November 12, 2019

Review Find bases for the fundamental subspaces of a matrix $\underset{m\times n}{A}$:

$$[A \mid I_m] \xrightarrow{G-J} E[A \mid I_m] = [U \mid E] = \begin{array}{c} \left[\begin{array}{c|c} V & E_1 \\ \hline 0 & E_2 \end{array}\right] \begin{array}{l} k \\ m-k \end{array} \\ \begin{array}{cc} n & m \end{array} \end{array}$$

The row vectors of V form a basis for $\text{row}(A)$.

The column vectors of V " " " " $\text{col}(A)$.

A basis for $\text{null}(V)$ can be a basis for $\text{null}(A)$.

The row vectors of E_2 form a basis for $\text{null}(A^T)$.

Thm 7.6.4 (CR factorization)

$\underset{m\times n}{A} = \underset{m\times k}{C}\,\underset{k\times n}{R}$, C : $m\times k$ matrix whose column vectors are the pivot columns of A

R : $k\times n$ " " row " " " nonzero vectors in the reduced row echelon form.

pf) Let $R_0 = EA$, reduced row echelon form of A.

Since $\text{rank}(A) = k$, $R_0 = \left[\begin{array}{c} R \\ \hline 0 \end{array}\right]\begin{array}{l} \}k \\ \}m-k \end{array}$

Partition E^{-1} as $[\underbrace{C}_{k} \mid \underbrace{D}_{m-k}]$. Then

$$A = E^{-1}R_0 = [C \mid D]\begin{bmatrix} R \\ 0 \end{bmatrix} = CR + D0 = CR.$$

ETS : The columns of C are the pivot columns of A.

Note $\mathbf{c}_j(R) = \mathbf{e}_j$ in $\mathbb{R}^k$, $1 \le j \le k$. Then

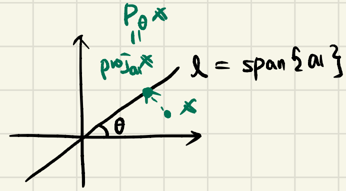
$$\mathbf{c}_j(C) = C\mathbf{e}_j = C\,\mathbf{c}_j(R) = \mathbf{c}_j(CR) = \mathbf{c}_j(A). \quad \blacksquare$$

Hence $A = CR = \begin{bmatrix} | & & | \\ \mathbf{c}_1 & \cdots & \mathbf{c}_k \\ | & & | \end{bmatrix}\begin{bmatrix} - \mathbf{r}_1 - \\ - \mathbf{r}_2 - \\ \vdots \\ - \mathbf{r}_k - \end{bmatrix} = \mathbf{c}_1\mathbf{r}_1 + \mathbf{c}_2\mathbf{r}_2 + \cdots + \mathbf{c}_k\mathbf{r}_k$

$R = [r_{ij}]_{k\times n} \Rightarrow CR = \left[\, r_{11}\mathbf{c}_1 + \cdots + r_{k1}\mathbf{c}_k \;\middle|\; r_{21}\mathbf{c}_1 + \cdots + r_{2k}\mathbf{c}_k \;\middle|\; \cdots \;\middle|\; r_{n1}\mathbf{c}_1 + \cdots + r_{nk}\mathbf{c}_k \,\right]$

$= \mathbf{c}_1[r_{11} \cdots r_{n1}] + \mathbf{c}_2[r_{12} \cdots r_{n2}] + \cdots + \mathbf{c}_k[r_{k1} \cdots r_{nk}]$

§ 7.7. Projection theorem



Recall that $P_\theta = \begin{bmatrix} \cos^2\theta & \sin\theta\cos\theta \\ \sin\theta\cos\theta & \sin^2\theta \end{bmatrix}$

Let $a \in \mathbb{R}^n - \{0\}$.

The <u>orthogonal projection</u> of x onto the line $\text{span}\{a\}$ is

$$\text{proj}_a x = \frac{x \cdot a}{\|a\|^2} a$$

$\left(\because\right.$ Since $\text{proj}_a x = k a$ for some $k \in \mathbb{R}$, we need to find k.

Since $x - \text{proj}_a x$ is orthogonal to a,

$$(x - k a) \cdot a = 0$$

$$\therefore k = \frac{x \cdot a}{a \cdot a} \quad \left.\right)$$

Using the above formula, we can recover the matrix P_θ as follows:

Let $u := (\cos\theta, \sin\theta)$. i.e., u is a unit vector spanning the straight line l. Then

$$\begin{bmatrix} | & | \\ \text{proj}_u e_1 & \text{proj}_u e_2 \\ | & | \end{bmatrix} = \begin{bmatrix} | & | \\ \cos\theta\, u & \sin\theta\, u \\ | & | \end{bmatrix} = \begin{bmatrix} \cos^2\theta & \cos\theta\sin\theta \\ \cos\theta\sin\theta & \sin^2\theta \end{bmatrix}.$$

<u>Theorem 7.7.1</u>

Given $a \in \mathbb{R}^n - \{0\}$, every $x \in \mathbb{R}^n$ can be expressed in exactly one way as $x = x_1 + x_2$, where $x_1 \in \text{span}\{a\}$ and $x_2 \in a^\perp$. Moreover,

$$x_1 = \frac{x \cdot a}{\|a\|^2} a \quad \& \quad x_2 = x - x_1 = x - \frac{x \cdot a}{\|a\|^2} a.$$

pf) ① Existence : Clearly, $x_1 \in \text{span}\{a\}$ and

$$x_2 \cdot a = x \cdot a - x_1 \cdot a = x \cdot a - \frac{x \cdot a}{\cancel{\|a\|^2}} \cancel{a \cdot a} = 0$$

$$\therefore x_2 \in a^\perp.$$

② Uniqueness : Suppose that $\mathbf{x}$ can also be written as

$$\mathbf{x} = \mathbf{x}_1' + \mathbf{x}_2',$$

where $\mathbf{x}_1' \in \text{span}\{\mathbf{a}\}$ & $\mathbf{x}_2' \in \mathbf{a}^\perp$.

Then $\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2 = \mathbf{x}_1' + \mathbf{x}_2'$

$$\Rightarrow \mathbf{x}_1 - \mathbf{x}_1' = \mathbf{x}_2' - \mathbf{x}_2 \in \text{span}\{\mathbf{a}\} \cap \mathbf{a}^\perp = \{\mathbf{0}\}.$$

$$\therefore \mathbf{x}_1 = \mathbf{x}_1' \quad \& \quad \mathbf{x}_2' = \mathbf{x}_2.$$

$\blacksquare$

Def. Given $\mathbf{a} \in \mathbb{R}^n - \{\mathbf{0}\}$, for every $\mathbf{x} \in \mathbb{R}^n$, the orthogonal projection of $\mathbf{x}$ onto $\text{span}\{\mathbf{a}\}$ is denoted by $\text{proj}_{\mathbf{a}}\mathbf{x}$ and is defined by $\text{proj}_{\mathbf{a}}\mathbf{x} = \dfrac{\mathbf{x}\cdot\mathbf{a}}{\|\mathbf{a}\|^2}\mathbf{a}$.

We call $\text{proj}_{\mathbf{a}}\mathbf{x}$ the vector component of $\mathbf{x}$ along $\mathbf{a}$, and $\mathbf{x} - \text{proj}_{\mathbf{a}}\mathbf{x}$ the vector component of $\mathbf{x}$ orthogonal to $\mathbf{a}$.

Note that the length of $\text{proj}_{\mathbf{a}}\mathbf{x}$ is $\dfrac{|\mathbf{x}\cdot\mathbf{a}|}{\|\mathbf{a}\|}$.

Define $T: \mathbb{R}^n \longrightarrow \mathbb{R}^n$ by $T(\mathbf{x}) = \text{proj}_{\mathbf{a}}\mathbf{x}$. Then T is called the orthogonal projection of $\mathbb{R}^n$ onto $\text{span}\{\mathbf{a}\}$.

Thm 7.7.3 For a nonzero column vector $\mathbf{a}$ in $\mathbb{R}^n$, the standar matrix for the orthogonal projection T of $\mathbb{R}^n$ onto $\text{span}\{\mathbf{a}\}$ is

$$P = \frac{1}{\mathbf{a}^T\mathbf{a}}\mathbf{a}\mathbf{a}^T.$$

The matrix P is symmetric and has rank 1.

pf) Let $\mathbf{a} = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} \in \mathbb{R}^n - \{\mathbf{0}\}$. Then $T(\mathbf{e}_j) = \dfrac{\mathbf{e}_j \cdot \mathbf{a}}{\mathbf{a} \cdot \mathbf{a}} \mathbf{a} = \dfrac{a_j}{\mathbf{a}^T \mathbf{a}} \mathbf{a}$.

Hence $[T] = \begin{bmatrix} | & & | \\ T(\mathbf{e}_1) & \cdots & T(\mathbf{e}_n) \\ | & & | \end{bmatrix} = \begin{bmatrix} \dfrac{a_1}{\mathbf{a}^T\mathbf{a}} \mathbf{a} & \dfrac{a_2}{\mathbf{a}^T\mathbf{a}} \mathbf{a} & \cdots & \dfrac{a_n}{\mathbf{a}^T\mathbf{a}} \mathbf{a} \end{bmatrix}$

$= \dfrac{\mathbf{a}}{\mathbf{a}^T \mathbf{a}} [a_1\ a_2 \cdots a_n] = \dfrac{1}{\mathbf{a}^T \mathbf{a}} \mathbf{a}\mathbf{a}^T$

Since $\mathbf{a}\mathbf{a}^T$ is symmetric and has rank 1,

P is also symmetric and has rank 1. ∎

If we take $\mathbf{u} := \dfrac{\mathbf{a}}{\|\mathbf{a}\|}$ in the previous theorem, then $P = \mathbf{u}\mathbf{u}^T$.

Thm 7.7.4 (Projection theorem for subspaces)

Let W be a subspace of $\mathbb{R}^n$. Then every $\mathbf{x} \in \mathbb{R}^n$ can be expressed in exactly one way as $\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2$, where $\mathbf{x}_1 \in W$ and $\mathbf{x}_2 \in W^\perp$.

※ In the above theorem, $\mathbf{x}_1$ is denoted by $\text{proj}_W \mathbf{x}$ and it is called the orthogonal projection of $\mathbf{x}$ on W, and $\mathbf{x}_2$ is denoted by $\text{proj}_{W^\perp} \mathbf{x}$ and it is called the orthogonal projection of $\mathbf{x}$ on $W^\perp$.

proof) ① If $W = \{\mathbf{0}\}$, then $W^\perp = \mathbb{R}^n$. $\therefore \forall \mathbf{x} \in \mathbb{R}^n$, $\mathbf{x} = \mathbf{0} + \mathbf{x}$. o.k.

② Now assume $W \neq \{\mathbf{0}\}$. There there is a basis for W. Let $\{\mathbf{w}_1, \cdots, \mathbf{w}_k\}$ be a basis for W and form a matrix $M = [\mathbf{w}_1 \cdots \mathbf{w}_k]$. Then $W = \text{col}(M)$ & $W^\perp = \text{null}(M^T)$.

Note that
$$\begin{cases} \bullet\ \mathbf{x}_1 \in W \iff \mathbf{x}_1 = M\mathbf{v} \text{ for some } \mathbf{v} \in \mathbb{R}^k \\ \bullet\ \mathbf{x}_2 \in W^\perp \iff M^T \mathbf{x}_2 = \mathbf{0} \\ \bullet\ \text{Uniqueness} \iff M^T(\mathbf{x} - M\mathbf{v}) = \mathbf{0} \text{ has a unique sol. for each } \mathbf{x} \in \mathbb{R}^k \end{cases}$$

Since M has full column rank, $M^T M$ is invertible.

$$\therefore \quad M^T(x - Mw) = 0 \iff w = (M^T M)^{-1} M^T x.$$

Hence $$x = \underbrace{Mw}_{\in W} + \underbrace{(x - Mw)}_{\in W^{\perp}}, \quad \text{where} \quad w = (M^T M)^{-1} M^T x.$$

<u>Thm 7.7.5</u>

Let W be a subspace of $\mathbb{R}^n$.
Let M be any matrix whose column vectors form a basis for W.

Then $\quad \mathrm{proj}_W\, x = M(M^T M)^{-1} M\, x.$

Hence the standard matrix for a linear operator $T: \mathbb{R}^n \to \mathbb{R}^n$ given by $T(x) = \mathrm{proj}_W\, x$ is $P = M(M^T M)^{-1} M.$ Moreover, $I - P$ is the standard matrix for the linear operator $U: \mathbb{R}^n \to \mathbb{R}^n$ given by $U(x) = \mathrm{proj}_{W^{\perp}}\, x$.

※ M is not unique, but we get the same matrix P.

If the column vectors of M form an orthonormal basis for W, then $P = MM^T$.

November 14, 2019

Projection theorem for subspaces

W : a subspace for $\mathbb{R}^n$

$\Rightarrow \forall \mathbf{x} \in \mathbb{R}^n, \quad \mathbf{x} = \underset{W}{\underset{\cap}{\mathbf{x}_1}} + \underset{W^\perp}{\underset{\cap}{\mathbf{x}_2}}$

From the above theorem, we can find two projections

$\text{proj}_W : \mathbb{R}^n \longrightarrow W \subseteq \mathbb{R}^n$ and $\text{proj}_{W^\perp} : \mathbb{R}^n \longrightarrow W^\perp \subseteq \mathbb{R}^n$

so that $\mathbf{x} = \text{proj}_W \mathbf{x} + \text{proj}_{W^\perp} \mathbf{x}$.

For a nonzero subspace W of $\mathbb{R}^n$,

if M is a matrix whose columns form a basis for W,

then $\text{proj}_W \mathbf{x} = M(M^TM)^{-1}M^T\mathbf{x}$.

If M is a matrix whose columns form an orthonormal basis for W,

then $M^TM = I_n$, and hence $\text{proj}_W \mathbf{x} = MM^T\mathbf{x}$.

Thm 7.7.6 Let P be an $n \times n$ matrix.

P is the standard matrix for an orthogonal projection of $\mathbb{R}^n$ onto a k-dim'l subspace W of $\mathbb{R}^n$.

$\Leftrightarrow$ P is symmetric, idempotent, and has rank k.

The subspace W is $\text{col}(P)$

Proof) ($\Rightarrow$) $P = M(M^TM)^{-1}M^T$, $P^T = (M^T)^T((M^TM)^{-1})^TM^T = M((M^TM)^{-1})^TM^T$

Since $(M^TM)(M^TM)^{-1} = I_n$, $((M^TM)^{-1})^T \underbrace{(M^TM)^T}_{M^TM} = I_n$.

$\therefore ((M^TM)^{-1})^T = (M^TM)^{-1}$

$$= M(M^TM)^{-1}M^T = P$$

$P^2 = M(M^TM)^{-1}M^TM(M^TM)^{-1}M^T = M(M^TM)^{-1}M^T = P$

Since $\dim W = k$ and $\text{ran}(T) = \text{col}(P)$ has rank k. $\therefore \text{rank}(P) = k$.

$(\Rightarrow)$ Let $W = \mathrm{col}(P)$.

$\forall \mathbf{x} \in \mathbb{R}^n$, $\mathbf{x} = P\mathbf{x} + (I-P)\mathbf{x}$

<u>Note</u> $P\mathbf{x} \in \mathrm{col}(P)$ & $(I-P)\mathbf{x} \in \mathrm{null}(P)$

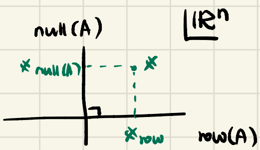
$(\because \; P(I-P)\mathbf{x} = P\mathbf{x} - P^2\mathbf{x} = P\mathbf{x} - P\mathbf{x} = \mathbf{0}.)$

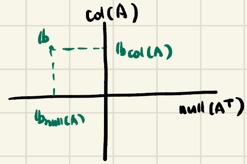
Since $P = P^T$, $\mathrm{null}(P) = \mathrm{null}(P^T) = W^\perp$.

$\boxed{}$

* Strang diagram

A: $m \times n$ matrix $\Rightarrow \begin{cases} \forall \mathbf{x} \in \mathbb{R}^n, & \mathbf{x} = \mathbf{x}_{\mathrm{row}(A)} + \mathbf{x}_{\mathrm{null}(A)} \\ \forall \mathbf{b} \in \mathbb{R}^m, & \mathbf{b} = \mathbf{b}_{\mathrm{col}(A)} + \mathbf{b}_{\mathrm{null}(A^T)} \end{cases}$





$\therefore$ $A\mathbf{x} = \mathbf{b}$ is consistent $\Leftrightarrow$ $\mathbf{b}_{\mathrm{null}(A^T)} = \mathbf{0}$.

<u>Thm 7.7.7</u>

A: $m \times n$ matrix & $\mathbf{b} \in \mathrm{col}(A)$ ($\rightsquigarrow$ $A\mathbf{x} = \mathbf{b}$ is consistent)

(a) If A has full column rank, then $A\mathbf{x} = \mathbf{b}$ has a unique solution and that sol. is in $\mathrm{row}(A)$.

(b) If A does not have full column rank, then $A\mathbf{x} = \mathbf{b}$ has infinitely many sol. but there is a unique sol. in $\mathrm{row}(A)$.

Moreover, among all the solutions of the system, the sol. in $\mathrm{row}(A)$ has the smallest norm.

pf) Let $\mathbf{x} = \mathbf{x}_{\mathrm{row}(A)} + \mathbf{x}_{\mathrm{null}(A)}$. Then $A\mathbf{x} = A\mathbf{x}_{\mathrm{row}(A)} + A\mathbf{x}_{\mathrm{null}(A)}$.

(a) If A has full column rank, then $A\mathbf{x} = \mathbf{b}$ has a unique sol. & $\mathrm{null}(A) = \{\mathbf{0}\}$. ($\because$ $\mathrm{rank}(A) = n$ & $\mathrm{nullity}(A) = 0$)

$\therefore$ $\mathbf{x} = \mathbf{x}_{\mathrm{row}(A)}$ is unique.

(b) If A does not have full col. rank, then $A\mathbf{x} = \mathbf{b}$ has infinitely many sol.

(i) $A\mathbf{x} = \mathbf{b}$ has a unique sol. in $\mathrm{row}(A)$.

$\Bigg($ $\because$ Suppose $\mathbf{x}_1, \mathbf{x}_2 \in \mathrm{row}(A)$ are sol. of $A\mathbf{x} = \mathbf{b}$.

Then $\mathbf{x}_1 - \mathbf{x}_2 \in \mathrm{row}(A) \cap \mathrm{null}(A) = \{\mathbf{0}\}$.

$\therefore$ $\mathbf{x}_1 = \mathbf{x}_2$ $\Bigg)$

(ii) Among all the sol's of the system, $\mathbf{x}_1$ has the smallest norm.

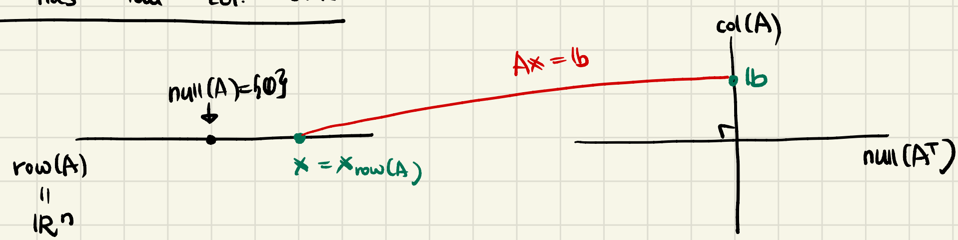
$\because$ If $\mathbf{x}$ is a sol. of $A\mathbf{x} = \mathbf{b}$, then $\mathbf{x} = \underset{\substack{\cap \\ \mathrm{row}(A)}}{\mathbf{x}_1} + \underset{\substack{\cap \\ \mathrm{null}(A)}}{(\mathbf{x} - \mathbf{x}_1)}$

By Pythagoras

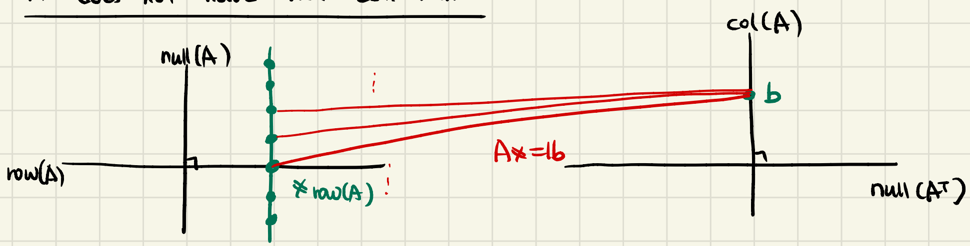
$$\|\mathbf{x}\| \geq \sqrt{\|\mathbf{x}_1\|^2 + \|\mathbf{x} - \mathbf{x}_1\|^2} \geq \|\mathbf{x}_1\|.$$

$\therefore$ $\mathbf{x}_1$ has the smallest norm among solutions. $\boxed{\,}$

A has full col. rank



A does not have full col. rank



The solution set of $A\mathbf{x} = \mathbf{b}$ is parallel to $\mathrm{null}(A)$.

Thm 7.7.8 (Double perp thm)

For every subspace W of $\mathbb{R}^n$, $(W^\perp)^\perp = W$.

Pf) ① $W \subseteq (W^\perp)^\perp$ (clear!) ② $W \supseteq (W^\perp)^\perp$. $\forall \mathbf{w} \in (W^\perp)^\perp \subseteq \mathbb{R}^n$, $\mathbf{w} = \underset{\substack{\cup \\ W}}{\mathbf{w}_1} + \underset{\substack{\cup \\ W^\perp}}{\mathbf{w}_2}$

Claim $\mathbf{w}_2 = \mathbf{0}$. Since $\mathbf{w} \in (W^\perp)^\perp$ & $\mathbf{w}_2 \in W^\perp$, $\underset{\substack{\| \\ \mathbf{w}_1 \cdot \mathbf{w}_2 + \mathbf{w}_2 \cdot \mathbf{w}_2}}{\mathbf{w} \cdot \mathbf{w}_2} = 0$.

$\therefore$ $\|\mathbf{w}_2\|^2 = 0$ $\Rightarrow$ $\mathbf{w}_2 = \mathbf{0}$. $\boxed{\,}$

§ 7.8. Best approximation & Least squares.

The minimum distance problem in $\mathbb{R}^n$

Given a subspace W of $\mathbb{R}^n$ and $\mathbb{b} \in \mathbb{R}^n$,

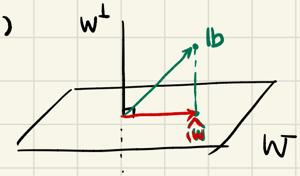
find a vector $\hat{\mathbb{w}}$ in W that is closest to $\mathbb{b}$ in the sence

that $\|\mathbb{b}-\hat{\mathbb{w}}\| > \|\mathbb{b}-\mathbb{w}\| \quad \forall\, \mathbb{w} \in W - \{\hat{\mathbb{w}}\}$.

Such a vector $\hat{\mathbb{w}}$, if it exists, is called a best approximation

to $\mathbb{b}$ from W.

e.g.)



$\hat{\mathbb{w}} = \text{proj}_W \mathbb{b}$ & $\|\mathbb{b}-\hat{\mathbb{w}}\| = \|\text{proj}_{W^\perp} \mathbb{b}\|$.

Thm 7.8.1 (Best approximation theorem)

If W is a subspace of $\mathbb{R}^n$ and $\mathbb{b}$ is a point in $\mathbb{R}^n$, then there is a unique best approximation to $\mathbb{b}$ from W, namely $\hat{\mathbb{w}} = \text{proj}_W \mathbb{b}$.

pf) $\forall\, \mathbb{w} \in W$, $\quad \mathbb{b}-\mathbb{w} = \underbrace{(\mathbb{b}-\text{proj}_W \mathbb{b})}_{\text{proj}_{W^\perp}\mathbb{b} \,\in\, W^\perp} + \underbrace{(\text{proj}_W \mathbb{b} - \mathbb{w})}_{\in W}$

$\Rightarrow \quad \|\mathbb{b}-\mathbb{w}\| \geq \sqrt{\|\mathbb{b}-\text{proj}_W \mathbb{b}\|^2 + \|\text{proj}_W \mathbb{b}-\mathbb{w}\|^2}$.

If $\mathbb{b} \neq \text{proj}_W \mathbb{b}$, then $\|\mathbb{b}-\mathbb{w}\| > \|\mathbb{b}-\text{proj}_W \mathbb{b}\|$. ⊛

$\therefore\ \underbrace{\text{proj}_W \mathbb{b}}_{=\, \hat{\mathbb{w}}}$ is a best approximation to $\mathbb{b}$ from W.

(Uniqueness)

By ⊛, any $\mathbb{w} (\neq \text{proj}_W \mathbb{b})$ in W, $\|\mathbb{b}-\mathbb{w}\| > \|\mathbb{b}-\hat{\mathbb{w}}\|$

$\therefore$ Such a vector $\mathbb{w}$ is not a best approximation to $\mathbb{b}$ from W.

Def. The <u>distance</u> from a point $\mathbb{b}$ to a subspace W is

$$d = \| \mathbb{b} - \text{proj}_W \mathbb{b} \| = \| \text{proj}_{W^\perp} \mathbb{b} \| .$$

Let A be an $m \times n$ matrix, $\mathbb{b} \in \mathbb{R}^m$.

$\hat{\mathbb{x}} \in \mathbb{R}^n$ is called a <u>best approximate sol.</u> or a least squares sol. of $A\mathbb{x} = \mathbb{b}$ if $\| \mathbb{b} - A\hat{\mathbb{x}} \| \leq \| \mathbb{b} - A\mathbb{x} \| \quad \forall \mathbb{x} \in \mathbb{R}^n$.

$\Rightarrow \begin{cases} \mathbb{b} - A\hat{\mathbb{x}} \text{ is called the least squares error vector} \\ \| \mathbb{b} - A\hat{\mathbb{x}} \| \text{ is called the least squares error.} \end{cases}$

Since $A\mathbb{x} \in \text{col}(A)$ and $\text{proj}_{\text{col}(A)} \mathbb{b}$ is the best approximation to $\mathbb{b}$ from $\text{col}(A)$, $\| \mathbb{b} - A\mathbb{x} \|$ is minimized when $A\mathbb{x} = \text{proj}_{\text{col}(A)} \mathbb{b}$.

Since $A\mathbb{x} = \text{proj}_{\text{col}(A)} \mathbb{b}$ is consistent, every $A\mathbb{x} = \mathbb{b}$ has at least one least squares solution. i.e., $\exists \hat{\mathbb{x}} \in \mathbb{R}^n$ s.t. $\| \mathbb{b} - A\hat{\mathbb{x}} \| = \| \mathbb{b} - \text{proj}_{\text{col}(A)} \mathbb{b} \| \leq \| \mathbb{b} - A\mathbb{x} \|$.

From $\mathbb{b} - A\mathbb{x} = \mathbb{b} - \text{proj}_{\text{col}(A)} \mathbb{b}$, we have

$$A^T(\mathbb{b} - A\mathbb{x}) = A^T(\underbrace{\mathbb{b} - \text{proj}_{\text{col}(A)} \mathbb{b}}_{\in \, \text{col}(A)^\perp = \text{null}(A^T)})$$

$$\therefore \underline{A^T A \mathbb{x} = A^T \mathbb{b}}.$$

This is called the <u>normal equation</u> or normal system associated with $A\mathbb{x} = \mathbb{b}$.

<u>Thm 7.8.3</u>

(a) $\{$ the least squares solutions of $A\mathbb{x} = \mathbb{b} \} = \{$ the solutions of $A^T A \mathbb{x} = A^T \mathbb{b} \}$

(b) If A has full col. rank, then $\exists! \; \hat{\mathbb{x}} = (A^T A)^{-1} A^T \mathbb{b}$.

(c) If A does not have full col. rank, $A^T A \mathbb{x} = A^T \mathbb{b}$ has ∞-many solutions but $\exists! \; \hat{\mathbb{x}} \in \text{row}(A)$ s.t. $A^T A \hat{\mathbb{x}} = A^T \mathbb{b}$. Among all the solutions of $A^T A \mathbb{x} = A^T \mathbb{b}$, $\hat{\mathbb{x}} \in \text{row}(A)$ has the smallest norm.

pf) ETS : "$\supseteq$" part of (a)

If x_0 is a solution of $A^TAx = A^T b$, then

$$b - Ax_0 \in \text{null}(A^T) = \text{col}(A)^\perp$$

Since $b = Ax_0 + (b - Ax_0)$, $Ax_0 \in \text{col}(A)$, and $b - Ax_0 \in \text{col}(A)^\perp$,

we have $Ax_0 = \text{proj}_{\text{col}(A)} b$

$\therefore \|b - Ax_0\| \leq \|b - Ax\| \quad \forall x \in \mathbb{R}^n.$

$\therefore$ x_0 is a least squares sol. of $Ax = b$. ∎

<u>Thm 7.8.4</u>

$\hat{x}$: a least squares sol. of $Ax = b$

$\Leftrightarrow$ $b - A\hat{x} \in \text{col}(A)^\perp$.

pf) Since $b = \text{proj}_{\text{col}(A)} b + \text{proj}_{\text{null}(A^T)} b$,

$$b - Ax = (\text{proj}_{\text{col}(A)} b - Ax) + \text{proj}_{\text{null}(A^T)} b.$$

Note that $\hat{x}$ is a least squares sol. of $Ax = b$ iff $\hat{x}$ is a sol. of $Ax = \text{proj}_{\text{col}(A)} b$.

$\therefore$ $\hat{x}$ is a least squares sol. of $Ax = b$ iff

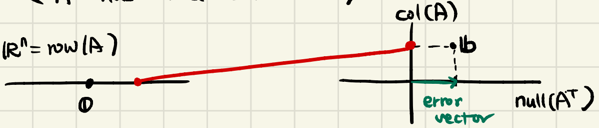
$$b - A\hat{x} = \text{proj}_{\text{null}(A^T)} b \in \text{null}(A^T) = \text{col}(A)^\perp.$$ ∎

Consequently, for $Ax = b$,

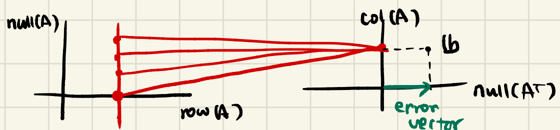
$$\begin{cases} \text{least squares error vector} = \text{proj}_{\text{null}(A^T)} b \\ \text{least squares error} = \|\text{proj}_{\text{null}(A^T)} b\| \end{cases}$$

<u>Strang diagram</u>

< A has full col. rank >



< A does not have full col. rank >



※ Read the textbook from p. 399 to p. 403.

Fitting a curve to experimental data : $(x_1, y_1), \cdots, (x_n, y_n)$

Mathematical model : $y = f(x) = a_0 + a_1 x + \cdots + a_m x^m$

Idea : Determine the coeff. $a_0, a_1, \cdots, a_m$ that make the graph $y = f(x)$ fit the data as closely as possible.

$$\underbrace{\begin{bmatrix} 1 & x_1 & \cdots & x_1^m \\ \vdots & & & \\ 1 & x_n & \cdots & x_n^m \end{bmatrix}}_{M} \underbrace{\begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_m \end{bmatrix}}_{w} = \underbrace{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}}_{y}$$

If $m = n-1$ and $x_1, \cdots, x_n$ are pairwise distinct, then we can always find the polynomial $y = f(x)$ passing through all the points $(x_1, y_1), \cdots, (x_n, y_n)$. (See p. 203)

In general, we find the sol. of the normal system $M^T M w = M^T y$ ass. with $Mw = y$.

If $m < n$ and at least $m+1$ of $x_1, \cdots, x_n$ are distinct, then M has full column rank and

$$w = (M^T M)^{-1} M^T y.$$