MATH 372
LECTURE 22

: Pseudoinverse

We are going to use the SVD of A to solve two previous problems of the course:

1) If $A\vec{x} = \vec{b}$ has infinite solutions, which one is the shortest length solution?

2) If A has linearly dependent columns, how to solve the least-squares problem associated to $A\vec{x} = \vec{b}$?

• Def: Pseudoinverse of A

$$A^+ = V \Sigma^+ U^T, \text{ where } \Sigma^+ = \begin{array}{cc} \begin{bmatrix} 1/\sigma_1 & & & \\ & \ddots & & \\ & & 1/\sigma_r & \\ & & & \\ \end{bmatrix} & \}\, n-r \\ \underbrace{\hphantom{xx}}_{m-r} & \end{array}.$$

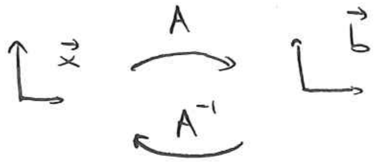
(Notice Σ^+ is $\underline{n \times m}$).

• Remark: If A is invertible, $A^+ = A^{-1}$.

↳ A invertible $\Rightarrow n = r \Rightarrow U, \Sigma, V$ square $n \times n$, and

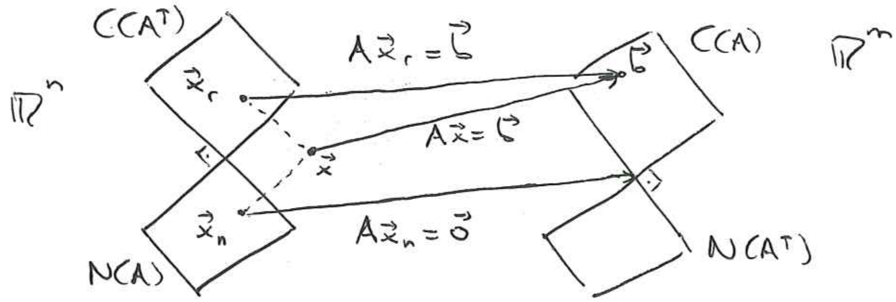
$$A^{-1} = (U \Sigma V^T)^{-1} = V \Sigma^{-1} U^T \equiv V \Sigma^+ U^T = A^+$$

One of the easiest ways to think about the inverse, A^{-1}, is as the solution of $A\vec{x}=\vec{b}$.



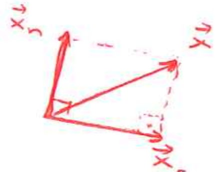
That is, knowing $\vec{b}$, we can find $\vec{x}$ as $A^{-1}\vec{b}$.

This is not possible if $A\vec{x}=\vec{b}$ has infinite solutions:



When $A\vec{x}=\vec{b}$ has infinite solutions, all the solutions are of the form $\vec{x}=\vec{x}_p+\vec{x}_n$, that is, we pick any solution and we can modify it by adding something in the nullspace.

How to decide which one to pick? It seems natural to choose the one with shortest length. That is the one (and only one) "living" in the row space, $\vec{x}_r$.



$\|\vec{x}\|^2=\|\vec{x}_r\|^2+\|\vec{x}_n\|^2$

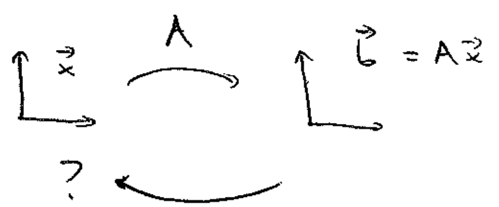
$$\left.\begin{array}{l}\vec{x}=\vec{x}_r+\vec{x}_n\\ \text{with}\\ \vec{x}_r\cdot\vec{x}_r=0\end{array}\right\} \Rightarrow \|\vec{x}\|^2=\|\vec{x}_r\|^2+\underbrace{\|\vec{x}_n\|^2}_{\text{choose this to be } \vec{0}\,!}$$

In all this, it is crucial the fact that $C(A^T)$ and $N(A)$ are orthogonal complements.

Since both spaces are orthogonal to each other, given $\vec{x} \in \mathbb{R}^n$, we can find $\vec{x}_r$ as the projection of $\vec{x}$ onto $C(A^T)$

(since $N(A) \perp C(A^T)$, the part of $\vec{x} \in N(A)$ is "killed" by this projection).

- In summary, it is natural to define the inverse of the process in the picture,



by choosing $\vec{x}_r$.

- In other words, we want the solution of $A\vec{x} = \vec{b}$ to be $\vec{x}_r$, when this system has infinite solutions.

We are now going to see that $\| \vec{x}_r = A^{+} \vec{b} \|$

- Is $A^+\vec{b} = \vec{x}_r$?

  That is, is $A^+\vec{b}$ the projection of $\vec{x}$ onto $C(A^T)$?

  Let's see if:

$$A\vec{x} = A\vec{b} \rightarrow A^+A\vec{x} = A^+\vec{b}\text{, so, is } A^+A\vec{x} = \text{proj}_{C(A^T)}\vec{x} \text{ ?}$$

- $A^+A = V\Sigma^+ U^T U \Sigma V^T = V\Sigma^+\Sigma V^T$,

$$\underset{(n\times m)}{\Sigma^+}\underset{(m\times n)}{\Sigma} = \begin{bmatrix} 1/\sigma_1 & & & \\ & \ddots & & \\ & & 1/\sigma_r & \\ & & & \end{bmatrix}\begin{bmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_r & \\ & & & \end{bmatrix} = \left[\begin{array}{ccc|c} 1 & & & \\ & \ddots & & \\ & & 1 & \\ \hline & & & \underbrace{}_{n-r} \end{array}\right]\}\, n-r$$

$$V\Sigma^+\Sigma V^T = \left[\begin{array}{ccc|ccc} \vec{v}_1 & \cdots & \vec{v}_r & \vec{v}_{r+1} & \cdots & \vec{v}_n \end{array}\right]\left[\begin{array}{ccc|c} 1 & & & \\ & \ddots & & \\ & & 1 & \\ \hline \underbrace{}_{r} & & & \end{array}\right]\left[\begin{array}{c} \vec{v}_1^T \\ \vdots \\ \vec{v}_r^T \\ \hline \vec{v}_{r+1}^T \\ \vdots \\ \vec{v}_n^T \end{array}\right] =$$

$$= \vec{v}_1\vec{v}_1^T + \dots + \vec{v}_r\vec{v}_r^T \equiv \text{Projection onto } \{\vec{v}_1, \dots, \vec{v}_r\}$$

the space spanned by

But recall that

$\{\vec{v}_1, \dots, \vec{v}_r\}$ was an orthonormal basis for $C(A^T)$!

• In summary, for $A\vec{x} = \vec{b}$, the solution $A^+\vec{b}$ gives

→ Shortest length solution.

→ Only solution in $C(A^T)$.

II.2) Least Squares.

Recall that if $A\vec{x} = \vec{b}$ doesn't have a solution (that is, $\vec{b}$ is not in the column space), then we solve instead

$$\left\| A\vec{x} = \text{proj}_{C(A)}\vec{b} \right\|.$$

We saw in chapter 4. that one way of solving this new system is by the normal equations:

$$A^T A\vec{x} = A^T\vec{b} \underset{\substack{\uparrow \\ \text{If } A^TA \text{ invertible}}}{\Longrightarrow} \vec{x} = (A^TA)^{-1}A^T\vec{b} \quad \left(\text{indeed, } A\vec{x} = \underbrace{A(A^TA)^{-1}A^T\vec{b}}_{\text{proj}_{C(A)}\vec{b}}\right)$$

What if A^TA is not invertible? That is, what if the system has infinite solutions? $\left(\text{always has at least one, as } \text{proj}_{C(A)}\vec{b} \in C(A) \text{ obviously}\right)$.

[A^TA is not invertible if cols. of A are lin. dependent].

Well, now we know that for systems with infinite solutions the pseudoinverse is concerned.

Let's see that $\left\| \vec{x} = A^+ \vec{b} \right.$ is also the least-squares solution $\|$
(with shortest length if there are many).

[] If $A\vec{x} = \vec{b}$ no solution $\to$ Solve instead $A\vec{x} = \text{proj}_{C(A)} \vec{b}$.

Proceeding as in -197-, one can find that

$$AA^+ = U \Sigma \Sigma^+ U^T = \vec{u}_1 \vec{u}_1^T + \dots + \vec{u}_r \vec{u}_r^T \equiv$$ Projection matrix onto subspace spanned by $\{\vec{u}_1, \dots, \vec{u}_r\}$

That is,

$$\left\| AA^+ \equiv \text{projection matrix onto } C(A) \right\|$$

So we want to solve $A\vec{x} = AA^+ \vec{b}$.

$\Rightarrow$ Solving $A\vec{x} = AA^+ \vec{b} \implies \vec{x} = A^+ AA^+ \vec{b}$

<u>Exercise</u>: Check that $A^+ = A^+ A A^+$.

Also check that $A = AA^+A$.

• <u>Summary</u>: For $A\vec{x} = \vec{b}$, the "solution" $\vec{x} = A^{+}\vec{b}$ gives:

1) Exact solution if $A\vec{x} = \vec{b}$ is solvable.

 1.1) If solution is unique, $A^{+}\vec{b} = A^{-1}\vec{b}$.

 1.2) If infinite solutions, $A^{+}\vec{b}$ is the shortest one.
 (<u>the</u> solution in $C(A^{T})$) ~~is the only solution~~

2) Least square solution to $A\vec{x} = \vec{b}$ if it is not solvable.

 2.1) If cols. of A are l.i., then $A^{+}\vec{b} = (A^{T}A)^{-1}A^{T}\vec{b}$.

 2.2) If cols. of A are l.i., then $A^{+}\vec{b}$ is the shortest least-squares solution.

Summary of SVD and pseudoinverse: A $m \times n$, $\operatorname{rank}(A) = r$

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

$$\underset{\substack{m\times m\\ \text{orthogonal}\\ \text{matrix}}}{U} = \Big[\underbrace{\underbrace{\vec{u}_1 \cdots \vec{u}_r}_{C(A)} \; \underbrace{\vec{u}_{r+1} \cdots \vec{u}_m}_{N(A^T)}}_{\substack{\text{orthonormal eigenvectors}\\ \text{of } AA^T}}\Big] \qquad V = \Big[\underbrace{\underbrace{\vec{v}_1 \cdots \vec{v}_r}_{C(A^T)} \; \underbrace{\vec{v}_{r+1} \cdots \vec{v}_n}_{N(A)}}_{\substack{\text{orthonormal eigenvectors}\\ \text{of } A^TA}}\Big] \begin{matrix} n \times n \\ \text{orthogonal} \\ \text{matrix}\end{matrix}$$

The order is important: $\vec{u}_i, \vec{v}_i$ follow the order of the singular values $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$.

- Reduced SVD: $\quad U_r = \begin{bmatrix} \vec{u}_1 \cdots \vec{u}_r \end{bmatrix} \quad V_r = \begin{bmatrix} \vec{v}_1 \cdots \vec{v}_r \end{bmatrix}$

$A = U_r \Sigma_r V_r^T$, where $\quad \Sigma_r = \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix}$ $r \times r$ diagonal (recall that Σ is $m \times n$)

or

$A = \sigma_1 \vec{u}_1 \vec{v}_1^T + \cdots + \sigma_r \vec{u}_r \vec{v}_r^T$.

• Pseudoinverse: $A^+ = V\Sigma^+ U^T$

1) A^+A is the projection matrix onto $C(A^T)$

$$A^+A = V\Sigma^+ U^T U\Sigma V^T = V\Sigma^+\Sigma V^T = \vec{v}_1\vec{v}_1^T + \dots + \vec{v}_r\vec{v}_r^T$$

(recall that the $\vec{v}$'s are orthonormal)

2) AA^+ is the projection matrix onto $C(A)$

$$AA^+ = U\Sigma V^T V\Sigma^+ U^T = U\Sigma\Sigma^+ U^T = \vec{u}_1\vec{u}_1^T + \dots + \vec{u}_r\vec{u}_r^T$$

(the $\vec{u}$'s are orthonormal)

Therefore:

1) If a system has infinite solutions, i.e.,

$A\vec{x} = \vec{b} \leadsto \vec{x} = \vec{x}_p + \vec{x}_n$ with $\vec{x}_p$ any particular solution (not necessarily in row space), $\vec{x}_n \in N(A)$

the shortest length solution (the only one living in the row space) is the projection of $\vec{x}$ onto the row space:

$$\vec{x}_r = A^+A\vec{x} = A^+A\vec{x}_p + \underbrace{A^+\cancel{A}\vec{x}_n}_{\vec{0}} = A^+\vec{b}.$$

(2) If $A\vec{x} = \vec{b}$ no sol. $\to$ Solve $A\vec{x} = P_{C(A)}\vec{b} = AA^+\vec{b} \Rightarrow \vec{x} = A^+AA^+\vec{b} = A^+\vec{b}$)
(shortest length least squares sol.

Example: Find the shortest length solution,

$$\left.\begin{aligned} x + y + z &= 1 \\ y - z &= 1 \end{aligned}\right\}$$

Sol:

Let's do it as if we didn't know about the pseudoinverse.

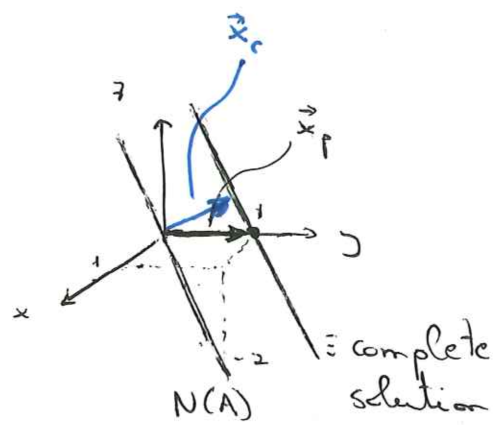
$A\vec{x} = \vec{b}$ with $A = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & -1 \end{bmatrix}$, $\vec{b} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$,

$$\left[\begin{array}{ccc|c} \boxed{1} & 1 & 1 & 1 \\ 0 & \boxed{1} & -1 & 1 \end{array}\right] \to \left[\begin{array}{ccc|c} \boxed{1} & 0 & 2 & 0 \\ 0 & \boxed{1} & -1 & 1 \end{array}\right] \qquad \begin{aligned} x_3 &= \alpha \\ x_2 &= 1 + \alpha \\ x_1 &= -2\alpha \end{aligned}$$

free

That is,

$$\vec{x} = \underbrace{\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}}_{\vec{x}_p} + \alpha \underbrace{\begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}}_{\text{basis for } N(A)} \quad \xrightarrow{\text{geometrically}}$$



We can see in the picture solutions different than $\vec{x}_p$ with shorter length. Indeed,

$\vec{x}_p = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$ is not in the row space, as it is not orthogonal to the nullspace.

If we find the projection matrix onto $C(A^T)$ then the shortest length solution (which we denote $\vec{x}_r$ because it is in the row space) is

$$\vec{x}_r = P_{C(A^T)}(\vec{x}_p + \vec{x}_n) = P_{C(A^T)}\vec{x}_p + \underbrace{\cancel{P_{C(A^T)}\vec{x}_n}}_{\vec{0} \text{ (as } C(A^T) \perp N(A))}.$$

Notice that we are lucky this time: the rows of A are orthogonal, so

$$P_{C(A^T)} = \frac{\vec{a}_1\vec{a}_1^T}{\|\vec{a}_1\|^2} + \frac{\vec{a}_2\vec{a}_2^T}{\|\vec{a}_2\|^2}, \text{ and so,}$$

$$\vec{x}_r = \vec{a}_1 \frac{1}{3}\vec{a}_1^T\vec{x}_p + \vec{a}_2\frac{1}{2}\vec{a}_2^T\vec{x}_p = \frac{1}{3}\begin{bmatrix}1\\1\\1\end{bmatrix} + \frac{1}{2}\begin{bmatrix}0\\1\\-1\end{bmatrix} = \begin{bmatrix}1/3\\5/6\\-1/6\end{bmatrix}$$

(Check: $\begin{bmatrix}1/3 & 5/6 & -1/6\end{bmatrix}\begin{bmatrix}-2\\1\\1\end{bmatrix} = 0$).

- Let's do it now using the pseudoinverse:

$\vec{x}_r = A^+\vec{b}$, where

$A^+ = V\Sigma^+U^T \rightarrow$ we need U, Σ, V

1) $\underline{\Sigma}$: A is $2\times 3 \rightarrow AA^T$ 2×2

$$AA^T = \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix} \quad \text{(A has orthogonal rows)}$$

$$\Downarrow$$

$$\lambda_1 = 3,\ \lambda_2 = 2 \rightarrow \sigma_1 = \sqrt{3},\ \sigma_2 = \sqrt{2}$$

$$\Sigma = \begin{bmatrix} \sqrt{3} & 0 & 0 \\ 0 & \sqrt{2} & 0 \end{bmatrix}$$

2) $\underline{V}$

$$A^T A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 0 \\ 1 & 0 & 2 \end{bmatrix}$$

$$A^T A - 3I = \begin{bmatrix} -2 & 1 & 1 \\ 1 & -1 & 0 \\ 1 & 0 & -1 \end{bmatrix} \rightarrow \begin{bmatrix} -2 & 1 & 1 \\ 0 & -1/2 & 1/2 \\ 0 & 1/2 & -1/2 \end{bmatrix} \rightarrow \vec{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \frac{1}{\sqrt{3}}$$

$$A^T A - 2I = \begin{bmatrix} -1 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} \rightarrow \begin{bmatrix} -1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix} \rightarrow \vec{v}_2 = \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix} \frac{1}{\sqrt{2}}$$

$$A^T A \rightarrow \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & -1 \\ 0 & -1 & 1 \end{bmatrix} \rightarrow \vec{v}_3 = \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix} \frac{1}{\sqrt{6}}$$

$$V = \begin{bmatrix} 1/\sqrt{3} & 0 & -2/\sqrt{6} \\ 1/\sqrt{3} & 1/\sqrt{2} & 1/\sqrt{6} \\ 1/\sqrt{3} & -1/\sqrt{2} & 1/\sqrt{6} \end{bmatrix}$$

3) $\underline{\underline{U}}$

$$\vec{u}_1 = \frac{A\vec{v}_1}{\sigma_1} = \frac{1}{\sqrt{3}}\begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & -1 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}\frac{1}{\sqrt{3}} = \frac{1}{3}\begin{bmatrix} 3 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix},$$

$$\vec{u}_2 = \frac{A\vec{v}_2}{\sigma_2} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & -1 \end{bmatrix}\begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}\frac{1}{\sqrt{2}} = \frac{1}{2}\begin{bmatrix} 0 \\ 2 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

$$U = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Check: $$U\begin{bmatrix} \sqrt{3} & 0 & 0 \\ 0 & \sqrt{2} & 0 \end{bmatrix}\begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3} \\ 0 & 1/\sqrt{2} & -1/\sqrt{2} \\ -2/\sqrt{6} & 1/\sqrt{6} & 1/\sqrt{6} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & -1 \end{bmatrix}$$ //

Thus,

$$A^+ = \begin{bmatrix} 1/\sqrt{3} & 0 & -2/\sqrt{6} \\ 1/\sqrt{3} & 1/\sqrt{2} & 1/\sqrt{6} \\ 1/\sqrt{3} & -1/\sqrt{2} & 1/\sqrt{6} \end{bmatrix}\begin{bmatrix} 1/\sqrt{3} & 0 \\ 0 & 1/\sqrt{2} \\ 0 & 0 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} =$$

$$= \begin{bmatrix} 1/3 & 0 \\ 1/3 & 1/2 \\ 1/3 & -1/2 \end{bmatrix} \rightsquigarrow \vec{x}_r = A^+\vec{b} = \begin{bmatrix} 1/3 \\ 5/6 \\ -1/6 \end{bmatrix}$$ //

$$AA^+ = \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}}_{\substack{\text{projection onto} \\ C(A) = \mathbb{R}^2}}, \quad A^+A = \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 1/3 & 5/6 & -1/6 \\ 1/3 & -1/6 & 5/6 \end{bmatrix} = \frac{1}{6}\begin{bmatrix} 2 & 2 & 2 \\ 2 & 5 & -1 \\ 2 & -1 & 5 \end{bmatrix}$$