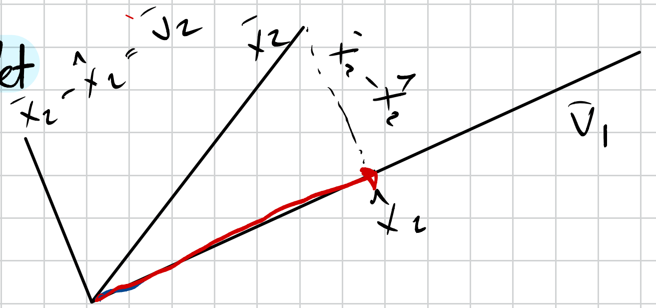
8. Orthogonality

Gram Schmidt Proces (GSP):

Given a basis, GSP provides a method for obtaining an orthogonal basis for any non-zero subspace W:

Assume a basis $\{\bar{x}_1 \dots \bar{x}_p\}$, let



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

$$\bar{v}_2 = \bar{x}_2 - \frac{\bar{x}_2 \cdot \bar{v}_1}{\bar{v}_1 \cdot \bar{v}_1} \cdot \bar{v}_1$$

$$\bar{v}_3 = \bar{x}_3 - \frac{\bar{x}_3 \cdot \bar{v}_1}{\bar{v}_1 \cdot \bar{v}_1} \bar{v}_1 - \frac{\bar{x}_3 \cdot \bar{v}_2}{\bar{v}_2 \cdot \bar{v}_2} \bar{v}_2$$

$$\vdots$$

$$\bar{v}_p = \bar{x}_p - \frac{x_p \cdot \bar{v}_1}{\bar{v}_1 \cdot \bar{v}_1} \cdot \bar{v}_1 - \frac{x_p \cdot \bar{v}_2}{\bar{v}_2 \cdot \bar{v}_2} \bar{v}_2 - \dots - \frac{x_p \cdot \bar{v}_{p-1}}{\bar{v}_{p-1} \cdot \bar{v}_{p-1}} \cdot \bar{v}_{p-1}$$

Then $\{\bar{v}_1 \dots \bar{v}_p\}$ is an orthogonal set!

$$\bar{v}_p = \bar{x}_p - \text{proj}_{\bar{v}_1}(\bar{x}_p) - \text{proj}_{\bar{v}_2}(\bar{x}_p) - \dots - \text{proj}_{\bar{v}_{p-1}}(\bar{x}_p)$$

$$\boxed{= \bar{x}_p - \left(\sum_{i=1}^{p-1} \text{proj}_{\bar{v}_i}(x_p)\right) = \bar{x}_p - \left(\sum_{i=1}^{p-1} \frac{\bar{x}_p \cdot \bar{v}_i}{\bar{v}_i \cdot \bar{v}_i} \bar{v}_i\right)}$$

Ex:

$$\bar{x}_1 = \begin{bmatrix} 3 \\ 6 \\ 0 \end{bmatrix}, \quad \bar{x}_2 = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$$

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

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

EX:

$$\bar{x}_1 = \begin{bmatrix}1\\1\\1\\1\end{bmatrix}, \quad \bar{x}_2 = \begin{bmatrix}0\\1\\1\\1\end{bmatrix}, \quad \bar{x}_3 = \begin{bmatrix}0\\0\\1\\1\end{bmatrix}$$

Using Python.

```
x1 = Matrix([[1],[1],[1],[1]])
x2 = Matrix([[0],[1],[1],[1]])
x3 = Matrix([[0],[0],[1],[1]])
```
✓ 0.0s

```
v1 = x1
v2 = x2 - x2.project(v1)
v3 = x3 - x3.project(v1) - x3.project(v2)
v1, v2, v3
```
✓ 0.0s

$$\left(\begin{bmatrix}1\\1\\1\\1\end{bmatrix}, \begin{bmatrix}-\frac{3}{4}\\\frac{1}{4}\\\frac{1}{4}\\\frac{1}{4}\end{bmatrix}, \begin{bmatrix}0\\-\frac{2}{3}\\\frac{1}{3}\\\frac{1}{3}\end{bmatrix}\right)$$

Ex

Assignment 4
Let

$$u_1 = \begin{bmatrix}1\\2\\1\\1\end{bmatrix}, u_2 = \begin{bmatrix}-2\\1\\-1\\1\end{bmatrix}, u_3 = \begin{bmatrix}1\\1\\-2\\-1\end{bmatrix}, u_4 = \begin{bmatrix}-1\\1\\1\\-1\end{bmatrix}, and\ v = \begin{bmatrix}4\\2\\-1\\0\end{bmatrix},$$

a) Show that $\{u_1, u_2, u_3, u_4\}$ is an orthogonal basis for $\mathbb{R}^4$
b) Write v as the sum of two vectors, one in $span\{u_1, u_2\}$ and the other in $span\{u_3, u_4\}$.
c) Determine the (shortest) distance between v and the subspace spanned by $\{u_1, u_2, u_3\}$

Using Pyton

```
u1 = Matrix([[1],[2],[1],[1]])
u2 = Matrix([[-2],[1],[-1],[1]])
u3 = Matrix([[1],[1],[-2],[-1]])
u4 = Matrix([[-1],[1],[1],[-1]])

v = Matrix([[4],[2],[-1],[0]])

U = Matrix.hstack(u1, u2, u3, u4)
U.T * U , U.rref()
```
✓ 0.0s

$$\left(\begin{bmatrix}7&0&0&1\\0&7&0&1\\0&0&7&-1\\1&1&-1&4\end{bmatrix}, \left(\begin{bmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1\end{bmatrix}, (0, 1, 2, 3)\right)\right)$$

```
# indpendent, not orthogonal
u1, u2, u3, u4 = GramSchmidt((u1, u2, u3, u4))
```
✓ 0.0s

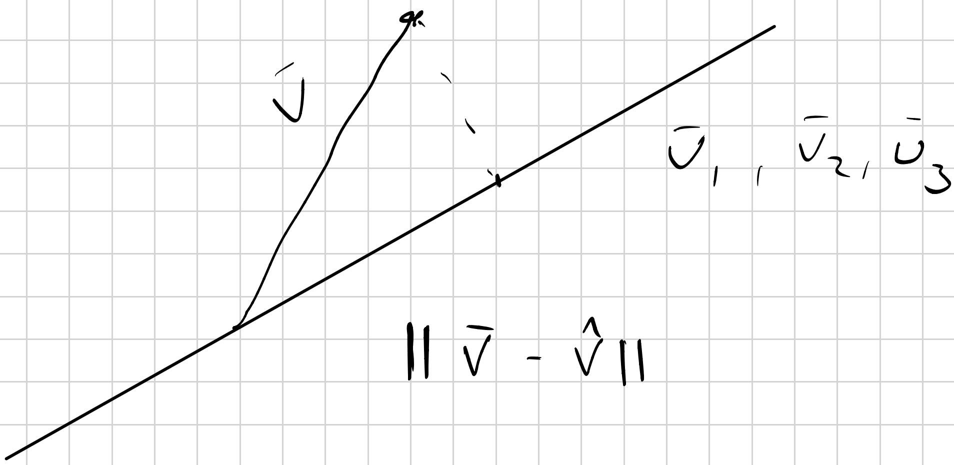
```
z1 = v.project(u1) + v.project(u2)
z2 = v.project(u3) + v.project(u4)

z1, z2
```
✓ 0.0s

[illegible]

```
((v - (v.project(u1) + v.project(u2) + v.project(u3))).norm(),
float((v - (v.project(u1) + v.project(u2) + v.project(u3))).norm()))
```
✓ 0.0s

$$\left(\frac{3\sqrt{7}}{7}, 1.13389341902768\right)$$

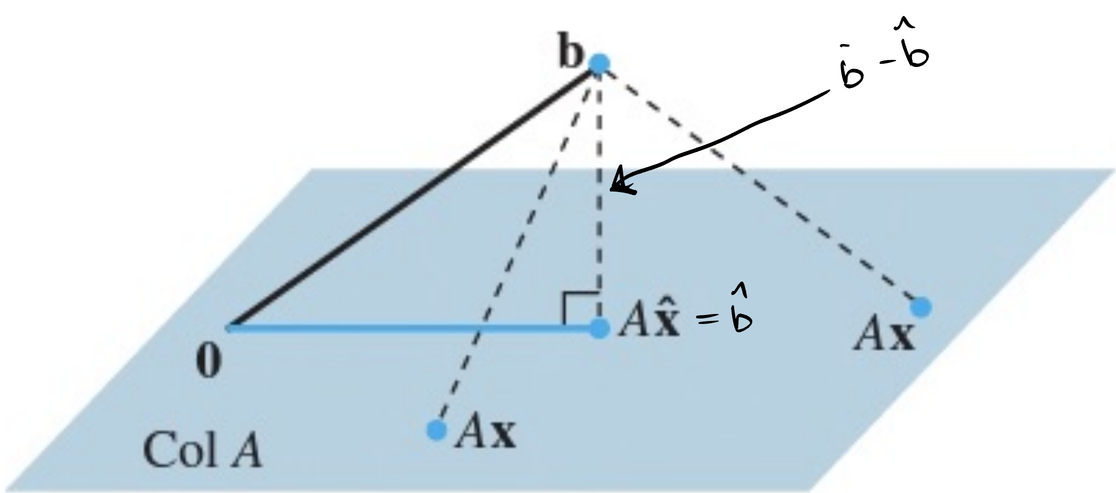


$$\|\bar{v} - \hat{v}\|$$

Least Squares Problem: Best Approximation

Problem: Not all $A\bar{x} = \bar{b}$ have solutions

Solution: Find the best approximation $\hat{b} \in \text{Col } A : \exists \hat{x}$ s.t. $A\hat{x} = \hat{b}$



$\hat{x}$ is called the least squares solution to $A\bar{x} = \bar{b}$:

$$\|\bar{b} - A\hat{x}\| \leq \|b - A\bar{x}\|$$

$A = [\bar{a}_1, \bar{a}_2, \dots \bar{a}_n]$

$\bar{a}_i \cdot (\bar{b} - \hat{b}) = 0$

$$A^T = \begin{bmatrix} — & \bar{a}_1 & — \\ — & a_2 & — \\ & \vdots & \\ — & a_n & — \end{bmatrix}, \quad \begin{aligned} A^T \cdot (\bar{b} - \hat{b}) &= \bar{0} \\ A^T \cdot (\bar{b} - A\hat{x}) &= \bar{0} \\ A^T \cdot \bar{b} - A^T \cdot A\hat{x} &= \bar{0} \\ A^T A \hat{x} &= A^T \bar{b} \end{aligned}$$

$$[A^T A \quad A^T \bar{b}] \overset{RREF}{\sim} [I \quad \hat{x}]$$

$$A^T A \hat{x} = A^T \bar{b} \Rightarrow (A^T A)^{-1} (A^T A) \hat{x} = (A^T A)^{-1} \cdot (A^T \bar{b})$$

$$\hat{x} = (A^T A)^{-1} (A^T \bar{b}) \Big\} \text{Normal Equations}$$

Method: Finding least squares

1) Find A^TA
2) Find A^Tb
3) Solve $[A^TA \;\; A^Tb]$
4) state solution, $\hat{x}$

Ex

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \\ 0 & 1 \end{bmatrix}, \quad \bar{b} = \begin{bmatrix} 3 \\ 3 \\ 2 \end{bmatrix}, \quad [A \;\; \bar{b}] \overset{RREF}{\sim} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

1) $$A^TA = \begin{bmatrix} 1 & 2 & 0 \\ 2 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 2 & 1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 5 & 4 \\ 4 & 6 \end{bmatrix}$$

2) $$A^T\bar{b} = \begin{bmatrix} 1 & 2 & 0 \\ 2 & 1 & 1 \end{bmatrix} \begin{bmatrix} 3 \\ 3 \\ 2 \end{bmatrix} = \begin{bmatrix} 9 \\ 11 \end{bmatrix}$$

3) $$\begin{bmatrix} 5 & 4 & 9 \\ 4 & 6 & 11 \end{bmatrix} \overset{RREF}{\sim} \begin{bmatrix} 1 & 0 & 5/7 \\ 0 & 1 & 19/14 \end{bmatrix}$$

4) $$\hat{x} = \begin{bmatrix} 5/7 \\ 19/14 \end{bmatrix}$$

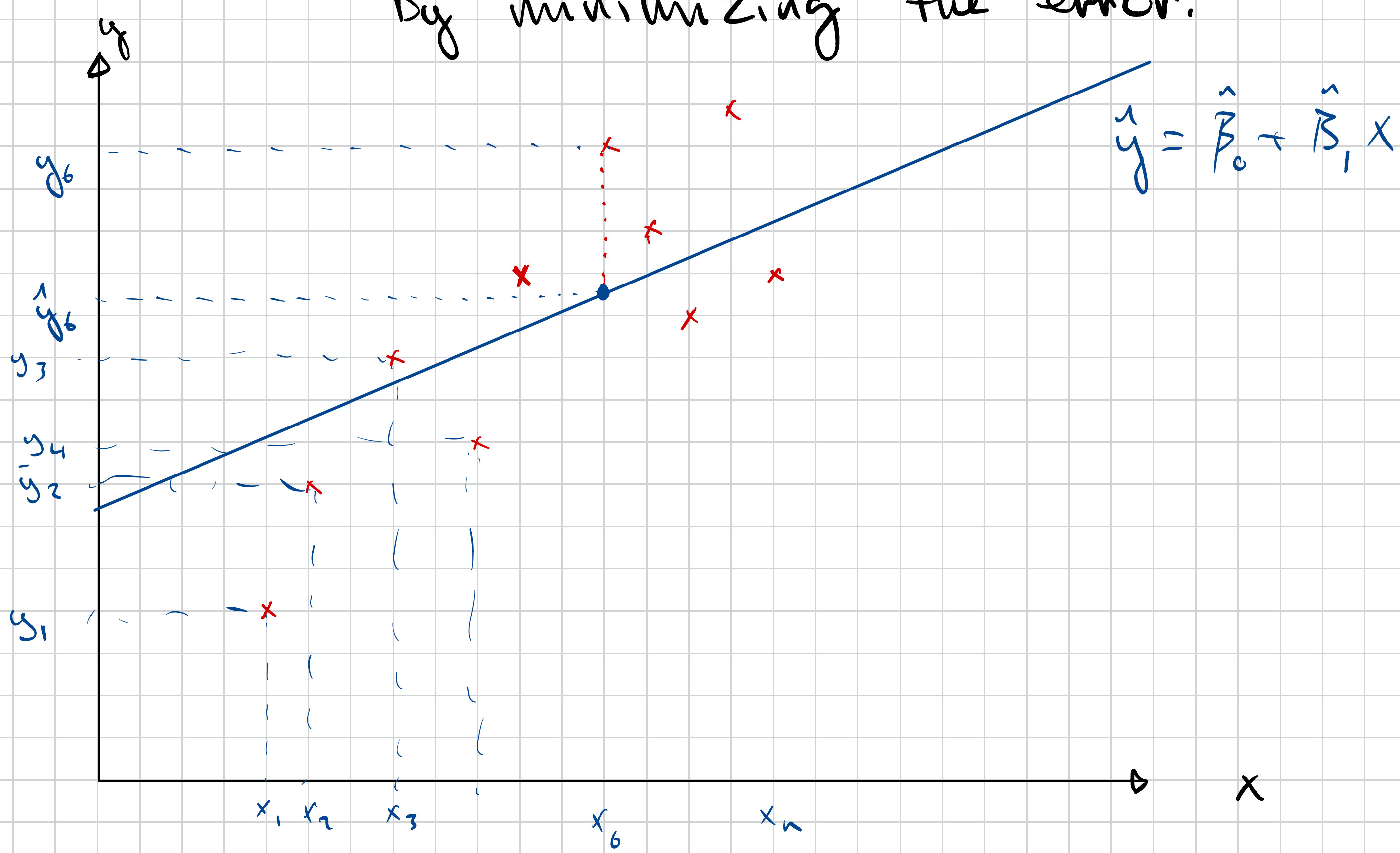
We know that $$\underbrace{\begin{bmatrix} 1 & 2 \\ 2 & 1 \\ 0 & 1 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} 5/7 \\ 19/14 \end{bmatrix}}_{\hat{x}} = \underbrace{\begin{bmatrix} 24/7 \\ 39/14 \\ 19/14 \end{bmatrix}}_{\hat{b}}$$ is

the closest point to $\bar{b}$ in Col A

Linear Models:

Problem: Real life problems rarely follow mathematical functions perfectly.

Solution: Use least square to find the function/model that yields the best approximation of the observation by minimizing the error.



Predicted y-value, $\hat{y}$	observed y-value, $\bar{y}$
$\hat{y}_1 = \hat{\beta}_0 + \hat{\beta}_1 x_1$	y_1
$\hat{y}_2 = \hat{\beta}_0 + \hat{\beta}_1 x_2$	y_2
$\vdots$	$\vdots$
$\hat{y}_n = \hat{\beta}_0 + \hat{\beta}_1 x_2$	y_n

$$X\hat{\beta} = \hat{y}$$

$$\boxed{X\beta \approx \bar{y}}$$

$$\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$$

Design Matrix

$$\hat{y} = \beta_0 + \beta_1 \bar{x} + \beta_2 \bar{x}^2 + \beta_3 \bar{x}^3$$

$$\hat{y} = \beta_1 \bar{x} + \beta \bar{x}^2$$

$$X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & \vdots \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}, \quad \hat{\vec{\beta}} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} \begin{matrix} \beta_2 \\ \beta_3 \end{matrix}, \quad \bar{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

To create a simple linear regression model means to find estimates of $\hat{\beta}_0$ and $\hat{\beta}_1$ that minimizes the error, i.e. find $\hat{\beta}_0$ and $\hat{\beta}_1$ s.t.

$$\|\bar{y} - X\hat{\beta}\| \leq \|\bar{y} - X\beta\|$$

Using least squares:

Normal Equation

3) $\hat{\beta} = (X^t X)^{-1}(X^t y)$

1) $X^t X$
2) $X^t y$
3) solve $[X^t X \quad X^t y]$
4) state $\hat{\beta}_0$ and $\hat{\beta}_1 \rightarrow \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix}$

Ex:

Fit a linear model of the form $\beta_0 + \beta_1 \cdot \ln w = p$

w	44	61	81	113	131
p	91	98	103	110	112

$$X = \begin{bmatrix} 1 & \ln 44 \\ 1 & \ln 61 \\ 1 & \ln 81 \\ 1 & \ln 113 \\ 1 & \ln 112 \end{bmatrix} = \begin{bmatrix} 1 & 3{,}78 \\ 1 & 4{,}11 \\ 1 & 4{,}39 \\ 1 & 4{,}73 \\ 1 & 4{,}88 \end{bmatrix} , \quad \bar{p} = \begin{bmatrix} 91 \\ 98 \\ 103 \\ 110 \\ 112 \end{bmatrix}$$

1) $$X^T X = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 3.78 & \cdots & & & 4.88 \end{bmatrix} \begin{bmatrix} 1 & 3.78 \\ \vdots & \vdots \\ 1 & 4.88 \end{bmatrix} = \begin{bmatrix} 5 & 21.74 \\ 21.74 & 95.14 \end{bmatrix}$$

2) $$X^T y = \begin{bmatrix} 514 \\ 2248.34 \end{bmatrix}$$

3) $$\begin{bmatrix} X^T X & X^T y \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 10.63 \\ 0 & 1 & 21.20 \end{bmatrix}$$

so $\hat{p} = 10.63 + 21.20 \cdot \ln w$

$$\begin{bmatrix} \vdots \end{bmatrix} \approx \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$$

$$X \hat{\beta} = \hat{p}$$

$$\overset{\bar{u}}{\bar{p}} - \overset{\bar{v}}{X \hat{\beta}} \rightarrow \sqrt{(u_1 - v_1)^2 + (u_2 - v_2)^2 \cdots}$$