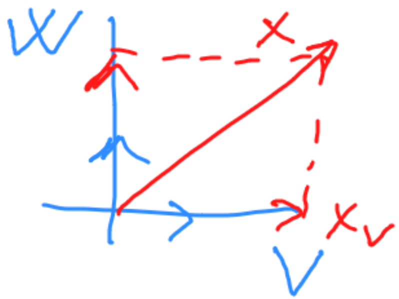
Projection. Least squares method.

If $V \perp W$ in $\mathbb{R}^m$ and if $\dim V + \dim W = m$, then every x in $\mathbb{R}^m$ can written (uniquely) as

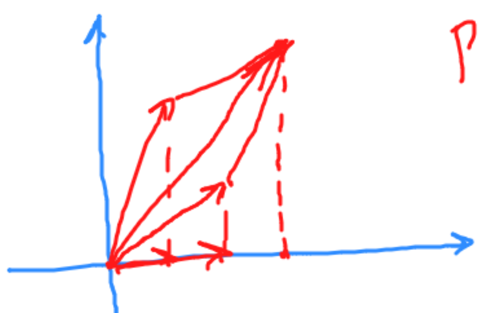
$$x = x_v + x_w,$$ where x_v is V, x_w in W.



Question: how to find x_v, x_w given x?

x_v is called the projection of x onto V.

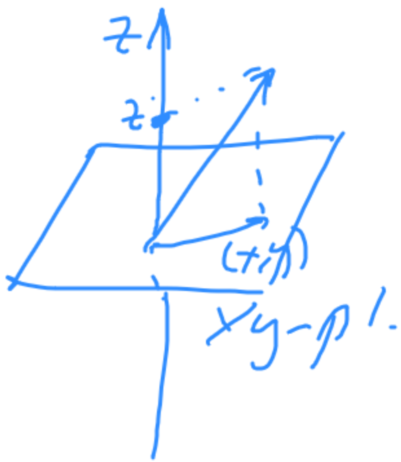
Main idea: projection is linear!



projection of $x+y$ =

= projection of x + proj. of y.

<u>Ex.</u> $\mathbb{R}^3$, take $V = z$-axis, $W = xy$-plane



$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} \to \begin{bmatrix} 0 \\ 0 \\ z \end{bmatrix} \qquad \begin{bmatrix} 0 \\ 0 \\ z \end{bmatrix} = \begin{bmatrix} 0 & & \\ & 0 & \\ & & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

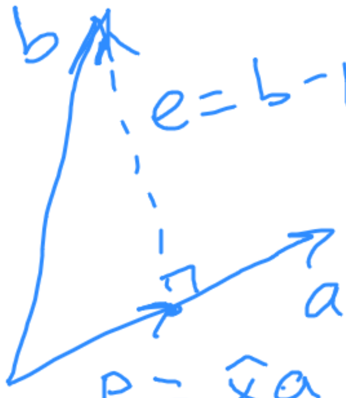
proj. onto V $\qquad p = P \quad b$

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} \to \begin{bmatrix} x \\ y \\ 0 \end{bmatrix} \qquad \begin{bmatrix} x \\ y \\ 0 \end{bmatrix} = \begin{bmatrix} 1 & & \\ & 1 & \\ & & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

proj. onto W

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

Start with projection onto $\langle a \rangle$. (1-dim subspace)
$\neq 0$



$e = b - p$ condition $e \perp a$:

$$a \cdot (b - p) = 0$$
$$a \cdot (b - \hat{x} a) = 0$$
$$a \cdot b = a \cdot a \, \hat{x}$$
$$\hat{x} = \frac{a \cdot b}{a \cdot a} = \frac{a^T b}{a^T a}$$

$$\begin{bmatrix} a^T \end{bmatrix} \begin{bmatrix} | \end{bmatrix}$$

$p = \hat{x} a$
scalar vector

Now: $p = \hat{x} a = \dfrac{a^T b}{a^T a} a$. Want: $p = P b$ ¿?

$$p = a \cdot \frac{a^T b}{a^T a} = \frac{a a^T b}{a^T a} = \frac{a a^T}{a^T a} b = P b$$ where

$$P = \frac{a a^T}{a^T a}.$$

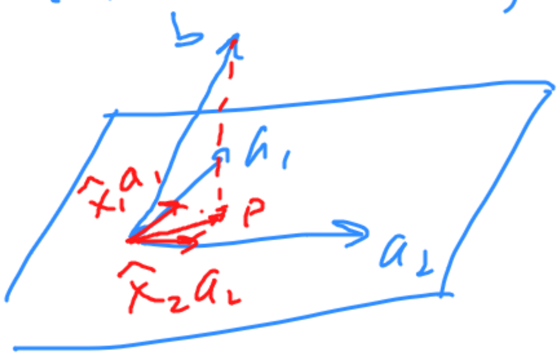
(rank 1 matrix)

number

$$\begin{bmatrix} \ \end{bmatrix} \begin{bmatrix} \ \end{bmatrix} \to \begin{bmatrix} m \times m \end{bmatrix}$$

Suppose $a_1, a_2, \dots, a_n$ are lin. indep. vectors in $\mathbb{R}^m$.

Find the projection matrix P onto $\langle a_1, a_2, \dots, a_n \rangle$.



$p = \hat{x}_1 a_1 + \hat{x}_2 a_2$

$$A = \begin{bmatrix} | & | & & | \\ a_1 & a_2 & \cdots & a_n \\ | & | & & | \end{bmatrix}$$

$$\hat{x} = \begin{bmatrix} \hat{x}_1 \\ \vdots \\ \hat{x}_n \end{bmatrix}$$

$p = \hat{x}_1 a_1 + \dots + \hat{x}_n a_n$

$= A\hat{x}$

condition:

$e \perp \underline{a_1, a_2, \dots, a_n}$

$e = b - A\hat{x}$

$$\begin{cases} a_1^T(b - A\hat{x}) = 0 \\ a_2^T(b - A\hat{x}) = 0 \\ \cdots \\ a_n^T(b - A\hat{x}) = 0 \end{cases}$$

$A^T(b - A\hat{x}) = 0$

so we want $\hat{x}$ s.t. $A^T(b - A\hat{x}) = 0$

$(n \times m)(m \times n)$

$$A^T b - A^T A \hat{x} = 0$$

$$A^T A \hat{x} = A^T b$$

$$\hat{x} = \underline{(A^T A)^{-1} A^T b}$$

if $a_1, \dots, a_n$ are indep. then matrix $A^T A$ has rank n (and it's $n \times n$ matrix), so it's <u>invertible</u>.

(recall: in 1-dim case $\hat{x} = \frac{a^T b}{a^T a}$)

$$p = A\hat{x} = \underbrace{A(A^T A)^{-1} A^T}_{P} b$$

and $p = Pb$.

Ex. $A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}$ $b = \begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix}$. (we are projecting $\begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix}$ onto the subspace $\left\langle \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix} \right\rangle$.)

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

$$A^TA = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 3 & 3 \\ 3 & 5 \end{bmatrix}$$

$$P = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 3 & 3 \\ 3 & 5 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix} =$$

$$= \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \cdot \frac{1}{6} \begin{bmatrix} 5 & -3 \\ -3 & 3 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix} =$$

$$= \frac{1}{6} \begin{bmatrix} 5 & 2 & -1 \\ 2 & 2 & 2 \\ -1 & 2 & 5 \end{bmatrix}$$

$$p = Pb = \begin{bmatrix} 5 \\ 2 \\ -1 \end{bmatrix}$$

$$A^TA\hat{x} = A^Tb$$

$$A^Tb = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 6 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} 3 & 3 \\ 3 & 5 \end{bmatrix} \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} = \begin{bmatrix} 6 \\ 6 \end{bmatrix}$$

$$\hat{x} = \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} = \begin{bmatrix} 5 \\ -3 \end{bmatrix}$$

$$p = A\hat{x} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 5 \\ -3 \end{bmatrix} = \begin{bmatrix} 5 \\ 2 \\ -1 \end{bmatrix}$$

$$\begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix} - \begin{bmatrix} 5 \\ 2 \\ -1 \end{bmatrix}$$

$$e = b - p = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} \perp \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}$$

what's projection of projection of b?

$$b \longrightarrow p \longrightarrow p$$

$$b \quad Pb \quad P^2 b$$

so $P^2 b = Pb$ for all b, so $P^2 = P$.

More interestingly, If any matrix P has the property $P^2 = P$ then it is some subspace's projection matrix.

which subspace? $C(P)$, orth. complement

$$C(P)^{\perp} = N(P).$$

If P is the ~~matrix~~ A projection onto some $C(A)$ then $I-P$ is the projection matrix onto $C(A)^{\perp}$.

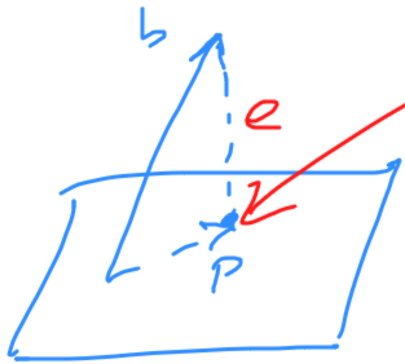
$$IX = \underbrace{x_A}_{\text{in } C(A)} + \underbrace{x_{\perp}}_{\text{in } C(A)^{\perp}} = Px + x_{\perp}$$

$$x_{\perp} = Ix - Px = (I-P)x$$

also: $(I-P)^2 = (I-P)(I-P) = I^2 - PI - IP + P^2$

$$= I - P - P + P = I - P$$

Geometrically:



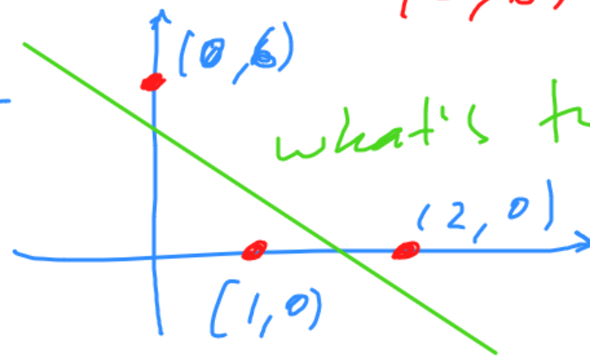
closest to b point of $\langle a_1, a_2, \ldots, a_n \rangle$

distance: $\|e\| = \|b - p\|$.

least squares ~~method~~ approximation (projection in disguise).

$(0,6)$, $(1,0)$, $(2,0)$ — approx. by $C+Dt$

Ex.



what's the "best" line?

Ideally:

$$C + D\cdot 0 = 6$$
$$C + D\cdot 1 = 0$$
$$C + D\cdot 2 = 0$$

$$\underbrace{\begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} C \\ D \end{bmatrix}}_{\hat{x}} = \underbrace{\begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix}}_{b}$$

want $A\hat{x} = b$ but cannot.

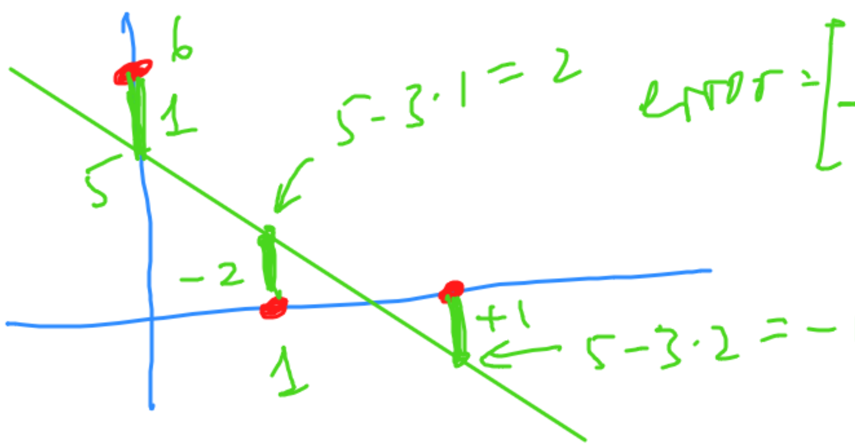
instead $\|A\hat{x} - b\|$ shortest possible

In other words: want a vector in $C(A)$ closest to b.

So we want: projection of b onto $C(A)$!

we computed: $\hat{x} = \begin{bmatrix} 5 \\ -3 \end{bmatrix} = \begin{bmatrix} C \\ D \end{bmatrix}$

so the best line is $5 - 3t$:



error: $\begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$

heights: $\begin{bmatrix} 5 \\ 2 \\ -1 \end{bmatrix}$

Also: we can minimize the "error" $\|(e_1, e_2, e_3)\|^2 =$

$$= e_1^2 + e_2^2 + e_3^2$$ by calculus:

This leads to $\begin{bmatrix} 3 & 3 \\ 3 & 5 \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} 6 \\ 0 \end{bmatrix}$.

In general: for data points (t_1, b_1), $(t_2, b_2), \ldots, (t_m, b_m)$ we want to minimize $\left\| A \begin{bmatrix} C \\ D \end{bmatrix} - b \right\|$

where $A = \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_m \end{bmatrix}$, $b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$

(want to fit $C + Dt_i = b_i$)

So we are doing projection of b onto $C(A) = \left\langle \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}, \begin{bmatrix} t_1 \\ t_2 \\ \vdots \\ t_m \end{bmatrix} \right\rangle$

$$A^TA = \begin{bmatrix} 1 & \cdots & 1 \\ t_1 & \cdots & t_m \end{bmatrix} \begin{bmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_m \end{bmatrix} = \begin{bmatrix} m & \sum t_i \\ \sum t_i & \sum t_i^2 \end{bmatrix}.$$

$$A^Tb = \begin{bmatrix} 1 & \cdots & 1 \\ t_1 & \cdots & t_m \end{bmatrix} \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix} = \begin{bmatrix} \sum b_i \\ \sum b_i t_i \end{bmatrix}$$

so eqn in C, D (for $C+Dt$) is:

$$\begin{bmatrix} m & \sum t_i \\ \sum t_i & \sum t_i^2 \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} \sum b_i \\ \sum b_i t_i \end{bmatrix}$$

especially easy case: when $\sum t_i = 0$.

useful: shift t_i by their average so that $\sum t_i = 0$.