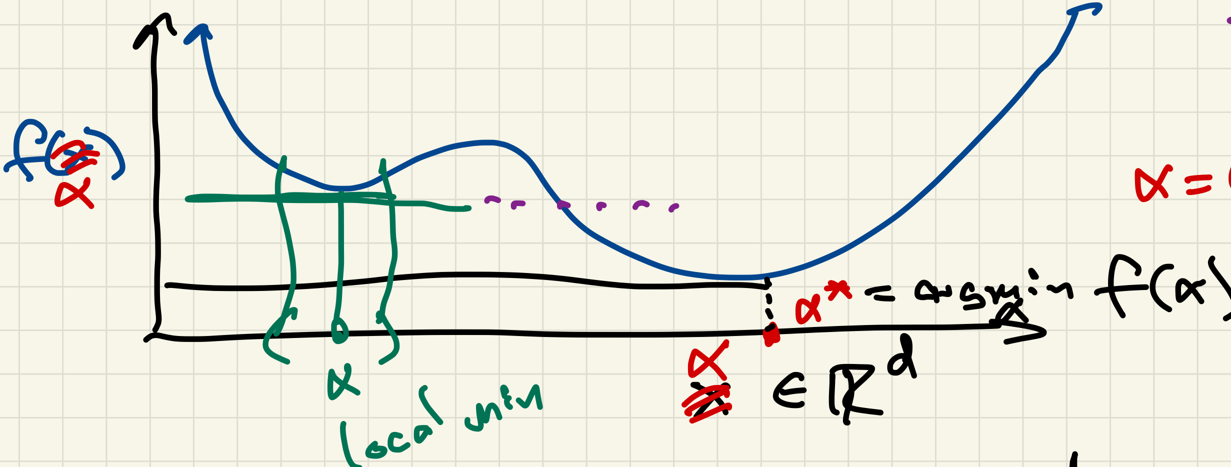
FoDA L15

Gradient Descent #1

Functions & Gradients

Oct 18, 2022

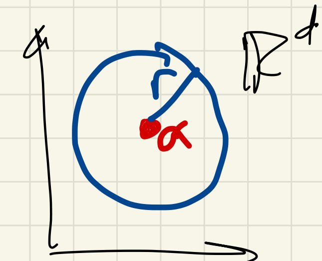
$\forall \equiv$ for all



$\alpha = (\alpha_1, \alpha_2, \ldots \alpha_d)$

<u>Neighborhood</u> $B_r(\alpha) = \{p \in \mathbb{R}^d \mid \|p-\alpha\| \leq r\}$

$r > 0$



<u>local Minimum</u> of f

$\alpha \in \mathbb{R}^d$ so all $p \in B_r(\alpha)$

$f(p) \geq f(\alpha)$ for some B_r

<u>global minimum of f</u> $\alpha \in \mathbb{R}^d$ $\forall p \in \mathbb{R}^d$ $f(p) \geq f(\alpha)$

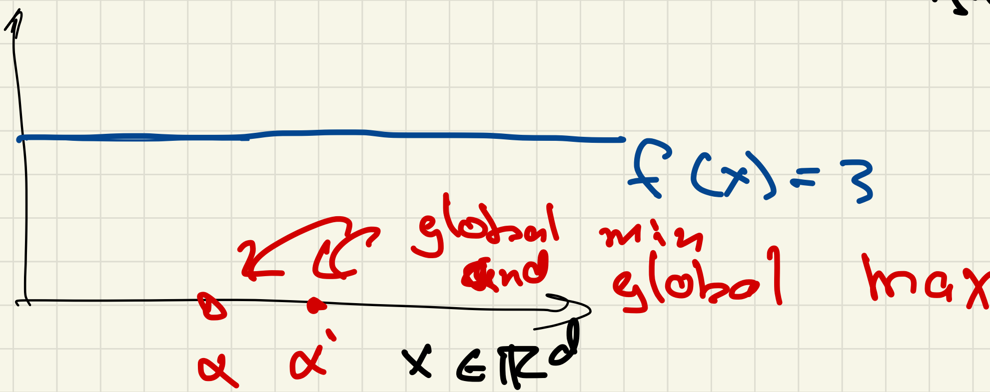
<u>maximum of f</u>

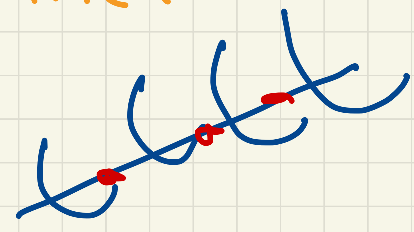
$\alpha \in \mathbb{R}^d$ s.t.

$\forall p \in \cancel{B_r(\alpha)}$ $\mathbb{R}^d \Rightarrow$ global

$f(p) \underset{\neq}{\le} f(\alpha)$

strict

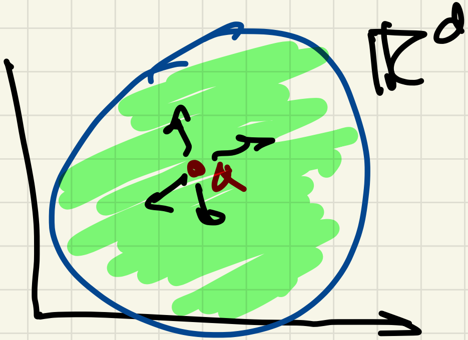




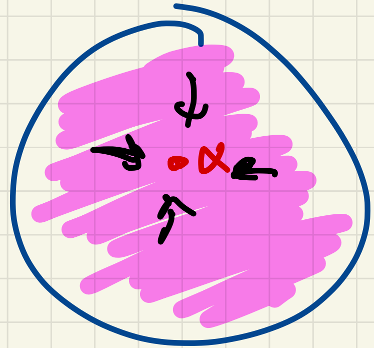
<u>strict minimum of f</u>

$\alpha \in \mathbb{R}^d$ s.t. $\forall p \in B_r(\alpha)$

$f(p) > f(\alpha)$ for $p \neq \alpha$

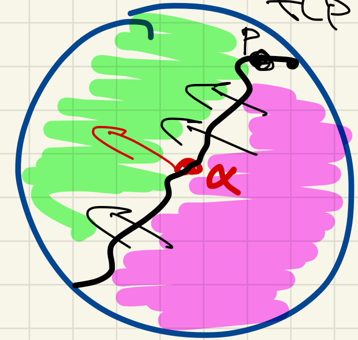


α max

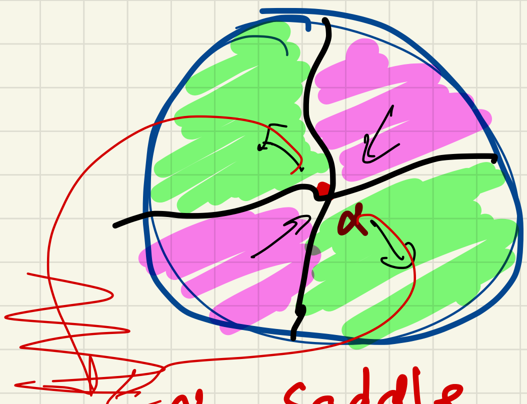


α min

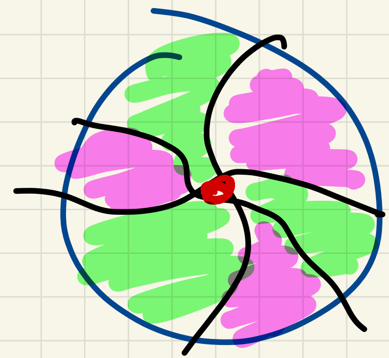
$f(p) = f(\alpha)$



α regular

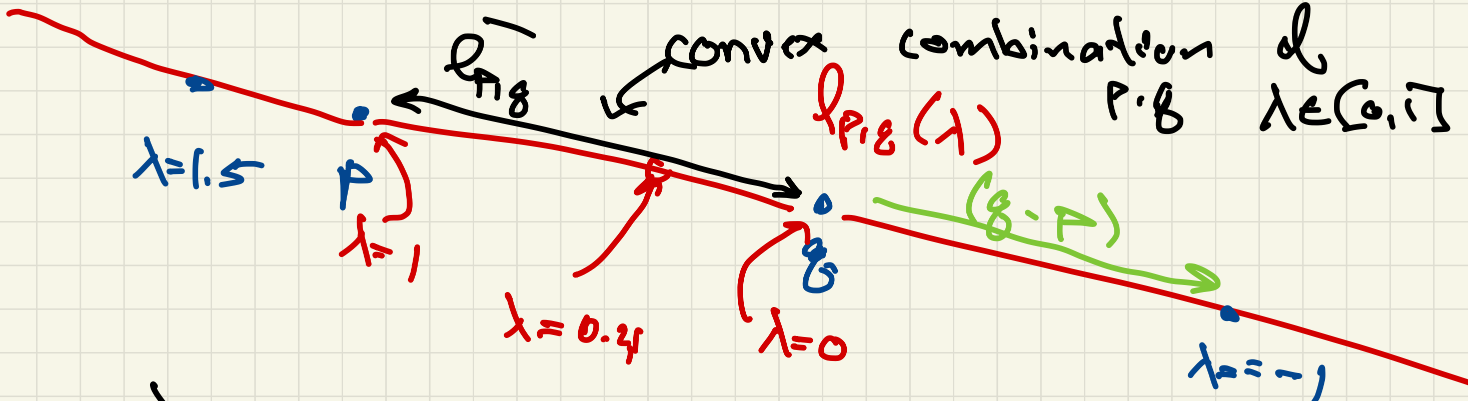


α saddle



Convex Functions

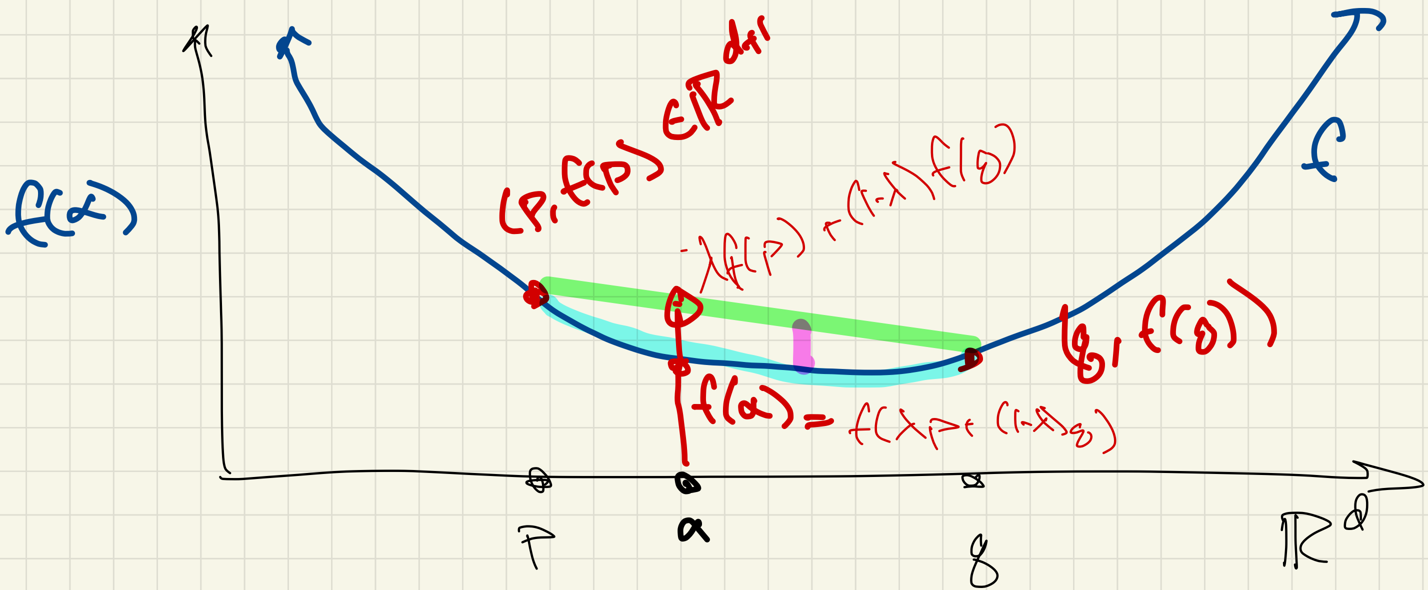
$\ell_{p,q}$ line going through $p, q \in \mathbb{R}^d$



$$\ell_{p,q}(\lambda) = \{ x = \lambda p + (1-\lambda) q \mid \lambda \in \mathbb{R} \}$$

$$-p + (1 - -1) q$$

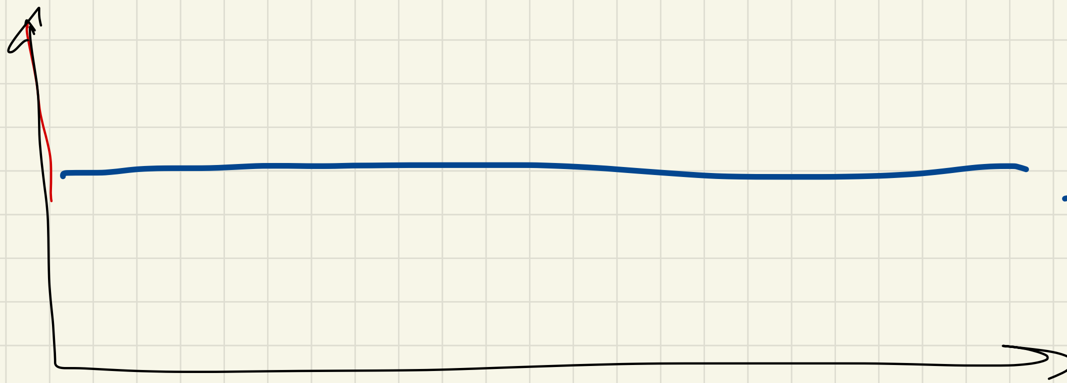
$$2q - p = q + (q - p)$$



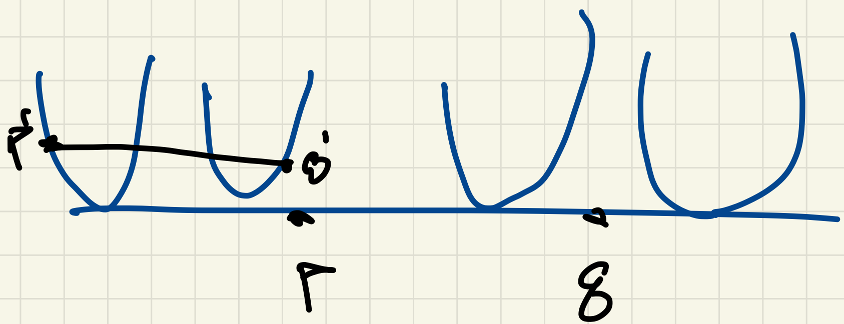
f is convex if $\forall\, p, q \in \mathbb{R}^d$ all $\lambda \in [0,1]$ $(0,1)$

$$f(\underbrace{\lambda p + (1-\lambda) q}_{\alpha}) \leq \lambda f(p) + (1-\lambda) f(q)$$

strict



Convex but not strictly convex



Convex function Properties

- f convex $\Rightarrow$ local min also global min

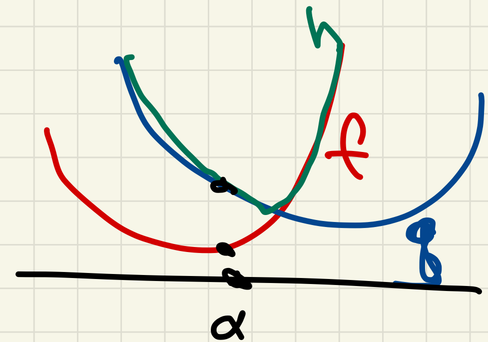
- f, g convex

  $h = g + f \rightarrow$ convex

  $h = \max\{f, g\} \rightarrow$ convex

  $h = f / \beta \rightarrow$ convex

  $\beta \leftarrow$ constant



$$f(x) = f(x_1, x_2, \ldots, x_d)$$

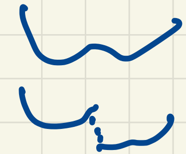
unit vector

$u = (u_1, u_2, \ldots u_d)$

$\|u\| = 1$

directional derivative

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

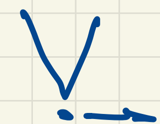


contin

not contin

if $\nabla_u f(x)$ is well-defined for all x, u

differentiable



∇ \ nabla

unit vectors $e_1, e_2, \ldots e_d$

$$e_i = (0, 0, \ldots 0, \underbrace{1}_{i\text{th coord}}, 0, 0 \ldots 0)$$

$$\nabla_i f(x) = \nabla_{e_i} f(x) = \frac{\partial}{\partial x_i} f(x)$$

gradient of f

$$\nabla f = \frac{\partial f}{\partial x_1} e_1 + \frac{\partial f}{\partial x_2} e_2 + \dots + \frac{\partial f}{\partial x_d} e_d$$

$$= \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \dots, \frac{\partial f}{\partial x_d} \right)$$

$$\nabla f : \mathbb{R}^d \to \mathbb{R}^d$$

direction (+ scale) steepest increase

Example

$$\alpha = (x, y, z) \in \mathbb{R}^3$$

$$f(x, y, z) = 3x^2 - 2y^3 - 2xe^z$$

$$\nabla f = (6x - 2e^z, -6y^2, -2xe^z)$$

$$\nabla f(3, -2, 1) = (18 - 2e, -24, -6e)$$