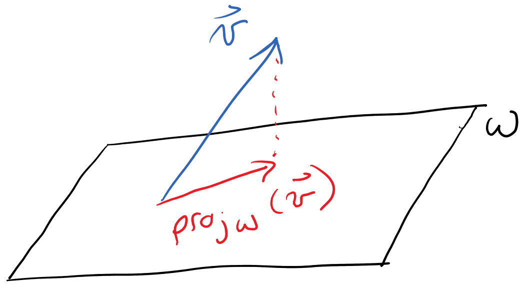
Section 7.3: Least-Squares Approximation

Friday, December 02, 2022 12:51 PM

Best Approximation Theorem

If W is a subspace of $\mathbb{R}^n$ and $\vec{v}$ is a vector in $\mathbb{R}^n$ which may or may not be in subspace W, then the best approximation to $\vec{v}$ in W is the projection of $\vec{v}$ onto W

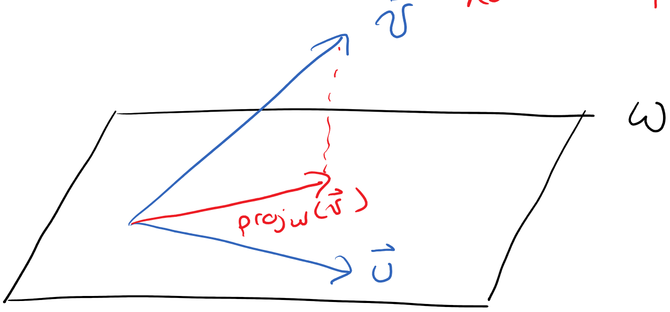


note: Euclidean distance from $\vec{v}$ to W is

$$\| \underbrace{\vec{v} - \text{proj}_W(\vec{v})}_{\text{perp}_W(\vec{v})} \|$$

for any vector $\vec{u}$ in W

$\vec{v}$ not in the plane W



$\vec{u}$ is in the plane W

$$\|\vec{v} - \text{proj}_W(\vec{v})\| \leq \|\vec{v} - \vec{u}\|$$

LHS is the "vertical" distance to the plane

RHS is the distance from tip of $\vec{v}$ to any point in the plane why do we care? we can use this idea to get approximate solutions to inconsistent systems.

example: let $\vec{w}_1 = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}$, $\vec{w}_2 = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$, $\vec{v} = \begin{bmatrix} 5 \\ 0 \\ 4 \end{bmatrix}$

Find the best approximation to $\vec{v}$ in the plane $W = \text{span}(\vec{w}_1, \vec{w}_2)$ and find the Euclidean distance from $\vec{v}$ to W.

answer: first observe that $\vec{w}_1 \perp \vec{w}_2$

so we can say

$$\text{proj}_W(\vec{v}) = \text{proj}_{\vec{w}_1}(\vec{v}) + \text{proj}_{\vec{w}_2}(\vec{v})$$

$$= \frac{\vec{w}_1 \cdot \vec{v}}{\vec{w}_1 \cdot \vec{w}_1} \vec{w}_1 + \frac{\vec{w}_2 \cdot \vec{v}}{\vec{w}_2 \cdot \vec{w}_2} \vec{w}_2$$

$$= \frac{10}{5} \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} + \frac{9}{6} \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$$

$$= 2 \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} + \frac{3}{2} \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} 11/2 \\ -1 \\ 3/2 \end{bmatrix} \leftarrow \text{best approximation of } \vec{v} \text{ in } W$$

Euclidean distance:

$$\vec{v} - \text{proj}_W(\vec{v}) = \begin{bmatrix} 5 \\ 0 \\ 4 \end{bmatrix} - \begin{bmatrix} 11/2 \\ -1 \\ 3/2 \end{bmatrix}$$

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

$$\| \vec{v} - \text{proj}_W(\vec{v}) \| = \sqrt{1/4 + 1 + 25/4} = \underline{\sqrt{30}}$$

$$\| \vec{v} - \text{proj}_W(\vec{v}) \| = \sqrt{1/4 + 1 + 25/4} = \frac{\sqrt{30}}{2}$$

least squares approximation

consider a system $A\vec{x} = \vec{b}$:

$$A = [\vec{a}_1 \mid \vec{a}_2 \mid \vec{a}_3 \mid \dots \mid \vec{a}_n]$$

$\Leftarrow$ A is made up of a bunch of column vectors $\vec{a}_i$

then $A\vec{x} = [\vec{a}_1 \mid \vec{a}_2 \mid \dots \ \vec{a}_n] \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$

$$= x_1 \vec{a}_1 + x_2 \vec{a}_2 + x_3 \vec{a}_3 + \dots \ x_n \vec{a}_n$$

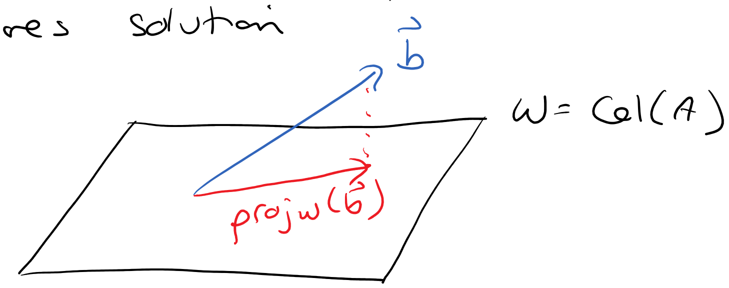
this is a linear combination of the columns of A

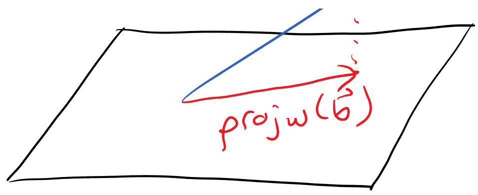
if $A\vec{x} = \vec{b}$ has a solution, the $\vec{b}$ belongs to Col(A)

if $A\vec{x} = \vec{b}$ has no solution, then $\vec{b}$ does not belong to Col(A) and the closest we can come is

$$A\vec{x} = \text{proj}_W(\vec{b})$$

which <u>will</u> have a solution, called the least squares solution





$W = \text{Col}(A)$

the reason this is called "least squares" is because it minimizes

$$\| \vec{b} - A\vec{x} \|^2$$

then

$$A\vec{x}_{LS} = \text{proj}_W(\vec{b})$$

the $\vec{x}$ that minimizes the distance between the actual $\vec{b}$ and the column space of A

but then

$$\underbrace{\vec{b} - \text{proj}_W(\vec{b})}_{\text{perp}_W(\vec{b})} = \underbrace{\vec{b} - A\vec{x}_{LS}}_{\text{must be orthogonal to } W \text{ where } W = \text{Col}(A)}$$

so that $\vec{b} - A\vec{x}_{LS}$ is in null (A^T)

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

$$A^T\vec{b} - A^TA\,\vec{x}_{LS} = \vec{0}$$

$$A^TA\,\vec{x}_{LS} = A^T\vec{b}$$

and if A has linearly independent columns, then (A^TA) is invertible and

$$\boxed{\vec{x}_{LS} = (A^T A)^{-1} A^T \vec{b}}$$

on final exam formula sheet

---

example: Find the least-squares solution for the inconsistent system $A\vec{x} = \vec{b}$ where

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

Also, calculate the least squares error.

answer: $\vec{x}_{LS} = (A^T A)^{-1} A^T \vec{b}$

$$A^T A = \underset{2\times 3}{\begin{bmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{bmatrix}} \underset{3\times 2}{\begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix}} = \begin{bmatrix} 17 & 1 \\ 1 & 5 \end{bmatrix}$$

$$(A^T A)^{-1} = \frac{1}{84} \begin{bmatrix} 5 & -1 \\ -1 & 17 \end{bmatrix}$$

$$A^T \vec{b} = \underset{2\times 3}{\begin{bmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{bmatrix}} \underset{3\times 1}{\begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix}} = \begin{bmatrix} 19 \\ 11 \end{bmatrix}$$

$$\text{so} \quad \vec{x}_{LS} = (A^T A)^{-1} A^T \vec{b}$$

$$= \frac{1}{84} \begin{bmatrix} 5 & -1 \\ -1 & 17 \end{bmatrix} \begin{bmatrix} 19 \\ 11 \end{bmatrix}$$

$$= \frac{1}{84} \begin{bmatrix} 84 \\ 168 \end{bmatrix}$$

$$= \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

$$= \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

now for the least squares error $\|\vec{b} - A\vec{x}_{LS}\|$

$$\vec{b} - A\vec{x}_{LS} = \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix} - \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

$$= \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix} - \begin{bmatrix} 4 \\ 4 \\ 3 \end{bmatrix}$$

$$= \begin{bmatrix} -2 \\ -4 \\ 8 \end{bmatrix}$$

$$\|\vec{b} - A\vec{x}_{LS}\| = \sqrt{(-2)^2 + (-4)^2 + 8^2} = \sqrt{84} = 2\sqrt{21}$$

Section 7.3: cont'd 2022/11/06

example: Find the least-squares solution to the following inconsistent system.

$$\begin{cases} x_1 - x_2 = 4 \\ 3x_1 + 2x_2 = 1 \\ -2x_1 + 4x_2 = 3 \end{cases}$$

answer: $A\vec{x} = \vec{b}$

$$\begin{bmatrix} 1 & -1 \\ 3 & 2 \\ -2 & 4 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 4 \\ 1 \\ 3 \end{bmatrix}$$

RREF of augmented matrix is

$$\left[\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{array}\right]$$

$$\vec{x}_{LS} = (A^TA)^{-1}A^T\vec{b}$$

$$A^TA = \underbrace{\begin{bmatrix} 1 & 3 & -2 \\ -1 & 2 & 4 \end{bmatrix}}_{2\times 3}\underbrace{\begin{bmatrix} 1 & -1 \\ 3 & 2 \\ -2 & 4 \end{bmatrix}}_{3\times 2} = \underbrace{\begin{bmatrix} 14 & -3 \\ -3 & 21 \end{bmatrix}}_{2\times 2}$$

$$A^T A = \begin{bmatrix} 1 & 3 & -2 \\ -1 & 2 & 4 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 3 & 2 \\ -2 & 4 \end{bmatrix} = \begin{bmatrix} 14 & -3 \\ -3 & 21 \end{bmatrix}$$

$$(A^T A)^{-1} = \frac{1}{285} \begin{bmatrix} 21 & 3 \\ 3 & 14 \end{bmatrix}$$

$$A^T \vec{b} = \begin{bmatrix} 1 & 3 & -2 \\ -1 & 2 & 4 \end{bmatrix} \begin{bmatrix} 4 \\ 1 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 \\ 10 \end{bmatrix}$$

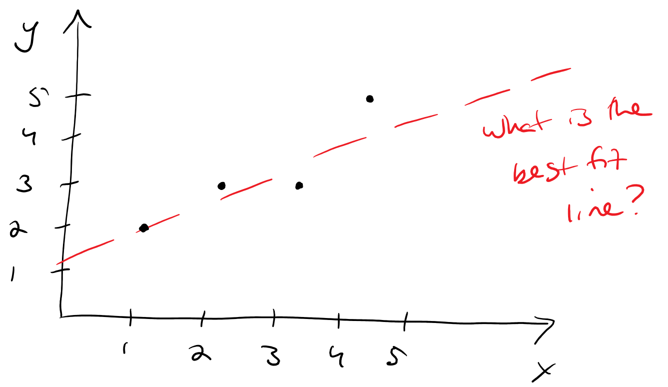
$$\vec{x}_{LS} = \frac{1}{285} \begin{bmatrix} 21 & 3 \\ 3 & 14 \end{bmatrix} \begin{bmatrix} 1 \\ 10 \end{bmatrix}$$

$$= \frac{1}{285} \begin{bmatrix} 51 \\ 143 \end{bmatrix} \approx \begin{bmatrix} 0.18 \\ 0.50 \end{bmatrix}$$

example: Find the least squares linear fit to the points $(1,2)$, $(2,3)$, $(3,3)$, $(4,5)$

answer:

x	y
1	2
2	3
3	3
4	5



$$\begin{cases} y_1 = m x_1 + b \\ y_2 = m x_2 + b \\ y_3 = m x_3 + b \\ y_4 = m x_4 + b \end{cases} \quad \Rightarrow \quad \begin{cases} 2 = 1m + b \\ 3 = 2m + b \\ 3 = 3m + b \\ 5 = 4m + b \end{cases}$$

$$\vec{y} = m\vec{x} + b$$

plug all pairs (x, y) into $y = mx + b$ to get these equations

then our system is

$$\begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} \overset{\text{slope}}{m} \\ b \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \\ 3 \end{bmatrix}$$

$$\underbrace{\begin{bmatrix} 2 & 1 \\ 3 & 1 \\ 4 & 1 \end{bmatrix}}_{A} \begin{bmatrix} \\ \underset{\uparrow}{b} \end{bmatrix} \qquad \underbrace{\begin{bmatrix} 3 \\ 3 \\ 5 \end{bmatrix}}_{\vec{b}}$$

y-int

$$\vec{x}_{LS} = (A^T A)^{-1} A^T \vec{b}$$

$$A^T A = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \\ 4 & 1 \end{bmatrix} = \begin{bmatrix} 30 & 10 \\ 10 & 4 \end{bmatrix}$$

$$(A^T A)^{-1} = \frac{1}{20} \begin{bmatrix} 4 & -10 \\ -10 & 30 \end{bmatrix} = \frac{1}{10} \begin{bmatrix} 2 & -5 \\ -5 & 15 \end{bmatrix}$$

$$A^T \vec{b} = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \\ 3 \\ 5 \end{bmatrix} = \begin{bmatrix} 37 \\ 13 \end{bmatrix}$$

$$\vec{x}_{LS} = \frac{1}{10} \begin{bmatrix} 2 & -5 \\ -5 & 15 \end{bmatrix} \begin{bmatrix} 37 \\ 13 \end{bmatrix}$$

$$= \frac{1}{10} \begin{bmatrix} 9 \\ 10 \end{bmatrix} = \begin{bmatrix} 9/10 \\ 1 \end{bmatrix} = \begin{bmatrix} \text{slope} \\ \text{y-int} \end{bmatrix}$$

so $\boxed{y = \frac{9}{10} x + 1}$

---

digression (not tested)

if you are fitting $y = mx + b$, then in general you set

$$\begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ x_3 & 1 \\ x_4 & 1 \end{bmatrix} \begin{bmatrix} m \\ b \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix}$$

but, if you want to fit $y = ax^2 + bx + c$

$$\begin{bmatrix} x_1^2 & x_1 & 1 \\ x_2^2 & x_2 & 1 \\ x_3^2 & x_3 & 1 \\ x_4^2 & x_4 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix}$$

then you'd solve the same way, but $(A^T A)$ is now a 3×3, and finding the inverse is more annoying