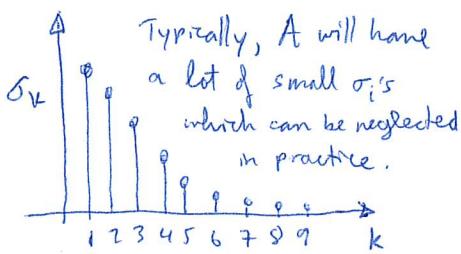


### low rank approximation

$$A = U, \Sigma, V^T = [U_1 \dots U_r] \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix} \begin{bmatrix} V_1^T \\ \vdots \\ V_r^T \end{bmatrix}$$



$$= \sigma_1 U_1 V_1^T + \dots + \sigma_r U_r V_r^T$$

sum of rank-1 matrices

\* idea: if we want to approximate  $A$  as a rank  $k < r$  matrix, we should keep the  $k$  most significant  $\sigma_i$ 's. i.e.

$$A \approx \sigma_1 U_1 V_1^T + \dots + \sigma_k U_k V_k^T$$

we can re-assemble  $A$  by writing:

$$A_k = [U_1 \dots U_k] \underbrace{\begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_k \\ & 0 & \\ & \vdots & \\ & 0 \end{bmatrix}}_{\text{only first } k \text{ sigma}_i \text{'s are kept; the rest are zeroed out.}} \begin{bmatrix} V_1^T \\ \vdots \\ V_k^T \end{bmatrix} = [U_1 \dots U_k] \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_k \end{bmatrix} \begin{bmatrix} V_1^T \\ \vdots \\ V_k^T \end{bmatrix}.$$

only first  $k$   $\sigma_i$ 's are kept;  
the rest are zeroed out.

ok, now why is this a good idea? i.e. in what sense is this a "good approximation to  $A$ "? It turns out  $A_k$  will be close to  $A$  in both the 2-norm and the Frobenius norm.

(2)

## Eckart - Young theorem (1936)

If  $A \in \mathbb{R}^{m \times n}$  has rank  $r > k$ , and  $\|\cdot\|$  is any unitarily invariant norm, then the solution to the optimization problem:

$$\begin{aligned} &\text{minimize } \|A - B\| \\ &\text{rank}(B) \leq k \end{aligned}$$

e.g. the induced 2-norm and the Frobenius norm.

is given by  $B = A_k = \sum_{i=1}^k \sigma_i u_i v_i^T$  where  $A = U \Sigma V^T$  is the svd.

proof for the induced 2-norm case: Let  $B^*$  be the optimal approximation.

Since  $\text{rank}(B^*) \leq k < r$ , then we have:

$$\dim N(B^*) = n - \text{rank}(B^*) \geq n - k.$$

Let  $[u_1 \dots u_r] \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix} \begin{bmatrix} v_1^T \\ \vdots \\ v_r^T \end{bmatrix}$  be the SVD of  $A$ . Consider the space

$R([v_1 \dots v_{k+1}])$ , which has dimension  $k+1$ .

Since  $\dim R([v_1 \dots v_{k+1}]) + \dim N(B^*) = n+1$ , there must exist some  $x \in \mathbb{R}^n$  that belongs to both sets. i.e. the basis vectors for both sets can't jointly be independent.

So let  $\underbrace{x \in R([v_1 \dots v_{k+1}])}_{\text{write } x = \sum_{i=1}^{k+1} \alpha_i v_i}$  and  $x \in N(B^*)$ . Also normalize so  $\|x\| = 1$ .

write  $x = \sum_{i=1}^{k+1} \alpha_i v_i$  where  $\alpha_1^2 + \dots + \alpha_{k+1}^2 = 1$ . We also have  $B^* x = 0$ .

$$\|A - B^*\|^2 \geq \|(A - B^*)x\|^2 = \|Ax\|^2 = \left\|U \Sigma V^T \left( \sum_{i=1}^{k+1} \alpha_i v_i \right) \right\|^2 = \sum_{i=1}^{k+1} (\alpha_i \sigma_i)^2 \geq \sigma_{k+1}^2 \sum_{i=1}^{k+1} \alpha_i^2 = \sigma_{k+1}^2$$

Therefore  $\|A - B^*\| \geq \sigma_{k+1}$ . But we can achieve this by setting  $B^* = \sum_{i=1}^k \sigma_i u_i v_i^T$ .

$$\text{then } \|A - \sum_{i=1}^k \sigma_i u_i v_i^T\| = \left\| \sum_{i=k+1}^r \sigma_i u_i v_i^T \right\| = \sigma_{k+1}.$$

Note: in Frobenius norm case;  
 $\|A - B^*\|_F^2 = (\sigma_{k+1}^2 + \sigma_{k+2}^2 + \dots + \sigma_r^2)$

■

(3)

revisiting  $Ax = b$

$$Ax = b$$

$$\begin{matrix} m \\ r \\ m-r \end{matrix} \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix} x = b$$

take SVD of  $A$ .

multiply both sides by  $U^T$

$$\begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_1^T x \\ V_2^T x \end{bmatrix} = \begin{bmatrix} U_1^T b \\ U_2^T b \end{bmatrix}$$

we can define  $\tilde{b}_1 = U_1^T b$  and  $\tilde{b}_2 = U_2^T b$ .

also, let  $\tilde{x}_1 = V_1^T x$  and  $\tilde{x}_2 = V_2^T x$ .

We can easily reconstruct  $x$  from  $\tilde{x}_1$  and  $\tilde{x}_2$

$$x = VV^T x = V_1 V_1^T x + V_2 V_2^T x = V_1 \tilde{x}_1 + V_2 \tilde{x}_2$$

$$\Rightarrow \begin{bmatrix} r & n-r \\ m-r & 0 \end{bmatrix} \begin{bmatrix} \tilde{x}_1 \\ \tilde{x}_2 \end{bmatrix} = \begin{bmatrix} \tilde{b}_1 \\ \tilde{b}_2 \end{bmatrix}$$

Note:  $\underbrace{U_1 \tilde{b}_1}_{\text{proj } b_{R(A)}} + \underbrace{U_2 \tilde{b}_2}_{\text{proj } b_{R(A)^\perp}} = b$

This has a solution iff  $\tilde{b}_2 = 0$ , i.e.  $\text{proj}_{R(A)^\perp} b = 0$ ;  $b \in R(A)$ .

In this case, we can pick  $\tilde{x}_2 = \text{anything}$ , and  $\tilde{x}_1 = \Sigma_1^{-1} \tilde{b}_1$

$$\text{So } x = V_1 \tilde{x}_1 + V_2 \tilde{x}_2 = \underbrace{V_1 \Sigma_1^{-1} U_1^T b}_{\text{particular solution.}} + \underbrace{V_2 \tilde{x}_2}_{\text{arbitrary element of } N(A)}.$$

(4)

Special case: underdetermined full-rank  $Ax = b$ .

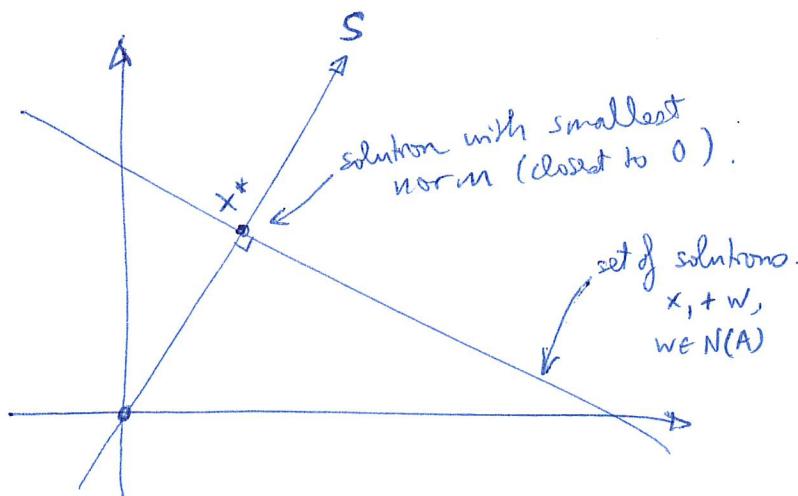
$$\begin{bmatrix} b \end{bmatrix} = \left[ \underbrace{A}_{m \times n \text{ with } m < n} \right] \begin{bmatrix} x \end{bmatrix}$$

where  $A$  has full row rank ( $\text{rank}(A) = m$ ).

In this case,  $\text{dom } R(A) = m$  so  $R(A) = \mathbb{R}^m$ . i.e.  $b \in R(A)$  is always true! so there is always a solution. Infinitely many, since  $N(A) \neq \{0\}$ . Which one should we choose?

One option: pick solution where  $x$  is small, i.e.  $\|x\|$  minimal.

geometry:



we want  $x^*$  such that

$$Ax^* = b$$

and  $\|x^*\|$  is minimal.

equivalently:

$$\text{proj}_S x^* = x^*$$

where  $S = N(A)^\perp$ .

but  $S = N(A)^\perp = R(A^\top)$ .

$$x^* = \text{proj}_{R(A^\top)} x^* = (A^\top) \left( (A^\top)^\top (A^\top) \right)^{-1} (A^\top)^\top x^*$$

$$= A^\top (A A^\top)^{-1} A x^*$$

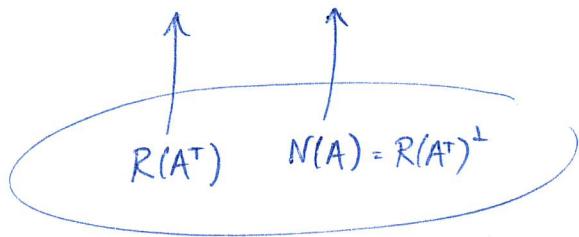
$$\boxed{x^* = A^\top (A A^\top)^{-1} b.}$$

minimum-norm  
solution when  $A$  has  
full row rank.

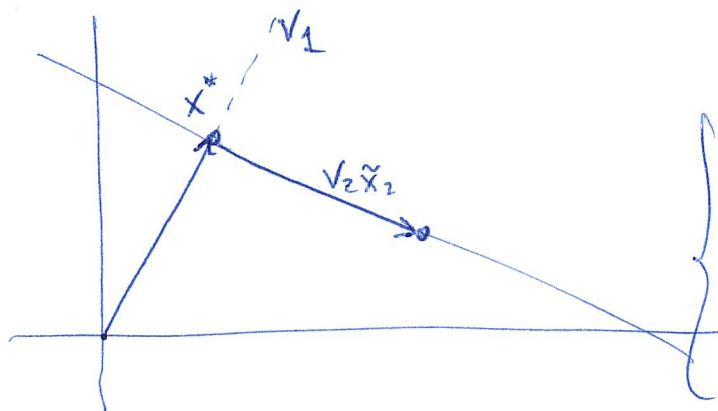
(5)

using the SVD:

$$x = V_1 \tilde{x}_1 + V_2 \tilde{x}_2 = V_1 \Sigma_1^{-1} U_1^T b + V_2 \tilde{x}_2.$$



these are precisely the subspaces we care about!



so to minimize  $\|x\|$ , choose  $\tilde{x}_2 = 0$ .

alternatively; solve:

$$\text{minimize}_{\tilde{x}_2} \|V_1 \Sigma_1^{-1} U_1^T b + V_2 \tilde{x}_2\|$$

This is a standard least squares problem with tall full-rank  $V_2$ .

$$\Rightarrow \tilde{x}_2^{\text{opt}} = (V_2^T V_2)^{-1} \underbrace{V_2^T (V_1 \Sigma_1^{-1} U_1^T b)}_{\text{zero}} = 0.$$

we can verify that  $V_1 \Sigma_1^{-1} U_1^T b = A^T (A A^T)^{-1} b$ .

$$A = U_1 [\Sigma_1 \ 0] \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix}. \quad A A^T = U_1 [\Sigma_1 \ 0] V^T V \begin{bmatrix} \Sigma_1 \\ 0 \end{bmatrix} U_1^T = U_1 \Sigma_1^2 U_1^T$$

$$A^T (A A^T)^{-1} = [V_1 \ V_2] \begin{bmatrix} \Sigma_1 \\ 0 \end{bmatrix} U_1^T U_1 \Sigma_1^{-2} U_1^T = V_1 \Sigma_1 \Sigma_1^{-2} U_1^T = V_1 \Sigma_1^{-1} U_1^T$$

(6)

If  $A$  has full column rank, then the LS solution is:

$$A = [U_1 U_2] \begin{bmatrix} \Sigma_1 \\ 0 \end{bmatrix} V_1^T$$

$$(A^T A)^{-1} A^T b = \left( V_1 [\Sigma_1 0] U^T U \begin{bmatrix} \Sigma_1 \\ 0 \end{bmatrix} V_1^T \right)^{-1} V_1 [\Sigma_1 0] \begin{bmatrix} U_1^T \\ U_2^T \end{bmatrix}$$

$$= (V_1 \Sigma_1^2 V_1^T)^{-1} V_1 \Sigma_1 U_1^T$$

$$= V_1 \Sigma_1^{-2} V_1^T V_1 \Sigma_1 U_1^T$$

$$= V_1 \Sigma_1^{-1} U_1^T. \quad \underline{\text{same as before!}}$$

If  $A$  is square + invertible:

$$A^{-1} b = V_1 \Sigma_1^{-1} U_1^T b.$$

(solution to  $Ax = b$ ),

If  $A$  tall + full-rank:

$$(A^T A)^{-1} A^T b = V_1 \Sigma_1^{-1} U_1^T b.$$

(solution to  $\min_x \|Ax - b\|$ ).

If  $A$  wide + full-rank:

$$A^T (A A^T)^{-1} b = V_1 \Sigma_1^{-1} U_1^T b$$

(solution to  $\min_x \|x\|$  s.t.  $Ax = b$ )

$A^+ = V_1 \Sigma_1^{-1} U_1^T$  is called the pseudo inverse of  $A$ .

and it solves all these problems. (a.k.a. "Moore-Penrose pseudo inverse")