3.3 Projections and Least Squares

- For more equations than the unknowns, there is usually no solution.
  → overconstrained cases.
- A: $m \times n$ matrix $(m > n)$.
- $Ax = b$ → b is not in the column space, $C(A)$
  ↳ no exact solution, but find the best one in the column space.
- How to find an optimal (best) solution? ↳ least squares.
  → minimizing errors, $\|Ax - b\|^2 = E^2$.

ⓔ 1 unknown,

$$\begin{aligned} 2x &= b_1 \\ 3x &= b_2 \\ 4x &= b_3 \end{aligned} \Rightarrow \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix} x = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}.$$

if $b = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} \in C(A) \rightarrow \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$ is multiple of $a = \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix}$

then there is a solution $x = \frac{b_1}{2} = \frac{b_2}{3} = \frac{b_3}{4}$.

$$E^2 = (2x - b_1)^2 + (3x - b_2)^2 + (4x - b_3)^2 \rightarrow \text{2-nd order equation of } x.$$

$$\frac{dE^2}{dx} = 0 \rightarrow \hat{x} = \frac{2b_1 + 3b_2 + 4b_3}{2^2 + 3^2 + 4^2} = \frac{a^T b}{a^T a}$$

→ Least square solution of 1 unknown, $\hat{x} = \dfrac{a^T b}{a^T a}$ → projection onto the line a.

- Orthogonality of a and e.
  - ⇒ The error vector e connecting b to p must be perpendicular to a,

$$a^T(b - \underbrace{\hat{x}a}_{p}) = a^Tb - a^T \frac{a^Tb}{a^Ta} a = 0.$$

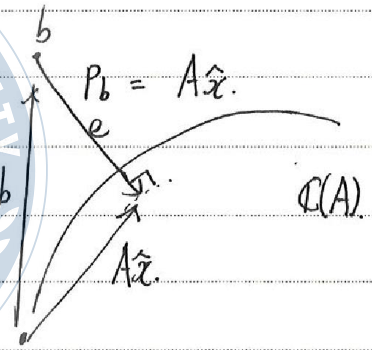
◎ Least square Problems with Several Variables

⇒ to project b onto a subspace.

⇒ $m \times n : A \qquad (m > n)$

$E = \| Ax - b \|$

$\hat{x}$: solution of least squares to minimize E.



$e = (b - A\hat{x}) \perp$ space.

1. Since column space is perpendicular to left nullspace, $b - A\hat{x}$ is in the left nullspace,

$$A^T(b - A\hat{x}) = 0 \qquad \text{or} \qquad A^TA\hat{x} = A^Tb.$$

2. The error vector must be orthogonal to each column vector,

$$a_i^T(b - A\hat{x}) = 0. \qquad \Rightarrow \begin{bmatrix} a_1^T \\ \vdots \\ a_n^T \end{bmatrix} \begin{bmatrix} b - A\hat{x} \end{bmatrix} = 0$$

$$\Rightarrow A^T[b - A\hat{x}] = A^Tb - A^TA\hat{x}.$$

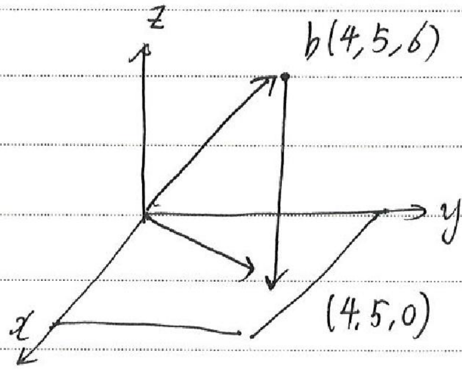
$$E^2 = \|Ax - b\|^2 = (Ax-b)^T (Ax-b)$$

$$\frac{dE^2}{dx} \rightarrow A^T(Ax-b) + (Ax-b)^T A^T = 0$$

$$\Rightarrow A^TAx = A^Tb \quad \cancel{+\ x^TA^TA = b^TA}$$

- Normal equation $\Rightarrow A^TA\hat{x} = A^Tb$
- Best estimate $\hat{x}$: for invertibility of A^TA. $\Rightarrow \hat{x} = (A^TA)^{-1}A^Tb$
- Projection $\Rightarrow p = A\hat{x} = A(A^TA)^{-1}A^Tb$.

ex) $A = \begin{bmatrix} 1 & 2 \\ 1 & 3 \\ 0 & 0 \end{bmatrix}$, $b = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}$



column space : p x-y plane in $\mathbb{R}^3$,
b : a 3-D point
p : (4, 5, 0).

- $A^TA = \begin{bmatrix} 2 & 5 \\ 5 & 13 \end{bmatrix}$, $\hat{x} = (A^TA)^{-1}A^Tb = \begin{bmatrix} 13 & -5 \\ -5 & 2 \end{bmatrix}\begin{bmatrix} 1 & 1 & 0 \\ 2 & 3 & 0 \end{bmatrix}\begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$
- projection $p = A\hat{x} = \begin{bmatrix} 1 & 2 \\ 1 & 3 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 5 \\ 0 \end{bmatrix}$
- $E = 0 \cdot x + 0 \cdot y + 6 \rightarrow 6.$

Remark 1.

If b is in $\mathbb{C}(A)$ $\rightarrow$ $b = Ax$.

$\rightarrow$ The projection of b is still b, $p = A(A^TA)^{-1}A^Tb$

$= A(A^TA)^{-1}A^TAx$

$= Ax = b$.

Remark 2.

If b is perpendicular to every column,

$A^Tb = 0$ $\rightarrow$ b in left nullspace,

$p = A(A^TA)^{-1}A^Tb = 0$.

$\Rightarrow$ projected onto zero

Remark 3.

A is square & invertible, $\rightarrow$ $\mathbb{C}(A)$ is the whole space.

$p = b$ $\Rightarrow$ $p = A(A^TA)^{-1}A^Tb = A(A^{-1})(A^T)^{-1}A^T \cdot b = b$.

~~Remark 4.~~

∘ A^TA has the same nullspace as A.

$Ax = 0$ $\rightarrow$ $A^T\underline{Ax} = 0$.

$A^TAx = 0$ $\rightarrow$ $x^TA^TAx = \|Ax\|^2 = 0$ $\rightarrow$ $A^TAx = Ax = 0$.

◎ Projection Matrix. $\mathbb{P}$

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

↳ b is projected onto column space of A.

$p = \mathbb{P}b \in \mathbb{C}(A)$

$e = b - \mathbb{P}b$ is orthogonal to $\mathbb{C}(A)$ $\rightarrow$ element of orthogonal complement.

↳ $\in \mathcal{N}(A^T)$: left nullspace.

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

$\left(\begin{array}{l} \mathbb{P}^2 = \mathbb{P} \\ \mathbb{P}^T = \mathbb{P} \end{array} \right) \Rightarrow$ projection matrix definition

$(b - \mathbb{P}b) \perp \mathbb{P}c \rightarrow (b-\mathbb{P}b)^T \mathbb{P}c = (b^T\mathbb{P} - b^T)\mathbb{P}c$

$= b^T\mathbb{P}c - b^T\mathbb{P}c = 0$ //

Ex 1) when A is square and invertible in $\mathbb{R}^n$.

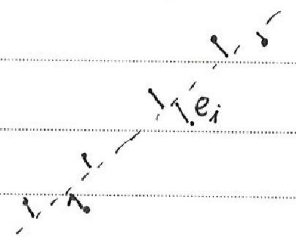
$\mathbb{P} = A(A^TA)^{-1}A^T = I$. $\rightarrow$ identity matrix

$\rightarrow$ projection onto whole space $\rightarrow I$.

◎ Least - Squares Fitting of Data.

$\rightarrow$ Since observation errors $\rightarrow$ not exactly fitted.

(ex) $\Rightarrow$ straight line fitting



$\min \left[E^2 = \Sigma e_i^2 \right]$.

$C + Dt = b$ $\quad$ ($\overset{\text{generally}}{\checkmark}$ $y = ax + b \rightarrow (x = t,\ y = b)$ $(a, b) \rightarrow (D, C)$

$\rightarrow$ $\begin{array}{l} C + Dt_1 = b_1 \\ C + Dt_2 = b_2 \\ \quad \vdots \\ C + Dt_m = b_m \end{array} \rightarrow \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \\ 1 & t_m \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix} \Rightarrow Ax = b$ //

$\hat{X} = \begin{bmatrix} \hat{C} \\ \hat{D} \end{bmatrix}$.