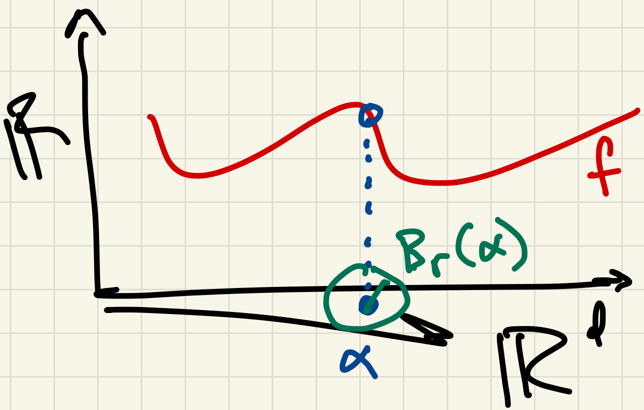
FoDA - L15

Gradient Descent

Functions

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

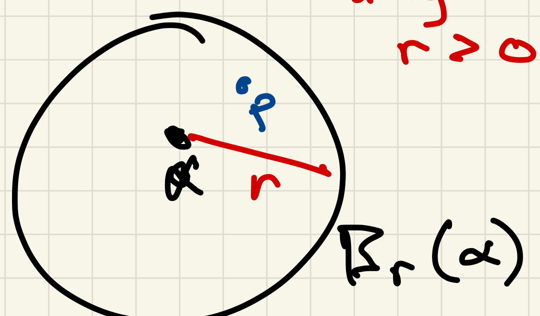
$f(\alpha)$ $\quad \alpha = (\alpha_1, \alpha_2, \ldots \alpha_d) \in \mathbb{R}^d$



Local Neighborhood

Ball $\quad B_r(\alpha) = \{p \in \mathbb{R}^d \mid \|\alpha - p\| \leq r\}$

any $r > 0$



minimum

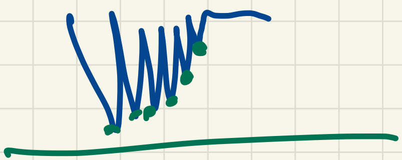
local maximum of f: a point $\alpha \in \mathbb{R}^d$

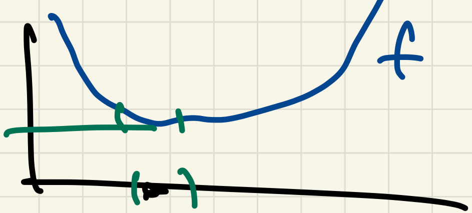
s.t. local nbhd $B_r(\alpha)$ all $p \in B_r(\alpha)$

$f(p) \geq f(\alpha)$

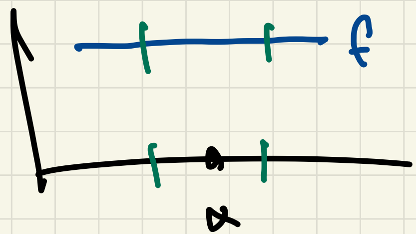
have $f(p) \leq f(\alpha)$

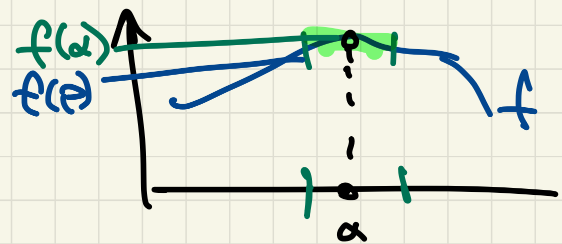
is strict $f(p) < f(\alpha)$

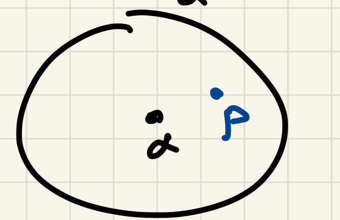




$\alpha \leftarrow$ local min







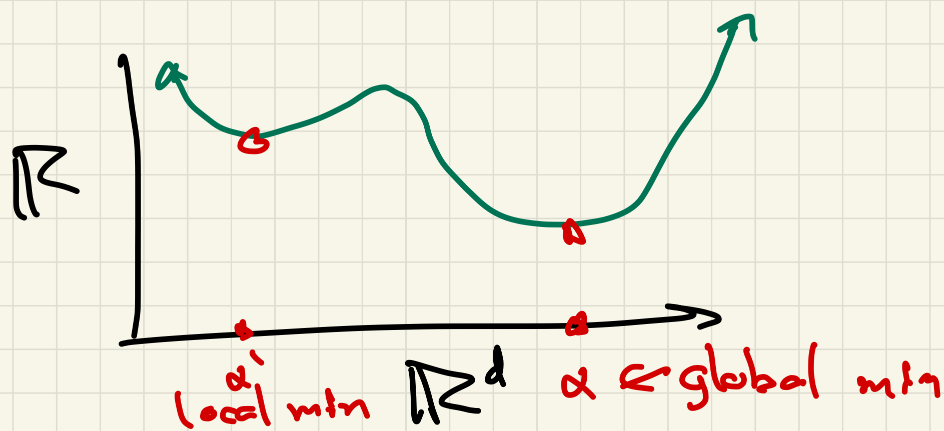
$B(\alpha)$

Global Maximum / Mininom

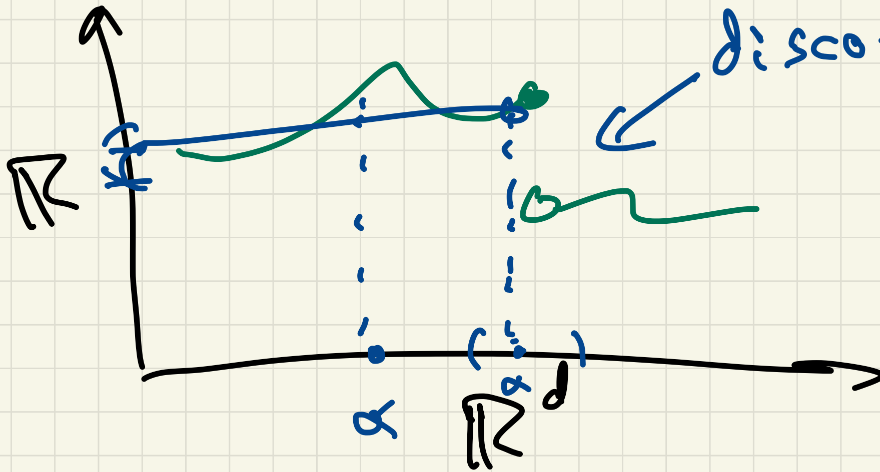
A point $\alpha \in \mathbb{R}^d$ s.t. for all $p \in \mathbb{R}^d$

$$f(p) \leq f(\alpha) \;/\; f(p) \geq f(\alpha)$$

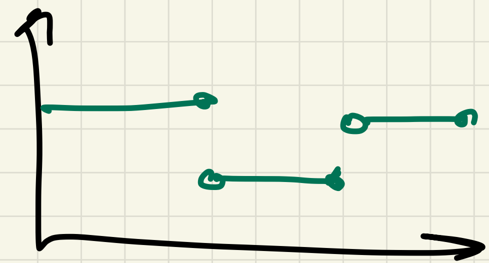
is strict $f(p) < f(\alpha)$ or $f(p) > f(\alpha)$

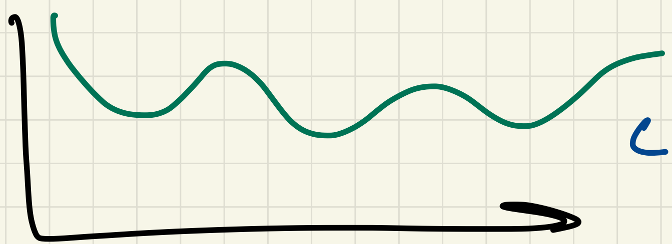


Continuous functions



discontinuity





continuous

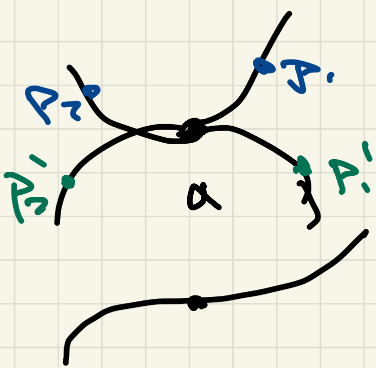
Saddle points

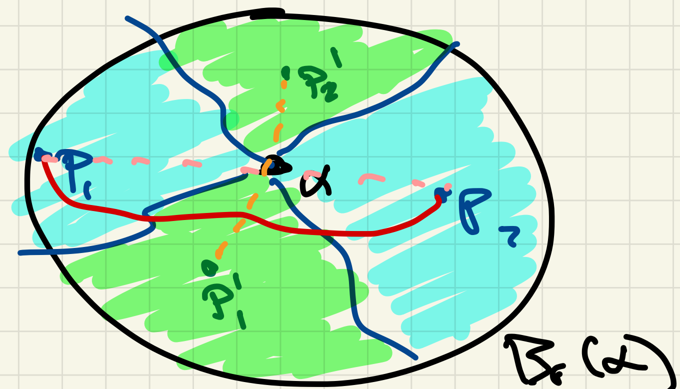
$\alpha \in \mathbb{R}^d$ s.t. for local nbhd $B_r(\alpha)$

some $p \in B_r(\alpha)$ $f(p) > f(\alpha)$

and some $p' \in B_r(\alpha)$ $f(p') < f(\alpha)$

and regions are not connected





$f(p_1) > f(\alpha)$

$f(p_2) > f(\alpha)$

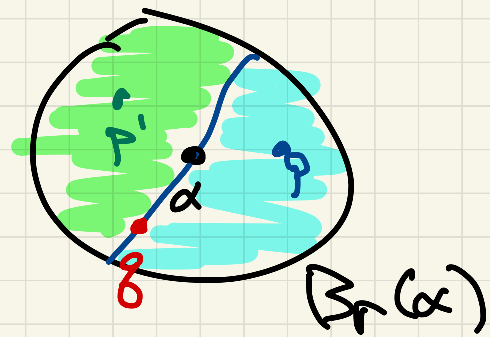
$f(p_1') < f(\alpha)$

$f(p_2') < f(\alpha)$

Regular point

$\alpha \in \mathbb{R}^d$ that is not a local min, max or saddle point.

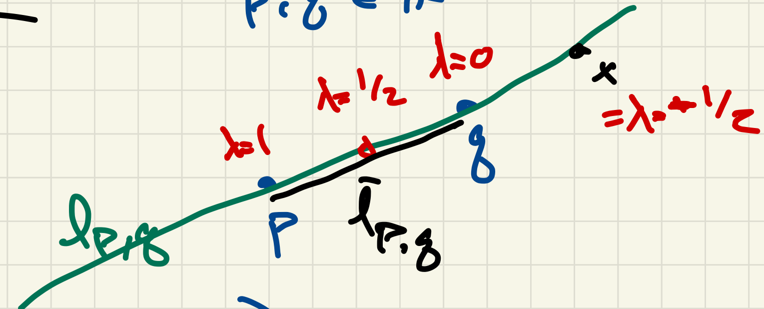
$$f(p') < f(\alpha) < f(p)$$



$f(q) = f(\alpha)$

Convex functions

$p, q \in \mathbb{R}^d$



line $\ell_{p,q}$

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

linear combination

$$\bar{\ell}_{p,q} = \{x = \lambda p + (1-\lambda) q \mid \lambda \in [0,1]\}$$

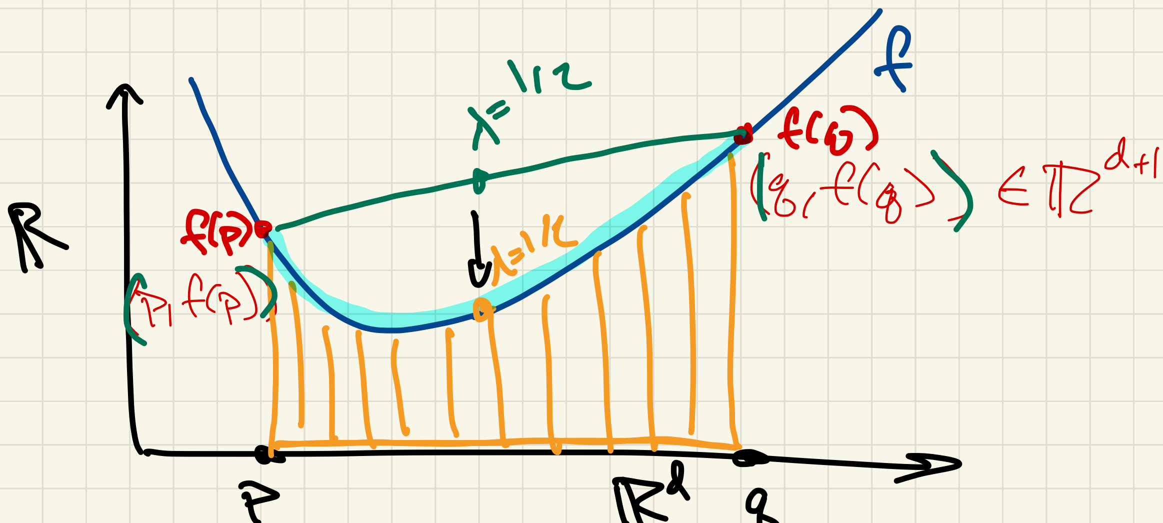
convex combination

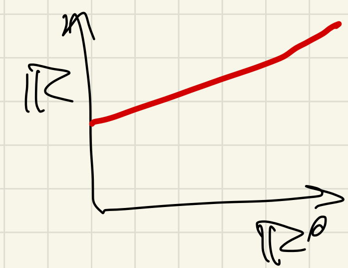
Convex function $f : \mathbb{R}^d \to \mathbb{R}$

s.t. $\forall p, q \in \mathbb{R}^d \quad \forall \lambda \in [0,1]$

$<$ ← strictly convex

$$f(\lambda p + (1-\lambda) q) \leq \lambda f(p) + (1-\lambda) f(q)$$





Convex properties

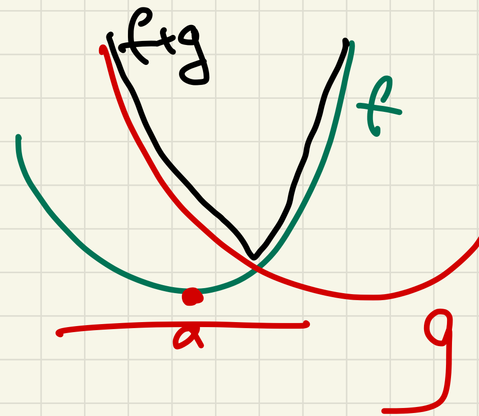
$f, g : \mathbb{R}^d \to \mathbb{R}$ convex

$h = f + g$ convex

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

$h = c \cdot f$ $\quad c \in \mathbb{R}^d \quad c > 0$

convex



---

f (convex), if $\alpha \in \mathbb{R}^d$ is local min

then α is a global min.

if strict

↳ exactly 1 global min

Gradients

\nabla

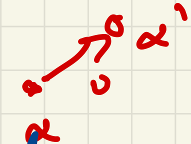
differentiable

$f(x)$ $\quad x=(x_1,x_2,\dots x_d) \in \mathbb{R}^d$

unit vector $\quad u=(u_1,u_2 \dots u_d)$

$\|u\| = 1$

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



directional derivative

$e_1, e_2 \dots e_d \in \mathbb{R}^d \qquad e_i = (0,0, \dots 0, \overset{\text{ith coord}}{1}, 0, \dots 0)$

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

gradient

$$\nabla f = \frac{\partial f}{\partial x_1} e_1 + \frac{\partial f}{\partial x_2} e_2 + \dots + \frac{\partial f}{\partial x_d} e_d = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \dots, \frac{\partial f}{\partial x_d} \right)$$

$x = (x, y, z) \in \mathbb{R}^3 \quad f(x,y,z) = 3x^2 - 2y^3 - 2xe^z$

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

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

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