

# ECE 532 - lecture 15 - trade-offs and regularization

## trading off multiple objectives

\* sometimes we want to minimize two different things  
but they are competing

→ regression:  $\|y - Ax\|^2$  vs  $\|x\|^2$

→ sparsity:  $\|y - Ax\|^2$  vs  $\text{nnz}(x)$  {# non zeroes}

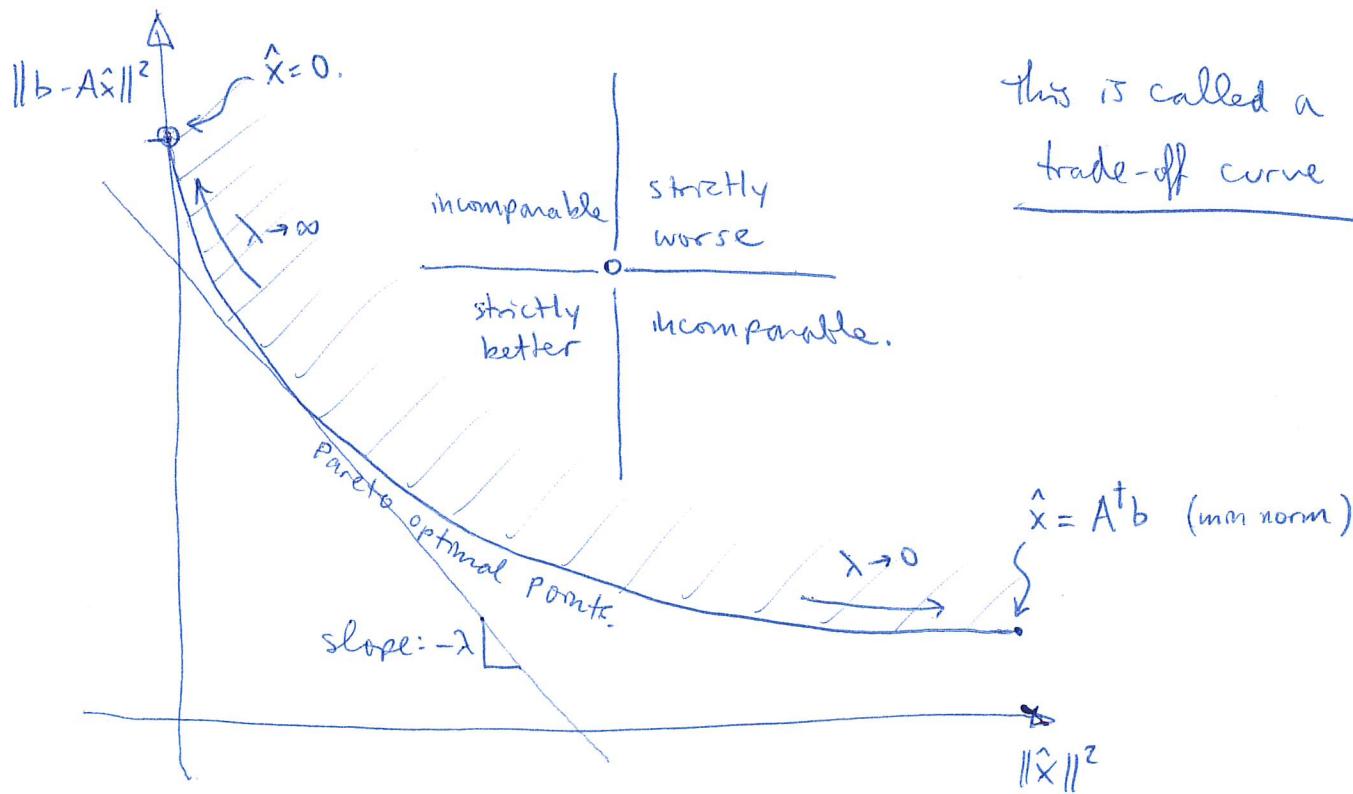
→ finance: expected return vs risk.

→ cars: top speed vs. fuel efficiency.

→ etc. ...

Visualization: ex: minimize  $\|b - Ax\|^2 + \lambda \|x\|^2 \Rightarrow \hat{x} = (A^T A + \lambda I)^{-1} A^T b$ .

\* for each  $\lambda > 0$ , find  $\hat{x}$ , compute  $\|b - A\hat{x}\|^2$  and  $\|\hat{x}\|^2$ ; plot them!

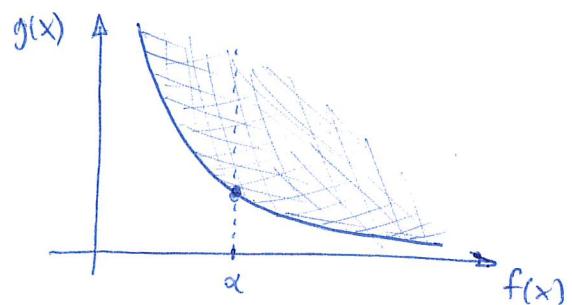


(2)

If we want to minimize  $f(x)$  and  $g(x)$ , three equivalent ways:

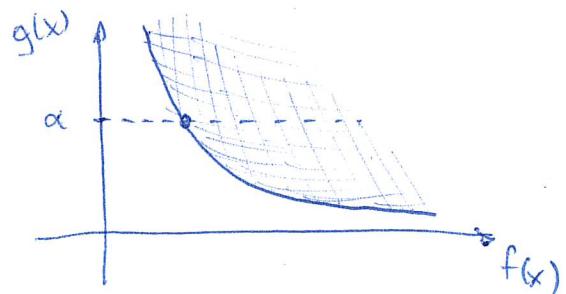
(1) fix  $f$ : minimize <sub>$x$</sub>   $g(x)$

subject to:  $f(x) = \alpha$

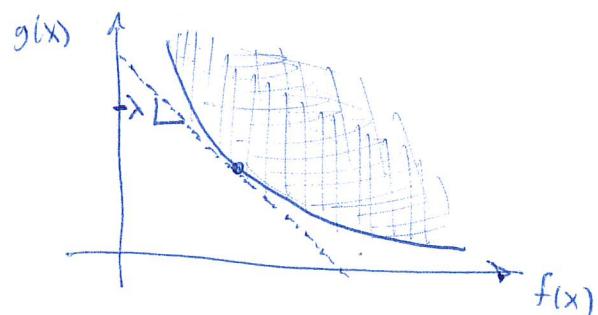


(2) fix  $g$ : minimize <sub>$x$</sub>   $f(x)$

subject to:  $g(x) = \alpha$



(3) linear combination:  
minimize <sub>$x$</sub>   $g(x) + \lambda f(x)$   
subject to: (no constraints)



### Notes

\* Formulation (3) is usually easiest to solve.

\* This only works if  $f$  and  $g$  are convex functions. e.g. norms, we'll talk more about this later.

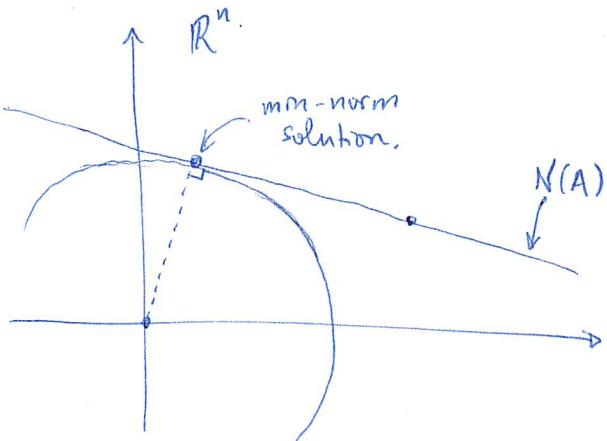
\* Extends to multiple functions. ex:

$$\text{minimize } f(x) + \lambda_1 g(x) + \lambda_2 h(x).$$

Then we have a Pareto surface with a tangent plane!

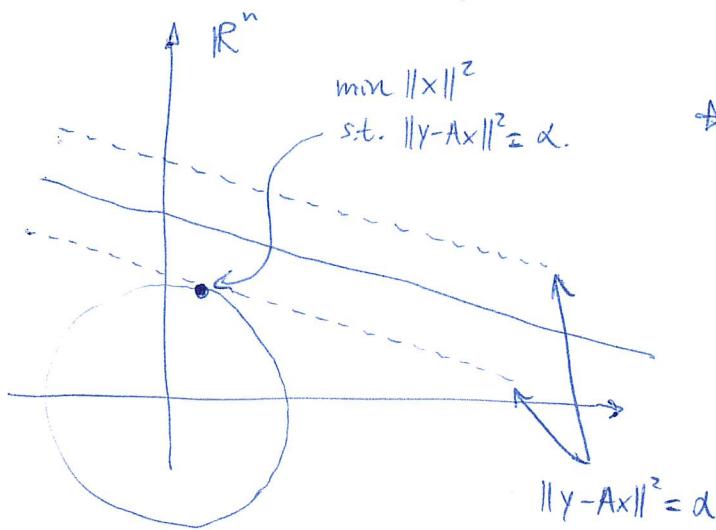
(3)

Underdetermined case : minimize  $\|y - Ax\|^2 + \lambda \|x\|^2$   $x \in \mathbb{R}^n$   $A \in \mathbb{R}^{m \times n}$  ( $m < n$ )



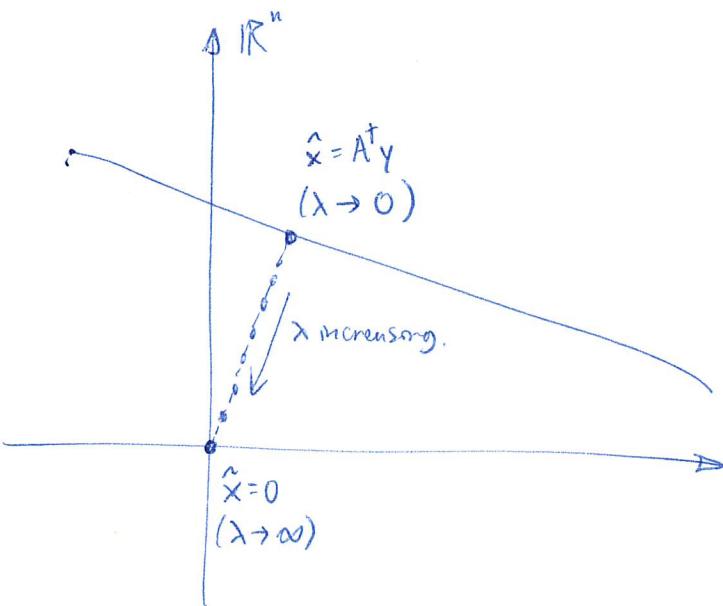
in limit  $\lambda \rightarrow 0$ , we have the minimum-norm solution to the problem.

\* Think of  $N(A)$  as the set of  $x$  such that  $\|y - Ax\|^2 = 0$  (shifted by  $A^T y$ , of course).



\* If instead we look at the set  $\|y - Ax\|^2 = d$ , we get parallel lines.

as  $d$  gets larger,  
solution gets pushed toward  
the origin.

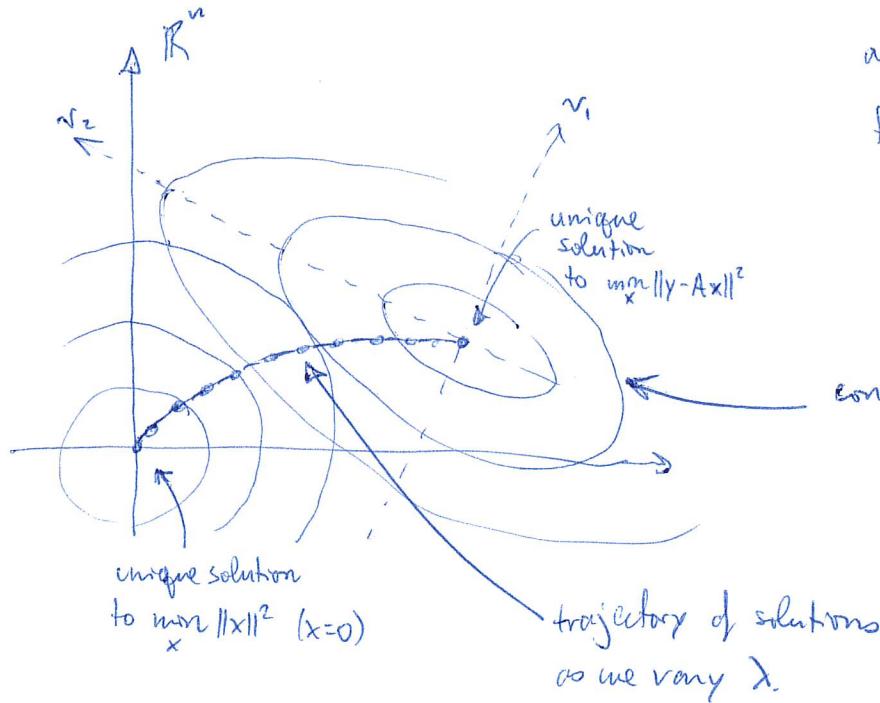


as  $\lambda$  goes from  $0 \rightarrow \infty$ ,  
the  $\hat{x}$  traces out a path  
(straight line in this case)  
from  $A^T y$  to 0.

(4)

## case with a unique solution

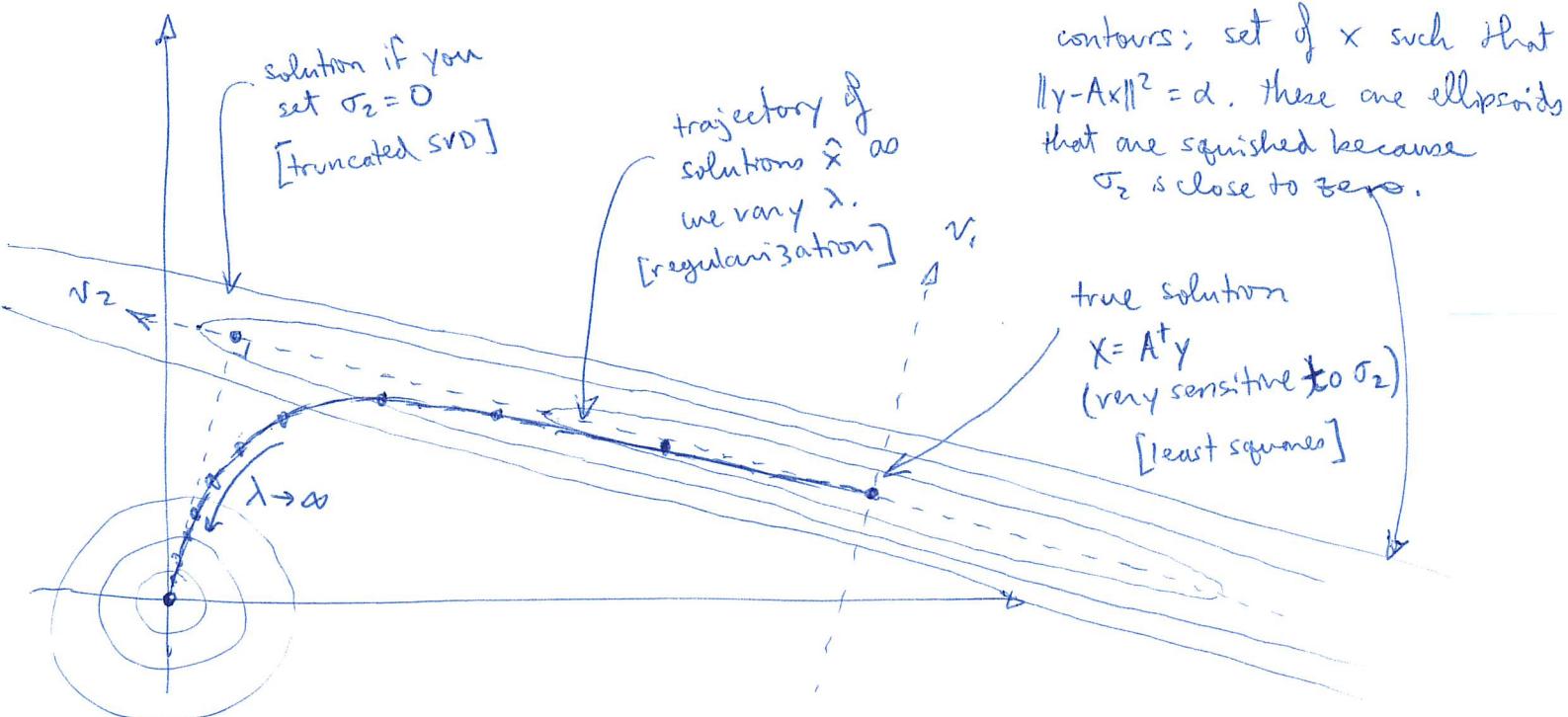
$$\text{minimize}_{x \in \mathbb{R}^n} \|y - Ax\|^2 + \lambda \|x\|^2 \quad (\text{rank}(A) = n).$$



as we change  $\lambda$ , we favor one objective or the other.

## regularization helps for badly conditioned cases

if  $\sigma_2 \approx 0$  (small positive number), picture looks like:



very common in machine learning:

$$\underset{x}{\text{minimize}} \quad l(x) + \lambda r(x),$$

"loss function"      regularizer

so far, we've seen:

$$l(x) = \|Ax - b\|_2^2 \quad \text{i.e. our data is a linear combination of features: } b_i \approx a_{i1}x_1 + \dots + a_{in}x_n \quad i=1, \dots, m.$$

Later in the semester, we will see other forms of loss functions.

e.g. nonlinear (Kernel methods) or different norms. This allows you to fit models that are not linear!

$$r(x) = \|x\|_2^2 \quad \text{also known as L}_2 \text{ regularization, ridge regression, and Tikhonov regularization.}$$

Now, we'll look at different types of regularizers.

Changing the regularizer affects the kind of bias we'd like to impose on our solution.

★ helpful if we know something else about  $x$  and there are many solutions to  $\underset{x}{\text{minimize}} l(x)$ .

★ also helpful for making solution less sensitive to noise and smoothing if necessary.

## The $L_1$ norm

$$\underset{x}{\text{minimize}} \quad \|Ax - b\|^2 + \lambda \|x\|_1$$

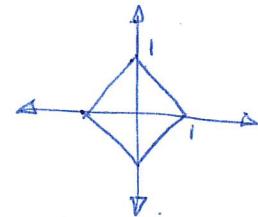
remember the definition:  $\|x\|_1 = \sum_{i=1}^n |x_i|$

this is also known as the LASSO. "least absolute shrinkage and selection operator".

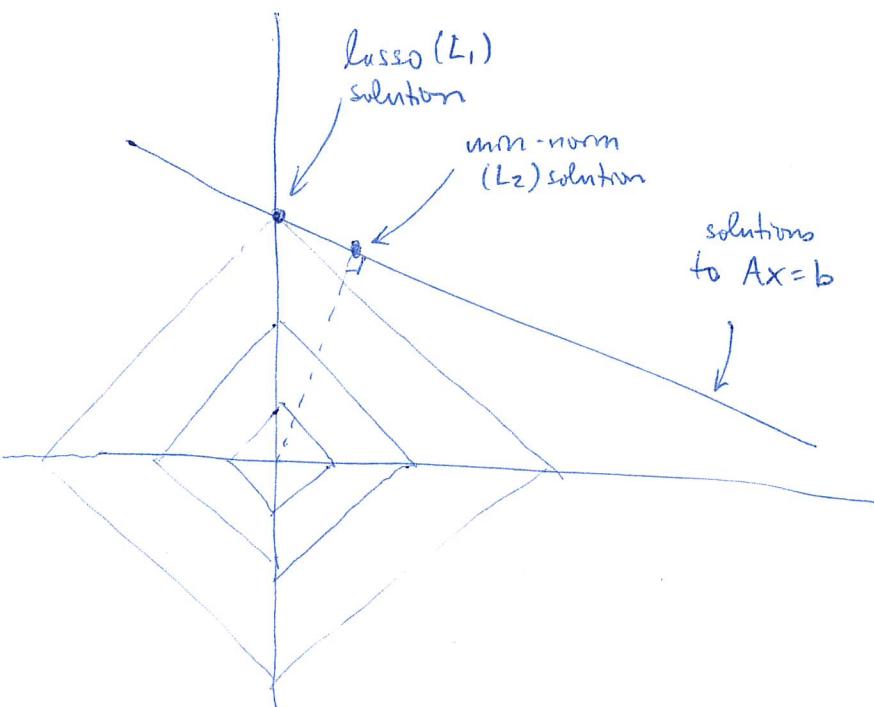
\* Using the lasso regularization encourages sparsity.

i.e. tends to find solutions where many of the components are zero.

geometric intuition: remember the  $L_1$  ball:



if we solve  $\underset{x}{\text{minimize}} \quad \|Ax - b\|^2 + \lambda \|x\|_1$  with  $A \in \mathbb{R}^{m \times n}$ ,  $m < n$   
(underdetermined)



Actually, for any  $\lambda > 0$ ,  
this finds a solution  
where either  $x_1 = 0$  or  $x_2 = 0$   
(a sparse solution!).

Note: in higher dimensions

(7)

**Note**: recall the  $L_2$  regularized least-squares problem  
always has a unique solution for any  $\lambda > 0$ .

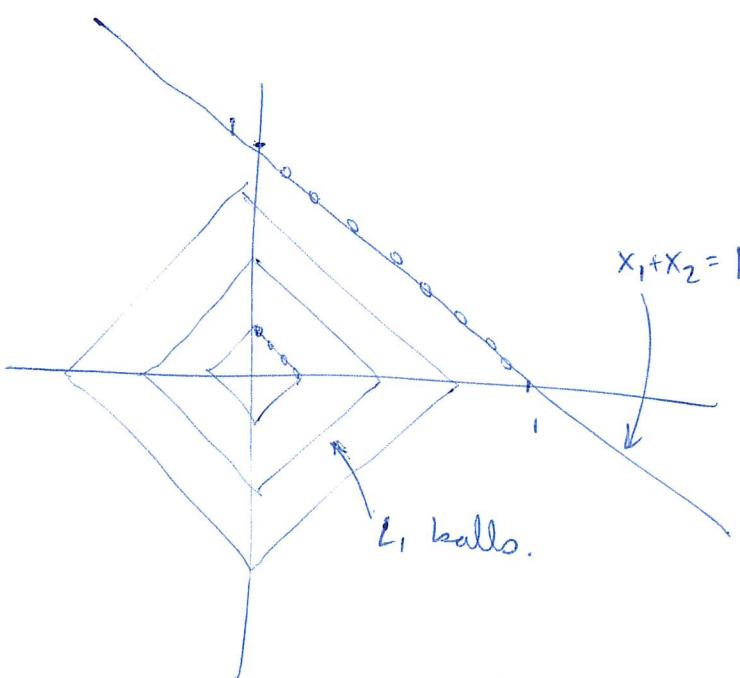
e.g. in general:  $\underset{x}{\operatorname{argmin}} \|Ax - b\|^2 + \lambda\|x\|^2 = (A^T A + \lambda I)^{-1} A^T b$ .

This is not the case for  $L_1$  regularization.

example: minimize  $\underset{x \in \mathbb{R}^2}{\| [1 \ 1]x - 1 \|^2 + \lambda \|x\|_1}$

$$\Rightarrow \underset{x_1, x_2}{\operatorname{min}} (x_1 + x_2 - 1)^2 + \lambda |x_1| + \lambda |x_2|.$$

draw a picture!



any point along the

$$\begin{cases} x_1 + x_2 = \alpha \\ x_1, x_2 \geq 0 \end{cases}$$

face of the ball  
 will be optimal!

Also important : there is no formula for the Lasso solution like in the  $L_2$  case.

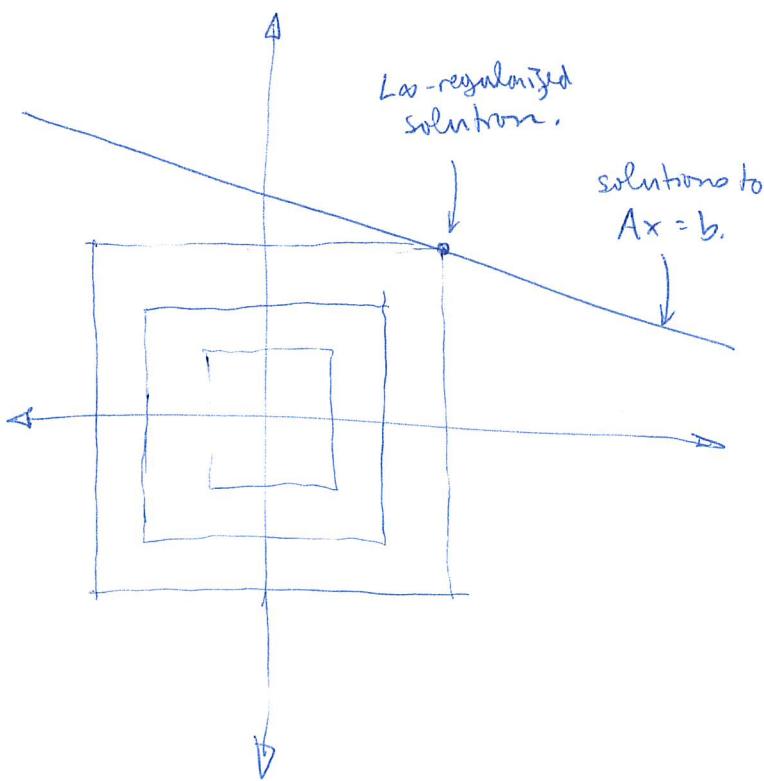
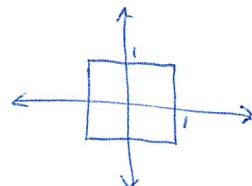
however, there are still efficient ways to compute it using iterative methods. More on this later.

### The $L_\infty$ -norm

$$\text{minimize } \|Ax - b\|^2 + \lambda \|x\|_\infty$$

remember the definition:  $\|x\|_\infty = \max_i |x_i|$

geometric intuition: the  $L_\infty$  ball:



$$\min_x \|Ax - b\|^2 + \lambda \|x\|_\infty$$

with  $m < n$  (underdetermined).

Finds a solution where equal components ( $x_i = x_j$ ) is encouraged.