Example) Let $W = \text{span}\left\{\begin{bmatrix}3\\0\\1\end{bmatrix}, \begin{bmatrix}0\\1\\0\end{bmatrix}\right\} \subseteq \mathbb{R}^3$ and $\vec{x} = \begin{bmatrix}0\\3\\10\end{bmatrix}$.

- Find the orthogonal projection of $\vec{x}$ onto W.

  Note that $\begin{bmatrix}3\\0\\1\end{bmatrix}, \begin{bmatrix}0\\1\\0\end{bmatrix}$ are an orthogonal basis of W. ($\because \begin{bmatrix}3\\0\\1\end{bmatrix}\cdot\begin{bmatrix}0\\1\\0\end{bmatrix}=0$)

  i) $\hat{x} = \dfrac{\vec{x}\cdot\vec{w_1}}{\vec{w_1}\cdot\vec{w_1}}\vec{w_1} + \dfrac{\vec{x}\cdot\vec{w_2}}{\vec{w_2}\cdot\vec{w_2}}\vec{w_2} = \dfrac{10}{10}\begin{bmatrix}3\\0\\1\end{bmatrix} + \dfrac{3}{1}\begin{bmatrix}0\\1\\0\end{bmatrix} = \begin{bmatrix}3\\3\\1\end{bmatrix}$

  ii) $\vec{x}^{\perp} = \vec{x} - \hat{x} = \begin{bmatrix}-3\\0\\9\end{bmatrix}$

  $\therefore \vec{x} = \hat{x} + \vec{x}^{\perp}$

- Definition Let $\vec{v_1}, \cdots, \vec{v_m}$ be an orthogonal basis of W, a subspace of $\mathbb{R}^n$.

  Then, the "projection map", $\pi_W : \mathbb{R}^n \to \mathbb{R}^n$ is given by

$$\vec{x} \mapsto \hat{x} = \left(\frac{\vec{x}\cdot\vec{v_1}}{\vec{v_1}\cdot\vec{v_1}}\right)\vec{v_1} + \cdots + \left(\frac{\vec{x}\cdot\vec{v_m}}{\vec{v_m}\cdot\vec{v_m}}\right)\vec{v_m}$$

  and π_W is linear!

  The matrix P representing π_W w.r.t. the standard basis is the corresponding "projection matrix".

$$\Rightarrow \pi_W(\alpha X + \beta Y) = \sum_{i=1}^{m} \frac{(\alpha X + \beta Y)\cdot V_i}{V_i \cdot V_i} V_i$$

$$= \sum_{i=1}^{m} \alpha \frac{X\cdot V_i}{V_i\cdot V_i} V_i + \beta \frac{Y\cdot V_i}{V_i \cdot V_i} V_i$$

$$= \alpha \pi_W(X) + \beta \pi_W(Y)$$

Example) Find the projection matrix P which corresponds to orthogonal projection onto $W = \text{span}\left\{\underset{w_1}{\begin{bmatrix}3\\0\\1\end{bmatrix}}, \underset{w_2}{\begin{bmatrix}0\\1\\0\end{bmatrix}}\right\}$ in $\mathbb{R}^3$.

$\Rightarrow$ The standard basis of $\mathbb{R}^3$ is $\left\{\begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\end{bmatrix}, \begin{bmatrix}0\\0\\1\end{bmatrix}\right\}$.

i) The first column of P encodes the projection of $\begin{bmatrix}1\\0\\0\end{bmatrix}$.

$$\frac{\begin{bmatrix}1\\0\\0\end{bmatrix}\cdot w_1}{w_1\cdot w_1} w_1 + \frac{\begin{bmatrix}1\\0\\0\end{bmatrix}\cdot w_2}{w_2\cdot w_2} w_2 = \begin{bmatrix}9/10\\0\\3/10\end{bmatrix}$$

ii) $$\frac{\begin{bmatrix}0\\1\\0\end{bmatrix}\cdot w_1}{w_1\cdot w_1} w_1 + \frac{\begin{bmatrix}0\\1\\0\end{bmatrix}\cdot w_2}{w_2\cdot w_2} w_2 = \begin{bmatrix}0\\1\\0\end{bmatrix}$$

iii) $$\frac{\begin{bmatrix}0\\0\\1\end{bmatrix}\cdot w_1}{w_1\cdot w_1} w_1 + \frac{\begin{bmatrix}0\\0\\1\end{bmatrix}\cdot w_2}{w_2\cdot w_2} w_2 = \begin{bmatrix}\frac{3}{10}\\0\\\frac{1}{10}\end{bmatrix}$$

$$\therefore P = \begin{bmatrix}9/10 & 0 & 3/10\\ 0 & 1 & 0\\ 3/10 & 0 & \frac{1}{10}\end{bmatrix}$$

Example) What is P^2? $= P$

$\Rightarrow$ $P^2 = P$ since projecting twice will not change the results.

Theorem. The orthogonal projection is unique.

$\rightarrow$ Let V be a vector space equipped with an innerproduct $\langle \cdot , \cdot \rangle$ and $W \subseteq V$ be a subspace of V. Given $\vec{x} \in V$,

$$\vec{x} = \underbrace{P\vec{x}}_{\in W} + \underbrace{\vec{u}}_{\in W^\perp} \quad \text{—(*)} \quad \text{where } P \text{ is the projection matrix.}$$

(*) is unique because if

$$\vec{x} = P\vec{x}_1 + \vec{u}_1$$
$$\vec{x} = P\vec{x}_2 + \vec{u}_2$$

then, by subtracting, we find that

$$P\vec{x}_1 - P\vec{x}_2 = -(\vec{u}_1 - \vec{u}_2)$$

Since $(P\vec{x}_1 - P\vec{x}_2) \in W$ and $-(\vec{u}_1 - \vec{u}_2) \in W^\perp$

It follows from

$$\langle P\vec{x}_1 - P\vec{x}_2 , P\vec{x}_1 - P\vec{x}_2 \rangle = \langle P\vec{x}_1 - P\vec{x}_2 , -(\vec{u}_1 - \vec{u}_2) \rangle = 0$$

$\therefore P\vec{x}_1 = P\vec{x}_2$ & $\vec{u}_1 = \vec{u}_2$

- Least squares

Definition $\hat{x}$ is a least square solution of the system $A\vec{x} = \vec{b}$ if $\hat{x}$ is such that $(A\vec{x} - \vec{b})$ is small as possible
(in an L2 sense)

Idea) $A\vec{x} = \vec{b}$ is consistent $\Longleftrightarrow$ $\vec{b} \in Col(A)$

If $A\vec{x} = \vec{b}$ is not consistent,

1) We replace $\vec{b}$ with its projection $\hat{b}$ onto $Col(A)$.
2) and solve $A\hat{x} = \hat{b}$ ← always consistent $\because$ $\hat{b} \in Col(A)$

Example) Find the least squares solution to $A\vec{x} = \vec{b}$

$$A = \begin{bmatrix} 1 & 1 \\ -1 & 1 \\ 0 & 0 \end{bmatrix}, \quad \vec{b} = \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}.$$

orthogonal

⇒ i) orthogonal projection of $\vec{b}$ onto $Col(A)$.

$$\hat{b} = \frac{\vec{b}\cdot\vec{a}_1}{\vec{a}_1\cdot\vec{a}_1}\vec{a}_1 + \frac{\vec{b}\cdot\vec{a}_2}{\vec{a}_2\cdot\vec{a}_2}\vec{a}_2 = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}$$

ii) Solve,

$$\left[\begin{array}{cc|c} 1 & 1 & 2 \\ -1 & 1 & 1 \\ 0 & 0 & 0 \end{array}\right] \Rightarrow \left[\begin{array}{cc|c} 1 & 0 & 1/2 \\ 0 & 1 & 3/2 \\ 0 & 0 & 0 \end{array}\right] \quad \therefore \begin{cases} x_1 = \frac{1}{2} \\ x_2 = 3/2 \end{cases}$$

(LS solution)

∘ The normal equation.

Theorem) $\hat{x}$ is a least square solution of $A\vec{x} = \vec{b}$
$$\Leftrightarrow A^TA\hat{x} = A^T\vec{b}$$

$\Rightarrow$ $\hat{x}$ is a least square solution of $A\vec{x} = b$

$\Leftrightarrow$ $(A\vec{x} - \vec{b})$ is as small as possible

$\Leftrightarrow$ $(A\vec{x} - \vec{b}) \perp \mathrm{Col}(A)$

$\Leftrightarrow$ $(A\vec{x} - \vec{b}) \in \mathrm{Nul}(A^T)$ FTLA

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

$\Leftrightarrow$ $A^TA\vec{x} = A^T\vec{b}$

Example) Find the LS solution:

$$A = \begin{bmatrix} 1 & 1 \\ -1 & 1 \\ 0 & 0 \end{bmatrix} \qquad \vec{b} = \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}$$

$$\Rightarrow A^TA = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}, \quad A^T\vec{b} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$$

$$\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}\vec{x} = \begin{bmatrix} 1 \\ 3 \end{bmatrix} \longrightarrow \therefore \hat{x} = \begin{bmatrix} 1/2 \\ 3/2 \end{bmatrix}$$

Example) i) Find the LS solution to $A\vec{x} = \vec{b}$ where

$$A = \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix}, \quad \vec{b} = \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix}$$

ii) What is the projection of $\vec{b}$ onto $\mathrm{Col}(A)$?

$$\Rightarrow A^TA = \begin{bmatrix} 17 & 1 \\ 1 & 5 \end{bmatrix}, \quad A^Tb = \begin{bmatrix} 19 \\ 11 \end{bmatrix}, \Rightarrow \begin{bmatrix} 17 & 1 \\ 1 & 5 \end{bmatrix}\hat{x} = \begin{bmatrix} 19 \\ 11 \end{bmatrix} \Rightarrow \hat{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

The projection of $\vec{b}$ onto $\mathrm{Col}(A)$ is $A\hat{x} = \begin{bmatrix} 4 \\ 4 \\ 3 \end{bmatrix}$.

* The projection $\hat{b}$ of $\vec{b}$ onto $\mathrm{Col}(A)$ is

$$\hat{b} = A\hat{x} \quad \text{where} \quad A^TA\hat{x} = A^Tb.$$

* If A is full rank, $\hat{b} = A(A^TA)^{-1}A^Tb$.

Hence, the projection matrix for projecting onto $\mathrm{Col}(A)$ is

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

Application : Least square lines

$$\begin{cases} \text{Data} = \{x_i, y_i\}_{i=1}^{N} \\ \text{Model}: \quad y_i = \beta_1 + \beta_2 x_i = f_\theta(x_i) \quad \text{where } \theta = \{\beta_1, \beta_2\} \\ \text{loss}: \quad \sum_i (y_i - f_\theta(x_i))^2 \end{cases}$$

$\Rightarrow$ Given 4 points $\{y_i = \beta_1 + \beta_2 x_i\}_{i=1}^{4}$

$$\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ 1 & x_4 \end{bmatrix} \underbrace{\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}}_{\text{optm. var.}} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix}$$

$(X\vec{\beta} = \vec{y})$ where the objective function $S(\beta) = \|\vec{y} - X\vec{\beta}\|^2$

$$\begin{aligned} \rightarrow \nabla_\beta \|\vec{y} - X\vec{\beta}\| &= \nabla_\beta (\vec{y} - X\vec{\beta})^T(\vec{y} - X\vec{\beta}) \\ &= \nabla_\beta \left(\vec{y}^T\vec{y} - \vec{\beta}^T X^T \vec{y} - \vec{y}^T X\vec{\beta} + \vec{\beta}^T X^T X \vec{\beta} \right) \\ &= -2X^T\vec{y} + 2X^T X\vec{\beta} \\ &= 0 \end{aligned}$$

$\therefore \hat{\beta} = (X^T X)^{-1} X^T \vec{y}$ $\Leftarrow$ least square solution.

Example) $y_i = \beta_1 + \beta_2 x_i + \beta_3 x_i^2$, $\theta = \{\beta_1, \beta_2, \beta_3\}$

$$\downarrow$$

$$\underbrace{\begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \vdots & \vdots & \vdots \end{bmatrix}}_{\text{Design mtx } X} \underbrace{\begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}}_{\text{optm. var.}} = \underbrace{\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \end{bmatrix}}_{\text{Observation vector } Y}$$

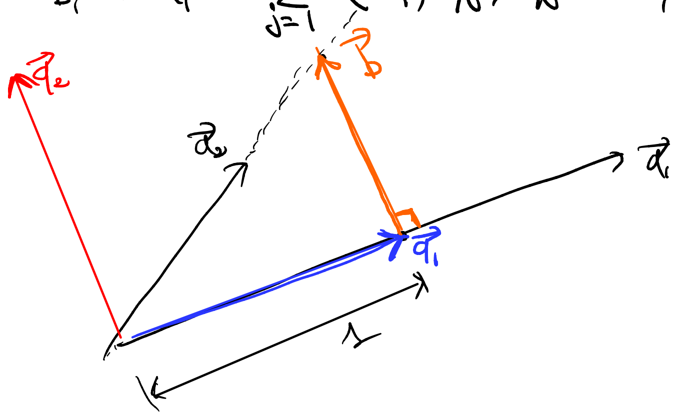
$\rightarrow X\vec{\beta} = \vec{y}$, $\quad \therefore \hat{\beta} = (X^T X)^{-1} X^T y$

- Gram–Schmidt

Given a basis $\{\vec{a}_1, \dots, \vec{a}_n\}$, produce an orthonormal basis $\vec{q}_1, \dots, \vec{q}_n$

- $\vec{b}_1 = \vec{a}_1$, $\vec{q}_1 = \dfrac{\vec{b}_1}{\|\vec{b}_1\|}$
- $\vec{b}_2 = \vec{a}_2 - \langle \vec{a}_2, \vec{q}_1 \rangle \vec{q}_1$, $\vec{q}_2 = \dfrac{\vec{b}_2}{\|\vec{b}_2\|}$
- $\vec{b}_3 = \vec{a}_3 - \langle \vec{a}_3, \vec{q}_1 \rangle \vec{q}_1 - \langle \vec{a}_3, \vec{q}_2 \rangle \vec{q}_2$, $\vec{q}_3 = \dfrac{\vec{b}_3}{\|\vec{b}_3\|}$

  $\vdots$
- $\vec{b}_i = \vec{a}_i - \sum_{j=1}^{i-1} \langle \vec{a}_i, \vec{q}_j \rangle \vec{q}_j$, $\vec{q}_i = \dfrac{\vec{b}_i}{\|\vec{b}_i\|}$



* Sparse coding: OMP (orthogonal matching pursuit)?