10/15/2021

Last time:

→ Proof of Fundamental Thm of Linear algebra

Thm (FTLA) Let $\underline{A} \in \mathbb{R}^{m \times n}$.

$\mathbb{R}^n = R(\underline{A}^T) \oplus N(\underline{A})$, $R(\underline{A}^T) \perp N(\underline{A})$

$\mathbb{R}^m = R(\underline{A}) \oplus N(\underline{A}^T)$, $R(\underline{A}) \perp N(\underline{A}^T)$.

$\dim R(\underline{A}^T) = \dim R(\underline{A}) = r \leftarrow$ "rank".

$\dim N(\underline{A}) = n - r$, $\dim N(\underline{A}^T) = m - r$.

Quick remark: There was a question if a direct sum of subspaces $\mathcal{U} \oplus \mathcal{V}$ necessarily implies that $\mathcal{U} \perp \mathcal{V}$...

Answer: You can consider a direct →

Sum of <u>NON-orthogonal</u> subspaces.

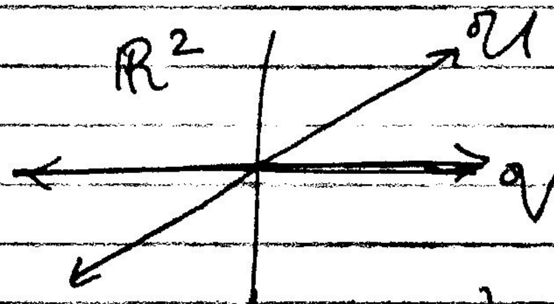
$\mathcal{U}$ and $\mathcal{V}$ and still have:

if $\underline{u}_1 + \underline{v}_1 = \underline{u}_2 + \underline{v}_2$, then $\underline{u}_1 = \underline{u}_2$ & $\underline{v}_1 = \underline{v}_2$,

for $\underline{u}_1, \underline{u}_2 \in \mathcal{U}$ and $\underline{v}_1, \underline{v}_2 \in \mathcal{V}$.

<u>Ex</u>: Consider $\mathcal{U} = \text{span}\left(\begin{pmatrix}1\\1\end{pmatrix}\right)$ and

$\mathcal{V} = \text{span}\left(\begin{pmatrix}1\\0\end{pmatrix}\right)$.



Suppose we have $\underline{u}_1, \underline{u}_2 \in \text{span}\left(\begin{pmatrix}1\\1\end{pmatrix}\right)$

and $\underline{v}_1, \underline{v}_2 \in \text{span}\left(\begin{pmatrix}1\\0\end{pmatrix}\right)$.

$\Rightarrow \underline{u}_1 = \gamma_1 \begin{pmatrix}1\\1\end{pmatrix}, \quad \underline{u}_2 = \gamma_2 \begin{pmatrix}1\\1\end{pmatrix}$

$\underline{v}_1 = \beta_1 \begin{pmatrix}1\\0\end{pmatrix}, \quad \underline{v}_2 = \beta_2 \begin{pmatrix}1\\0\end{pmatrix},$

and $\underline{u}_1 + \underline{v}_1 = \underline{u}_2 + \underline{v}_2$.

$$\Rightarrow \gamma_1 \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \beta_1 \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \gamma_2 \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \beta_2 \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

$$\Rightarrow$$

$$(\gamma_1 - \gamma_2) \begin{pmatrix} 1 \\ 1 \end{pmatrix} + (\beta_1 - \beta_2) \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Since $\left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\}$ are linearly independent, we must have:

$$\gamma_1 - \gamma_2 = \beta_1 - \beta_2 = 0$$

$\Rightarrow$ $\underline{u}_1 = \underline{u}_2$ and $\underline{v}_1 = \underline{v}_2$.

Let's continue with the proof (sketch) of FTLA:

want to see $\mathbb{R}^m = R(\underline{\underline{A}}) \oplus N(\underline{\underline{A}}^T)$.

Since $R(\underline{\underline{A}})$ and $N(\underline{\underline{A}}^T)$ are separately subspaces of $\mathbb{R}^m$, we must have

$$\mathbb{R}^m \supset R(\underline{\underline{A}}) \oplus N(\underline{\underline{A}}^T).$$

To see the reverse containment,

take $\underline{w} \in \mathbb{R}^m$.

Since $\dim R(\underline{A}) = r$ and $\dim N(\underline{A}^T) = m - r$,

we have bases:

$\{y_1, y_2, \ldots, y_r\} \leftarrow$ basis for $R(\underline{A})$.

$\{\underline{z}_1, \underline{z}_2, \ldots, \underline{z}_{m-r}\} \leftarrow$ basis for $N(\underline{A}^T)$

Note, the collection of all of these vectors:

$\{y_1, \ldots, y_r, \underline{z}_1, \ldots, \underline{z}_{m-r}\}$ are

linearly independent since $y_i^T \underline{z}_j = 0$,

for $i = 1, \ldots, r$ and $j = 1, \ldots, m-r$.

Question: Can we express $\underline{w} \in \mathbb{R}^m$ as:

$\underline{w} = \gamma_1 y_1 + \cdots + \gamma_r y_r + \beta_1 \underline{z}_1 + \cdots + \beta_{m-r} \underline{z}_{m-r}$

?

Yes! Because this is equivalent to the linear system:

$$\underbrace{\begin{pmatrix} | & & | & | & & | \\ \underline{y}_1 & \cdots & \underline{y}_r & \underline{z}_1 & \cdots & \underline{z}_{m-r} \\ | & & | & | & & | \end{pmatrix} \begin{pmatrix} \gamma_1 \\ \vdots \\ \gamma_r \\ \beta_1 \\ \vdots \\ \beta_{m-r} \end{pmatrix}} = \underbrace{\begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_m \end{pmatrix}}_{= \underline{w}}$$

$\rightarrow$ and this matrix is invertible.

So, we have $\underline{w} = \underline{y} + \underline{z}$ where

$\underline{y} \in R(\underline{A})$ and $\underline{z} \in N(\underline{A}^T)$.

Last thing to check: Suppose we have $\tilde{\underline{y}} \in R(\underline{A})$ and $\tilde{\underline{z}} \in N(\underline{A}^T)$ so that $\underline{w} = \tilde{\underline{y}} + \tilde{\underline{z}}$.

Then, $\underline{y} + \underline{z} = \underline{w} = \tilde{\underline{y}} + \tilde{\underline{z}}$

$\Leftrightarrow \quad \underline{y} - \tilde{\underline{y}} = \underline{z} - \tilde{\underline{z}}$.

Since we are working with subspaces,

$\underline{y} - \tilde{\underline{y}} \in R(\underline{A})$ and $\underline{z} - \tilde{\underline{z}} \in N(\underline{A}^T)$, so

$(\underline{y}-\tilde{\underline{y}})^T(\underline{z}-\tilde{\underline{z}}) = 0$ because $R(\underline{A}) \perp N(\underline{A}^T)$.

Also, $\|\underline{y}-\tilde{\underline{y}}\|_2^2 = (\underline{y}-\tilde{\underline{y}})^T(\underline{y}-\tilde{\underline{y}})$

$$= (\underline{y}-\tilde{\underline{y}})^T(\underline{z}-\tilde{\underline{z}}) = 0$$

(from above).

So, $\underline{y}=\tilde{\underline{y}}$ and $\underline{z}=\tilde{\underline{z}}$.

Thus, $\underline{w}=\underline{y}+\underline{z}$ is the unique way to express $\underline{w}$ as a sum of a vector in $R(\underline{A})$ and $N(\underline{A}^T)$.

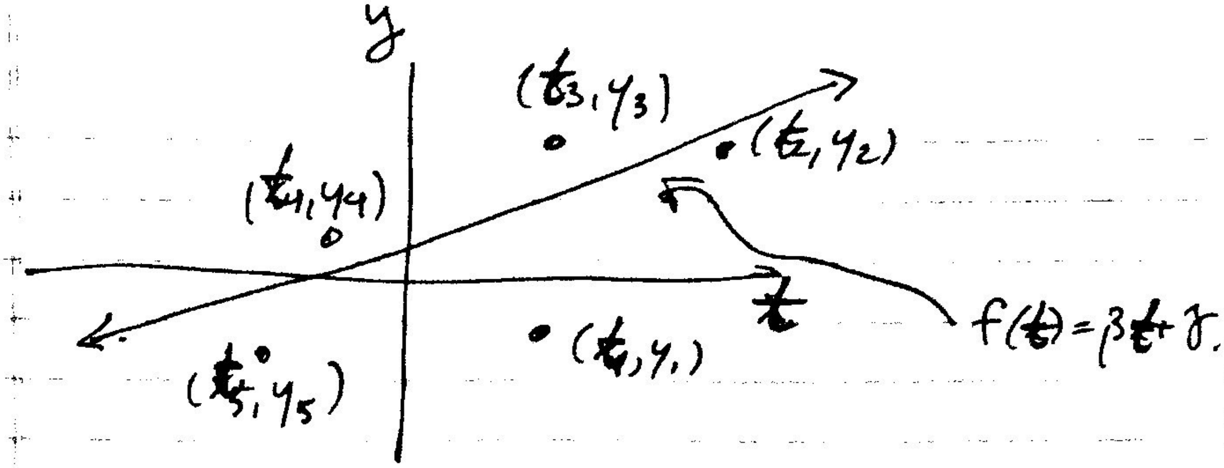
$\Rightarrow \underline{w} \in R(\underline{A}) \oplus N(\underline{A}^T)$

$\to \mathbb{R}^m \subset R(\underline{A}) \oplus N(\underline{A}^T)$.

Important application (of FTLA)

(Linear) Least Squares.

Setup: We are given data (t_i, y_i), $i=1,\ldots,m$.



Goal: Find the line $f(t) = \beta t + \gamma$ that "best fits" the data (t_i, y_i), $i = 1, \dots, m$.

For each $i = 1, \dots, m$, we want

$$y_i \approx f(t_i) = \beta t_i + \gamma,$$

which can be written in matrix form as.

$$\underbrace{\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}}_{\underline{b}} \approx \underbrace{\begin{pmatrix} t_1 & 1 \\ t_2 & 1 \\ \vdots & \vdots \\ t_m & 1 \end{pmatrix}}_{\underline{\underline{A}}} \underbrace{\begin{pmatrix} \beta \\ \gamma \end{pmatrix}}_{= \underline{x}}.$$

We can make "$\approx$" more formal by formulating as an optimization problem:

$$\min_{\underline{x}} \| \underline{b} - \underline{\underline{A}}\, \underline{x} \|_2^2 \doteq \min_{\underline{x}} r(\underline{x}).$$

At a minimum $\underline{x}^* = \begin{pmatrix} \beta^* \\ \gamma^* \end{pmatrix}$, we will have $\nabla r(\underline{x}^*) = \underline{0}$.

We can compute:

$$r(\underline{x}) = \underline{x}^T \underline{\underline{A}}^T \underline{\underline{A}}\, \underline{x} - 2\, \underline{x}^T \underline{\underline{A}}^T \underline{b} + \underline{b}^T \underline{b}$$

$$\nabla r(\underline{x}) = 2 \underline{\underline{A}}^T \underline{\underline{A}}\, \underline{x} - 2 \underline{\underline{A}}^T \underline{b},$$ so

setting $\nabla r(\underline{x}^*) = 0$ gives:

$$\underline{\underline{A}}^T \underline{\underline{A}}\, \underline{x}^* = \underline{\underline{A}}^T \underline{b}.$$

These are called the <u>normal</u> equations.