①

EE 381K – Convex Optimization

Constantine Caramanis

What this class is & is not about...

Not a class about the modeling power of (convex) optimization.

Is a class about algorithms.

$$\boxed{\begin{array}{l} \min:\ f(x), \\ \text{st}:\ x \in \mathcal{X} \end{array}} \qquad x \in \mathbb{R}^n$$

For what structure of f, $\mathcal{X}$ can we solve this "$\underline{\text{quickly}}$"

When $n \longrightarrow \infty$

$\varepsilon \longrightarrow 0$

$\varepsilon = 10^{-2}$

$\varepsilon = 10^{-16}$

Ex: $\cancel{10^{20}} \cdot n \cdot 1/\varepsilon^2$

$\cancel{2} \cdot n^3 \cdot \log(1/\varepsilon)$

This course:

Linear algebra

Singular Value Decomposition

Eigenvalues

Analysis

Ex:

$f(x) = ax^2 + bx + c$

$$f(x) = \frac{1}{2}x^T Q x + g^T x + c$$

Q $n \times n$

g $n \times 1$

c 1×1

If Q has e-values are ≥ 0

$$\nabla\left(\tfrac{1}{2}x^T Q x\right) + g$$

$$= \boxed{Qx + g = 0}$$

"$Ax = b$"

"Hands on"

HW: weekly or every other week

25%

MT: 25% Oct. 26th or Nov. 14th

F: 35%

<u>Project:</u> 15% "Interesting" comput'l exploration

a. Explore application

b. Explore comp'l issues of "standard methods"

c. Less cleanly understood yet popular alg., ex: ADMM.

d. sth else?

Team: Nov 2

Proposal: Nov 16

Final proj: Dec 5

Simple but fundamental example:

$$\min_{\beta} \|X\beta - y\|_2^2$$

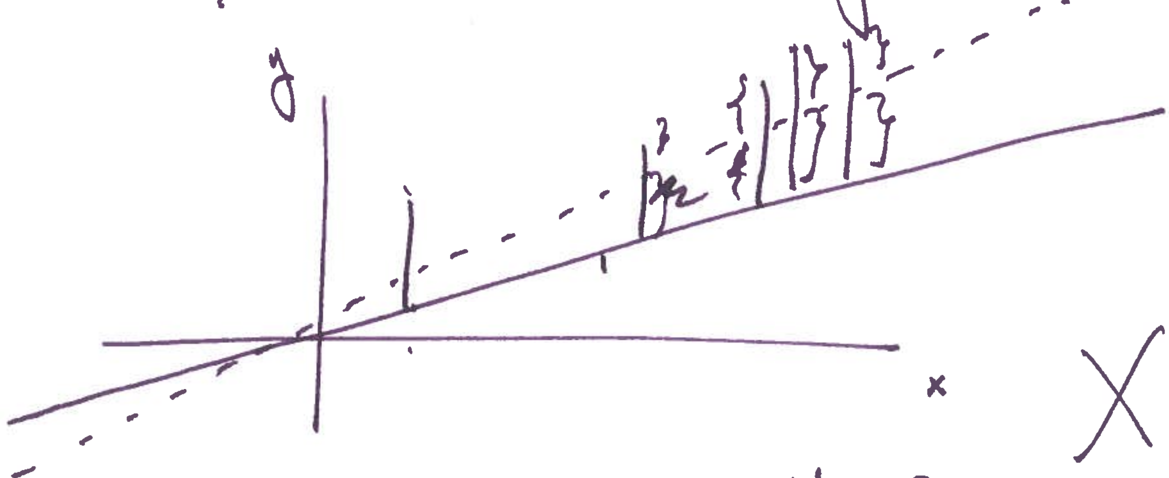
$\|\cdot\|_2$ - Euclidean norm. $\|x\|_2 = \left(\sum x_i^2\right)^{1/2}$

$$X = \begin{pmatrix} - x_1 - \\ \vdots \\ - x_n - \end{pmatrix} \in \mathbb{R}^{n \times p}$$

$$\min_{\beta} \sum \left(\langle x_i, \beta \rangle - y_i \right)^2$$

sum of squares.

Least squares regression: $y_i = \langle x_i, \underline{\beta} \rangle + e_i$



X $n \times p$

$n > p$

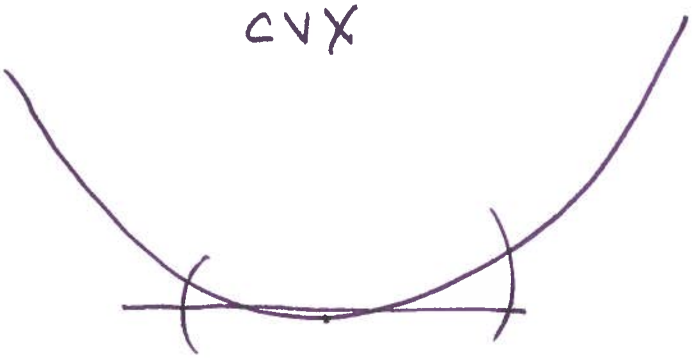
$$\hat{\beta} = (X^T X)^{-1} X^T y$$

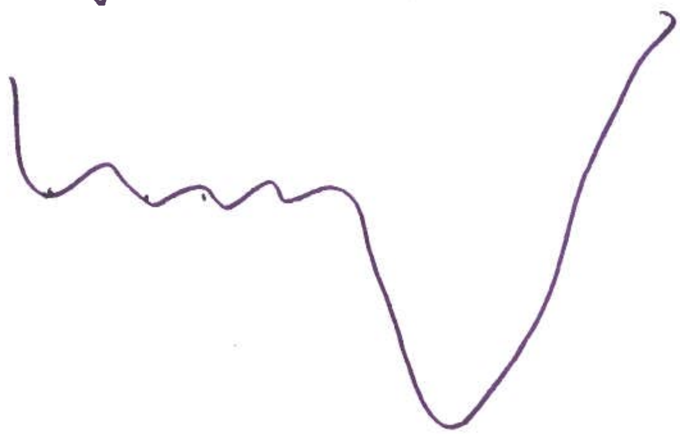
Lin. Alg: why?

Comp side: $O(p^3)$ bottleneck inversion.

Gradient Descent - "rolling down hill"

CVX





$\hat{\beta} = (X^TX)^{-1}X^Ty$ why?

1. X^TX is invertible

then: min: $\|X^T\beta - y\|_2^2$

$$\beta^TX^TX\beta - y^TX\beta - \beta^TX^Ty + y^Ty$$

$$-2\beta^TX^Ty$$

$(x\beta - y)^2$

$x^2\beta^2 - 2x\beta y + y^2$

$\cancel{2}x^2\beta - \cancel{2}xy = 0$

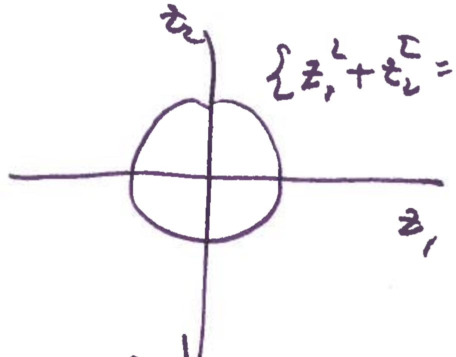
$\beta = \frac{xy}{x^2}$

$$\nabla_X \left(\qquad \right)$$

$$= \cancel{2}X^TX\beta - \cancel{2}X^Ty = 0$$

$$X^TX\beta = X^Ty \Rightarrow \boxed{\beta = (X^TX)^{-1}X^Ty}$$

$$\min_{\beta} \; \|X\beta - y\|_2^2$$



X n rows, p columns $n > p$

y $n \times 1$

β $p \times 1$

$n = 2$

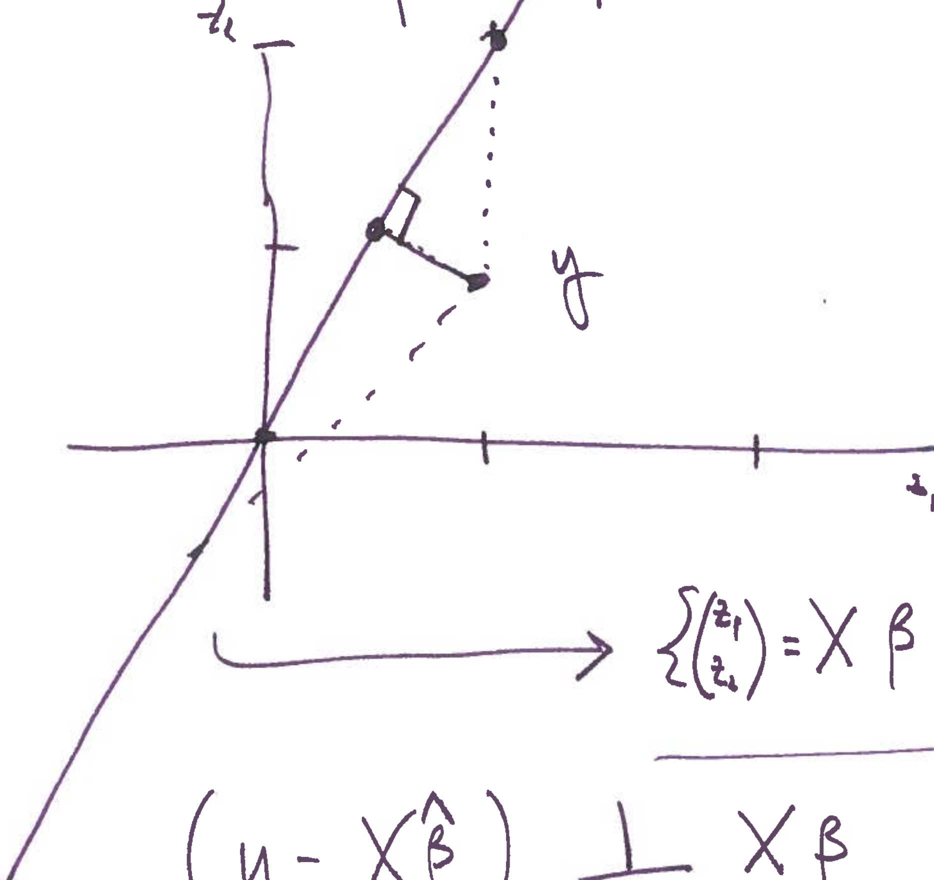
$p = 1$

$X = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$

$y = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

$\beta = 0: \begin{bmatrix} 1 \\ 2 \end{bmatrix} 0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

$\beta = 1: \begin{bmatrix} 1 \\ 2 \end{bmatrix} 1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$



$\left\{ \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = X\beta : \beta \in \mathbb{R} \right\}$

$$(y - X\hat{\beta}) \perp X\beta \quad \forall \beta \qquad (*)$$

<u>Exercise</u>: How does $(*)$ yield

$$\hat{\beta} = (X^T X)^{-1} X^T y$$

Gradient Descent:

min: $\|X\beta - y\|_2^2$

= min: $\sum (\langle x_i, \beta \rangle - y_i)^2$

Ex: $p = 1$, n

$x_1 = 1$ $\quad y_1 = -1$

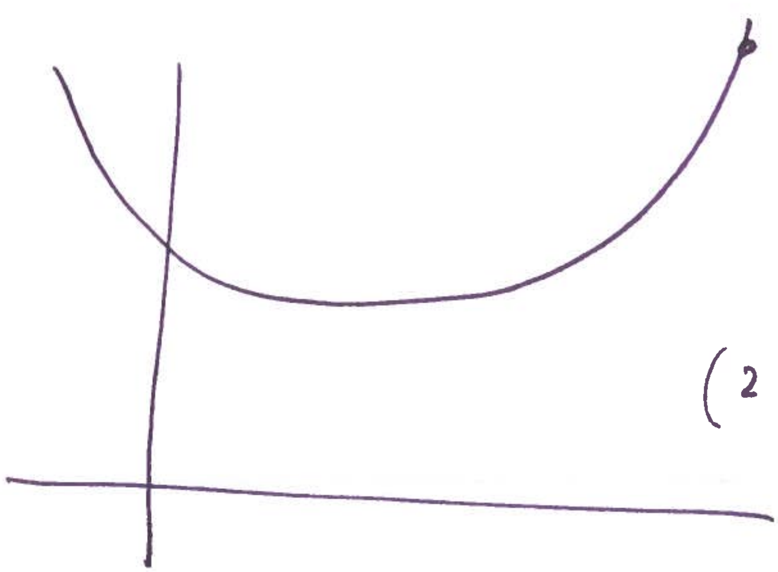
$x_2 = 2$ $\quad y_2 = -2$ $\qquad (\beta = -1)$

$x_3 = 3$ $\quad y_3 = -3$

Plug in: $(\beta + 1)^2 + (2\beta + 2)^2 + (3\beta + 3)^2$

$= 14\beta^2 + 28\beta + 14$

$= a\beta^2 + b\beta + c$. $\quad \cancel{p^3 \quad p=1}$

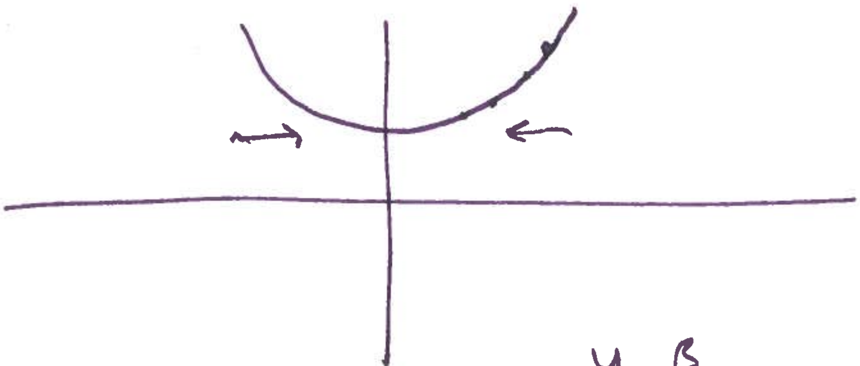


Which direction to move in?

By how much?

$(2a\beta + b)$

min: $2\beta^2 + 1$



4β

"move in dir. of neg. gradient"

$$\beta_{t+1} = \beta_t - \alpha \nabla f(\beta)$$

$$f(\beta) = \|X\beta - y\|_2^2$$

$$\beta^T X^T X \beta - 2\beta^T X^T y + \|y\|_2^2$$

$$\nabla f(\beta) = 2(X^T X)\beta - \underline{X^T y}$$

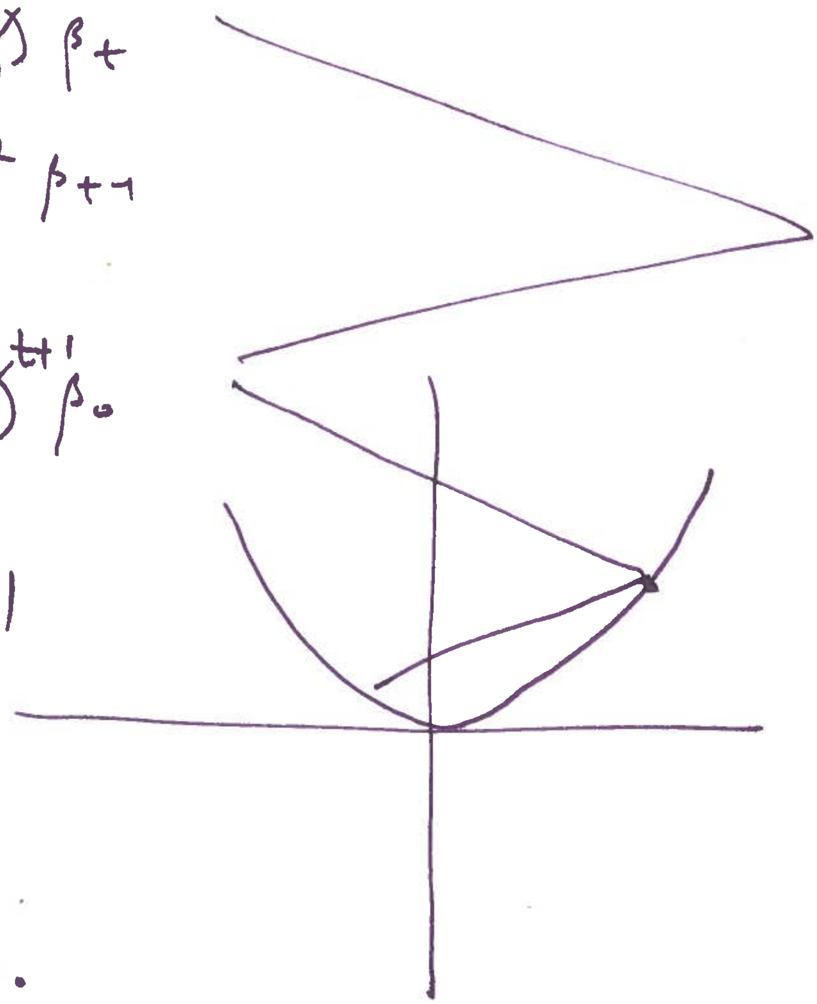
$$f(\beta) = a\beta^2 + \cancel{b\beta} + c \qquad a > 0$$

$$\cancel{\beta} \quad \beta_{t+1} = \beta_t - \cancel{a}\,\gamma \nabla f(\beta)$$

$$2a\gamma\beta + \cancel{b}$$

$$\beta_{t+1} = \beta_t - 2\alpha\beta_t\gamma$$

$$\begin{aligned}\beta_{t+1} &= (1-2\alpha\gamma)\beta_t \\ &= (1-2\alpha\gamma)^2 \beta_{t-1} \\ &\vdots \\ &= (1-2\alpha\gamma)^{t+1}\beta_0\end{aligned}$$



$$|1-2\alpha\gamma| < 1$$

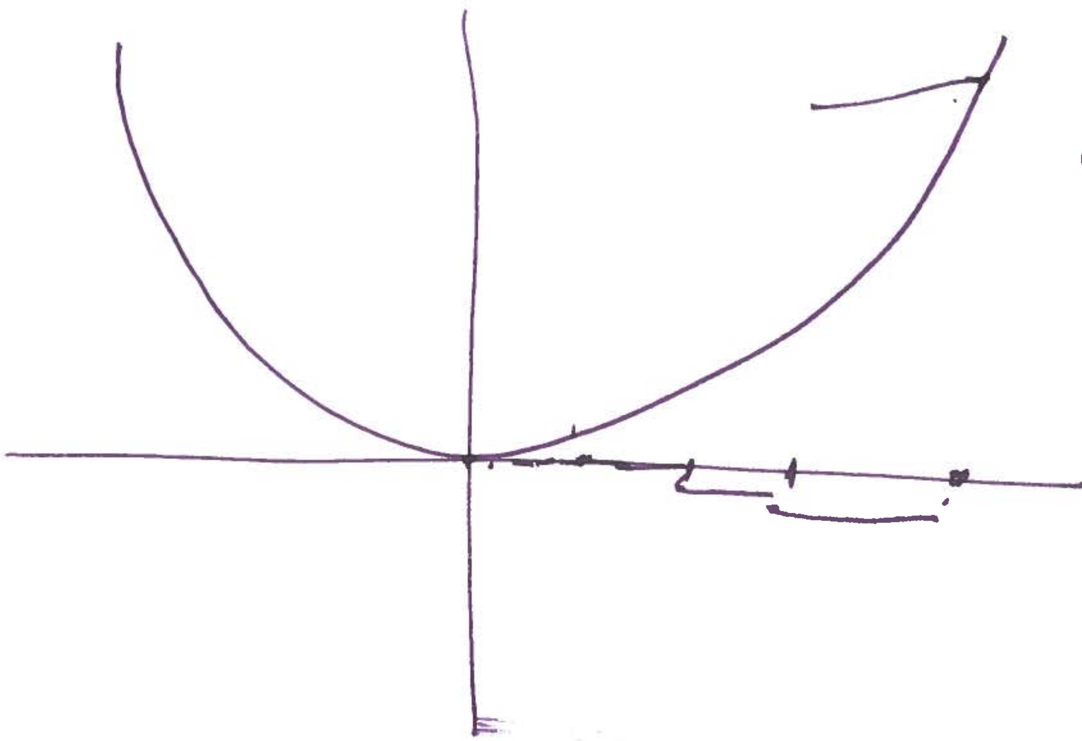
$$\gamma < 1/2\alpha$$

$$\beta_{t+1} = (\kappa)^{t+1} \cdot \beta_0$$

$$\kappa < 1$$

Picture: Curvature: 1-d ex

"$1/q$"



q = curvature

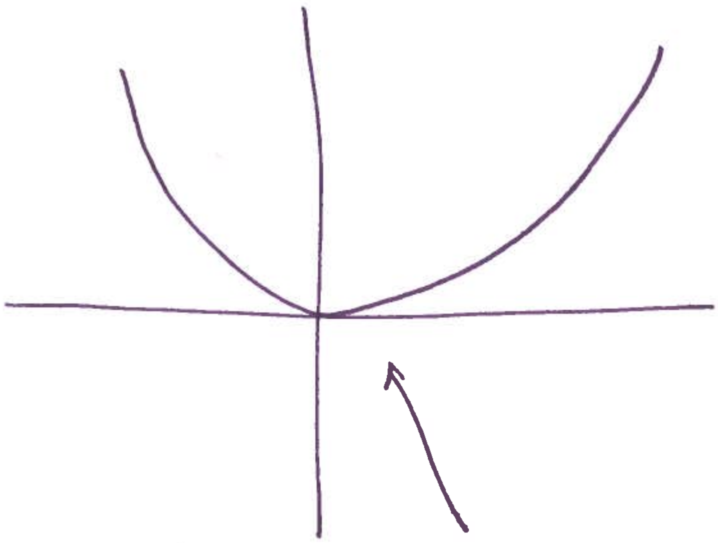
$$f(\beta) = \frac{1}{2}\beta^T Q \beta + q^T \beta$$

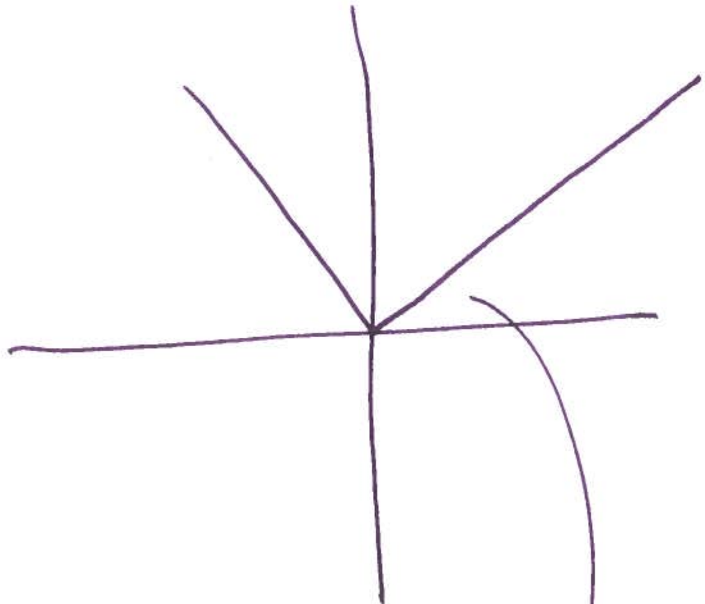
A different example: $\beta_{t+1} = \beta_t - \gamma \beta_t$

min: $|\beta|$ $\beta_{t+1} = \beta_t - \gamma\, \text{sign}(\beta)$

For this problem, step size must be very small for near-convergence

Two themes:





Needed an upper bound on curvature to make sure we don't diverge.

~~Ma~~ The gradient $\not\to 0$ near opt sol'n

$\Rightarrow$ very small step size $\sim \varepsilon$

$\Rightarrow$ slower convergence.

Gradient Descent: Upper & Lower bounds on curvature.