Linear algebra for AI and ML

(November- 12)

LS problem:

$$Ax = b$$

given   given

find x.

$$\min_x \|Ax - b\|_2$$

$A \in \mathbb{R}^{m \times n}$

and i) $m \geq n$

ii) columns of A are linearly independent.

$$\boxed{Ax = b}$$

$m << n$ (m is much less than n)

$A = \begin{bmatrix} & & \end{bmatrix} \in \mathbb{R}^{m \times n}$

will have non-trivial nullspace / Kernel.

$\mathcal{N} = \{ x_N \in \mathbb{R}^n \mid Ax_N = 0 \}$

$x + \mathcal{N}$

Suppose $x \in \mathbb{R}^n$ is such that $Ax = b$,

then for every $x_N \in \mathcal{N}$, $x + x_N$ also is a solution

$A(x + x_N) = b$

Least norm solution $\|\hat{x}\|_2$ is the least.

A sparse solution is a solution with "many zeros".

Compressive sensing:

Basis pursuit:

$$\min \|x\|_1 = |x_1| + \dots + |x_n|$$
$$\text{s.t.} \quad Ax = b$$

$\min (F_1(x) + F_2(x))$

LASSO (in statistics)

$$\boxed{\min \frac{1}{2}\|Ax-b\|_2^2 + \lambda\|x\|_1}$$

$$\boxed{\begin{array}{l} \min \frac{1}{2}\|Ax-b\|_2^2 \\ \text{s.t.} \ \|x\|_1 \le t \end{array}}$$

ℓ^0

$\|x\|_0$ = number of components which are nonzero.

$$x = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} \qquad \|x\|_0 = 2$$

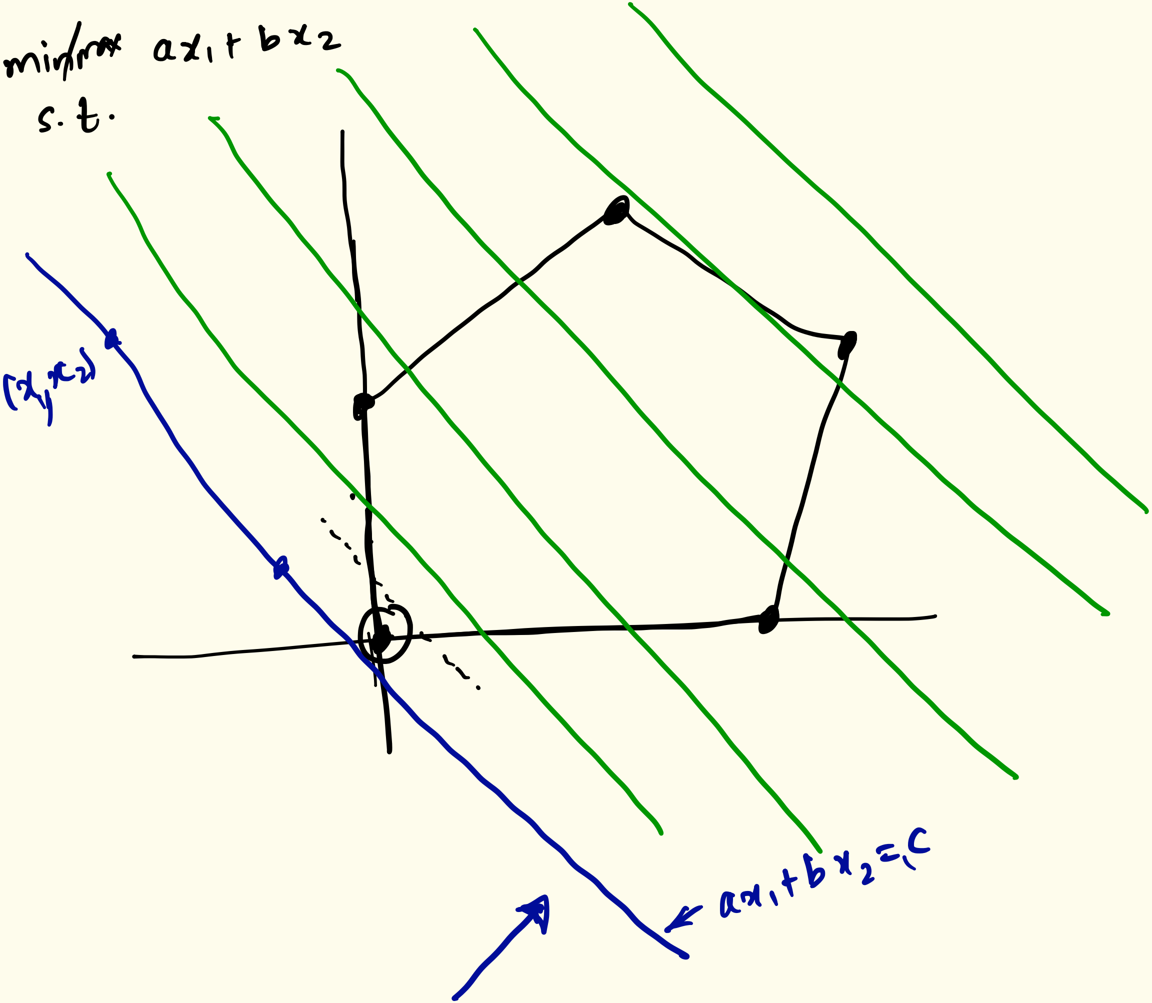
$$2x = \begin{bmatrix} 2 \\ 0 \\ 4 \end{bmatrix} \qquad \|2x\|_0 = 2$$

$\downarrow$ after convexification

ℓ^1 - norm

min/max $ax_1 + bx_2$

s.t.



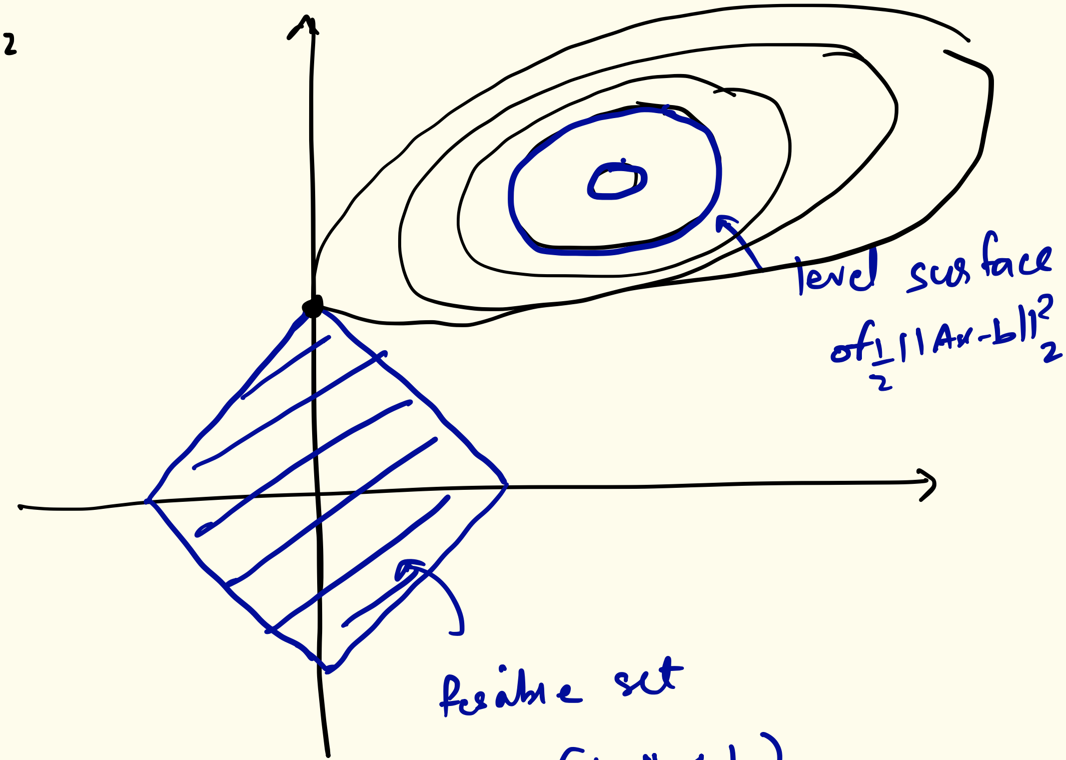
$ax_1 + bx_2 = c$

↓

different values of c the line will shift.

(level surfaces of the cost fn!)

$\frac{1}{2}\|Ax-b\|_2^2$



(Tibshirani : sparse learning)

$$A = U \Sigma V^T$$

$$A = U \begin{bmatrix} \sigma_1 & & & & & \\ & \sigma_2 & & & 0 & \\ & & \ddots & & & \\ & 0 & & \sigma_r & & \\ & & & & 0 & \\ & & & & & 0 \end{bmatrix} V^T$$

Let $A \in \mathbb{R}^{n \times n}$; A is invertible.

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n > 0$$

$$A = U \begin{bmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_n \end{bmatrix} V^T$$

$$0 < \frac{1}{\sigma_1} \leq \frac{1}{\sigma_2} \leq \cdots \leq \frac{1}{\sigma_n}$$

$$A^{-1} = \left(U \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{bmatrix} V^T \right)^{-1} = \underbrace{V \begin{bmatrix} 1/\sigma_1 & & & \\ & 1/\sigma_2 & & 0 \\ & & \ddots & \\ 0 & & & 1/\sigma_n \end{bmatrix} U^T}$$

$$A = \begin{bmatrix} & & \\ & & \\ & & \end{bmatrix}_{m\times n}$$

$m > n$

$\text{rank}(A) = n$

(columns of A are linearly indep.)

$$\underset{m\times n}{A} = U\Sigma V^T = U \underbrace{\begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \\ \hline & 0 & \end{bmatrix}_{m\times n}}_{\Sigma} V^T$$

$$A^T = V\Sigma^T U^T = V \left[\begin{array}{cccc|c} \sigma_1 & & & & \\ & \sigma_2 & & & 0 \\ & & \ddots & & \\ & & & \sigma_n & \end{array}\right]_{n\times m} U^T$$

$$(A^TA)^{-1}A^T = (V\Sigma^T U^T U\Sigma V^T)^{-1} V\Sigma^T U^T$$

$$= V \begin{bmatrix} 1/\sigma_1^2 & & \\ & \ddots & \\ & & 1/\sigma_n^2 \end{bmatrix} V^T V \Sigma^T U^T$$

$(A^TA)^{-1}A^T \rightarrow$ pesudo-inverse

$$= V \begin{bmatrix} 1/\sigma_1 & & & \vdots & \\ & \ddots & & \vdots & 0 \\ & & 1/\sigma_n & \vdots & \end{bmatrix} U^T$$