

LECTURE I

1/21/20

170 vs 270



TOPICS

- 1) Continuous Optimization →
- 2) Linear Algebraic Tools
PCA, SVD
- 3) Embeddings: Dimension reduction
Tree embeddings

4) Spectral Graph Theory

- Analysing graphs via
eigenvectors

5) Semidefinite programming

Max Cut, Mean estimation

6) Differential Privacy

7) Other topics!!



- Biweekly homeworks

- Peer grading

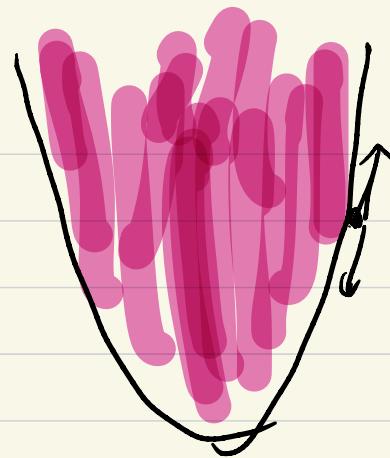
- Scribe notes

- Exam -

- Project \Rightarrow

GRADIENT DESCENT:

Minimize $f(x)$



Convexity:

A set $K \subseteq \mathbb{R}^n$ K is convex

iff $\forall x, y \in K$ line segment joining $\in K$
 x or y

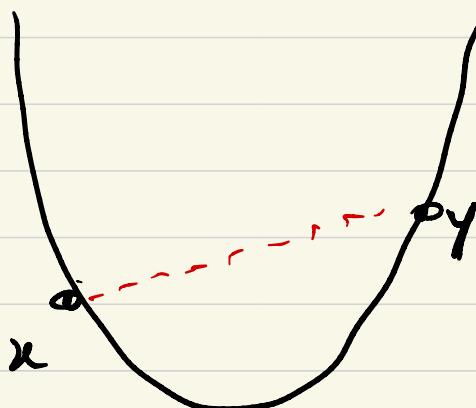
$$\forall \lambda \in [0, 1] \quad \lambda x + (1 - \lambda)y \in K.$$

Convex Function

Convex Function $f: \mathbb{R}^n \rightarrow \mathbb{R}$
(Zeroth Order)

$\forall x, y \quad \forall \lambda \in [0, 1]$

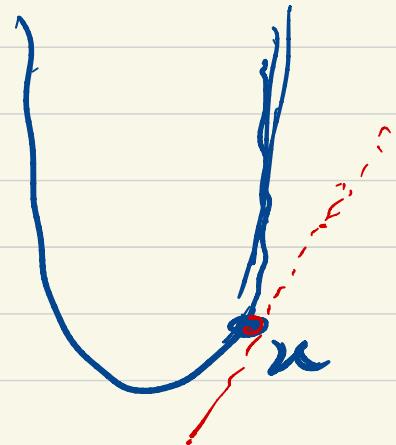
$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y)$$



(First Order)

$\forall x, y$

$$f(y) \geq f(x) + \langle \nabla f(x), y-x \rangle$$



(Second Order)

$$\text{Hessian}(f) = H_f(x) \quad \forall x$$

is positive
semidefinite
(all eigen values are > 0)

Notation

\mathbb{R}^n

$$\langle u, v \rangle = \sum_{i=1}^n u_i v_i$$

$$\|u\|_2 = \sqrt{\sum u_i^2}$$

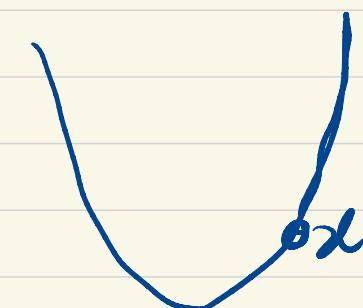
$$\nabla f(u) = \left(\frac{\partial}{\partial x_i} f \right)_{i=1}^n$$

$$H_f(u) = \left(\frac{\partial^2 f}{\partial u_i \partial u_j} \right)$$

$f(u+\delta) \approx$

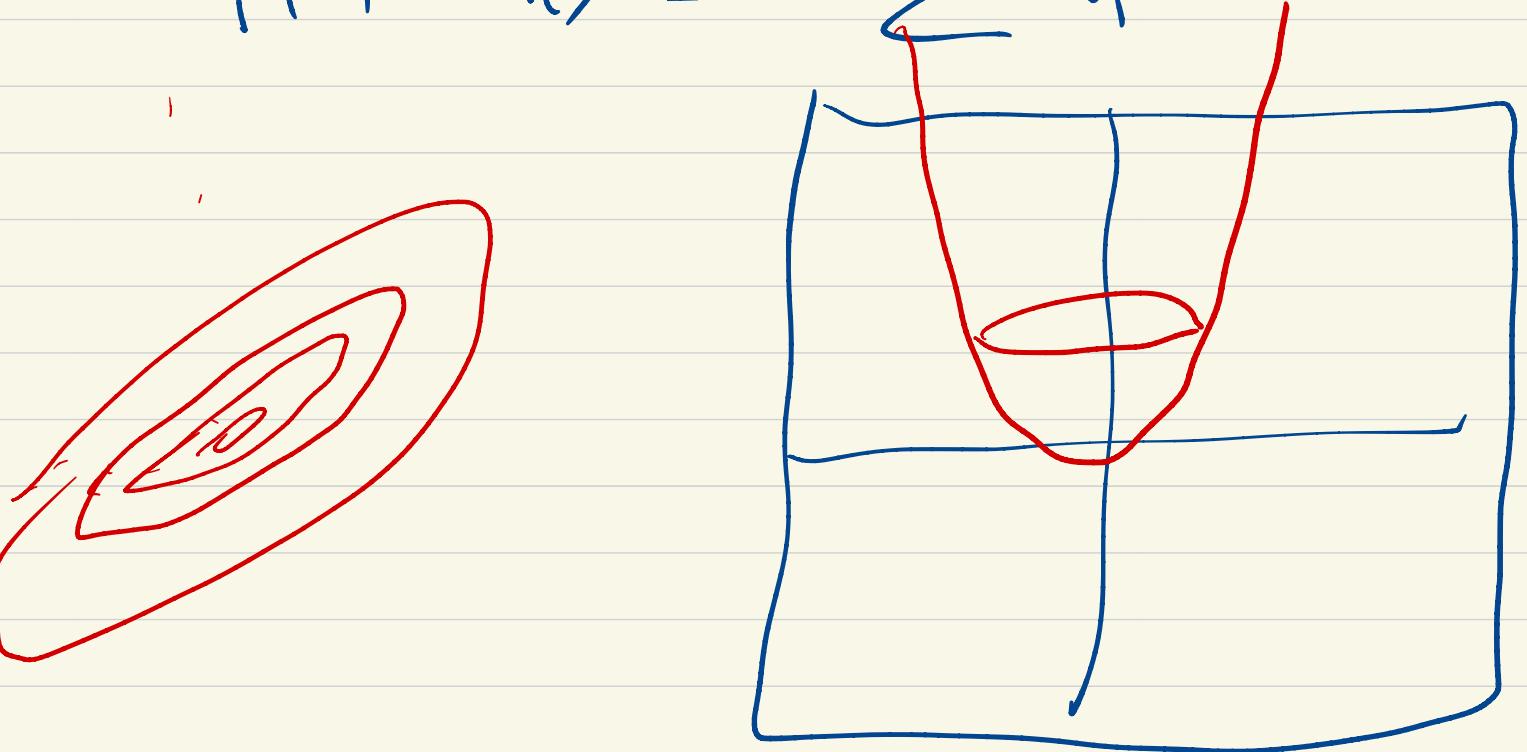
$$a + b\delta + c\delta^2 +$$

$c > 0$



$$P(x_1, \dots, x_n) = \underbrace{[u, x]}_{+ \sum a_{ij} x_i u_j}$$

$$P(x_1, \dots, x_n) = \sum x_i^2$$



$$\sum w_i x_i^2$$

Unconstrained Optimisation

$$\underset{x \in \mathbb{R}^n}{\text{Min}} \quad f(x)$$

f is convex

Algo:

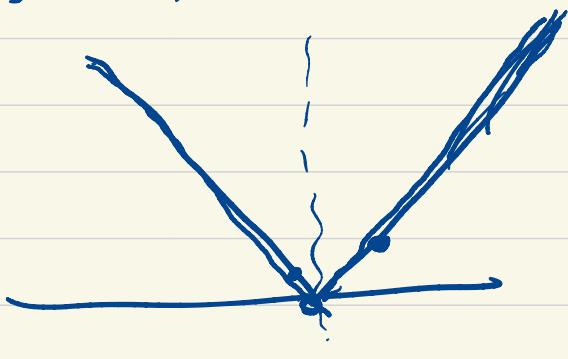
$x_0 \leftarrow$ initial point

$$x_{t+1} \leftarrow x_t - \boxed{\eta} \downarrow \sqrt{f(x_t)}$$

parameter

Output : $\frac{1}{T} \sum_{i=1}^T x_i$

Observation: (Intuitively) if you sufficiently small
 $f(x_{t+1}) < f(x_t)$,





Fact 1: $-\nabla f(u)$ points towards the optimum x^* .

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle$$

If $x^* = \text{optimum} = \arg \min_x f(x)$

$$y = x^*$$

$$f(x^*) \geq f(x) + \langle \nabla f(x), x^* - x \rangle$$

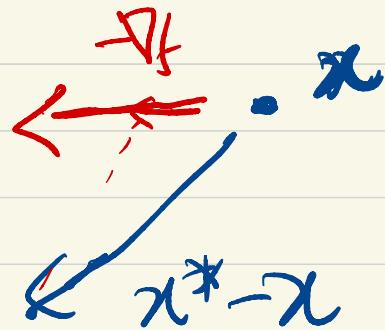
$$\Rightarrow \langle -\nabla f(x), x^* - x \rangle \geq f(x) - f(x^*)$$



↓

≥ 0

$$\boxed{\langle -\nabla f(x), x^* - x \rangle \geq 0}$$



x^*

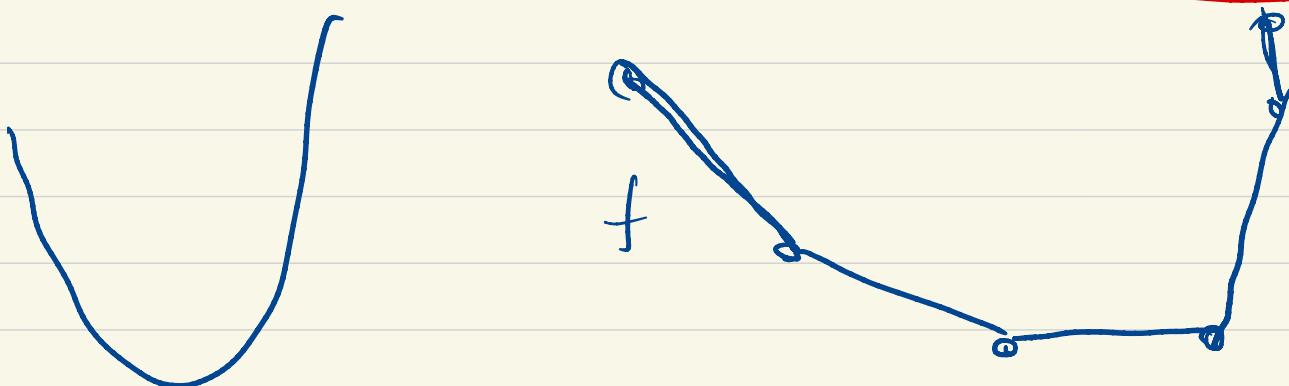
Theorem: After t steps (for appropriate $n = \frac{R}{L \cdot \delta t}$)

$$\frac{1}{t} \sum_{i=1}^t f(x_i) \leq f(x^*) + O\left(\frac{RL}{\sqrt{t}}\right)$$

R = diameter of set / distance $\|x_0 - x^*\|$

$L \in f$ is L -Lipschitz $\|\nabla f\| \leq L$

$$\boxed{\forall x, y \quad |f(x) - f(y)| \leq L \cdot \|x - y\|}$$



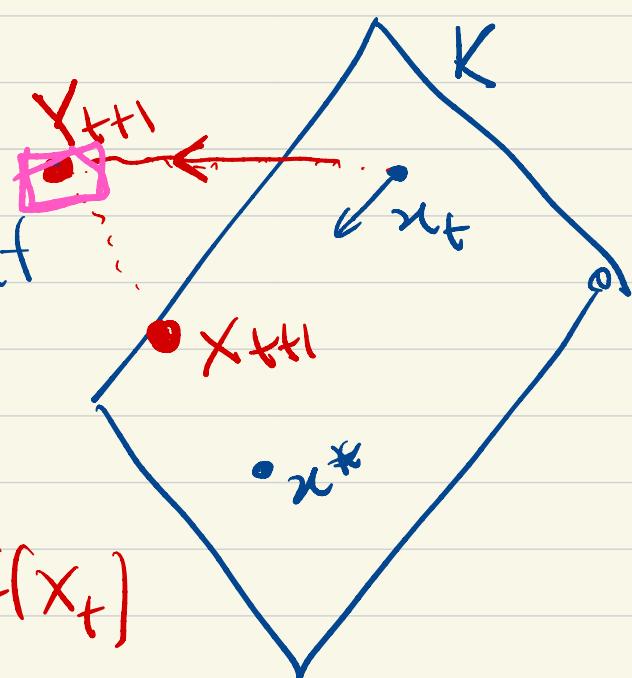
$K = \text{convex set} \in \mathbb{R}^n$

$$\boxed{\begin{array}{ll} \text{Min} & f(x) \\ x \in K & \end{array}}$$

Projected Gradient Descent

$x_0 \leftarrow \text{initial}$

$$y_{t+1} = x_t - \eta \nabla f(x_t)$$



$$x_{t+1} = \Pi_K(y_{t+1})$$

\downarrow $\ell_2 \text{ norm}$

$$\text{Projection}_{\text{on } K} = \arg \min_{z \in K} \|z - y_{t+1}\|$$

(Same Theorem holds)

$$\|y_{t+1} - x_*\| \geq \|\Pi_K(y_{t+1}) - x_*\|$$

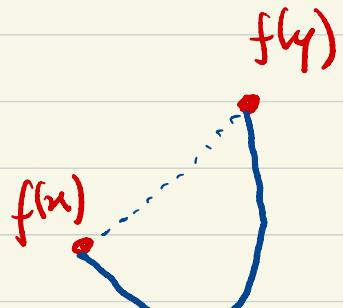


"FLATNESS" \leftarrow "Not too flat"

[Strong Convexity]

α -strongly convex iff $\forall x, y$

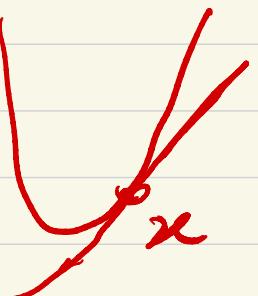
$$= f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y)$$



$$-\frac{\alpha(\lambda(1-\lambda))}{\|x-y\|^2}$$

$$f(y) \geq f(x) + \langle \nabla f(x), y-x \rangle$$

$$+ \underbrace{\alpha \|y-x\|^2}_{\text{red wavy line}}$$



β -Smoothness

$$f(\lambda x + (1-\lambda)y) \geq \lambda f(x) + (1-\lambda)f(y)$$

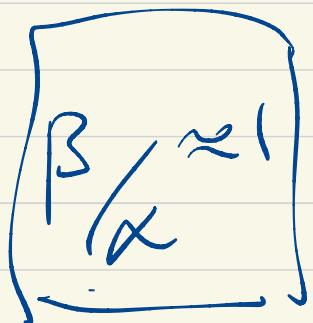
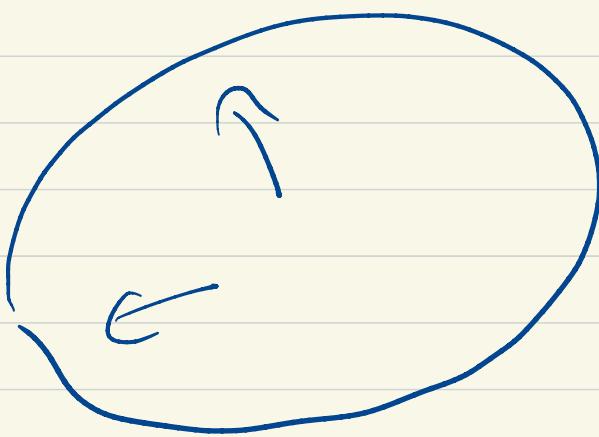
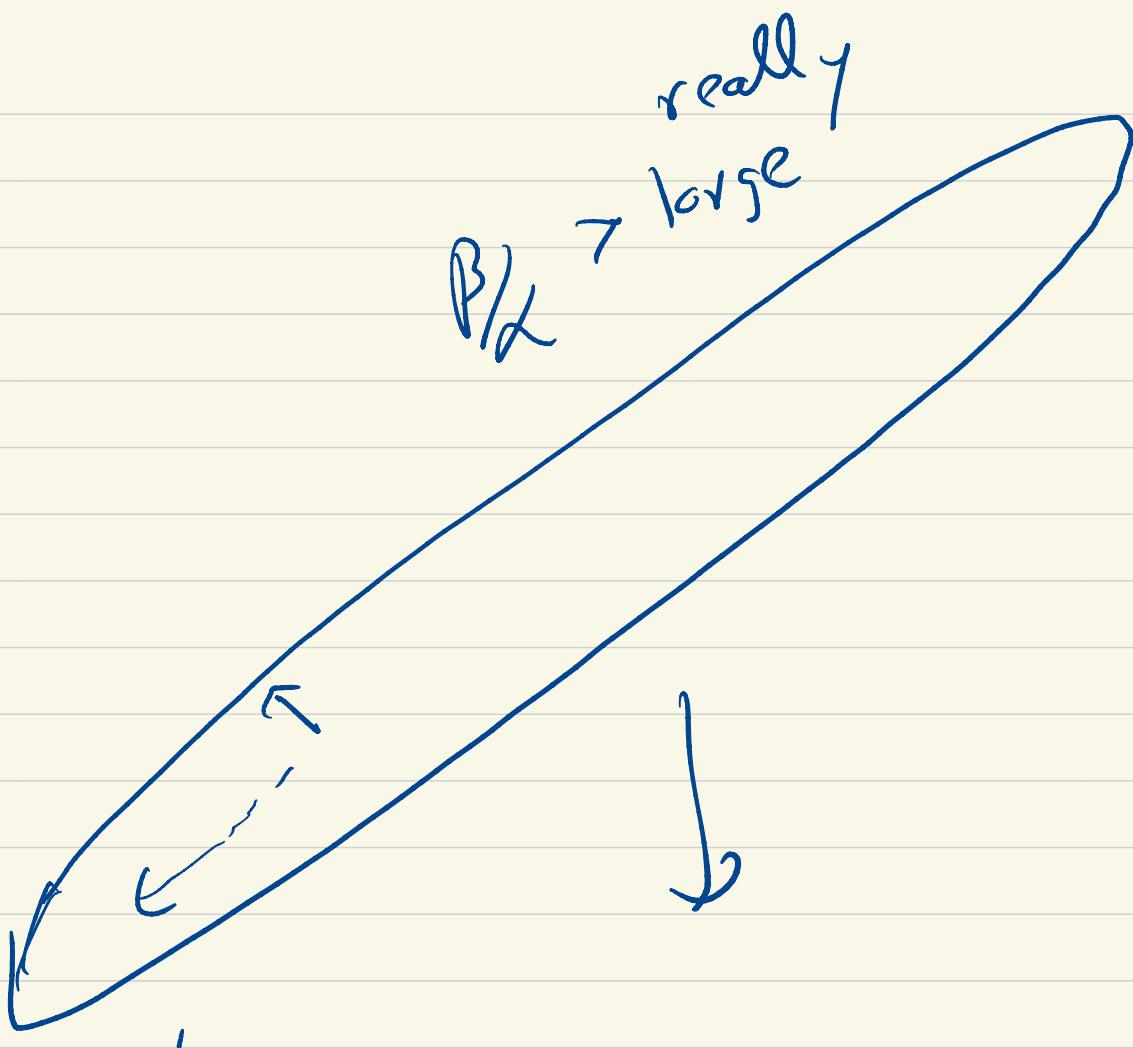
$$-\frac{\beta}{2} (\lambda(1-\lambda)) \|x-y\|^2$$

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

Thm: After t steps [appropriate]

$$\frac{1}{t} \sum f(u_i) - f(x^*) \leq \exp\left(-\frac{t}{\frac{\beta}{\alpha}}\right)$$

Condition number = $\frac{\beta}{\alpha}$



Precondition