

# ECE 532 - lecture 14 - geometry & sensitivity

①

Last time, we introduced the pseudoinverse.

If  $A = U, \Sigma, V^T$  is the thin svd, then  
the pseudoinverse is  $A^+ = V \Sigma^{-1} U^T$ .

if  $\left\{ \begin{array}{l} A \text{ is square, invertible, } A^+ = A^{-1} \quad (\text{soln to } Ax = b) \\ A \text{ is tall, full-rank, } A^+ = (A^T A)^{-1} A^T \quad (\text{LS solution}) \\ A \text{ is wide, full (row) rank, } A^+ = A^T (A A^T)^{-1} \quad (\text{min-norm solution}) \end{array} \right.$

some properties hold for the more general  $A^+ = V \Sigma^{-1} U^T$  in cases where  $A$  is not full rank. In this case,  $A^T b$  is the solution to  $\min_x \|Ax - b\|$  with minimum norm  $\|x\|$ .

## \* Some useful properties:

$A^+$  is a "left inverse" of  $A$ .

$$* A^{++} = A$$

\*  $A^+$  is same size as  $A^T$

\* if  $A$  tall, full-rank, then  $(A^T A)^{-1}$ ,  $A A^T$  is projection onto  $R(A)$

\* if  $A$  wide, full-rank, then  $(A^T A)^{-1}$ ,  $A^T A$  is projection onto  $R(A^T)$

(2)

in general★ If  $X$  is a matrix satisfying

- i)  $XAX = X$
  - ii)  $AXA = A$
  - iii)  $AX$  is symmetric
  - iv)  $XA$  is symmetric
- then  $X = A^+$ . These properties uniquely define  $A^+$ .

Note, for example, that if  $A$  is tall, full rank, then  $A^+$  is not the only left inverse of  $A$ ! (think about it).

Connection to regularization with  $L_2$  norm (Tikhonov).

$$\underset{X}{\text{minimize}} \quad \|Ax - b\|^2 + \lambda \|x\|^2 \quad \text{for some } \lambda > 0.$$

We saw this always has a unique solution. Why?take SVD of  $A$ , let  $\tilde{x}_1 = V_1^T x$ ,  $\tilde{x}_2 = V_2^T x$ ,  $\tilde{b}_1 = U_1^T b$ ,  $\tilde{b}_2 = U_2^T b$ .

$$\Rightarrow \underset{\tilde{x}_1, \tilde{x}_2}{\text{minimize}} \quad \left\| \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix} x - b \right\|^2 + \lambda \|x\|^2$$

$$= \underset{\tilde{x}_1, \tilde{x}_2}{\text{minimize}} \quad \left\| \sum_i \tilde{x}_i - \tilde{b}_1 \right\|^2 + \lambda \|\tilde{x}_1\|^2 + \lambda \|\tilde{x}_2\|^2$$

$$= \underset{\tilde{x}_1}{\text{minimize}} \quad \sum_{i=1}^r \left[ (\sigma_i \tilde{x}_i - \tilde{b}_i)^2 + \lambda \tilde{x}_i^2 \right]$$

clearly choose this to be zero.

(3)

$$= \min_{\tilde{x}_i} \sum_{i=1}^r \left[ (\sigma_i \tilde{x}_i - \tilde{b}_i)^2 + \lambda \tilde{x}_i^2 \right]$$

$$= \min_{\tilde{x}_i} \sum_{i=1}^r (\sigma_i^2 + \lambda) \tilde{x}_i^2 - 2\sigma_i \tilde{b}_i \tilde{x}_i + \tilde{b}_i^2$$

{ minimized when  $2(\sigma_i^2 + \lambda) \tilde{x}_i - 2\sigma_i \tilde{b}_i = 0$ ,

$$\text{or equivalently: } \tilde{x}_i = \frac{\sigma_i}{\sigma_i^2 + \lambda} \tilde{b}_i$$

In the limit  $\lambda \rightarrow 0$ , we recover  $\tilde{x}_i = \sigma_i^{-1} \tilde{b}_i$ . i.e.  $x = A^+ b$ .

Therefore: we can say that:

$$\begin{matrix} \text{minimize} \\ \|Ax - b\|^2 \end{matrix}$$

x

such that  $\|x\|$  is small as possible

(equivalent)



$$\lim_{\lambda \rightarrow 0} \min_x \|Ax - b\|^2 + \lambda \|x\|^2$$

$$\hat{x} = A^T (A A^T)^{-1} b \quad (\text{when } A \text{ full row rank})$$

$$= A^+ b \quad (\text{in general})$$

$$\hat{x} = \lim_{\lambda \rightarrow 0} (A^T A + \lambda I)^{-1} A^T b$$



we can't evaluate this limit by substituting  
 $\lambda = 0$ , because  $A^T A$  is not necessarily invertible!

Trick:

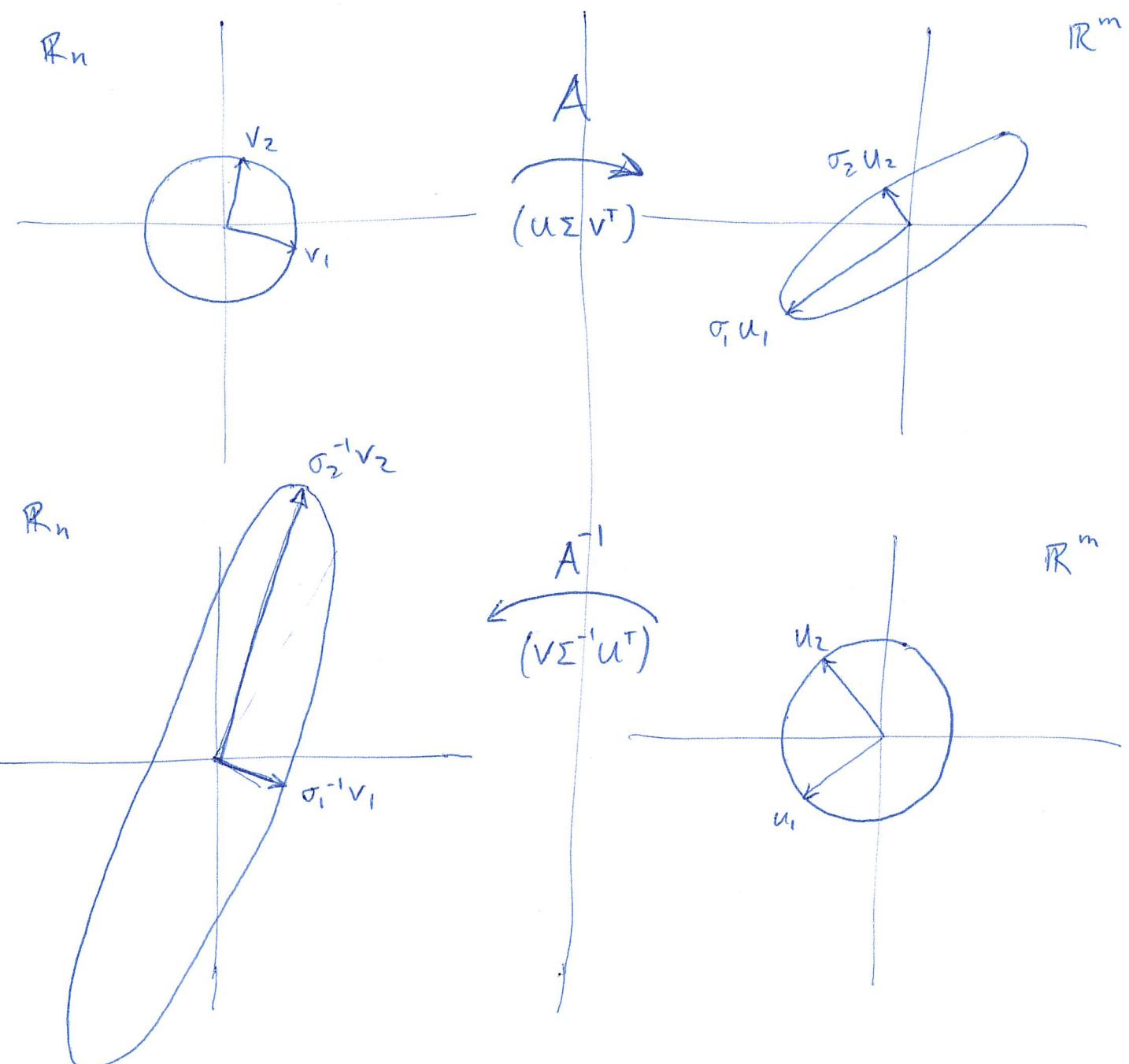
$$A^T (A A^T + \lambda I) = A^T A A^T + \lambda A^T = (A^T A + \lambda I) A^T$$

$$\Rightarrow (A^T A + \lambda I)^{-1} A^T = A^T (A A^T + \lambda I)^{-1} \rightarrow A^T (A A^T)^{-1}$$

if  $A$  full row rank.

## Sensitivity demo

(4)



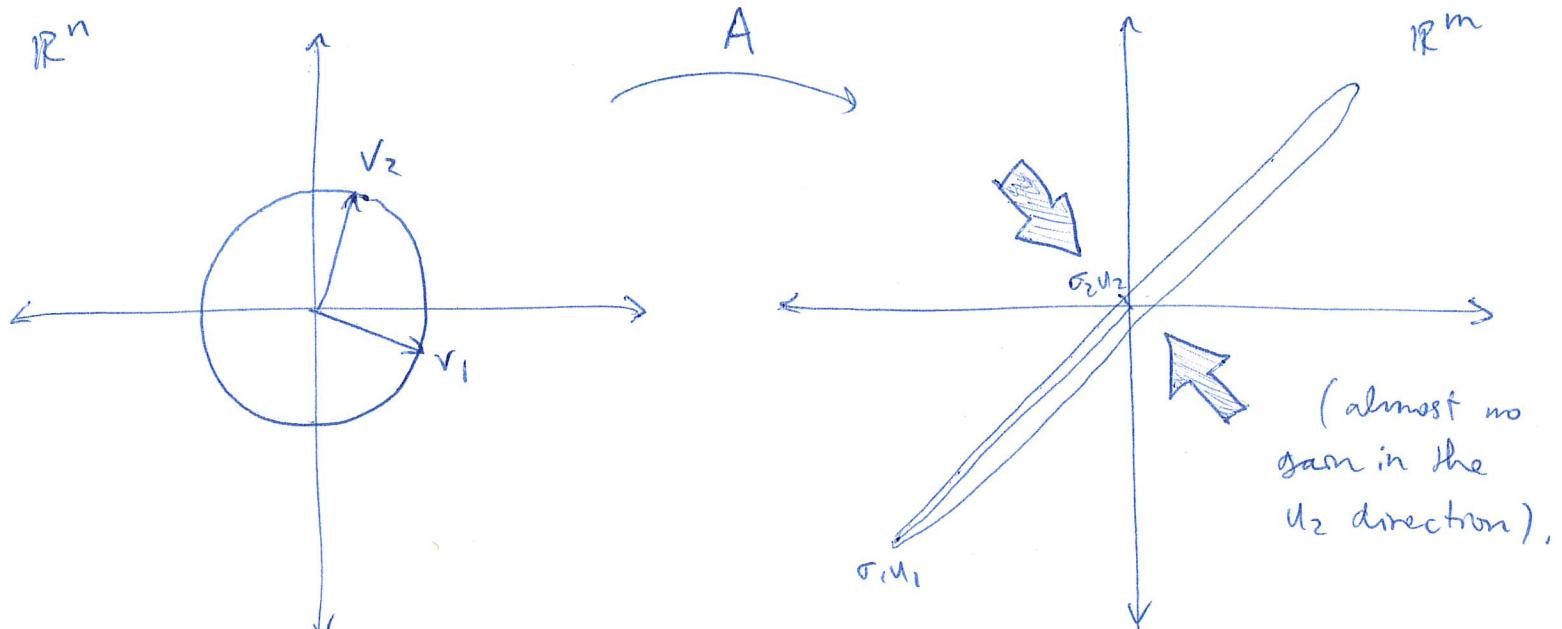
\* mapping  $\xrightarrow{A}$ ,  $v_i \rightarrow \sigma_i u_i$ , (max amplification).

\* mapping  $\xrightarrow{A^{-1}}$ ,  $u_2 \rightarrow \frac{1}{\sigma_2} v_1$  (max amplification).

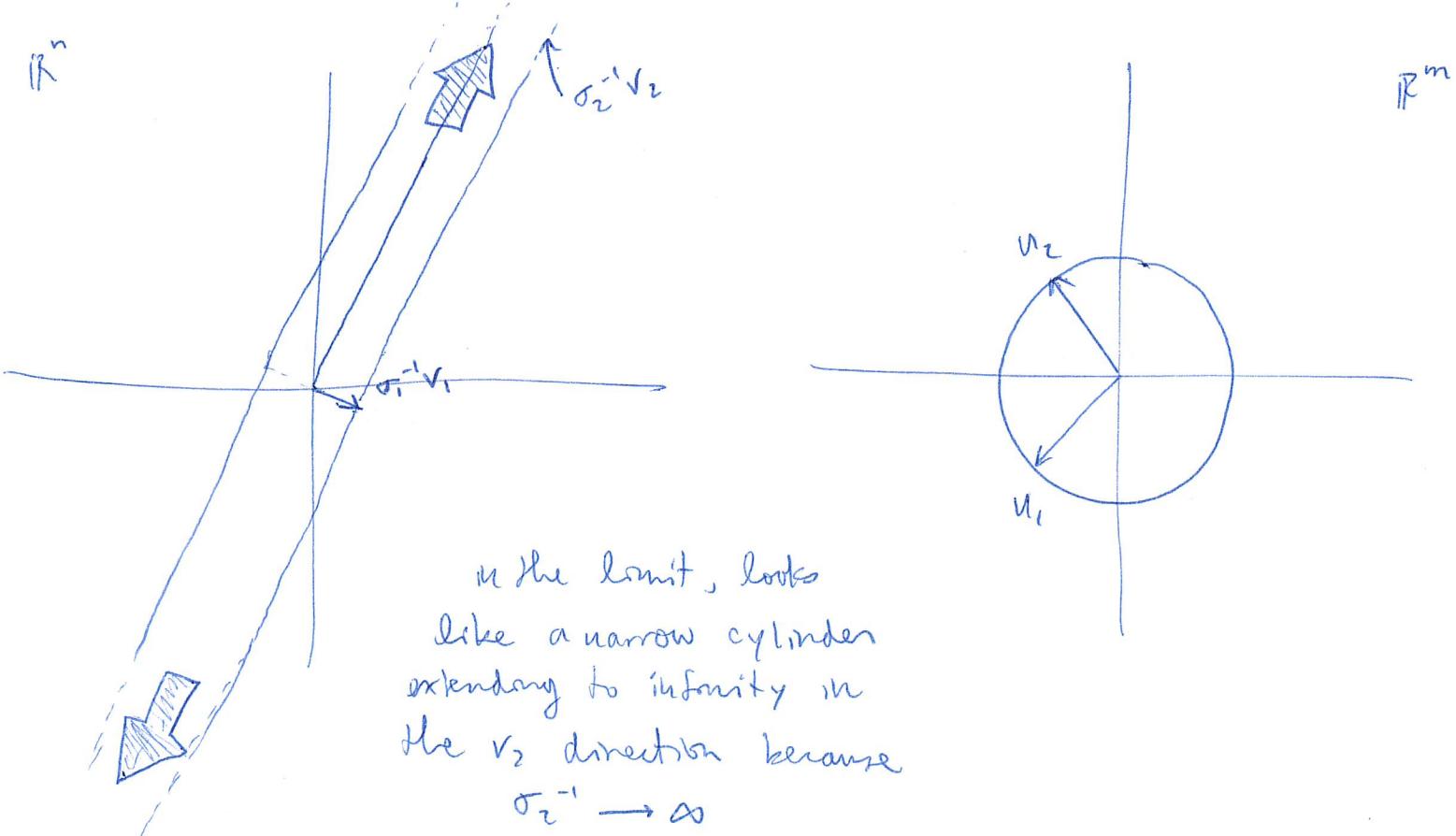
5

## More geometry: degeneracy

If  $\sigma_2$  is very small, we get:



i.e.  $v_2$  is "almost" in the null space of  $A$ ,  
and  $R(A)$  is "almost" equal to  $\text{span}\{u_1\}$ ,



if we care about solving  $Ax = b$

⑥

and we perturb  $b$ ; say  $b_0 \rightarrow b_0 + \Delta b$ .

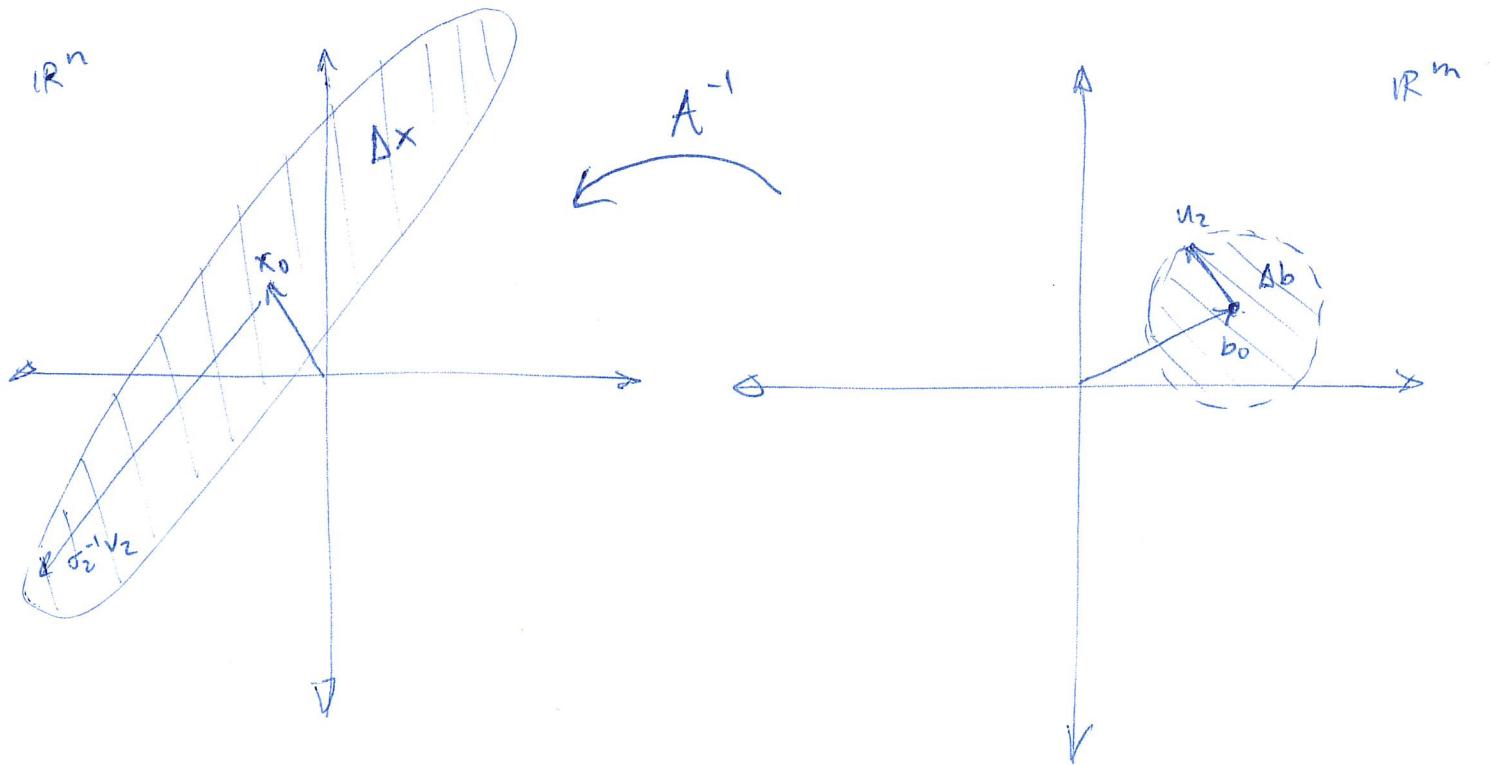
then  $x_0$  will change to  $x_0 + \Delta x$ .

$$\left. \begin{array}{l} Ax_0 = b_0 \\ A(x_0 + \Delta x) = b_0 + \Delta b \end{array} \right\} \Rightarrow A \Delta x = \Delta b.$$

if  $\Delta b$  is a ball of uncertainty,  $\Delta x = A^{-1}\Delta b$  is  
the ball mapped through  $A^{-1} = V\Sigma^{-1}U^T$ ,

so if  $\Delta b$  is in direction  $u_2$ , then

$\Delta x$  will be maximized.



(7)

## Condition number

suppose we want to solve  $Ax = b$ , and we want to characterize "worst-case" error.

suppose  $b \rightarrow b + \Delta b$ . relative error =  $\frac{\|\Delta b\|}{\|b\|}$

and  $x \rightarrow x + \Delta x$ . relative error =  $\frac{\|Ax\|}{\|x\|}$ .

so we assume  $Ax = b$  and  $A(x + \Delta x) = (b + \Delta b)$ ,  $\Rightarrow A\Delta x = \Delta b$ .

$$\begin{aligned} \max_{x, b} \frac{\text{rel. error in } x}{\text{rel. error in } b} &= \max_{x, b} \frac{\|Ax\|}{\|x\|} \cdot \frac{\|b\|}{\|\Delta b\|} = \max_{x, b} \frac{\|A^{-1}\Delta b\|}{\|x\|} \cdot \frac{\|Ax\|}{\|\Delta b\|} \\ &= \max_{x, \Delta b} \left( \frac{\|Ax\|}{\|x\|} \right) \left( \frac{\|A^{-1}\Delta b\|}{\|\Delta b\|} \right) = \|A\| \cdot \|A^{-1}\| = K(A), \end{aligned}$$

Note:  $\|A\| = \sigma_1$  and  $A^{-1}$  has singular values  $\left[ \begin{smallmatrix} \sigma_r^{-1} & & \\ & \sigma_{r-1}^{-1} & \\ & & \ddots & \\ & & & \sigma_1^{-1} \end{smallmatrix} \right]$

$$\text{so } \|A^{-1}\| = \sigma_r^{-1}.$$

$$\text{i.e. } K(A) = \|A\| \cdot \|A^{-1}\| = \frac{\sigma_1}{\sigma_r}.$$

Note:  $K(A) \geq 1$  always, and if not full rank,  $K(A) = \infty$ .

ex: if  $K(A) = 50$  and I change  $b$  by 1%, then

the solution to  $Ax = b$  will change by at most 50%.