Linear Algebra Review

Fundamental subspaces of a matrix

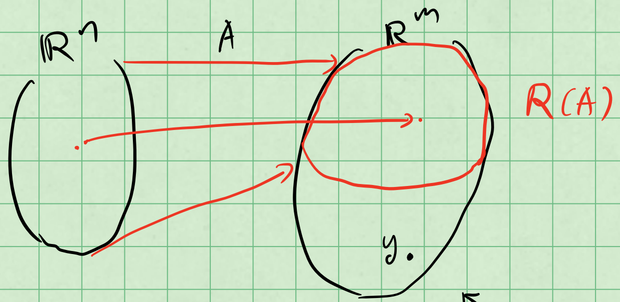
A : $m \times n$ matrix

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

$$A : \mathbb{R}^m \longrightarrow \mathbb{R}^n$$
$$x \longmapsto Ax$$

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

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



There is no $x \in \mathbb{R}^m$ s.t. $Ax = y$

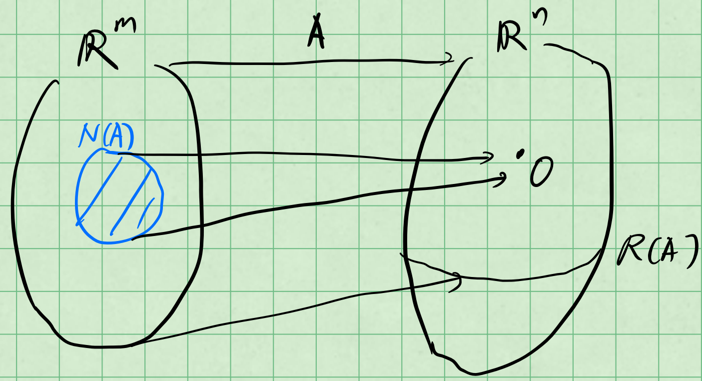
Using matrix algebra, we can easily check

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

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

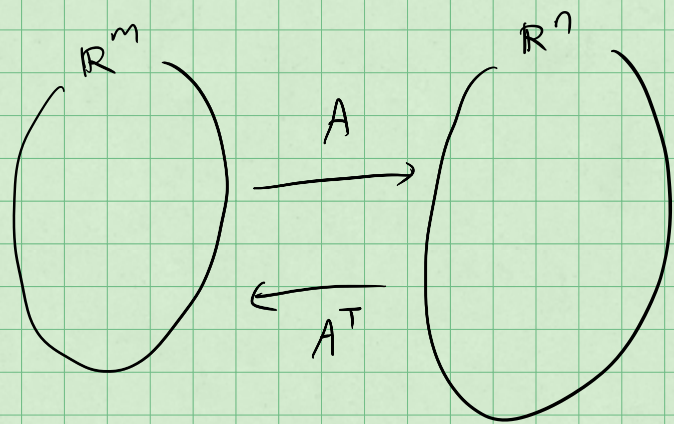
(2) We are also interested in the "kernel" of A, or "zeros" of A:

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



(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:

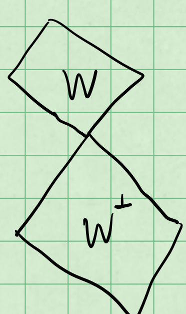
$W \subseteq \mathbb{R}^m$, $\quad W^\perp = \{ x \in \mathbb{R}^m : \langle x, y \rangle = 0 \ \forall y \in W \}$

subspace

x is orthogonal to W

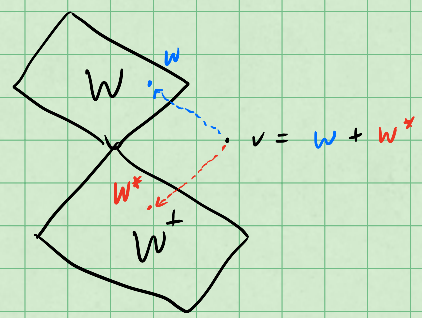
collect all such x's

$\mathbb{R}^m$ =



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

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



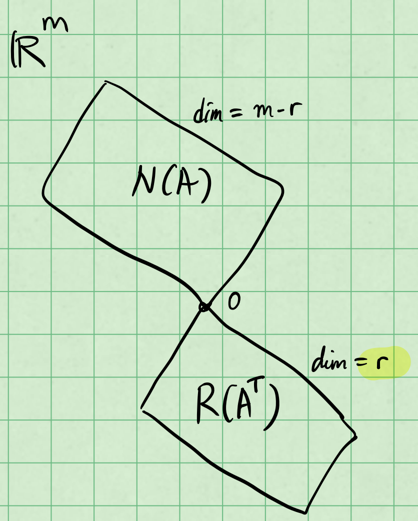
Interestingly, the four "fundamental" subspaces

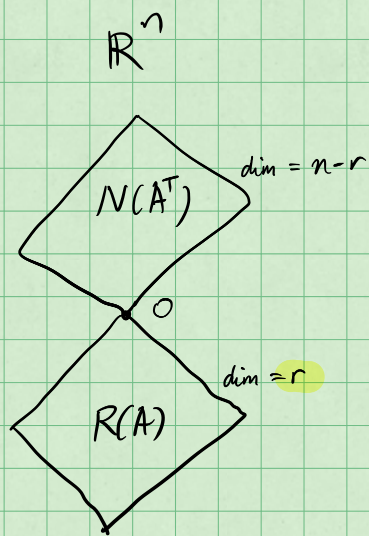
$R(A)$, $N(A)$, $R(A^T)$ and $N(A^T)$ have some orthogonality relations

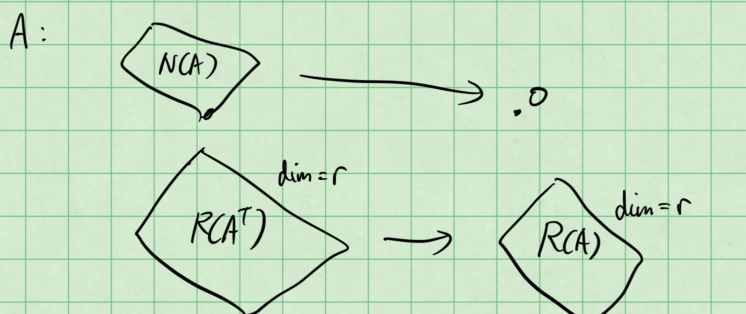
$$R(A^T) = N(A)^\perp$$

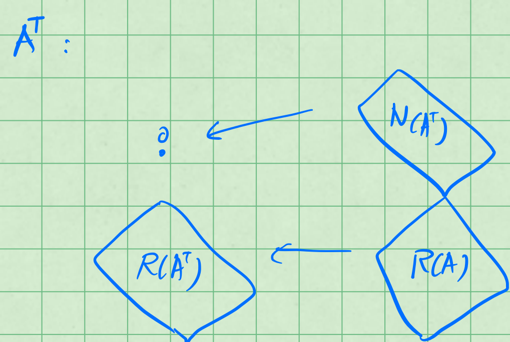
$$R(A)^\perp = N(A^T)$$

(The former is equiv. to the latter.)
Consider $A \leftarrow A^T$





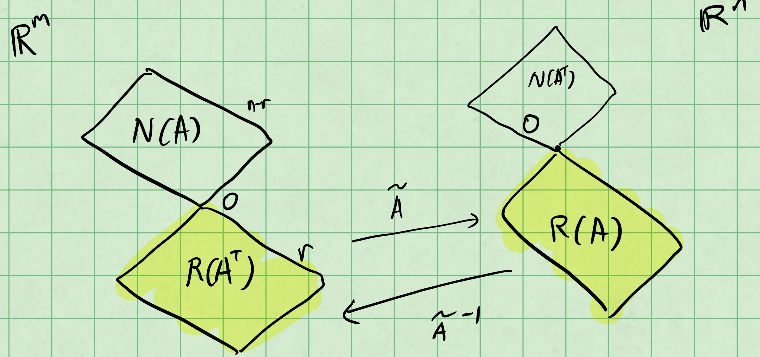




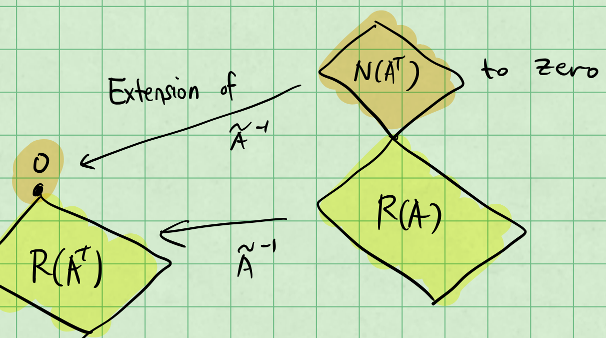
A is one-to-one and onto

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

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



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



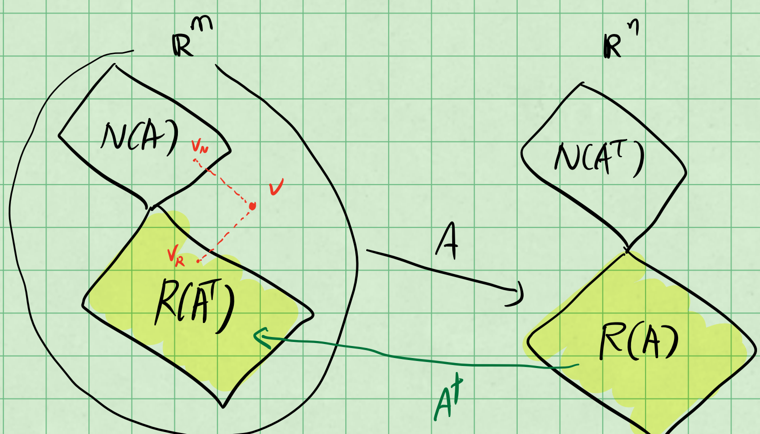
We call this extension of $\tilde{A}^{-1}$, a pseudo inverse of A

(Moore - Penrose inverse)

and denote by A^{+}.

(5) A^{+} has many interesting properties.

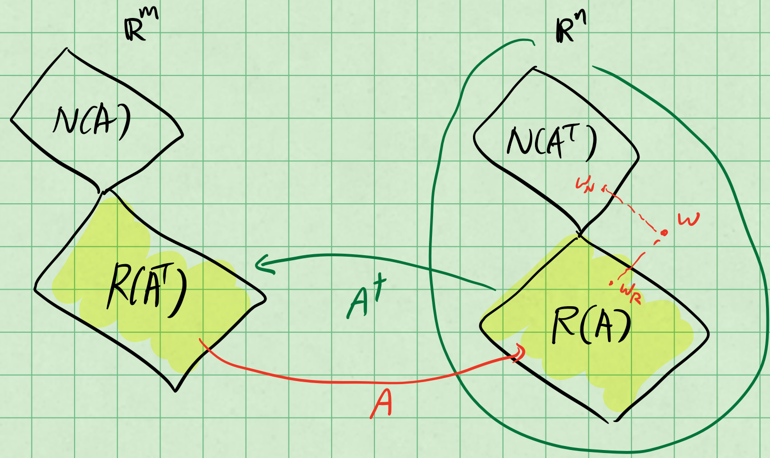
Most of them can be understood using the previous diagram.



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

$$A^{+}Av = A^{+}(Av) = A^{+}(Av_R) = v_R$$

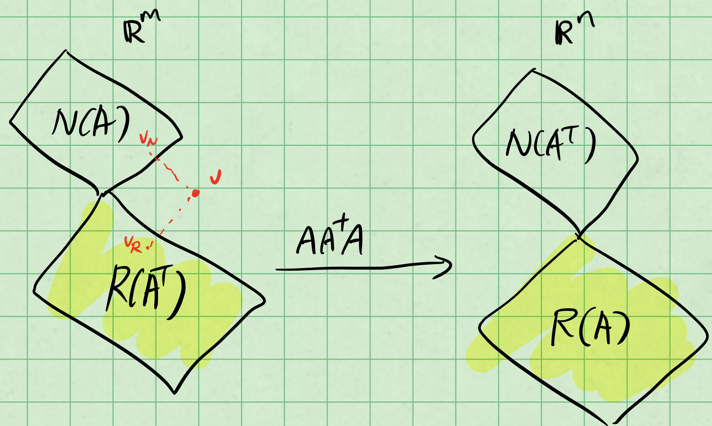
i.e., $A^{+}A$ is a projection onto $R(A^T)$ (in $\mathbb{R}^m$)



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

(NOTE: AA^+ is not the identity matrix!)

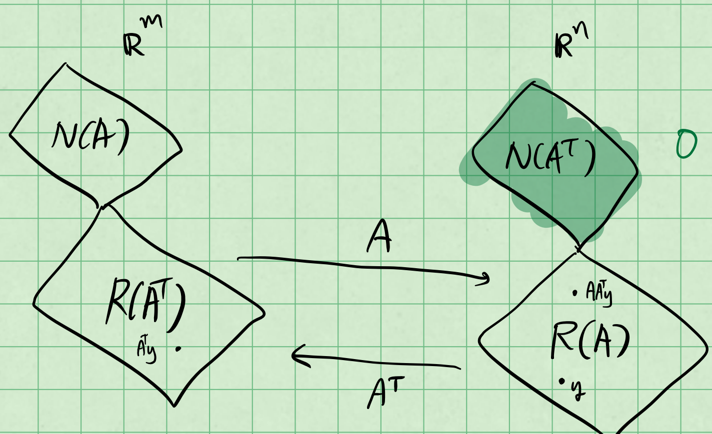
What about AA^+A ?



$$AA^+A = A \circ \mathrm{Proj}_{R(A^T)} = A$$

similarly, $A^+AA^+ = A^+$.

Finally, consider AA^T and A^TA:



$$\begin{cases} AA^T\big|_{R(A)} = A\big|_{R(A^T)} \, A^T\big|_{R(A)} \\ AA^T\big|_{N(A^T)} = 0 \end{cases}$$

$$\Downarrow$$

$(AA^T)^{-1}$ exists if $N(A^T) = 0$

$\Leftrightarrow$ A lin. indep. rows

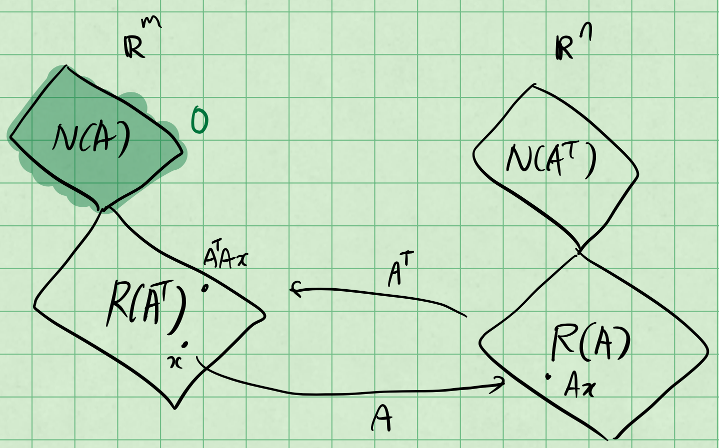
Then,

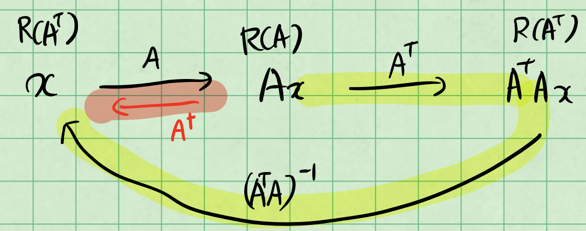
$$\begin{tikzcd} R(A) & R(A^T) & R(A) \\ y \arrow[r, "A^T"] & A^Ty \arrow[r, "A"] & AA^Ty \arrow[l, shift left=2, "A^+"] \arrow[ll, bend left=60, "(AA^T)^{-1}"] \end{tikzcd}$$

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

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

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

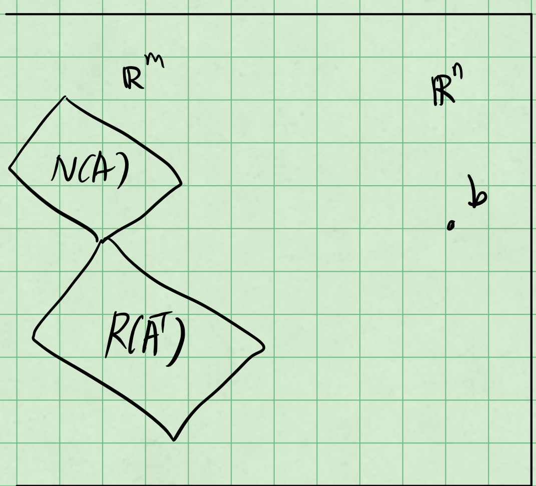




(A^TA) has an inverse!

☆ 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 :

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

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

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

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

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

Calculus and Optimization

- When a function f is convex,

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

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

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

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

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

Let's do ∂_{x_1} for instance:

$$\partial_{x_1}\left(\sum_{i,j} x_i A_{ij} x_j\right) = \partial_{x_1}\left(\sum_{j>1} x_1 A_{1j} x_j + \sum_{i>2} x_i A_{i1} x_1 + x_1 A_{11} x_1\right)$$

$$= \sum_{j>1} A_{1j} x_j + \sum_{i>2} A_{i1} x_i + 2A_{11} x_1$$

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

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

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

- Hence, $\nabla(\|Ax-b\|_2^2) = \nabla(x^TA^TAx - 2b^TAx)$ $\quad (b^TAx = (A^Tb)^Tx)$

$$= 2A^TAx - 2A^Tb \quad \text{——} \quad (*)$$

Zeros of $(*)$ satisfy

$$A^TAx^* = A^Tb$$

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

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

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

$$\nabla f = 2A^TAx_* - 2A^Tb + 2\lambda x_* = 0$$

$$\Rightarrow \quad (A^TA+\lambda I)\,x_* = A^Tb.$$

$$\Rightarrow \quad x_* = (A^TA+\lambda I)^{-1}A^Tb.$$

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

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

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

and $x^* = (A^TA+\lambda\Sigma)^{-1}A^Tb$ $\square$

- Differentiating $tr(AX)$

Now consider $\begin{cases} A : & m \times n \\ X : & n \times m \end{cases}$ $\qquad AX : m \times m$, $\quad XA : n \times n$

- Differentiation w.r.t. X

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

- $$tr(AX) = \sum_{i=1}^{m} (AX)_{ii} = \sum_{i=1}^{m} \left(\sum_{j=1}^{n} A_{ij} X_{ji} \right)$$
$$= \sum_{i=1}^{m} \sum_{j=1}^{n} A_{ij} X_{ji}$$

- $$\nabla_X tr(AX) = \left[\partial_{X_{ij}} tr(AX) \right]_{ij}$$
$$= \left[A_{ji} \right]_{ij} = A^T \quad \square$$