

Linear Algebra Review

Fundamental subspaces of a matrix

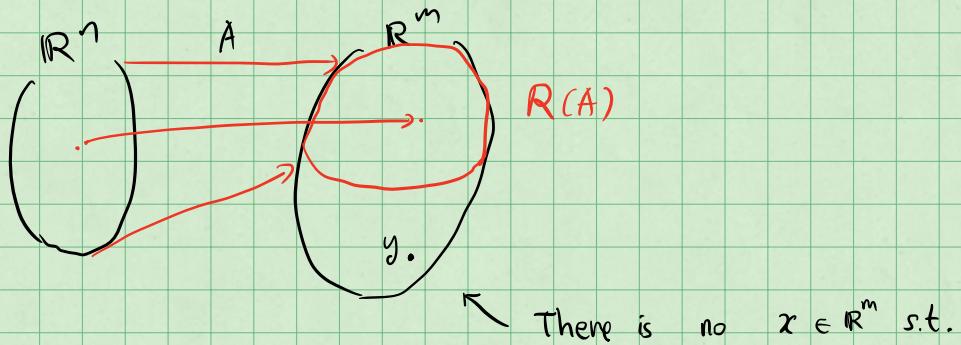
A : $m \times n$ matrix

As a linear operator, A maps \mathbb{R}^m to \mathbb{R}^n

$$\begin{aligned} A : \mathbb{R}^m &\longrightarrow \mathbb{R}^n \\ x &\longmapsto Ax \end{aligned}$$

(1) Then we can think of the "Range" of an operator A :

$$R(A) = \text{Range}(A) = \{Ax \in \mathbb{R}^n : x \in \mathbb{R}^m\}$$



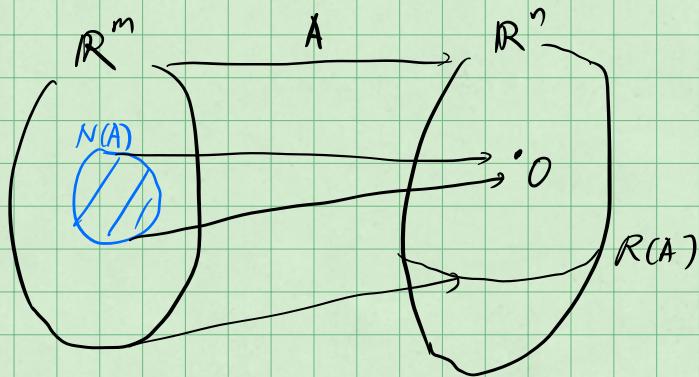
Using matrix algebra, we can easily check

$$R(A) = \text{col}(A) = \text{span}\{\text{columns of } A\}$$

Because $Ax = \begin{bmatrix} a_1 & \cdots & a_n \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \sum_{i=1}^n x_i a_i$

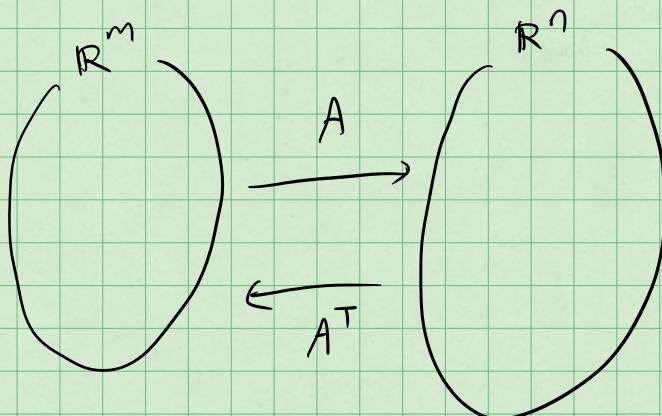
(2) We are also interested in the "kernel" of A , or "zeros" of A :

$$\ker(A) = \text{null}(A) = N(A) = \{x \in \mathbb{R}^m : Ax = 0\}$$



(3) Similarly, we can think of $R(A^T)$ and $N(A^T)$.

Note that A^T maps \mathbb{R}^n to \mathbb{R}^m



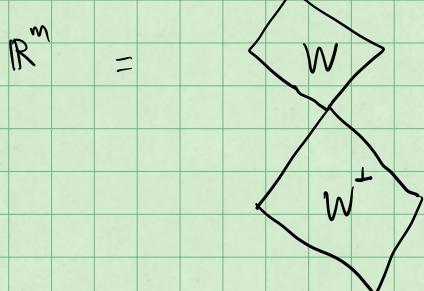
(4) Recall "orthogonal complement" of a space:

$$W \subseteq \mathbb{R}^m, \quad W^\perp = \{x \in \mathbb{R}^m : \langle x, y \rangle = 0 \quad \forall y \in W\}$$

↑
Subspace

↑
 x is orthogonal to W

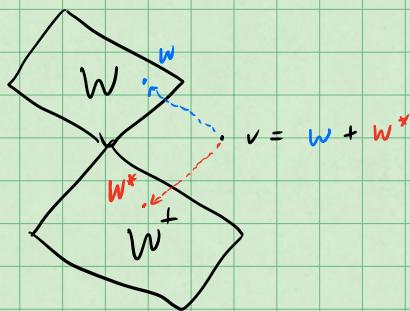
collect all such x 's



Note: any vector $v \in \mathbb{R}^m$ admits a "unique" decomposition $v = w + w^*$

$$\begin{matrix} w \\ \cap \\ w^* \end{matrix}$$

$$w = \text{Proj}_W(v), \quad w^* = \text{Proj}_{W^\perp}(v)$$



Interestingly, the four "fundamental" subspaces

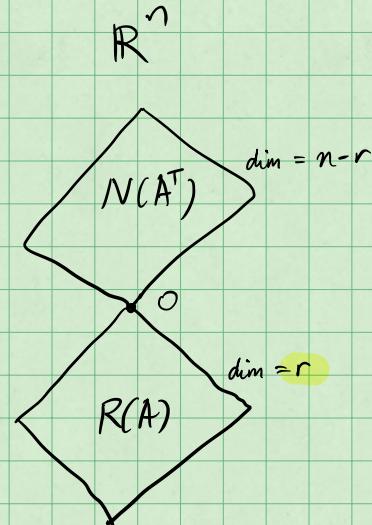
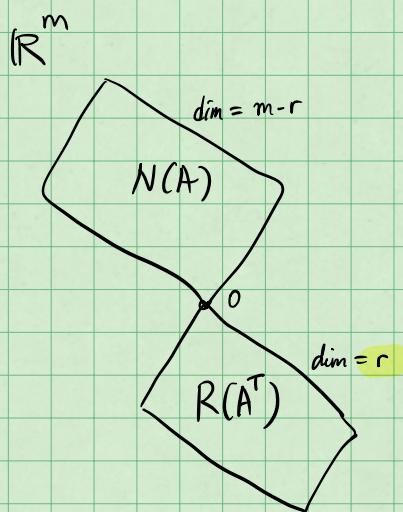
$R(A)$, $N(A)$, $R(A^T)$ and $N(A^T)$ have some orthogonality relations

$$R(A^T) = N(A)^\perp$$

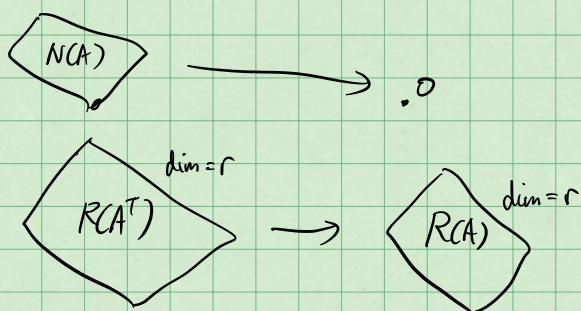
$$R(A)^\perp = N(A^T)$$

(The former is equiv. to the latter.)

Consider $A \leftarrow A^T$



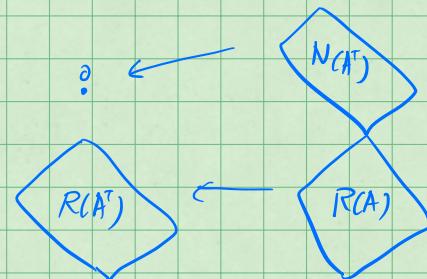
A:



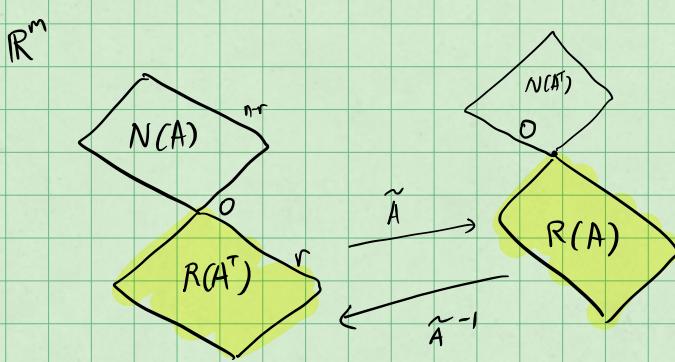
A is one-to-one and onto

from $R(A^T)$ to $R(A)$, and so is A^T from $R(A^T)$ to $R(A)$.

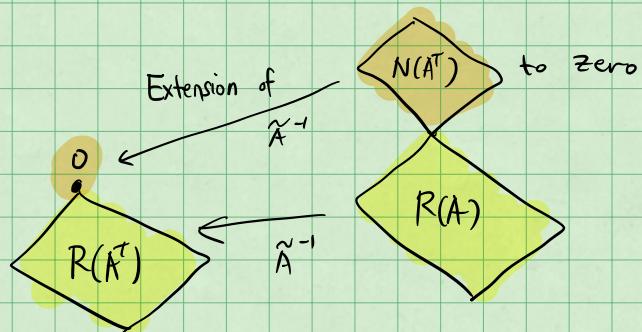
A^T :



$\Rightarrow \tilde{A}: R(A^T) \rightarrow R(A)$ has an inverse operator!



Let's extend $\tilde{A}^{-1} : R(A) \rightarrow R(A^T)$ to the whole space :



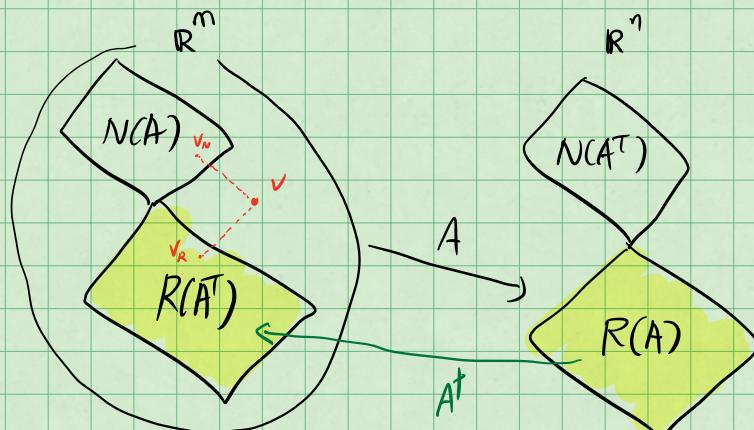
We call this extension of \tilde{A}^{-1} , a **pseudo inverse** of A

and denote by A^+ .

(Moore-Penrose inverse)

(5) A^+ has many interesting properties.

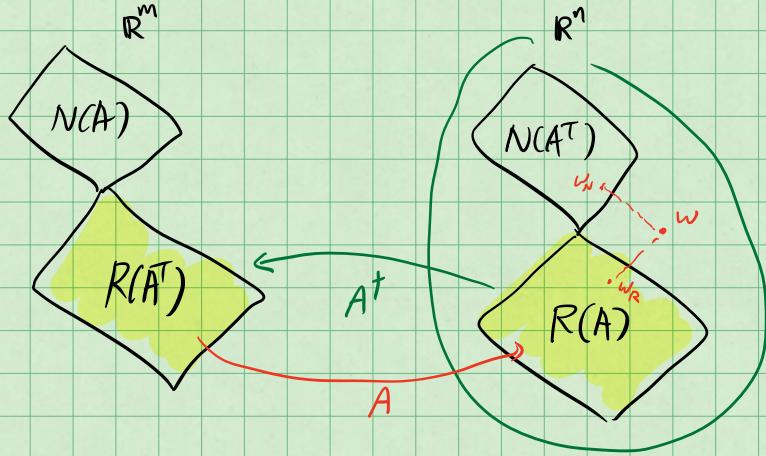
Most of them can be understood using the previous diagram.



$$v = v_N + v_R = \text{Proj}_{N(A)}(v) + \text{Proj}_{R(A^T)}(v)$$

$$A^+ A v = A^+(A v) = A^+(A v_R) = v_R$$

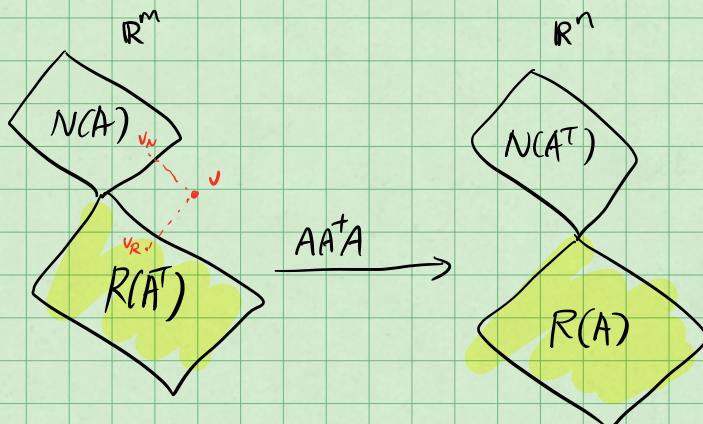
i.e., $A^+ A$ is a projection onto $R(A^T)$ ($\in R^m$)



Similarly, AA^T is a projection onto $R(A)$ in \mathbb{R}^m

(NOTE: A^T is not the identity matrix!)

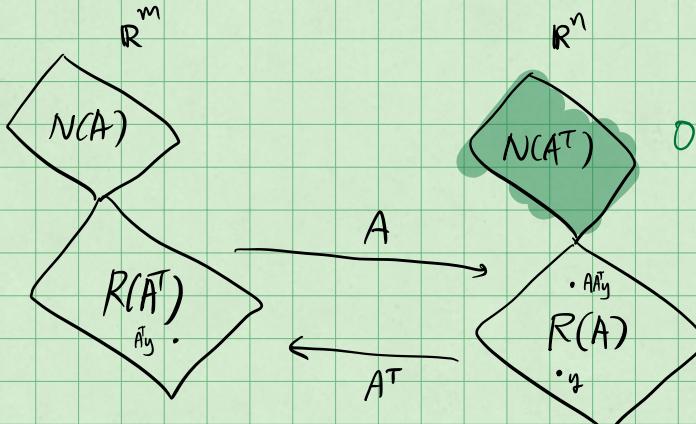
What about AA^TA ?



$$AA^TA = A \circ \text{Proj}_{R(AT)} = A$$

similarly, $A^TAA^T = A^T$.

Finally, consider AA^T and A^TA :

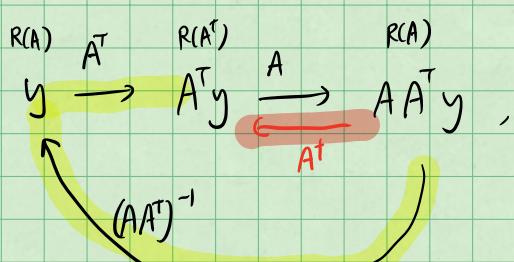


$$\left\{ \begin{array}{l} AA^T|_{R(A)} = A|_{R(A)}, \\ AA^T|_{R(AT)} = 0 \end{array} \right.$$

↓

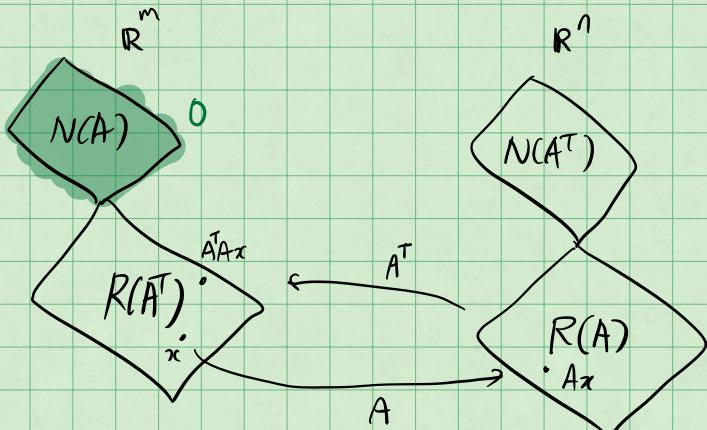
$(AA^T)^{-1}$ exists if $N(A^T) = 0$
 $\Leftrightarrow A$ lin. indep. rows

Then,

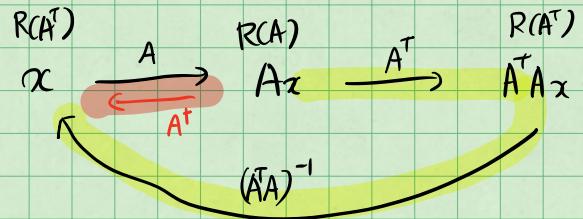


thus $A^T(AA^T)^{-1} = A^T$

Similarly, when A has lin. indep. columns, we have



$$(A^T A)^{-1} A^T = A^+$$

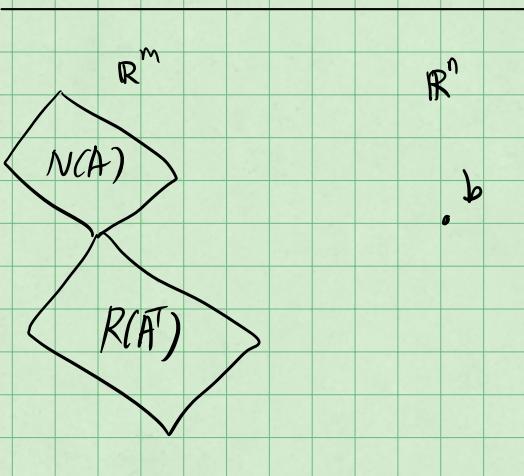


$(A^T A)$ has an inverse!

★ Minimum Norm Solution through A^+

Recall that, when $N(A)$ is non-trivial,

$Ax = b$ may have infinitely many solutions
(or zero solutions)



Namely, if x_0 satisfies $Ax_0 = b$,

and $x_n \in N(A)$, $A(x_0 + x_n) = b$.

In fact, the solution set is $x_0 + N(A)$.

In this situation, A^+ provides the minimum norm solution :

$$\|A^T b\|_2 \leq \|x_0 + x_n\| \quad \forall x_n \in N(A)$$

$$(\text{or, } A^T b = \underset{x \in x_0 + N(A)}{\operatorname{argmin}} \|x\|_2)$$

This is because $A^T b = \operatorname{proj}_{R(A^T)}(x_0 + N(A))$ In other words,

$$\forall x^* \in x_0 + N(A), \quad x^* = A^T b + z, \quad \text{where } z \in N(A)$$

unique representation of the form $R(A^T) + N(A)$

Calculus and Optimization

- When a function f is convex,

$$\underset{x \in \mathbb{R}^n}{\operatorname{argmin}} f(x) = \text{zero } (\nabla f)$$

Thus, to minimize $\|Ax - b\|_2^2$, only need to find the zeros of $\nabla(\|Ax - b\|_2^2)$.

- Note that $\|x\|_2^2 = x^T x = \sum_{i=1}^n x_i^2 \quad \forall x \in \mathbb{R}^n$

$$\text{Therefore, } \|Ax - b\|_2^2 = (Ax - b)^T (Ax - b) = x^T A^T A x - 2b^T A x + b^T b$$

- Differentiation Rules

$$(1) \quad \nabla_x (v^T x) = \nabla_x \left(\sum_{i=1}^n v_i x_i \right) = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} = v$$

$$(2) \quad \nabla_x (x^T A x) = \nabla_x \left(\sum_{i=1}^n \sum_{j=1}^n x_i A_{ij} x_j \right)$$

Let's do ∂_{x_i} for instance:

$$\partial_{x_i} \left(\sum_{i,j} x_i A_{ij} x_j \right) = \partial_{x_i} \left(\sum_{j \geq 1} x_i A_{ij} x_j + \sum_{i \geq 2} x_i A_{i1} x_i + x_i A_{ii} x_i \right)$$

$$= \sum_{j \geq 1} A_{ij} x_j + \sum_{i \geq 2} A_{i1} x_i + 2A_{ii} x_i$$

$$= \sum_{j=1}^n A_{ij} x_j + \sum_{i=1}^n A_{i1} x_i$$

$$= Ax + A^T x = (A + A^T)x$$

In particular, when A is symmetric, $\nabla_x (x^T A x) = 2Ax$.

o Hence, $\nabla(\|Ax - b\|_2^2) = \nabla(x^T A^T A x - 2b^T A x)$ ($b^T A x = (A^T b)^T x$)

$$= 2A^T A x - 2A^T b \quad \text{--- (*)}$$

Zeros of (*) satisfy

$$A^T A x^* = A^T b$$

Recall that when A has lin. indep. columns, $A^T A$ is invertible
and $(A^T A)^{-1} A^T = A^T$

$\Rightarrow x^* = (A^T A)^{-1} A^T b = A^T b$ is the (unique) minimizer.

o What about $f(x) = \|Ax - b\|_2^2 + \lambda \|x\|_2^2$?

$$\nabla f = 2A^T A x_* - 2A^T b + 2\lambda x_* = 0$$

$$\Rightarrow (A^T A + \lambda I) x_* = A^T b.$$

$$\Rightarrow x_* = (A^T A + \lambda I)^{-1} A^T b.$$

* (Advanced) if $\|x\|_{\Sigma} := \sqrt{\sum_{i=1}^n \sigma_i x_i^2}$ (weighted norm)

$$\text{and } f_{\sigma}(x) = \|Ax - b\|_2^2 + \lambda \|x\|_{\sigma}^2,$$

$$\text{then } \frac{1}{2} \nabla_x f_{\sigma}(x) = A^T A x - A^T b + \lambda \Sigma x, \text{ where } \Sigma = \begin{pmatrix} \sigma_1 & & 0 \\ & \sigma_2 & \\ 0 & \ddots & \sigma_n \end{pmatrix}$$

$$\text{and } x^* = (A^T A + \lambda \Sigma)^{-1} A^T b$$

• Differentiating $\text{tr}(AX)$

Now consider $\begin{cases} A : m \times n \\ X : n \times m \end{cases}$

$$AX : m \times m$$

$$XA : n \times n$$

• Differentiation w.r.t. X

$$\nabla_X f(X) := \left[\partial_{X_{ij}} f(X) \right]_{ij} \quad (\text{similar to } \nabla_{\text{vector}})$$

$$\begin{aligned} \cdot \quad \text{tr}(AX) &= \sum_{i=1}^m (AX)_{ii} = \sum_{i=1}^m \left(\sum_{j=1}^n A_{ij} X_{ji} \right) \\ &= \sum_{i=1}^m \sum_{j=1}^n A_{ij} X_{ji} \end{aligned}$$

$$\begin{aligned} \cdot \quad \nabla_X \text{tr}(AX) &= \left[\partial_{X_{ij}} \text{tr}(AX) \right]_{ij} \\ &= \left[A_{ji} \right]_{ij} = A^T \quad \square \end{aligned}$$