Chapter 6: Orthogonality and least squares

Goal: section 6.1 - 6.3
if time permits: sections 6.4 - 6.5

1) Inner product, length, and orthogonality

Definition: Let $\vec{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ and $\vec{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$ be vectors in $\mathbb{R}^n$.

The inner product (or dot product) of $\vec{x}$ and $\vec{y}$ is:

$$\vec{x} \cdot \vec{y} = \vec{x}^T \vec{y} = x_1 y_1 + \cdots + x_n y_n$$

Example: $u = \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}$ and $\vec{v} = \begin{pmatrix} -3 \\ -4 \\ 5 \end{pmatrix}$. Compute $u \cdot \vec{v}$, $\vec{v} \cdot \vec{u}$, $\vec{u} \cdot \vec{u}$ and $\vec{v} \cdot \vec{v}$

Solution

$$\vec{u} \cdot \vec{v} = (1\ 1\ 2)\begin{pmatrix} -3 \\ -4 \\ 5 \end{pmatrix} = (1)(-3) + (1)(-4) + (2)(5) = -3 - 4 + 10 = 3$$

$$\vec{v} \cdot \vec{u} = (-3\ -4\ 5)\begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix} = (-3)(1) + (-4)(1) + (5)(2) = -3 - 4 + 10 = 3$$

$$\vec{u} \cdot \vec{u} = (1\ 1\ 2)\begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix} = 1^2 + 1^2 + 2^2 = 6$$

$$\vec{v} \cdot \vec{v} = (-3\ -4\ 5)\begin{pmatrix} -3 \\ -4 \\ 5 \end{pmatrix} = (-3)^2 + (-4)^2 + 5^2 = 9 + 16 + 25 = 50$$

Theorem

Let $\vec{u}, \vec{v}, \vec{w} \in \mathbb{R}^n$. Let $c \in \mathbb{R}$. We have:

(a) $\vec{u} \cdot \vec{v} = \vec{v} \cdot \vec{u}$

(b) $(\vec{u} + \vec{v}) \cdot \vec{w} = \vec{u} \cdot \vec{w} + \vec{v} \cdot \vec{w}$

(c) $(c\vec{u}) \cdot \vec{v} = c(\vec{u} \cdot \vec{v}) = \vec{u} \cdot (c\vec{v})$

(d) $\vec{u} \cdot \vec{u} \geq 0$
$\vec{u} \cdot \vec{u} = 0$ if and only if $\vec{u} = \vec{0}$

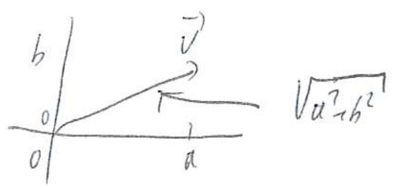
(e) $\vec{u} \cdot \vec{0} = 0$

Definition:

Let $\vec{v} \in \mathbb{R}^n$. The length (or norm) of $\vec{v}$ is the nonnegative scalar $\|\vec{v}\|$ defined by:

$$\|\vec{v}\| = \sqrt{\vec{v} \cdot \vec{v}} = \sqrt{v_1^2 + \dots + v_n^2}$$

For $\vec{v} = \begin{pmatrix} a \\ b \end{pmatrix}$, $\|\vec{v}\| = \sqrt{a^2 + b^2}$ is the standard definition of length of a segment from $(0,0)$ to (a,b)

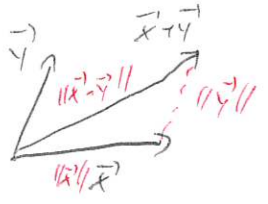


Theorem: Let $\vec{u}, \vec{v} \in \mathbb{R}^n$ and $c \in \mathbb{R}$.

(a) $\|\vec{u}\| \geq 0$ and $\|\vec{u}\| = 0$ if and only if $\vec{u} = \vec{0}$

(b) $\|c\vec{u}\| = |c| \, \|\vec{u}\|$

(c) $\|\vec{u} + \vec{v}\| \leq \|\vec{u}\| + \|\vec{v}\|$ (triangle inequality)



length of any side of triangle cannot exceed the sum of lengths of the two other sides

Definition:

For $\vec{u}, \vec{v} \in \mathbb{R}^n$, the distance between $\vec{u}$ and $\vec{v}$ is the length of $\vec{u}-\vec{v}$.

$$\text{dist}(\vec{u},\vec{v}) = \|\vec{u}-\vec{v}\| = \|\vec{v}-\vec{u}\| = \text{dist}(\vec{v},\vec{u})$$

Example:

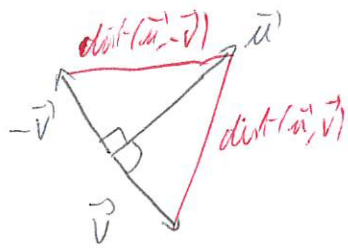
Compute the distance between $\vec{u} = \begin{pmatrix} 1 \\ -2 \\ 3 \end{pmatrix}$ and $\vec{v} = \begin{pmatrix} 0 \\ 4 \\ 5 \end{pmatrix}$

Solution.

$$\text{dist}(\vec{u},\vec{v}) = \|\vec{u}-\vec{v}\| = \sqrt{(1-0)^2+(-2-4)^2+(3-5)^2} = \sqrt{1^2+(-6)^2+(-2)^2} = \sqrt{41}$$

Orthogonal vectors

When are two vectors orthogonal / perpendicular?



$\vec{u}$ and $\vec{v}$ are $\perp$ $\iff$ $\text{dist}(u,v) = \text{dist}(u,-v)$

$\iff$ $\text{dist}(u,v)^2 = \text{dist}(u,-v)^2$

$$\begin{aligned}\text{dist}(\vec{u},-\vec{v})^2 &= \|\vec{u}-(-\vec{v})\|^2 = \|\vec{u}+\vec{v}\|^2 \\ &= (\vec{u}+\vec{v})\cdot(\vec{u}+\vec{v}) = \vec{u}\cdot\vec{u} + \vec{u}\cdot\vec{v} + \vec{v}\cdot\vec{u} + \vec{v}\cdot\vec{v} \\ &= \|\vec{u}\|^2 + 2\vec{u}\cdot\vec{v} + \|\vec{v}\|^2\end{aligned}$$

Similarly,

$$\text{dist}(\vec{u},\vec{v})^2 = \|\vec{u}-\vec{v}\|^2 = (\vec{u}-\vec{v})\cdot(\vec{u}-\vec{v}) = \cdots = \|\vec{u}\|^2 - 2\vec{u}\cdot\vec{v} + \|\vec{v}\|^2$$

Thus,

$$\text{dist}(\vec{u},\vec{v}) = \text{dist}(\vec{u},-\vec{v}) \iff 2\vec{u}\cdot\vec{v} = -2\vec{u}\cdot\vec{v}$$

$$\iff \vec{u}\cdot\vec{v} = 0$$

Definition

Two vectors $\vec{u}, \vec{v} \in \mathbb{R}^2$ are orthogonal if $\vec{u} \cdot \vec{v} = 0$

Theorem (The Pythagorean theorem)

Two vectors $\vec{u}, \vec{v} \in \mathbb{R}^n$ are orthogonal if and only if $\|\vec{u} + \vec{v}\|^2 = \|\vec{u}\|^2 + \|\vec{v}\|^2$

2) Orthogonal sets

Section 6-2 not covered.

Definitions

- $S = \{\vec{u}_1, \dots, \vec{u}_p\}$ is an orthogonal set if $\vec{u}_i \cdot \vec{u}_j = 0$ for $i \neq j$
- $S = \{\vec{u}_1, \dots, \vec{u}_p\}$ is an orthogonal basis if (for a subspace W) $\left|\begin{array}{l} S \text{ basis for } W \\ \text{and} \\ S \text{ orthogonal set.} \end{array}\right.$
- $W^\perp = \{z \in V \mid z \cdot w = 0 \text{ for all } w \in W\}$ with W subspace of V

3) Orthogonal projections

$\perp$ (notation)

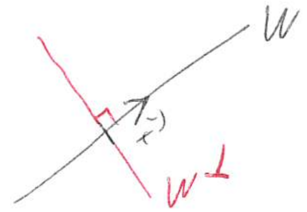
Definition

Let W be a subspace of $\mathbb{R}^n$. A vector $\vec{z} \in \mathbb{R}^n$ is said to be orthogonal to W if $\vec{z}$ is $\perp$ to every vector in W. The set of all vectors $\vec{z}$ that are $\perp$ to W is called the orthogonal complement of W and is denoted by $W^\perp$.

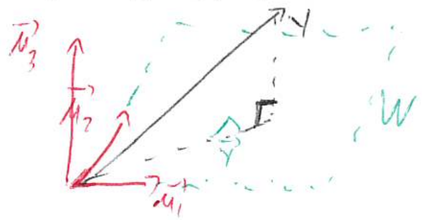
$$W^\perp = \{\vec{z} \in \mathbb{R}^n \mid \vec{z} \cdot \vec{x} = 0, \ \forall \vec{x} \in W\}$$

Examples:

(*) Let $\vec{x} \in \mathbb{R}^2$ and let $W = \operatorname{span}\{\vec{x}\}$



(*) Let $\vec{u}_1, \vec{u}_2, \vec{u}_3$ be an orthogonal basis for $\mathbb{R}^3$ and let $W = \operatorname{span}\{\vec{u}_1, \vec{u}_2\}$



For any $\vec{y} \in \mathbb{R}^3$, we can write:

$$\vec{y} = \underbrace{\left(\frac{\vec{y}\cdot\vec{u}_1}{\vec{u}_1\cdot\vec{u}_1}\right)\vec{u}_1 + \left(\frac{\vec{y}\cdot\vec{u}_2}{\vec{u}_2\cdot\vec{u}_2}\right)\vec{u}_2}_{(\text{in } W) \;=\; \hat{\vec{y}}} + \underbrace{\left(\frac{\vec{y}\cdot\vec{u}_3}{\vec{u}_3\cdot\vec{u}_3}\right)\vec{u}_3}_{(\text{in } W^\perp) \;=\; \vec{z}}$$

Theorem: (The orthogonal Decomposition Theorem)

Let W be a subspace of $\mathbb{R}^n$. Then each $\vec{y} \in \mathbb{R}^n$ can be uniquely decomposed as:

$$\vec{y} = \hat{\vec{y}} + \vec{z} \quad \text{where } \hat{\vec{y}} \in W \text{ and } \vec{z} \in W^\perp.$$

If $\{\vec{u}_1, \dots, \vec{u}_p\}$ is an orthogonal basis for W, then:

$$\hat{\vec{y}} = \operatorname{proj}_W \vec{y} = \left(\frac{\vec{y}\cdot\vec{u}_1}{\vec{u}_1\cdot\vec{u}_1}\right)\vec{u}_1 + \dots + \left(\frac{\vec{y}\cdot\vec{u}_p}{\vec{u}_p\cdot\vec{u}_p}\right)\vec{u}_p$$

(orthogonal projection of y onto W)

and

$$\vec{z} = \vec{y} - \hat{\vec{y}}$$

<u>Example:</u>

Let $\vec{u}_1 = \begin{pmatrix} 3 \\ 0 \\ 1 \end{pmatrix}$, $\vec{u}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$ and $\vec{y} = \begin{pmatrix} 0 \\ 3 \\ 10 \end{pmatrix}$

Note that $\{\vec{u}_1, \vec{u}_2\}$ is an $\perp$ basis for $W = \text{span}\{\vec{u}_1, \vec{u}_2\}$.

Write $\vec{y}$ as a sum of a vector in W and a vector in $W^{\perp}$.

<u>Solution:</u>

$$\rightarrow \hat{\vec{y}} = \text{proj}_W(\vec{y}) = \frac{\vec{y}\cdot\vec{u}_1}{\vec{u}_1\cdot\vec{u}_1}\vec{u}_1 + \frac{\vec{y}\cdot\vec{u}_2}{\vec{u}_2\cdot\vec{u}_2}\vec{u}_2$$

$$= \frac{0+0+10}{9+0+1}\begin{pmatrix} 3 \\ 0 \\ 1 \end{pmatrix} + \frac{0+3+0}{0+1+0}\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$$

$$= \begin{pmatrix} 3 \\ 0 \\ 1 \end{pmatrix} + \begin{pmatrix} 0 \\ 3 \\ 0 \end{pmatrix} = \begin{pmatrix} 3 \\ 3 \\ 1 \end{pmatrix} \checkmark$$

$$\rightarrow \vec{z} = \vec{y} - \hat{\vec{y}} = \begin{pmatrix} 0 \\ 3 \\ 10 \end{pmatrix} - \begin{pmatrix} 3 \\ 3 \\ 1 \end{pmatrix} = \begin{pmatrix} -3 \\ 0 \\ 9 \end{pmatrix} \checkmark$$

<u>Theorem</u> (The Best Approximation Theorem)

Let W be a subspace of $\mathbb{R}^n$, $\vec{y}$ any vector in $\mathbb{R}^n$ and $\hat{y}$ the orthogonal projection of $\vec{y}$ onto W. Then $\hat{\vec{y}}$ is the point in W closest to $\vec{y}$ in the sense that:

$$\text{dist}(\vec{y}, \hat{\vec{y}}) = \|\vec{y} - \hat{\vec{y}}\| \leq \|\vec{y} - \vec{v}\| = \text{dist}(\vec{y}, \vec{v}) \text{ for any } \vec{v} \text{ in } W.$$

Example:

Find the closest point to $\vec{y}$ in $W = \text{span}\{\vec{u}_1, \vec{u}_2\}$ where

$$\vec{y} = \begin{pmatrix} 2 \\ 4 \\ 0 \\ -2 \end{pmatrix}, \quad \vec{u}_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} \quad \text{and} \quad \vec{u}_2 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \end{pmatrix}$$

What is the shortest distance from $\vec{y}$ to W?

Solution:

Closest point to $\vec{y}$ in W is:

$$\hat{\vec{y}} = \text{proj}_W \vec{y} = \frac{\vec{y} \cdot \vec{u}_1}{\vec{u}_1 \cdot \vec{u}_1} \vec{u}_1 + \frac{\vec{y} \cdot \vec{u}_2}{\vec{u}_2 \cdot \vec{u}_2} \vec{u}_2$$

$$= \frac{2+4+0+0}{1+1+0+0} \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} + \frac{0+0+0-2}{0+0+1+1} \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \end{pmatrix}$$

$$= \begin{pmatrix} 3 \\ 3 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ -1 \\ -1 \end{pmatrix} = \begin{pmatrix} 3 \\ 3 \\ -1 \\ -1 \end{pmatrix}$$

The shortest distance is:

$$\|\vec{y} - \hat{\vec{y}}\| = \sqrt{(-1)^2 + 1^2 + 1^2 + (-1)^2} = \sqrt{4} = 2$$

$$\shortparallel$$

$$\sqrt{(2-3)^2 + (4-3)^2 + (0-(-1))^2 + (-2-(-1))^2}$$

Least Squares

Consider a $m \times n$ matrix A. The linear system $A\vec{x}=\vec{b}$ is consistent if and only if $\vec{b} \in Col(A)$

If $A\vec{x}=\vec{b}$ is inconsistent, we may want to find $\vec{x}$ such that $A\vec{x}$ is as close to $\vec{b}$ as possible.

Definition:

The least square solution of $A\vec{x}=\vec{b}$ is a vector $\hat{\vec{x}}$ such that:

$$dist(A\hat{\vec{x}}, \vec{b}) \leq dist(A\vec{x}, \vec{b}) \quad \text{for all } \vec{x} \text{ in } \mathbb{R}^n$$

The least square solution is given by:

$$\hat{x} = proj_{Col(A)} \vec{b} \iff \text{equivalent to } b - A\hat{x} \text{ being orthogonal to } Col(A)$$

$$\iff \vec{a}_i \cdot (\vec{b} - A\hat{\vec{x}}) = 0 \quad i=1,\dots,n \quad A=(\vec{a}_1 \cdots \vec{a}_n)$$

$$\iff \vec{a}_i^{\,T}(\vec{b} - A\hat{\vec{x}}) = 0 \quad i=1,\dots,n$$

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

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

(least squares not included in final)

4) The Gram-Schmidt Process

This process is a simple algorithm that produces an orthogonal basis for any nonzero subspace of $\mathbb{R}^n$.

Example:

Let $W = \text{span}\{\vec{x}_1, \vec{x}_2\}$ with $\vec{x}_1 = \begin{pmatrix} 3 \\ 6 \\ 0 \end{pmatrix}$ and $\vec{x}_2 = \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix}$ so $\{\vec{x}_1, \vec{x}_2\}$ basis for W.

goal: construct orthogonal basis $\{\vec{v}_1, \vec{v}_2\}$

idea: keep first vector $\vec{x}_1$ then construct a vector $\vec{v}_2$ orthogonal to $\vec{v}_1 = \vec{x}_1$ such that $\text{span}\{\vec{v}_1, \vec{v}_2\} = \text{span}\{\vec{v}_1, \vec{x}_2\}$ $(= W)$

$\Rightarrow$ set $\vec{v}_2 = \vec{x}_2 - \text{proj}_{V_1} \vec{x}_2$ with $V_1 = \text{span}\{\vec{v}_1\}$

$\Rightarrow$ $\vec{v}_2$ in $V_1^{\perp}$ so $\vec{v}_1 \cdot \vec{v}_2 = 0$ $\Rightarrow$ $(\vec{v}_1, \vec{v}_2)$ basis for W ($\dim W = 2$)

and $\vec{v}_2$ lin. indep with $\vec{v}_1$ (if $\vec{v}_2 \neq \vec{0}$) $\Rightarrow$ $(\vec{v}_1, \vec{v}_2)$ $\perp$ basis for W.

$$\vec{v}_2 = \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix} - \frac{\vec{x}_2 \cdot \vec{x}_1}{\vec{x}_1 \cdot \vec{x}_1} x_1 = \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix} - \frac{15}{45} \begin{pmatrix} 3 \\ 6 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 2 \end{pmatrix}$$

⚠ Same idea can be applied to subspace of dimension n, i.e. subspace with a basis that contains n vectors.

Theorem (The Gram-Schmidt Process)

Let W be a nonzero subspace of $\mathbb{R}^n$ and let $\{x_1, \dots, x_p\}$ be a basis for W.

Define:

$$\vec{v}_1 = \vec{x}_1$$

$$\vec{v}_2 = \vec{x}_2 - \frac{\vec{x}_2 \cdot \vec{v}_1}{\vec{v}_1 \cdot \vec{v}_1}\vec{v}_1 = \vec{x}_2 - \text{proj}_{\text{span}\{\vec{v}_1\}}\vec{x}_2$$

$$\vec{v}_3 = \vec{x}_3 - \left(\frac{\vec{x}_3 \cdot \vec{v}_1}{\vec{v}_1 \cdot \vec{v}_1}\vec{v}_1 + \frac{\vec{x}_3 \cdot \vec{v}_2}{\vec{v}_2 \cdot \vec{v}_2}\vec{v}_2\right) = \vec{x}_3 - \text{proj}_{\text{span}\{\vec{v}_1, \vec{v}_2\}}\vec{x}_3$$

$$\vdots$$

$$\vec{v}_p = \vec{x}_p - \left(\frac{\vec{x}_p \cdot \vec{v}_1}{\vec{v}_1 \cdot \vec{v}_1}\vec{v}_1 + \frac{\vec{x}_p \cdot \vec{v}_2}{\vec{v}_2 \cdot \vec{v}_2}\vec{v}_2 + \cdots + \frac{\vec{x}_p \cdot \vec{v}_{p-1}}{\vec{v}_{p-1} \cdot \vec{v}_{p-1}}\vec{v}_{p-1}\right) = \vec{x}_p - \text{proj}_{\text{span}\{\vec{v}_1, \dots, \vec{v}_{p-1}\}}\vec{x}_p$$

Then $\{v_1, \dots, v_p\}$ is an orthogonal basis for W. In addition, we have:

$$\text{span}\{\vec{v}_1, \dots, \vec{v}_k\} = \text{span}\{\vec{x}_1, \dots, \vec{x}_k\} \text{ for any } 1 \le k \le p.$$

Example:

Let $\vec{x}_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$, $\vec{x}_2 = \begin{pmatrix} 0 \\ 1 \\ 1 \\ 1 \end{pmatrix}$, $\vec{x}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \end{pmatrix}$ and $W = \text{span}\{\vec{x}_1, \vec{x}_2, \vec{x}_3\}$ subspace of $\mathbb{R}^4$.

x_1, x_2, x_3 lin. indep. $\Rightarrow \{\vec{x}_1, \vec{x}_2, \vec{x}_3\}$ basis for W.

Construct an orthogonal basis for W using Gram-Schmidt process.

Solution:

Step 1: $\vec{v_1} = \vec{x_1} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$ and $W_1 = \text{span}\{\vec{v_1}\}$

Step 2: $\vec{v_2} = \vec{x_2} - \text{proj}_{W_1} \vec{x_2} = \vec{x_2} - \dfrac{\vec{x_2} \cdot \vec{v_1}}{\vec{v_1} \cdot \vec{v_1}} \vec{v_1}$

$$= \begin{pmatrix} 0 \\ 1 \\ 1 \\ 1 \end{pmatrix} - \frac{0+1+1+1}{1+1+1+1} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 1 \\ 1 \end{pmatrix} - \frac{3}{4} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} -3/4 \\ 1/4 \\ 1/4 \\ 1/4 \end{pmatrix}$$

(can replace $\vec{v_2}$ by $\begin{pmatrix} -3 \\ 1 \\ 1 \\ 1 \end{pmatrix}$ if we want).

$\rightarrow \vec{v_2} = \begin{pmatrix} -3 \\ 1 \\ 1 \\ 1 \end{pmatrix}$ and $W_2 = \text{span}\{\vec{v_1}, \vec{v_2}\}$ ($\vec{v_1}, \vec{v_2}$ $\perp$ basis of W_2.)

Step 3. $\vec{v_3} = \vec{x_3} - \text{proj}_{W_2} \vec{x_3} = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \end{pmatrix} - \dfrac{\vec{x_3} \cdot \vec{v_1}}{\vec{v_1} \cdot \vec{v_1}} \vec{v_1} - \dfrac{\vec{x_3} \cdot \vec{v_2}}{\vec{v_2} \cdot \vec{v_2}} \vec{v_2}$

$$= \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \end{pmatrix} - \frac{2}{4} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} - \frac{2}{12} \begin{pmatrix} -3 \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

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

$\{\vec{v_1}, \vec{v_2}, \vec{v_3}\}$ $\perp$ basis for W.

⊲ changing $\begin{pmatrix} -\frac{3}{4} \\ \frac{1}{4} \\ \frac{1}{4} \\ \frac{1}{4} \end{pmatrix}$ into $\begin{pmatrix} -3 \\ 1 \\ 1 \\ 1 \end{pmatrix}$ does not change the span $\{v_1, v_2\}$

and does not change that $\vec{v}_1 \cdot \vec{v}_2 = 0$ (dot product = 0 multiplied by 4 so still 0)

5) Least Squares

see page 8 of lecture notes.

Not included in final.