17. Prop 1.2: $\boxed{\text{proj}_V b = \underbrace{A(A^TA)^{-1}A^T}_{P_V \text{ projection matrix}} b}$ where $A = \begin{bmatrix} | & & | \\ v_1 & \cdots & v_n \\ | & & | \end{bmatrix}$ for a basis $v_1, \dots, v_n$ of V.

Pf: $p = \text{proj}_V b \Rightarrow p = A\bar{x}$, $\quad \bar{x}$ = coeffs. on $v_1,\dots,v_n$ needed to express $p \in V$

where $A^TA\bar{x} = A^Tb$ uniquely by previous Prop.

$$\Updownarrow$$

$$\Rightarrow A\bar{x} = A(A^TA)^{-1}A^Tb. \quad \square$$

E.g. Find a point $p = Ax$ closest to $b = \begin{bmatrix} 2 \\ 1 \\ 1 \\ -1 \end{bmatrix}$ for $A = \begin{bmatrix} 2 & 1 \\ 1 & 1 \\ 0 & 1 \\ 1 & -1 \end{bmatrix}$.

How? It's $A(A^TA)^{-1}A^Tb = \text{proj}_{C(A)} b$.

$$A^TA = \begin{bmatrix} 2 & 1 & 0 & 1 \\ 1 & 1 & 1 & -1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 1 \\ 0 & 1 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 6 & 2 \\ 2 & 4 \end{bmatrix} \Rightarrow (A^TA)^{-1} = \frac{1}{20}\begin{bmatrix} 4 & -2 \\ -2 & 6 \end{bmatrix}$$

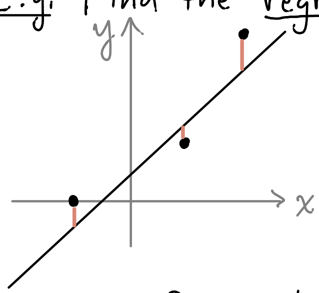
don't need to compute; why? $(A^TA)^T = A^TA^{TT}$ symmetric

$$A^Tb = \begin{bmatrix} 2 & 1 & 0 & 1 \\ 1 & 1 & 1 & -1 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \\ 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 4 \\ 5 \end{bmatrix}$$

$$(A^TA)^{-1}A^Tb = \frac{1}{10}\begin{bmatrix} 2 & -1 \\ -1 & 3 \end{bmatrix}\begin{bmatrix} 4 \\ 5 \end{bmatrix} = \frac{1}{10}\begin{bmatrix} 3 \\ 11 \end{bmatrix}$$

$$A\left(\qquad \right) = \frac{1}{10}\begin{bmatrix} 2 & 1 \\ 1 & 1 \\ 0 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} 3 \\ 11 \end{bmatrix} = \frac{1}{10}\left(\begin{bmatrix} 6 \\ 3 \\ 0 \\ 3 \end{bmatrix} + \begin{bmatrix} 11 \\ 11 \\ 11 \\ -11 \end{bmatrix}\right) = \frac{1}{10}\begin{bmatrix} 17 \\ 14 \\ 11 \\ -8 \end{bmatrix}$$

E.g. Find the regression line (least squares line) for the (data) points $\begin{bmatrix} -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \end{bmatrix}$.



Want (but can't have!) $y = ax + b$ with

$$\begin{aligned} 0 &= -1a + b \\ 1 &= 1a + b \\ 3 &= 2a + b \end{aligned} \iff \underbrace{\begin{bmatrix} -1 & 1 \\ 1 & 1 \\ 2 & 1 \end{bmatrix}}_{A} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 3 \end{bmatrix}.$$

Can get $y = \bar{a}x + \bar{b}$ so that $\bar{y} = \begin{bmatrix} \bar{y}_1 \\ \bar{y}_2 \\ \bar{y}_3 \end{bmatrix} = A\begin{bmatrix} \bar{a} \\ \bar{b} \end{bmatrix}$ as close to as possible.

looking for $\begin{bmatrix} \bar{a} \\ \bar{b} \end{bmatrix}$ but closeness condition is on the y's.

$\bar{y} = A\begin{bmatrix} \bar{a} \\ \bar{b} \end{bmatrix}$ is the least squares solution,

minimizes $\|\varepsilon\|$ for $\varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{bmatrix} = \begin{bmatrix} \bar{y}_1 - 0 \\ \bar{y}_2 - 1 \\ \bar{y}_3 - 3 \end{bmatrix}$ error vector

we don't actually care about these!

no leading "A"

Solve: $\begin{bmatrix} \bar{a} \\ \bar{b} \end{bmatrix} = (A^TA)^{-1}A^T\begin{bmatrix} 0 \\ 1 \\ 3 \end{bmatrix}$.

$$(A^TA)^{-1} = \left(\begin{bmatrix}-1 & 1 & 2\\ 1 & 1 & 1\end{bmatrix}\begin{bmatrix}-1 & 1\\ 1 & 1\\ 2 & 1\end{bmatrix}\right)^{-1} = \begin{bmatrix}6 & 2\\ 2 & 3\end{bmatrix}^{-1} = \frac{1}{14}\begin{bmatrix}3 & -2\\ -2 & 6\end{bmatrix}$$

$$(A^TA)^{-1}A^T\begin{bmatrix}0\\1\\3\end{bmatrix} = \frac{1}{14}\begin{bmatrix}3 & -2\\ -2 & 6\end{bmatrix}\begin{bmatrix}-1 & 1 & 2\\ 1 & 1 & 1\end{bmatrix}\begin{bmatrix}0\\1\\3\end{bmatrix}$$

$$= \frac{1}{14}\begin{bmatrix}3 & -2\\ -2 & 6\end{bmatrix}\begin{bmatrix}7\\4\end{bmatrix}$$

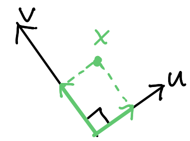
$$= \frac{1}{14}\begin{bmatrix}13\\10\end{bmatrix} \Rightarrow \text{regression line is } y = \frac{13}{14}x + \frac{5}{7}$$ (check against reality)

<u>E.g.</u> $\dim V = 1 \Rightarrow A = a \in \mathbb{R}^m$

$$\Rightarrow P_V = a(a^Ta)^{-1}a^T = a\frac{1}{\|a\|^2}a^T = \frac{1}{\|a\|^2}aa^T = \frac{aa^T}{a^Ta}.$$

from early HW

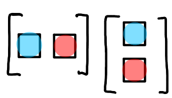
<u>Recall</u> $u \perp v$ under $\cdot$ and $x \in \operatorname{span}(u,v) \Rightarrow x = \operatorname{proj}_u x + \operatorname{proj}_v x$



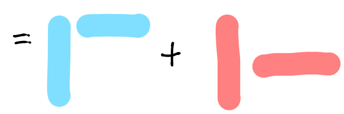
Rewrite: $u \perp v$ and $V = \operatorname{span}(u,v) \Rightarrow \operatorname{proj}_V = \operatorname{proj}_u + \operatorname{proj}_v$

$$= \frac{1}{\|u\|^2}uu^T + \frac{1}{\|v\|^2}vv^T$$

special case of product of block matrices:



$$= \begin{bmatrix} | & | \\ \frac{u}{\|u\|} & \frac{v}{\|v\|} \\ | & | \end{bmatrix}\begin{bmatrix} — u^T/\|u\| — \\ — v^T/\|v\| — \end{bmatrix}$$



<u>General</u>: $v_1, \dots, v_n$ orthogonal $\Rightarrow P_V = \underbrace{\begin{bmatrix} | & & | \\ v_1 & \cdots & v_n \\ | & & | \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} \frac{1}{\|v_1\|^2} & & \\ & \ddots & \\ & & \frac{1}{\|v_n\|^2} \end{bmatrix}}_{(A^TA)^{-1}} \underbrace{\begin{bmatrix} — v_1^T — \\ \vdots \\ — v_n^T — \end{bmatrix}}_{A^T}$

can be taken as def. of "orthogonal set"

Q. $(A^TA)^{-1} = \cancel{?}\begin{bmatrix} \|v_1\|^2 & & \\ & \ddots & \\ & & \|v_n\|^2 \end{bmatrix}^{-1}$

<u>Algorithm 2.4</u> (Gram-Schmidt)

Input: basis $v_1, \dots, v_n$ for inner product space V

Output: orthogonal basis of V

Initialize: $w_1 = v_1$

$W_1 = \operatorname{span}(w_1)$

$i = 1$

While: $i < n$

Do: $w_{i+1} = v_{i+1} - \mathrm{proj}_{W_i} v_{i+1}$

$W_{i+1} = \mathrm{span}(w_1, \dots, w_{i+1})$ $\left(= \mathrm{span}(v_1, \dots, v_{i+1})!\right)$ but think in terms of the w_i

$i \leftarrow i+1$

Return: $w_1, \dots, w_n$

To get orthonormal basis $q_1, \dots, q_n$ set $q_i = \frac{w_i}{\|w_i\|}$

Return: $q_1, \dots, q_n$

E.g. $v_1 = \begin{bmatrix}1\\1\\1\\1\end{bmatrix}, v_2 = \begin{bmatrix}3\\1\\-1\\1\end{bmatrix}, v_3 = \begin{bmatrix}1\\1\\3\\3\end{bmatrix} \Rightarrow w_1 = \begin{bmatrix}1\\1\\1\\1\end{bmatrix}$

$$w_2 = \begin{bmatrix}3\\1\\-1\\1\end{bmatrix} - \frac{(3,1,-1,1)\cdot(1,1,1,1)}{\|(1,1,1,1)\|^2}\begin{bmatrix}1\\1\\1\\1\end{bmatrix} = \begin{bmatrix}3\\1\\-1\\1\end{bmatrix} - \frac{4}{4}\begin{bmatrix}1\\1\\1\\1\end{bmatrix} = \begin{bmatrix}2\\0\\-2\\0\end{bmatrix}$$

$$w_3 = v_3 - \mathrm{proj}_{W_2} v_3 = v_3 - \left(\mathrm{proj}_{w_1} v_3 + \mathrm{proj}_{w_2} v_3\right)$$
$$= v_3 - \left(\frac{v_3 \cdot w_1}{\|w_1\|^2} w_1 + \frac{v_3 \cdot w_2}{\|w_2\|^2} w_2\right)$$
$$= \begin{bmatrix}1\\1\\3\\3\end{bmatrix} - \frac{8}{4}\begin{bmatrix}1\\1\\1\\1\end{bmatrix} - \frac{-4}{8}\begin{bmatrix}2\\0\\-2\\0\end{bmatrix}$$
$$= \begin{bmatrix}1\\1\\3\\3\end{bmatrix} - \begin{bmatrix}2\\2\\2\\2\end{bmatrix} + \begin{bmatrix}1\\0\\-1\\0\end{bmatrix} = \begin{bmatrix}0\\-1\\0\\1\end{bmatrix}$$

orthonormal basis: columns of

$$\underset{A}{\begin{bmatrix}1&3&1\\1&1&1\\1&-1&3\\1&1&3\end{bmatrix}} = \underset{Q}{\begin{bmatrix}1/2&1/\sqrt{2}&0\\1/2&0&-1/\sqrt{2}\\1/2&-1/\sqrt{2}&0\\1/2&0&1/\sqrt{2}\end{bmatrix}} \underset{R}{\begin{bmatrix}2&2&4\\0&2\sqrt{2}&-\sqrt{2}\\0&0&\sqrt{2}\end{bmatrix}}$$

$q_1 \quad q_2 \quad q_3$

By construction, each w_i (or q_i) lies in span of $v_1, \dots, v_i$. Thus

$$\underset{Q}{\begin{bmatrix}|&&|\\q_1&\cdots&q_n\\|&&|\end{bmatrix}} = \underset{A}{\begin{bmatrix}|&&|\\v_1&\cdots&v_n\\|&&|\end{bmatrix}} \underset{U}{\begin{bmatrix}*&*&\cdots&*\\0&*&&*\\\vdots&0&\ddots&\vdots\\0&0&\cdots&*\end{bmatrix}}. \quad U^{-1} = R \Rightarrow QR = A$$

U upper-triangular

<u>QR decomposition</u> expresses A as product of (matrix with orthonormal cols) and (upper-triangular matrix)

E.g. $A = QR$

Questions: 1. $r_{ij} = $ ~~?~~ coeff. of q_i in $v_j = q_i \cdot v_j$ since $q_1, \dots, q_n$ orthonormal

2. What's P_V in terms of Q? $P_V = QQ^T = \mathrm{proj}_{q_1} + \cdots + \mathrm{proj}_{q_n}$!

Pf: $P_V = A(A^TA)^{-1}A^T = QR(R^T\cancel{Q^T}\cancel{Q}R)^{-1}R^TQ^T$

$= QR(R^TR)^{-1}R^TQ^T = Q\cancel{R}\cancel{R^{-1}}\cancel{(R^T)^{-1}}\cancel{R^T}Q^T = QQ^T$. $\square$

Note: $R^{-1} = U$