Steven Labalme

(H) Linear Algebra, 3

Chapter 4: Orthogonality (Notes 1)

Forsyth

11/13: <u>Orthogonality of Subspaces</u>

- $C(A) \oplus N(A^T) = \mathbb{R}^m \longrightarrow \dim C(A) = r, \ \dim N(A^T) = m - r$
- $C(A^T) \oplus N(A) = \mathbb{R}^n \longrightarrow \dim C(A^T) = r, \ \dim N(A) = n - r$
  - where A is $m \times n$ of rank r.
- Orthogonal complement $W^{\perp}$[1] of W.
  - $W^{\perp} = \{v \in V : v^T w = 0 \ \forall \ w \in W\}$
- Let $w = \text{span } C(A)$, $W^{\perp} = \text{span } N(A^T)$
  - $A^T y = 0, \ y^T A = 0$

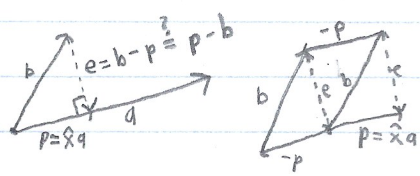
$$-\begin{bmatrix} - & y_i & - \end{bmatrix} \begin{bmatrix} | & | & & | \\ a_1 & a_2 & \cdots & a_n \\ | & | & & | \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$$

  - $Ax = 0$

$$-\begin{bmatrix} - & a_1 & - \\ & \vdots & \\ - & a_n & - \end{bmatrix} \begin{bmatrix} | \\ x_i \\ | \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$$

  - $C(A) = N(A^T)^{\perp}, \ C(A)^{\perp} = N(A^T), \ C(A^T) = N(A)^{\perp}, \ C(A^T)^{\perp} = N(A)$

11/14: <u>Projections</u>



- $e \perp a, \ b - p \perp a, \ b - \hat{x}a \perp a$
- $a \cdot (b - \hat{x}a) = 0, \ a^T(b - \hat{x}a) = 0, \ a^Tb - \hat{x}a^Ta = 0, \ \hat{x}a^Ta = a^Tb, \ \boxed{\hat{x} = \frac{a^Tb}{a^Ta}}$
- $a \cdot (\hat{x}a - b) = 0, \ a^T(\hat{x}a - b) = 0, \ \hat{x}a^Ta - a^Tb = 0, \ \hat{x}a^Ta = a^Tb, \ \hat{x} = \frac{a^Tb}{a^Ta}$

1 "double you perp"

- Project $b = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ onto $a = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$.

- $\hat{x} = \dfrac{\begin{bmatrix} 1 & 1 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}}{\begin{bmatrix} 1 & 1 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}} = \dfrac{6}{3} = 2$, $\quad p = 2\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix}$

- Projecting onto a subspace (multiple vectors).
  - $p = \hat{x}_1 a_1 + \cdots + \hat{x}_n a_n$
  - $p = A\hat{x}$
  - $\begin{bmatrix} | \\ p \\ | \end{bmatrix} = \begin{bmatrix} | & & | \\ a_1 & \cdots & a_n \\ | & & | \end{bmatrix}\begin{bmatrix} \hat{x}_1 \\ \vdots \\ \hat{x}_n \end{bmatrix}$
  - $e = b - A\hat{x}$
  - $A^T(b - A\hat{x}) = 0$, $A^Tb - A^TA\hat{x} = 0$, $A^TA\hat{x} = A^Tb$, $\hat{x} = (A^TA)^{-1}A^Tb$

11/15:

$$A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \quad b = \begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix}$$

- Solve $Ax = b$
  - $\left[\begin{array}{cc|c} 1 & 0 & 6 \\ 1 & 1 & 0 \\ 1 & 2 & 0 \end{array}\right] \rightarrow \left[\begin{array}{cc|c} 1 & 0 & 6 \\ 0 & 1 & -6 \\ 0 & 2 & -6 \end{array}\right] \rightarrow \left[\begin{array}{cc|c} 1 & 0 & 6 \\ 0 & 1 & -6 \\ 0 & 0 & 6 \end{array}\right]$
  - $b \notin C(A)$.

Chapter 4: Orthogonality (Notes 2) Forsyth

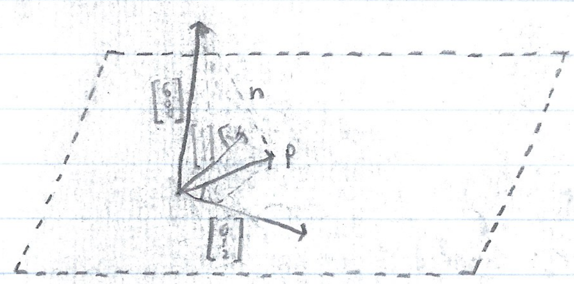
— $A^T A\hat{x} = A^T b$

— $\begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}\hat{x} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix}\begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix}$

— $\begin{bmatrix} 3 & 3 \\ 3 & 5 \end{bmatrix}\begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} = \begin{bmatrix} 6 \\ 0 \end{bmatrix} \longrightarrow \left[\begin{array}{cc|c} 3 & 3 & 6 \\ 3 & 5 & 0 \end{array}\right] \longrightarrow \left[\begin{array}{cc|c} 3 & 3 & 6 \\ 0 & 2 & -6 \end{array}\right] \longrightarrow \left[\begin{array}{cc|c} 3 & 6 & 15 \\ 0 & 1 & -3 \end{array}\right]$

— $\hat{x} = \begin{bmatrix} 5 \\ -3 \end{bmatrix}$

— $5\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - 3\begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 5 \\ 2 \\ -1 \end{bmatrix} = p$



- $\hat{x} = \frac{a^T b}{a^T a}$
- $p = a\hat{x}$
- $p = a\frac{a^T b}{a^T a}$
- $p = \frac{aa^T}{a^T a} b$
- $P = \frac{aa^T}{a^T a}$
- $p = Pb$

- $\hat{x} = (A^T A)^{-1} A^T b$
- $p = A\hat{x}$
- $p = A\left((A^T A)^{-1} A^T b\right)$
- $p = \left(A(A^T A)^{-1} A^T\right) b$
- $P = A(A^T A)^{-1} A^T$
- $p = Pb$

- Properties of Projection matrices:

① $P = P^T$

$$A(A^T A)^{-1} A^T \stackrel{?}{=} \left(A(A^T A)^{-1} A^T\right)^T$$
$$\stackrel{?}{=} A^{TT}\left[(A^T A)^{-1}\right]^T A^T$$
$$\stackrel{\checkmark}{=} A(A^T A)^{-1} A^T$$

11/18

② $P = P^2$

$$A(A^TA)^{-1}A^T \stackrel{?}{=} A(A^TA)^{-1}[A^TA][(A^TA)^{-1}]A^T$$
$$\stackrel{\checkmark}{=} A(A^TA)^{-1}A^T$$

11/19: **Partial Derivatives**

- $z = f(x,y) = xy^2$
- Partial derivative with respect to x:
  - $f_x(x,y) = f_x = \frac{\partial f}{\partial x} = \frac{\partial}{\partial x} f(x,y) = f_1 = D_1 f = D_x f$
- $\frac{\partial f}{\partial x} = \lim_{h \to 0} \frac{f(x+h,y) - f(x,y)}{h}$
- $\frac{\partial f}{\partial y} = \lim_{h \to 0} \frac{f(x,y+h) - f(x,y)}{h}$
- Ex: $f(x,y) = xy^2$
  - $\frac{\partial f}{\partial x} = \lim_{h \to 0} \frac{(x+h)y^2 - xy^2}{h} = \lim_{h \to 0} \frac{\cancel{xy^2} + hy^2 - \cancel{xy^2}}{h} = y^2$
  - $\frac{\partial f}{\partial y} = \lim_{h \to 0} \frac{x(y+h)^2 - xy^2}{h} = \lim_{h \to 0} \frac{x(y^2 + 2yh + h^2) - xy^2}{h} = 2y$
- Ex: $f(x,y) = x^3 + x^2y^3 - 2y^2$
  - $\frac{\partial f}{\partial x} = 3x^2 + 2xy^3$
  - $\frac{\partial f}{\partial y} = 3x^2y^2 - 4y$
  - Evaluate at $(2,1)$
    - $(16, 8)$
- $\nabla f(x,y) = \langle \frac{\partial}{\partial x}, \frac{\partial}{\partial y} \rangle$
- Del $f(x,y)$ – vector-valued function whose components are the partial derivatives of f.
  - Direction and rate of fastest increase at a point.

11/20:

- Find the directional derivative for $v = \langle 3, 4 \rangle$.
  - $u = \langle \frac{3}{5}, \frac{4}{5} \rangle$
  - $D_u f(2,1) = \nabla f(2,1) \cdot u = \langle 16, 8 \rangle \cdot \langle \frac{3}{5}, \frac{4}{5} \rangle = \frac{48}{5} + \frac{32}{5} = \frac{80}{5} = 16$
  - 16 is the rate of change of f at (2,1) pointing in the direction $\langle 3, 4 \rangle$.
- Ex: $f(x,y) = \sin\left(\frac{x}{1+y}\right)$
  - $\frac{\partial}{\partial x}\left(\sin\left(\frac{x}{1+y}\right)\right) = \frac{\partial}{\partial u}(\sin(u)) \cdot \frac{\partial}{\partial x}\left(\frac{x}{1+y}\right) = \frac{1}{1+y}\cos\left(\frac{x}{1+y}\right)$
  - $\frac{\partial}{\partial y}\left(\sin\left(\frac{x}{1+y}\right)\right) = \frac{\partial}{\partial u}(\sin(u)) \cdot \frac{\partial}{\partial y}\left(x(1+y)^{-1}\right) = \cos\left(\frac{x}{1+y}\right) \cdot \frac{-x}{(1+y)^2}$
- Ex: $x^3 + y^3 + z^3 + 6yz = 1$
  - $3x^2 + 0 + 3z^2 \frac{\partial z}{\partial x} + 6y \frac{\partial z}{\partial x} = 0$, $(3z^2 + 6y)\frac{\partial z}{\partial x} = -3x^2$, $\frac{\partial z}{\partial x} = \frac{-x^2}{z^2 + 2y}$

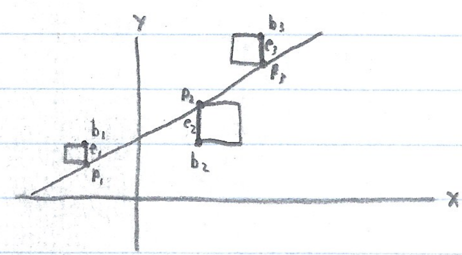
Chapter 4: Orthogonality (Notes 3) Forsyth

11/25: Least Squares Regression

- Statistics: $\hat{y} = a + bx \quad (x, y)$
- Strang: $b = C + Dt \quad (t, b)$

— $$\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix} = \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_m \end{bmatrix} \begin{bmatrix} \hat{C} \\ \hat{D} \end{bmatrix}$$

- $(t_1, b_1) = (-1, 1)$; $(t_2, b_2) = (1, 1)$; $(t_3, b_3) = (2, 3)$



- $e = b - p$ — a residual value
- Goal: $e^2 = \min$
- $E(x) = e_1^2 + e_2^2 + e_3^2$ — error function in Strang.
  $= \|Ax - b\|^2$ — $(Ax = b = p + e)$

- $$\begin{array}{c} C+Dt \\ \begin{bmatrix} 1 & -1 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \end{array} \begin{bmatrix} \hat{C} \\ \hat{D} \end{bmatrix} = \begin{array}{c} b \\ \begin{bmatrix} 1 \\ 1 \\ 3 \end{bmatrix} \end{array}$$

  — $b \notin C(A)$
  — $A^T A \hat{x} = A^T b$

— $$\begin{bmatrix} 1 & 1 & 1 \\ -1 & 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} \hat{C} \\ \hat{D} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ -1 & 1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 3 \end{bmatrix}$$

— $\begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix} \begin{bmatrix} \hat{C} \\ \hat{D} \end{bmatrix} = \begin{bmatrix} 5 \\ 6 \end{bmatrix}$

— $\begin{bmatrix} \hat{C} \\ \hat{D} \end{bmatrix} = \begin{bmatrix} 9/7 \\ 4/7 \end{bmatrix}$

— $\begin{bmatrix} 1 & 1 & \cdots & 1 \\ t_1 & t_2 & \cdots & t_m \end{bmatrix} \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_m \end{bmatrix} = \begin{bmatrix} m & \sum t_m \\ \sum t_m & \sum t_m^2 \end{bmatrix} = A^T A$

— $A^T b = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ t_1 & t_2 & \cdots & t_m \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix} = \begin{bmatrix} \sum b_m \\ \sum (b_m t_m) \end{bmatrix}$

• $(0,6)$, $(1,0)$, $(2,0)$

— $\begin{bmatrix} 3 & 3 \\ 3 & 5 \end{bmatrix} \begin{bmatrix} \hat{C} \\ \hat{D} \end{bmatrix} = \begin{bmatrix} 6 \\ 0 \end{bmatrix}$, $\begin{bmatrix} \hat{C} \\ \hat{D} \end{bmatrix} = \begin{bmatrix} 5 \\ -3 \end{bmatrix}$, $\hat{b} = 5 - 3t$

— $b_1 = 6$ $\quad t_1 = 0$ $\quad p_1 = 5$ $\quad e_1 = 1$

$b_2 = 0$ $\quad t_2 = 1$ $\quad p_2 = 2$ $\quad e_2 = -2$

$b_3 = 0$ $\quad t_3 = 2$ $\quad p_3 = -1$ $\quad e_3 = 1$

— $E = e_1^2 + e_2^2 + e_3^2 = \|Ax - b\|^2 = (C + Dt_1 - b_1)^2 + (C + Dt_2 - b_2)^2 + (C + Dt_3 - b_3)^2$

$= (C-6)^2 + (C+D)^2 + (C+2D)^2$

11/26: — $\frac{\partial E}{\partial C} = 0 = 2(C-6) + 2(C+D) + 2(C+2D) = 6C + 6D - 12$, $6 = 3C + 3D$

— $\frac{\partial E}{\partial D} = 0 = 2(C+D) + 4(C+2D) = 6C + 10D$, $0 = 3C + 5D$

— $\begin{bmatrix} 3 & 3 \\ 3 & 5 \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} 6 \\ 0 \end{bmatrix}$

Chapter 4: Orthogonality (Notes 4) Forsyth

12/2: **Benefits of an Orthonormal Basis**

- Orthonormal: Orthogonal ($v^T w = 0$) normal ($\|v\| = 1$)
- $q_1, \dots, q_n$ is a set of orthonormal vectors.
  - $$Q = \begin{bmatrix} | & & | \\ q_1 & \cdots & q_n \\ | & & | \end{bmatrix}$$
- $Q^{-1} = Q^T \rightarrow Q^T Q = I$
  - $$\begin{bmatrix} — q_1^T — \\ \vdots \\ — q_n^T — \end{bmatrix} \begin{bmatrix} | & & | \\ q_1 & \cdots & q_n \\ | & & | \end{bmatrix} = \begin{bmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{bmatrix}$$
  - $q_i^T q_i = 1$ (normal)
  - $q_i^T q_j = 0$ (orthogonal)
- $Ax = b$, $x = A^{-1} b$; $Qx = b$, $x = Q^T b$
- Multiplication by Q presents $\|x\|$.
  - $\|b\|^2 = \|Qx\|^2 = (Qx)^T (Qx) = x^T Q^T Q x = x \cdot x = \|x\|^2$
- Simplifies projections
  - $A^T A \hat{x} = A^T b$, $\hat{x} = (A^T A)^{-1} A^T b$
  - $Q^T Q \hat{x} = Q^T b$, $\hat{x} = Q^T b$
- Simplifies projection matrices
  - $P = A(A^T A)^{-1} A^T = Q Q^T$
- Gram-Schmidt algorithm (see print out)

12/3: **Gram-Schmidt Example**

$$a = \begin{bmatrix} 1 \\ -1 \\ -1 \\ 1 \end{bmatrix} \quad b = \begin{bmatrix} 2 \\ 1 \\ 0 \\ 1 \end{bmatrix} \quad c = \begin{bmatrix} 2 \\ 2 \\ 1 \\ 2 \end{bmatrix}$$

- $b^{\perp} = e = b - p = b - \hat{x}a = b - \frac{a^T b}{a^T a} a$

- $$\begin{bmatrix} 2 \\ 1 \\ 0 \\ 1 \end{bmatrix} - \frac{2}{4}\begin{bmatrix} 1 \\ -1 \\ -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 3/2 \\ 3/2 \\ 1/2 \\ 1/2 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 3 \\ 3 \\ 1 \\ 1 \end{bmatrix} \rightarrow b^{\perp} = \begin{bmatrix} 3 \\ 3 \\ 1 \\ 1 \end{bmatrix}$$

- $c^{\perp} = c - \left(\frac{a^T c}{a^T a} a + \frac{b^{\perp T} c}{b^{\perp T} b^{\perp}} b^{\perp}\right) = c - \frac{a^T c}{a^T a} a - \frac{b^{\perp T} c}{b^{\perp T} b^{\perp}} b^{\perp}$

- $$\begin{bmatrix} 2 \\ 2 \\ 1 \\ 2 \end{bmatrix} - \frac{1}{4}\begin{bmatrix} 1 \\ -1 \\ -1 \\ 1 \end{bmatrix} - \frac{15}{20}\begin{bmatrix} 3 \\ 3 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -1/2 \\ 0 \\ 1/2 \\ 1 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} -1 \\ 0 \\ 1 \\ 2 \end{bmatrix} \rightarrow c^{\perp} = \begin{bmatrix} -1 \\ 0 \\ 1 \\ 2 \end{bmatrix}$$

- Orthogonal basis $= \left\{ \begin{bmatrix} 1 \\ -1 \\ -1 \\ 1 \end{bmatrix}, \begin{bmatrix} 3 \\ 3 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \\ 2 \end{bmatrix} \right\}$

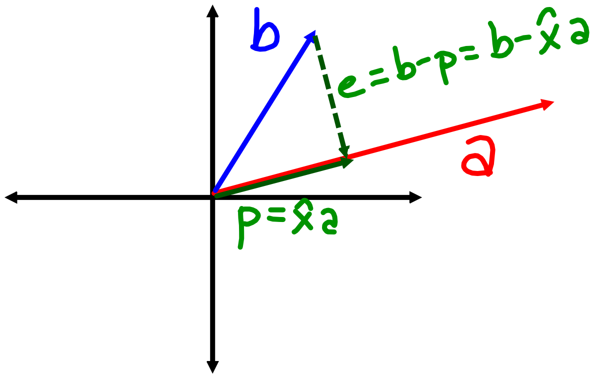
- $\frac{v}{\|v\|} = n$

- Orthonormal basis $= \left\{ \begin{bmatrix} 1/2 \\ -1/2 \\ -1/2 \\ 1/2 \end{bmatrix}, \begin{bmatrix} 3/\sqrt{20} \\ 3/\sqrt{20} \\ 1/\sqrt{20} \\ 1/\sqrt{20} \end{bmatrix}, \begin{bmatrix} -1/\sqrt{6} \\ 0 \\ 1/\sqrt{6} \\ 2/\sqrt{6} \end{bmatrix} \right\}$

- $$Q = \begin{bmatrix} 1/2 & 3/\sqrt{20} & -1/\sqrt{6} \\ -1/2 & 3/\sqrt{20} & 0 \\ -1/2 & 1/\sqrt{20} & 1/\sqrt{6} \\ 1/2 & 1/\sqrt{20} & 2/\sqrt{6} \end{bmatrix}$$

has a, b, and c vectors!

- $A = QR$ factorization, $Q^T A = R$



$$a^T(b-\hat{x}a) = 0$$

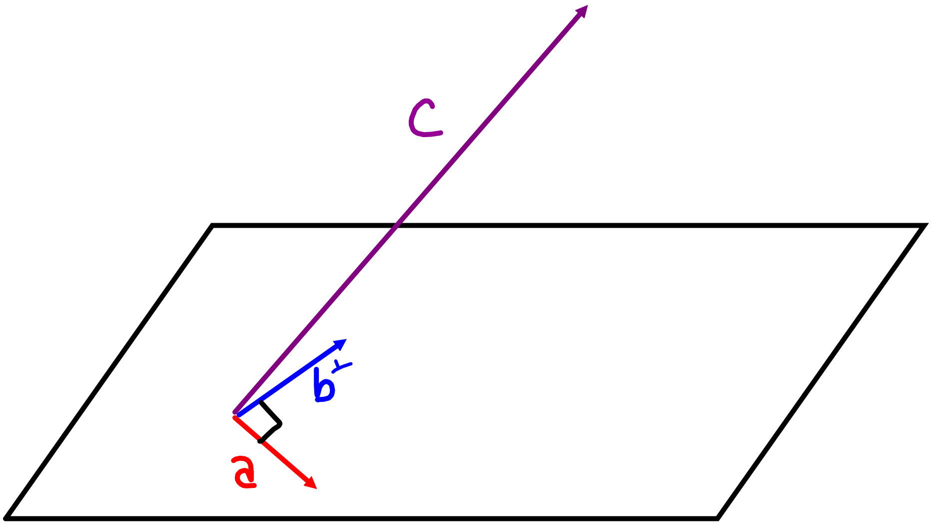
$$a^Tb - \hat{x}a^Ta = 0$$

$$\hat{x} = \frac{a^Tb}{a^Ta}$$

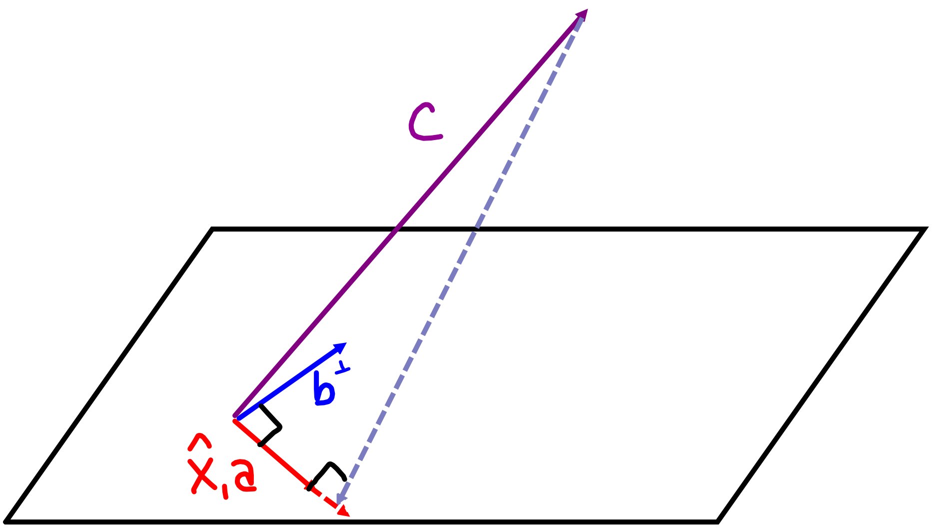
$$p = \hat{x}a \therefore p = \frac{a^Tb}{a^Ta}a$$

Gram-Schmidt

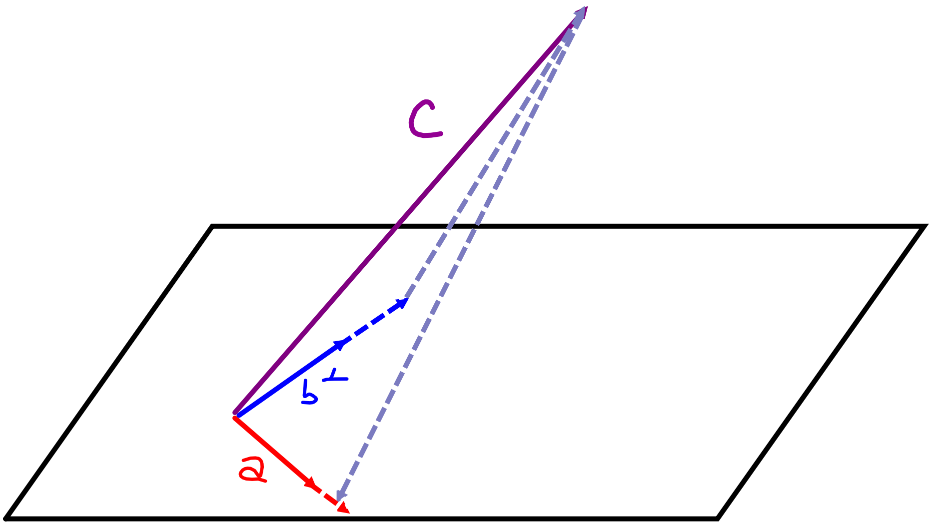
1. Keep a
2. Find $b^\perp = e = b - p = b - \hat{x}a = \boxed{b - \frac{a^Tb}{a^Ta}a}$



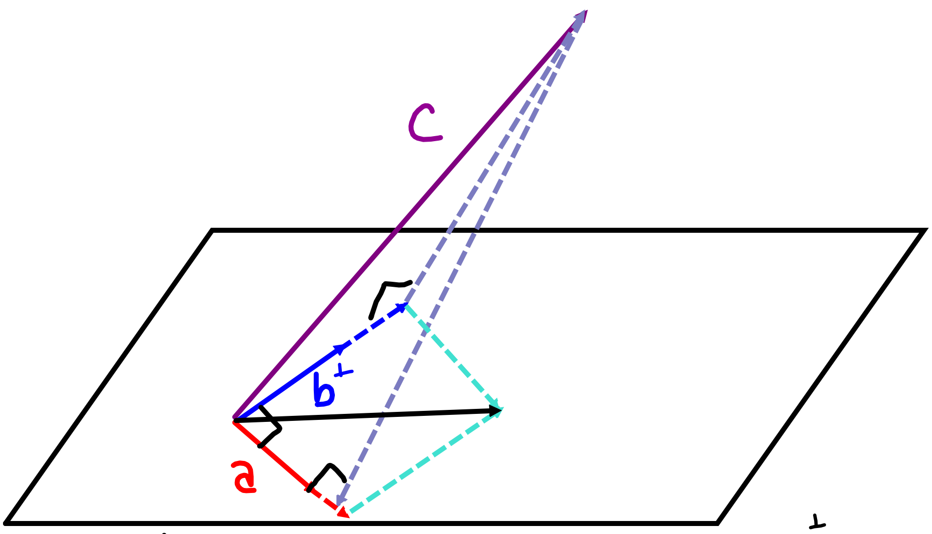
$a^T b^\perp = 0$



project c onto a $\frac{a^T c}{a^T a} a$



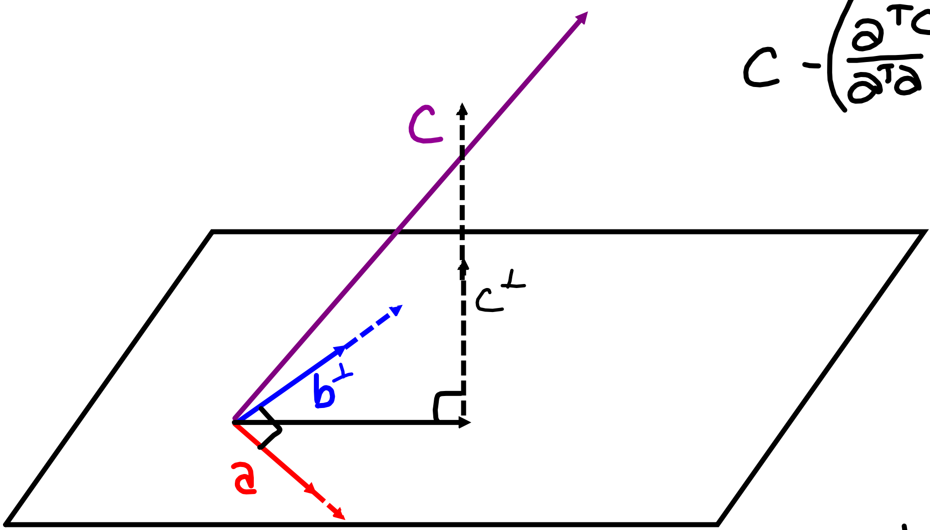
project c onto $b^{\perp}$ $\quad \dfrac{b^{\perp T} c}{b^{\perp T} b^{\perp}} \, b^{\perp}$



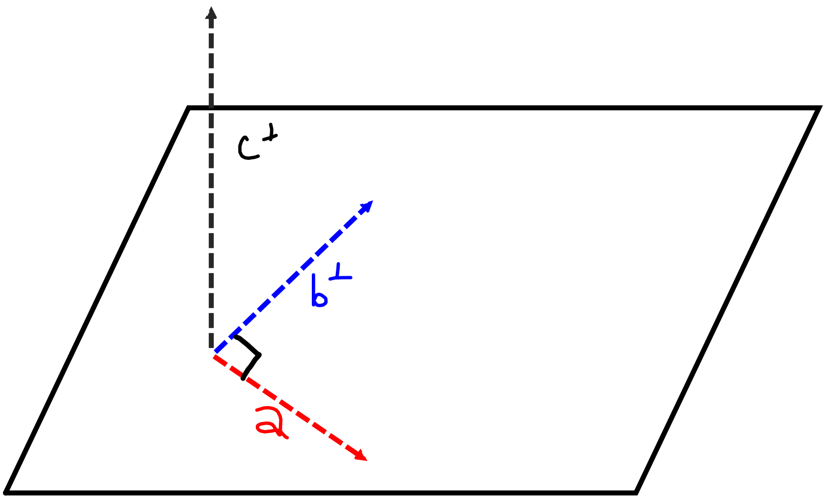
Combine the two projections

$$\frac{a^T c}{a^T a} a + \frac{b^{\perp T} c}{b^{\perp T} b^{\perp}} b^{\perp}$$

$$c - \left(\frac{a^T c}{a^T a} a + \frac{b^{\perp T} c}{b^{\perp T} b^{\perp}} b^{\perp}\right)$$



we want the vector perpendicular
to the sum of both projections
(the error vector)



Normalize the vectors

$$\frac{a}{\|a\|} \quad \frac{b^{\perp}}{\|b^{\perp}\|} \quad \frac{c^{\perp}}{\|c^{\perp}\|}$$