FoDA L16

Gradient Descent #2

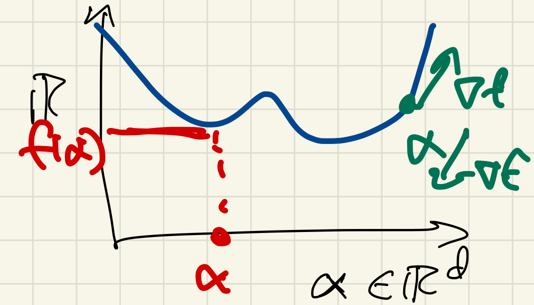
Algorithm & Convergence

Oct 20, 2022

Gradients

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

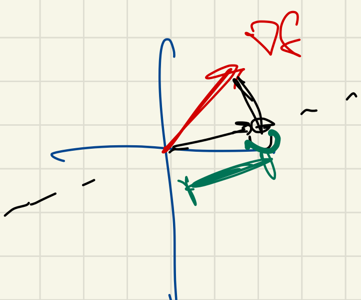
$$\nabla f = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_d} \right)$$



Directional Derivative

$u \in \mathbb{R}^d \quad \|u\| = 1$

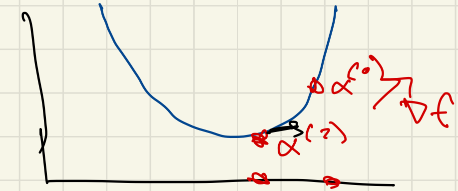
$$\nabla_u f(x) = \langle \nabla f, u \rangle \in \mathbb{R}$$



$u = \frac{\nabla f}{\|\nabla f\|}$

$$\nabla_u f(x) = \left\langle \nabla f, \frac{\nabla f}{\|\nabla f\|} \right\rangle = \|\nabla f\|$$

Gradient Descent



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

Goal $\min_{\alpha \in \mathbb{R}^d} f(\alpha)$ $\qquad \alpha^* = \arg\min_{\alpha \in \mathbb{R}^d} f(\alpha)$

0. Initialize $\alpha^{(0)} = \alpha_{start} \in \mathbb{R}^d$ $\qquad k = 0$

1. repeat

learning rate

$$\alpha^{(k+1)} := \alpha^{(k)} - \gamma_k \nabla f(\alpha^{(k)}) \qquad (k = k+1)$$

until $\left(k = T \quad \text{or} \quad \|\nabla f(\alpha^{(k)})\| \le T \right)$

2. return $\alpha^{(k)}$

Simple GD

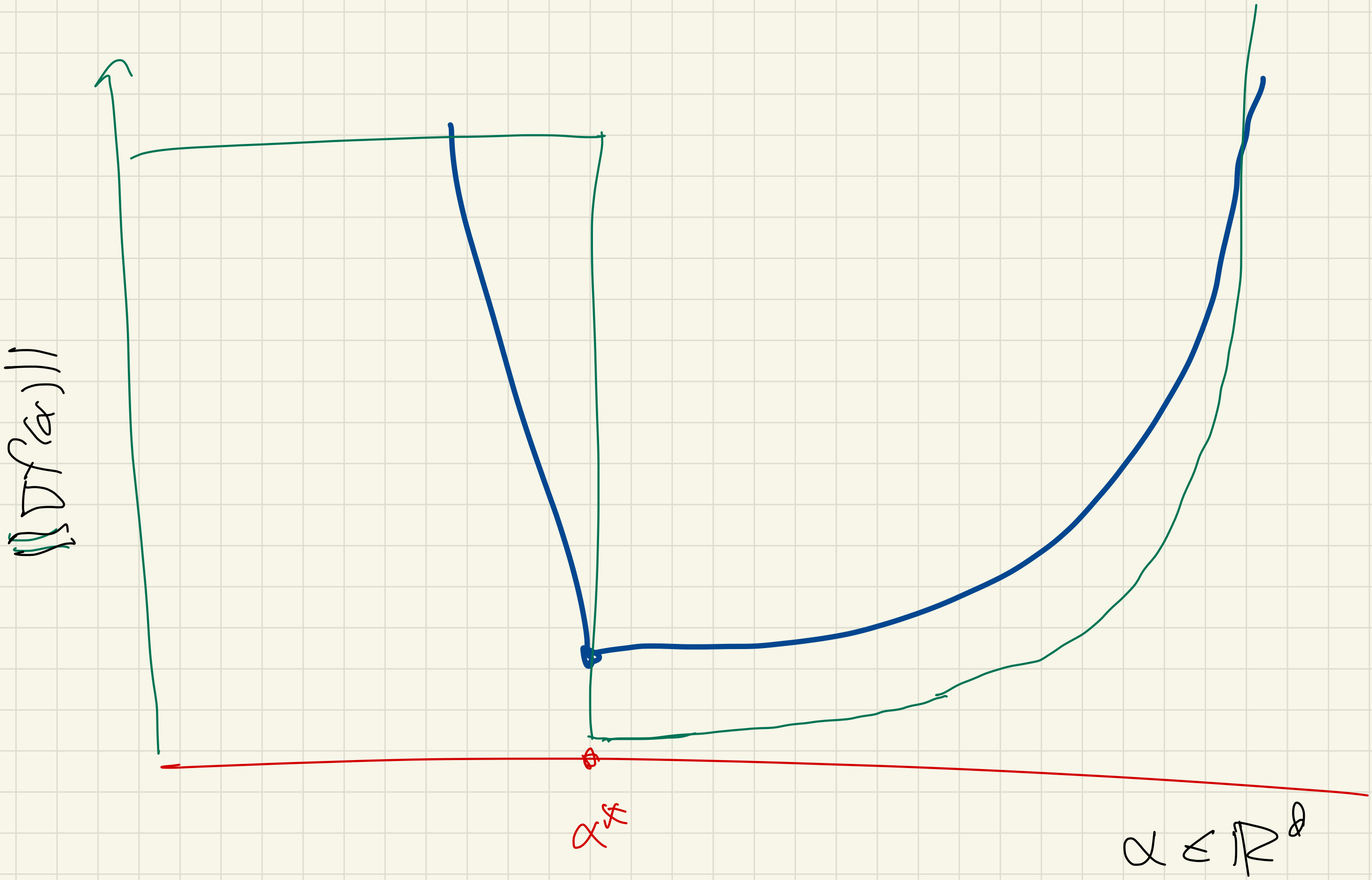
$\alpha = \alpha_{start}$

repeat

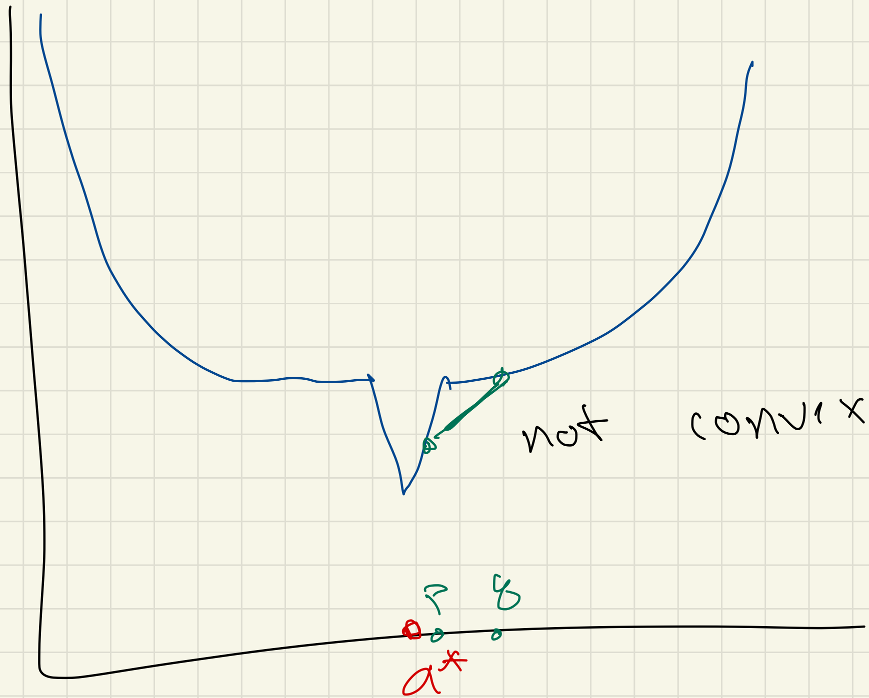
$\alpha = \alpha - \gamma \nabla f(\alpha)$

until $(\|\nabla f(\alpha)\| \leq \tau)$

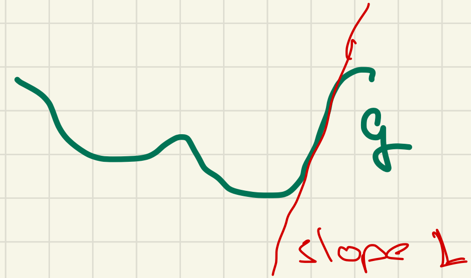
return α



$\|\nabla f(x)\|$
x^*
$x \in \mathbb{R}^d$


not convex
a^*

Learning Rate? γ



- $g : \mathbb{R}^d \to \mathbb{R}^k$ **L-Lipschitz**

  $\forall p,q \in \mathbb{R}^d$ $\quad \|g(p) - g(q)\| \le L\|p-q\|$

if $g = \nabla f$ is L-Lipschitz set $\gamma \le \frac{1}{L}$

↳ GD of f will converge to stationary point.

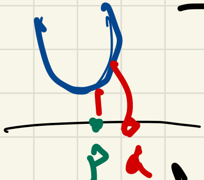
if f convex GD $\to$ global min

$t = \underbrace{C}_{\text{const.}}/\varepsilon$ steps then $\underbrace{f(x^{(t)})}_{\text{after } t \text{ steps}} - \underbrace{f(x^*)}_{\text{opt}} \le \varepsilon$

Strongly Convex Functions

η = \eta

$\eta > 0$



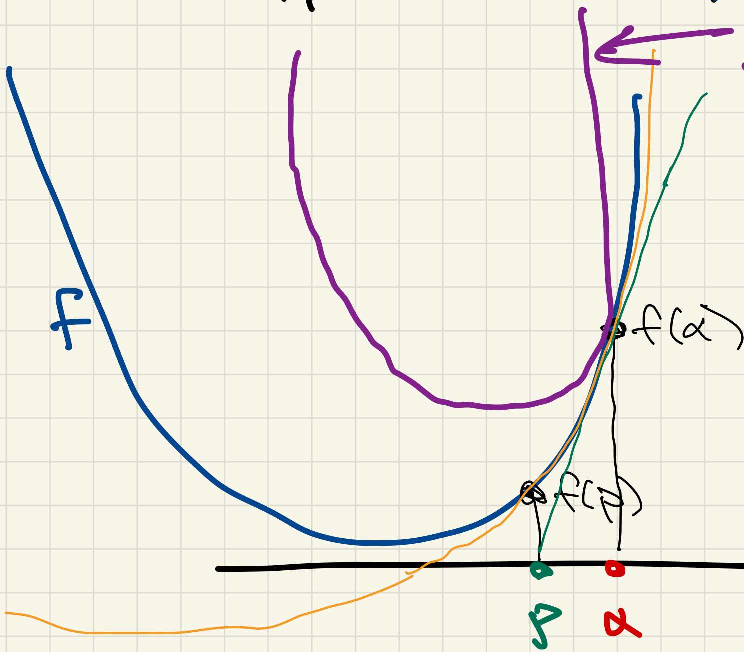
$f : \mathbb{R}^d \to \mathbb{R}$ η - strongly convex

$\forall \alpha, p \in \mathbb{R}^d$ $\quad f(p) \geq f(\alpha) + \langle \nabla f(\alpha), p - \alpha \rangle + \frac{\eta}{2} \|p - \alpha\|^2$

$f(\alpha) + \langle \nabla f(\alpha), p - \alpha \rangle + \frac{L}{2} \|p - \alpha\|^2$

$\geq f(p) \implies$

∇f L-Lipschitz



$f: \mathbb{R}^d \to \mathbb{R}$ $\quad \eta$-strongly convex

$\nabla f \quad L$-Lipschitz

$$\gamma \leq \frac{2}{L+\eta}$$

after $t = \underline{c \cdot \log(1/\varepsilon)}$ steps

$$f(x^{(t)}) - f(x^*) \leq \varepsilon$$

linear convergence