Mathematical Overview - L1 - 22/09/23 - Optimization

$$\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = x = (x_1, \dots x_n)^T \in \mathbb{R}^N \qquad y = (y_1, \dots, y_n)$$

$$x \cdot y = \langle x, y \rangle = x_1 y_1 + \dots + x_n y_n = \sum_{i=1}^{n} x_i y_i$$

dot product (scalar product)

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

$\| x \|_p = \left(\sqrt{x_1^p + \dots + x_n^p}\right)^{1/p}$ (mintorski norm) $p=1$ city block distance

$d(x,y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$ distance

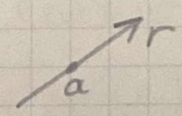
$\cos(x,y) = \frac{x \cdot y}{\|x\| \|y\|}$

$x \perp y \iff x \cdot y = 0$

$$x \times y = \begin{bmatrix} i & j & k \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{bmatrix} = \begin{bmatrix} x_2 y_3 - x_3 y_2 \\ x_1 y_3 - x_3 y_1 \\ x_1 y_2 - x_2 y_1 \end{bmatrix}$$ (cross product in $\mathbb{R}^3$)

<u>fact</u>: $x \times y \perp x$, $x \times y \perp y$

<u>area</u> $= \| x \times y \|$ <u>line in $\mathbb{R}^3$</u>:

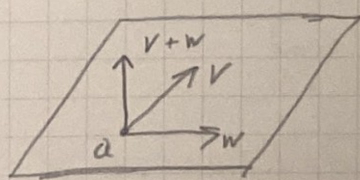


$a = (a_1, a_2, a_3)$, $v = (v_1, v_2, v_3)$

$x = d + tv$, $t \in \mathbb{R}$ (line through d, in direction v)

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} + t \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}$$

plane in $\mathbb{R}^3$:



$x = a + tv + sw$, $t, s \in \mathbb{R}$ (plane through a, spanned by v, w)

In other cordinates (x, y, z): $Ax + By + Cz + D = 0$, $A, B, C, D \in \mathbb{R}$

$(A, B, C)^T$ is a normal

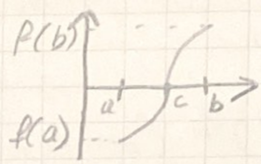
$f: D \to \mathbb{R}$, $D \subset \mathbb{R}^n$

f is continious at $a \in D$ iff $\exists \lim_{x \to a} f(x) = f(a)$

$\hookrightarrow \| x - a \| \to 0$

Bolzano Theorem

$f: [a,b] \to \mathbb{R}$ continious
assume $f(a) f(b) < 0$
then $\exists\, c \in (a,b)$ s.t. $f(c) = 0$



Weierstraß Theorem

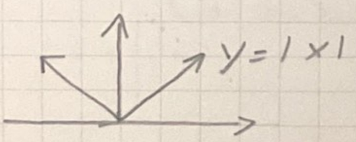
$f: K \to \mathbb{R}$ continious , K is compact ($\iff$ closed and $\mathbb{R}^n$ bounded)
then f is bounded and acheives it's maximum/minimum $([a,b])$

$$f'(a) = \lim_{h \to 0} \frac{f(h+a) - f(h)}{h} \quad \text{(if it exists)}$$

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

$$\approx f(a+h) \simeq f(a) + f'(a) \cdot h$$

f is diffentiable in open $D \subset \mathbb{R} \iff \forall a \in D, \exists f'(a)$



continious but
not differentiable
($\nexists f'(a)$ at $x = 0$)

$\mathbb{R}^n$, $f: \mathbb{R}^n \to \mathbb{R}$, $a = (a_1, \dots, a_n)^T$

Partial derivative at a in direction of x_i:

$$f_{x_i}(a) = \frac{\partial f}{\partial x_i}(a) = \lim_{h \to 0} \frac{f(a_1, \dots, a_{i-1}, a_i + h, a_{i+1}, \dots) - f(a_1, \dots, a_n)}{h}$$

Gradient of f:

$$\nabla f(a) = \left(\frac{\partial f}{\partial x_1}(a), \dots, \frac{\partial f}{\partial x_n}(a) \right)^T$$

f is differentiable at $a \in \mathbb{R}^n$ in the domain of f iff

$$\lim_{\|h\| \to 0} \frac{|f(a+h) - f(a) - \nabla f(a)^T \cdot h|}{\|h\|} = 0$$

Directional derivative: $r \in \mathbb{R}^n$, $a \in \mathbb{R}^n$

$$\nabla_r f(a) = \lim_{h \to 0} \frac{f(a+hr) - f(a)}{h} \quad \text{(assuming } \|r\| = 1\text{)}$$

Lemma: $\nabla_r f(a) = (\nabla f(a))^T \cdot v$

Proof: $\nabla_r f(a) = \frac{\partial}{\partial h} f(a+hv)\big|_{h=0}$

$$= \sum_{i=1}^{n} \left(\frac{\partial f}{\partial x_i}(a+hv) \frac{\partial}{\partial h}(a+hv) \right)$$

↳ in the direction of i

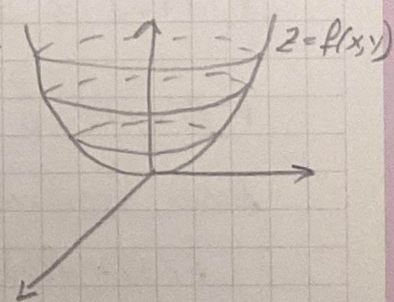
$a + hv = (a_1 + hv_1, a_2 + hv_2, \ldots, a_n + hv_n)$

$$= \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(a) \cdot v_i = \nabla f(a) \cdot v \quad \blacksquare$$

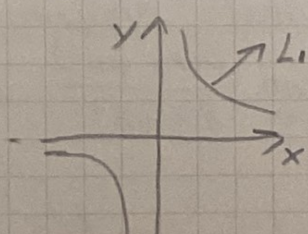
↳ by definition

For a function $f: \mathbb{R}^n \to \mathbb{R}$, it's level set $L_c = \{x \in \mathbb{R}^n : f(x) = c\}$, $c \in \mathbb{R}$

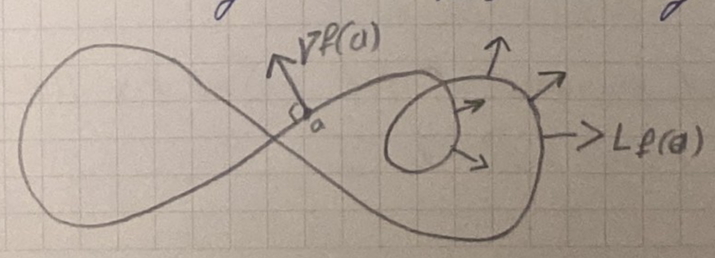
- $f(x, y) = x^2 + y^2 \to$ all L_c, for $c > 0$ are circles



- $f(x, y) = x^2 - y^2$; $L_1 = \{(x, y)^T \mid x^2 - y^2 = 1\}$



Lemma: $\forall$ a in the domain of definition of $f: \mathbb{R}^n \to \mathbb{R}$. $\nabla f(a)$ is orthogonal to $L_{f(a)}$ (assuming it is defined, $\nabla f(a) \neq 0$)



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

$L_0 =$  $\nabla f(0,0) = 0$

Partial derivatives of order 2:

$$f_{x_i x_j}(a) = \frac{\partial^2 f}{\partial x_i \partial x_j}(a) = \frac{\partial}{\partial x_i}\left(\frac{\partial f}{\partial x_j}\right)(a) \text{ or } \frac{\partial^2 f}{\partial x_j \partial x_i}(a)$$

Hessian of f at $a \in \mathbb{R}^n$

$$\nabla^2 f(a) = Hf(a) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2}(a) & \frac{\partial^2 f}{\partial x_2 \partial x_1}(a) & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_1} \\ \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_1} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n} & \frac{\partial^2 f}{\partial x_2 \partial x_n} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

in our setting (assuming f is smooth enough)

$\nabla^2 f(a)$ is a symmetric $n \times n$ matrix

$$\nabla^2 f(x,y) = \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix}$$

linear approximation (follows from the taylor formula)

$$f(x) \approx f(a) + (\nabla f(a))^T \cdot (x-a) + O(\|x-a\|^2)$$

quadratic approximation: $f(x) = f(a) + (\nabla f(a))^T \cdot (x-a) + \frac{1}{2}(x-a)^T \cdot \nabla^2 f(a) \cdot (x-a) + O(\|x-a\|^3)$

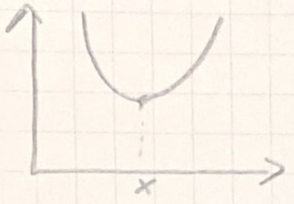
$\lim\limits_{x \to a} \frac{O(\|x-a\|^2)}{\|x-a\|^2}$ is bounded

Local optimality conditions

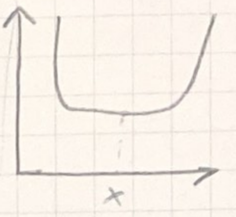
$f(x) \to \min$ $\quad -f(x) \to \max$

- x^* is a local minimizer iff $\exists \delta > 0 \ \forall x : \|x - x^*\| < \delta$ we have $f(x^*) \leq f(x)$

- x^* is a strict local minimizer iff $f(x^*) < f(x)$ for all $x : 0 < \|x - x^*\| < \delta$



strict local minimizer



not a strict local minimizer

Theorem: Necessity Conditions for local minima

(1) (first order) $\nabla f(x^*) = 0$

(2) (second order) $\nabla^2 f(x^*)$: positive semi-definite

$\nabla^2 f(x^*)$ is positive semi-definite iff

$x^T \nabla^2 f(x^*) \cdot x \geq 0 \quad \forall x \in \mathbb{R}^n$