^T(x - (-2,3)) = 0$$

$$b = -w^T(-2,3)$$

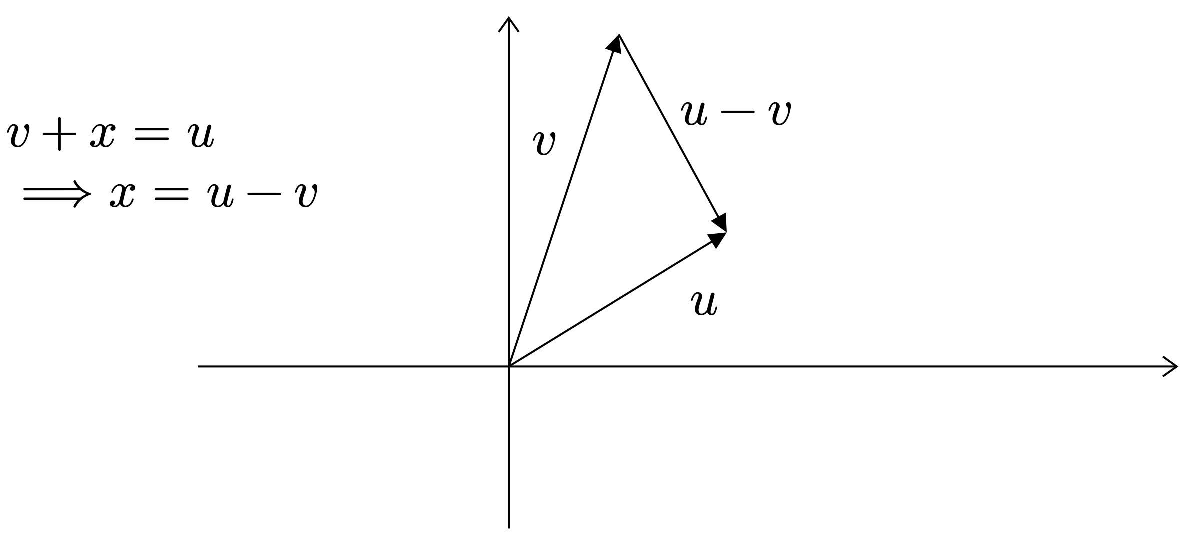
$$L = \left\{x : w^T x + b = 0\right\}$$

$$w^T x + b = 0$$

Here, $x - u$ is parallel to the line L. It represents the line segment starting from u and ending on x.



The following image should make the difference between two vectors more clear.



Caution: $w^T x + b = 0$ notation for a line works only if you are in $\mathbb{R}^2$. Lines in $\mathbb{R}^3$ can't be defined using $w^T x + b = 0$. You have to go back to the other method: specify a vector parallel to the line and a point on the line.

Planes

First we look at planes passing through the origin:

$$\left\{x : w^T x = 0\right\}$$

$$w_1 x_1 + w_2 x_2 + w_3 x_3 = 0$$

$$(1, 2, 3)$$

P is a plane passing through the origin whose normal is $(1, 2, 3)$.

$$P = \{(x_1, x_2, x_3) : x_1 + 2x_2 + 3x_3 = 0\}$$

For planes not necessarily passing through the origin. Let x, x_0 be two points on the plane. Then $x - x_0$ lies on the plane. w is perpendicular to the plane

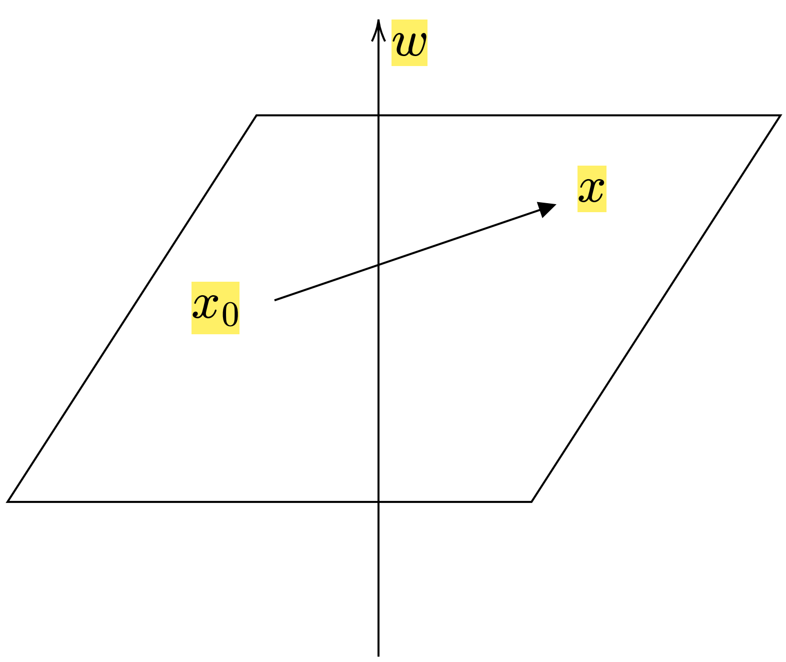
$$w^T(x - x_0) = 0$$

$$w^T x + b = 0$$

where,

$$b = -w^T x_0$$

Plane in $\mathbb{R}^3$



$P = \left\{x : w^T x + b = 0\right\}$ is in general called a hyperplane, especially, when you are in higher dimensions. Let us call the dimension d:

- If $d = 2$, P is a line

- If $d=3$, P is a plane
- If $d>3$, P is a hyperplane

A line has dimension 1

$$\{\alpha \cdot (x,y) : \alpha \in \mathbb{R}\}$$

span of just one vector, namely (x,y).

A plane has dimension 2

$$\{\alpha \cdot (a,b) + \beta \cdot (c,d) : \alpha, \beta \in \mathbb{R}\}$$

A hyperplane in $\mathbb{R}^d$ has dimension $d-1$.

Why is this the case? A basis for $\mathbb{R}^d$ has d vectors. If you have a vector $w \in \mathbb{R}^d$. A hyperplane is the collection of all vectors perpendicular to w. You can now form an orthogonal basis with w as the first element.

$$\{w, w_1, \cdots, w_{d-1}\}$$

It follows that span$\{w_1, \cdots, w_{d-1}\}$ is the set of all vectors orthogonal to w. This is the hyperplane and its dimension is $d-1$.

$$\text{span}\{w_1, \cdots, w_{d-1}\} \to d-1$$

Use of hyperplanes in ML

- Linear classifiers
  - supervised learning
    - classification
      - spam vs not-spam
      - dog vs cat
    - anything that falls on one side of the hyperplane belongs to class spam and the rest belong to not-spam

Functions of two variables

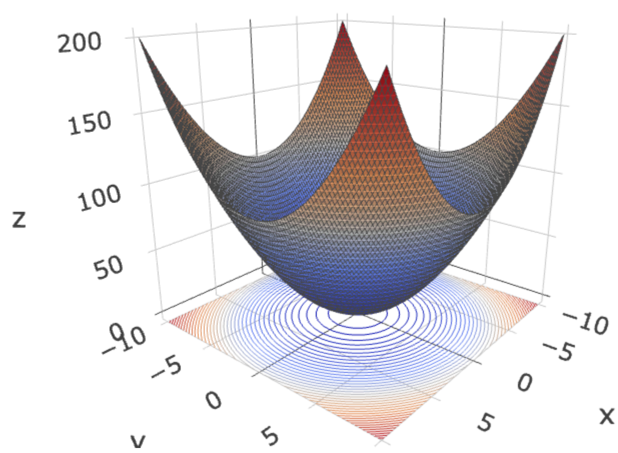
- ☑ Contours
- ☑ Partial derivatives, gradient
- ☑ Directional derivative
- ☑ Taylor series expansion
- ☑ Linear approximation
- ☐ Gradient $\rightarrow$ steepest ascent, orthogonality to contours

$$f(x,y) = x^2 + y^2$$

A function of two variables can be visualized as a surface in $\mathbb{R}^3$:

$$(x, y, f(x,y))$$

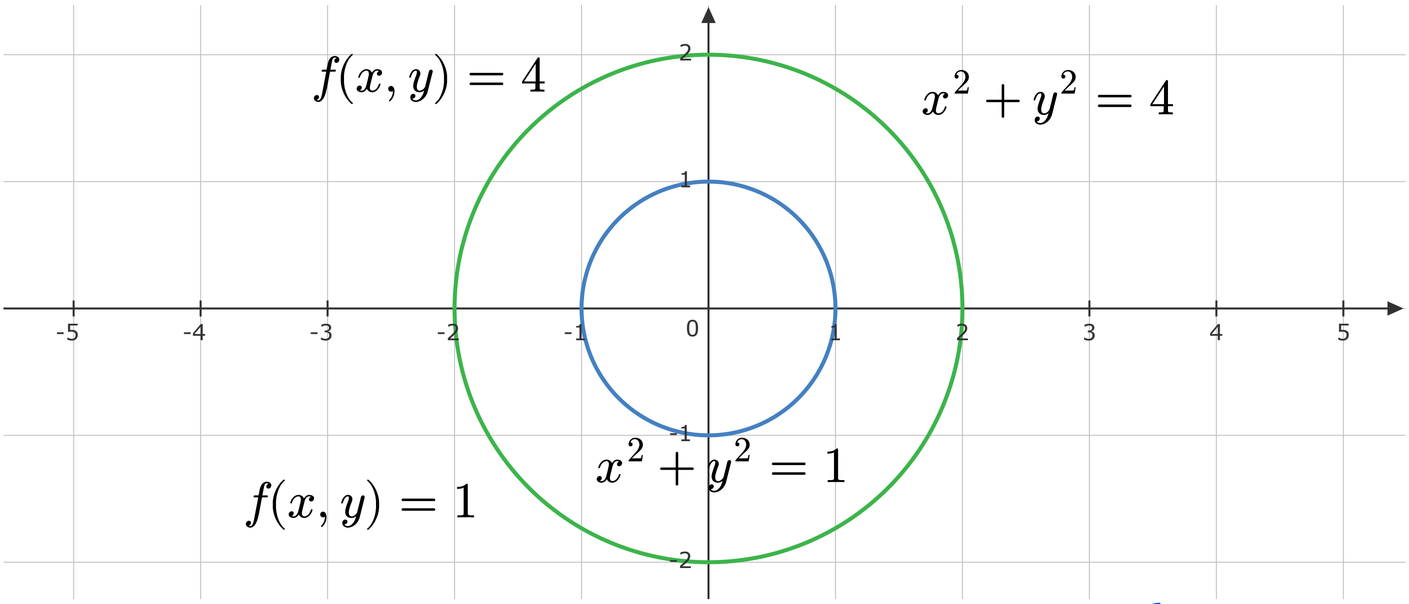
Function's value is the (signed) height of the surface above the XY plane.



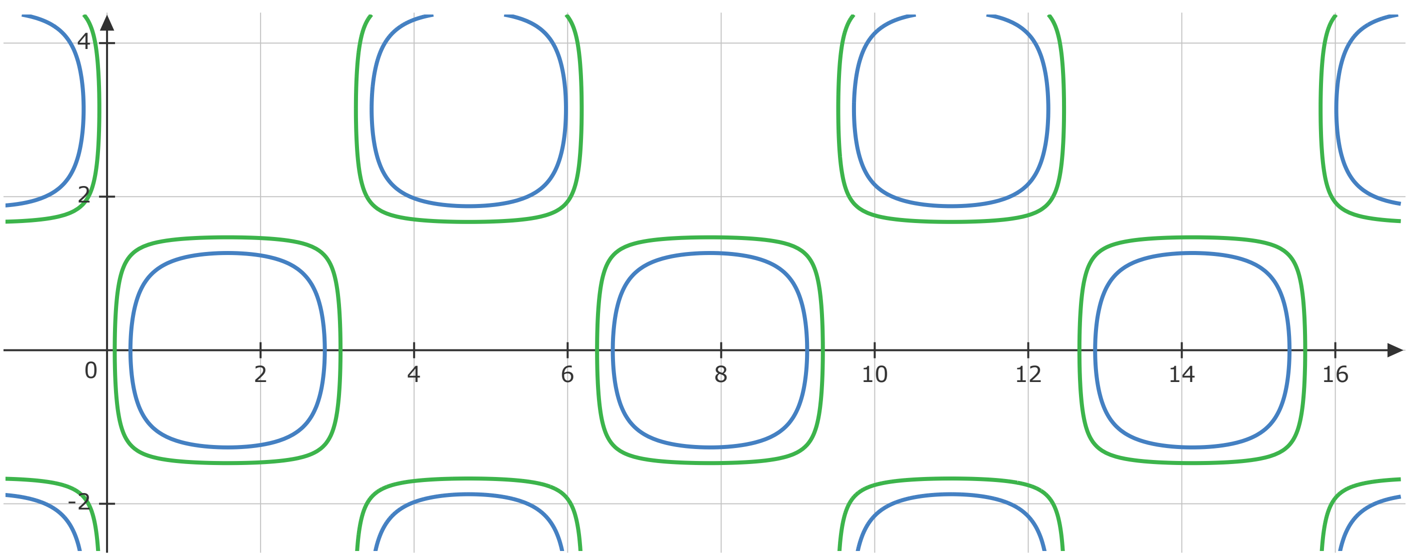
Contours, also called level curves, are obtained by slicing the surface with planes parallel to the XY plane:

$$f(x,y) = 1$$
$$f(x,y) = 2$$
$$f(x,y) = 3$$

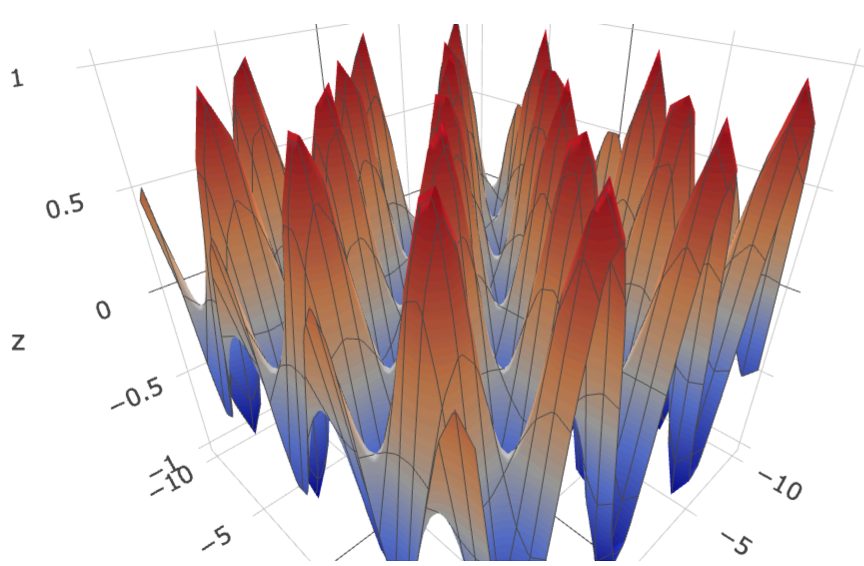
What is the value of $f(x,y)$ on any point on the green curve?



Contours for $f(x, y) = \sin(x)\cos(y)$



The surface plot for this function:



Partial derivatives and gradients

$$f(x, y)$$

$$\nabla f = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right)$$

Directional derivatives: the rate of change of f in some direction u, where u is a unit vector. If the function f is differentiable, then the directional derivative is given by:

$$D_u[f] = (\nabla f)^T u$$

$$D_u[f] = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix}^T \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$$

$$D_u[f] = \frac{\partial f}{\partial x} u_1 + \frac{\partial f}{\partial y} u_2$$

You should be able to recover the partial derivatives by setting $u = (1, 0)$ and $u = (0, 1)$. Partial derivatives are specific directional derivatives.

Linear approximation

$$f(x, y)$$

Linear approximation to f at (a, b)

$$L_{(a,b)}[f](x) = f(a, b) + (\nabla f)^T \begin{bmatrix} x - a \\ y - b \end{bmatrix}$$

$$\boxed{f(a,b) + \frac{\partial f}{\partial x}(x-a) + \frac{\partial f}{\partial y}(y-b)}$$

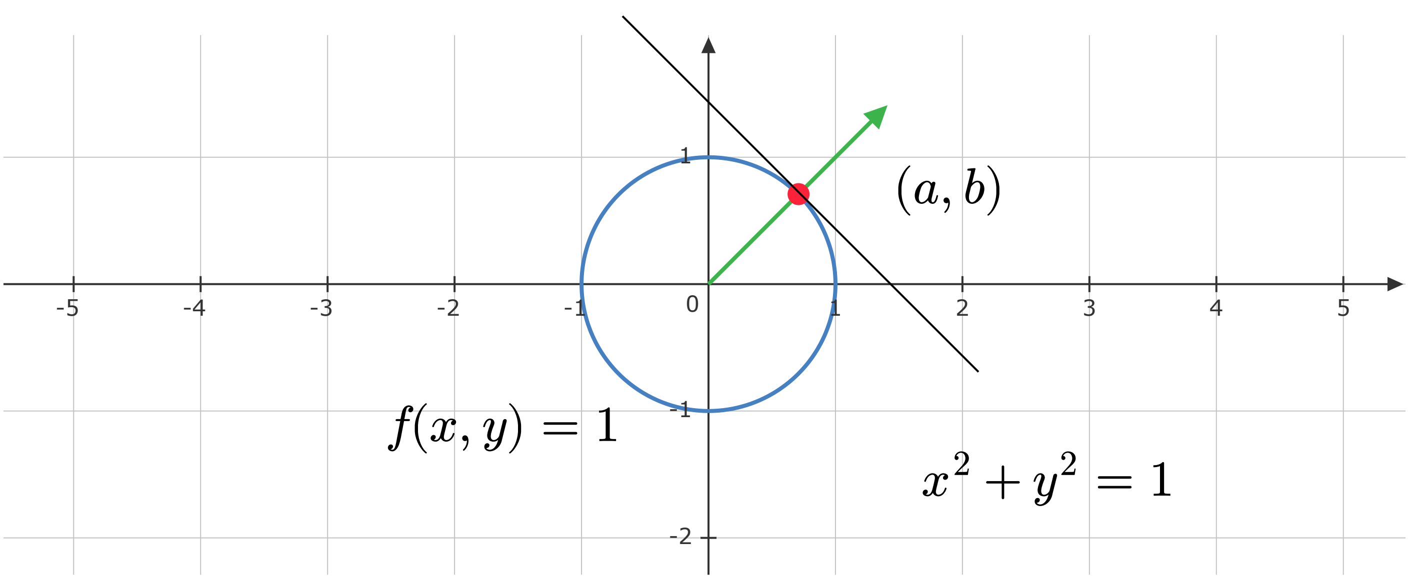
Gradient and contours

$f(x,y) = x^2 + y^2$

$$\nabla f = (2x, 2y)$$

$$\nabla f_{(1/\sqrt{2},1/\sqrt{2})} = \left(\frac{2}{\sqrt{2}}, \frac{2}{\sqrt{2}}\right) = (\sqrt{2}, \sqrt{2})$$

The gradient of f at the red point is the green vector.



$$f(x,y) \approx f(a,b) + \nabla f(a,b)^T \begin{bmatrix} x-a \\ y-b \end{bmatrix}$$

If you move by a small distance around (a,b) while remaining on the circle, the value of $f(x,y)$ will continue to remain $f(a,b)$.

$$f(a,b) \approx f(a,b) + \nabla f(a,b)^T \begin{bmatrix} x-a \\ y-b \end{bmatrix}$$

$$\nabla f(a,b)^T \begin{bmatrix} x-a \\ y-b \end{bmatrix} = 0$$

The vector $(x-a, y-b)$ is parallel to the tangent at (a, b):

$$(x, y) - (a, b)$$

Therefore, we have:

The gradient at a point on the contour is perpendicular to the (tangent to) contour at that point.

Gradient and steepest ascent

If I move away from (a, b), then the function's value will change by the following amount:

$$\boxed{f(x, y) - f(a, b)} \approx \nabla f(a, b)^T \begin{bmatrix} x-a \\ y-b \end{bmatrix}$$

You will get the maximum change when the RHS is maximum.

When will this quantity be maximum?

$$\nabla f(a, b)^T \begin{bmatrix} x-a \\ y-b \end{bmatrix}$$

When the two vectors are in the same direction, meaning $\theta = 0$.

The direction of steepest ascent is the gradient.