- Consider the vectors $x_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$, $x_2 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ $x_3 = \begin{bmatrix} 2 \\ 3 \\ 3 \end{bmatrix}$
- let M the set of all linear combinations of x_1, x_2, x_3
- $\{x_1, x_2, x_3\}$: spanning set
- We can easily check $3x_1 - x_2 = x_3 \Rightarrow x_1, x_2, x_3$ linearly dep.
  Any two of x_1, x_2 & x_3 form a basis for ~~the space~~ M
- M contains vectors with the form $\begin{bmatrix} a \\ b \\ b \end{bmatrix}$, $a, b \in \mathbb{R}$

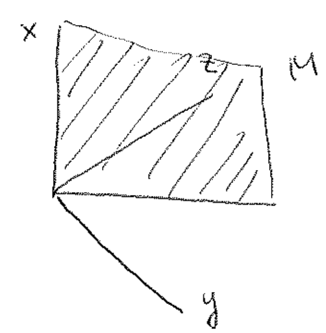
† Def [Rank of a subspace]: The *rank* of a subspace $\mathcal{M}$ is the number of elements in a basis for $\mathcal{M}$. The rank is written *rank*$(\mathcal{M})$, $r(\mathcal{M})$ or *dim*$(\mathcal{M})$.

If **A** is a $m \times n$ matrix, the rank of $C(\mathbf{A})$ is called the rank of **A** and is written as $r(\mathbf{A})$. $= n$ $< n$

★ Result: $r(C(\mathbf{A})) = dim(C(\mathbf{A})) = r(\mathbf{A})$: # of linearly independent columns of **A** (=# of basis vectors for $C(\mathbf{A})$).

$\Rightarrow$ When $r(\mathbf{X}) = r$, the r linearly independent columns of **X** span the r-dimensional estimation space $C(\mathbf{X})$

$\Rightarrow$ The space spanned by $\{x_1, x_2, x_3\}$ has rank 2.



OR consider $A = \begin{bmatrix} 1 & 1 & 2 \\ 1 & 0 & 3 \\ 1 & 0 & 3 \end{bmatrix}$ $r(A) = 2 = r(C(A))$

$\underline{X^T X}$

- **Q:** How to find a solution $\hat{\beta}$ if $r(\underset{n\times p}{\mathbf{X}}) = r(\underset{p\times p}{\mathbf{X}^T\mathbf{X}}) < p$ (singular)?

Here is our strategy!

△ There is no unique solution for β when $r(\mathbf{X}) = r(\mathbf{X}^T\mathbf{X}) < p$.

△ Instead we will find $\hat{\mathbf{y}} \in C(\mathbf{X})$ closest to $\mathbf{y}$ in the Euclidean distance.

△ Then we will find $\hat{\beta}$ such that $\hat{\mathbf{y}} = \mathbf{X}\hat{\beta}$.

- **Q:** How to find such $\hat{\mathbf{y}}$?

We will consider the perpendicular projection matrix!

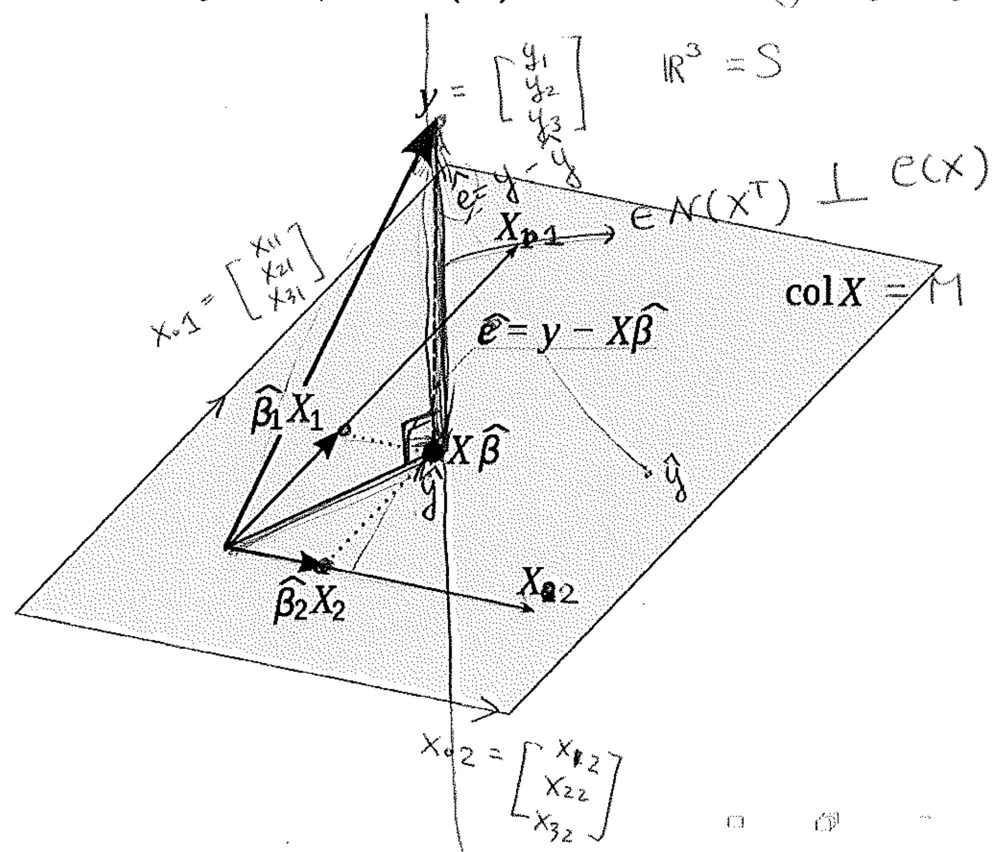
$$\begin{bmatrix} x_{11} \\ x_{21} \\ x_{31} \end{bmatrix} \quad \begin{bmatrix} x_{12} \\ x_{22} \\ x_{32} \end{bmatrix}$$

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} X_{\cdot 1} & X_{\cdot 2} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ e_3 \end{bmatrix}$$

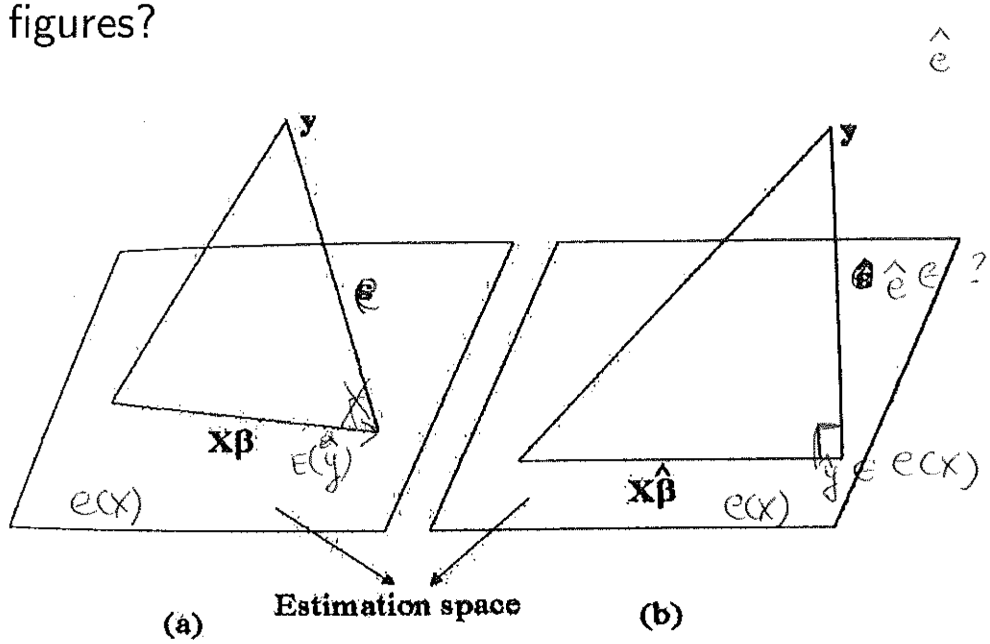
★ Recall the linear models: $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$ and $\mathbf{X}\boldsymbol{\beta} \in C(\mathbf{X})$.

★ Suppose $\hat{\boldsymbol{\beta}}$ is a solution for the NEs (that is, minimize $||\mathbf{y} - \mathbf{X}\boldsymbol{\beta}||^2$).

$\rightarrow$ Let fitted value $\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}} \in C(\mathbf{X})$ and residual $\hat{\mathbf{e}} = \mathbf{y} - \hat{\mathbf{y}}$.



★ More figures?



Q: Where does $\hat{\mathbf{e}}$ lie?

$Ay = 0$: homogeneous linear system

$\Rightarrow y$: a solution

$\Rightarrow N(A) =$ the set of solutions to the homogeneous linear system

† Def **[Nullsapce]:** The *null space* of a $m \times n$ matrix $\mathbf{A}$, denoted by $\mathcal{N}(\mathbf{A})$,

$$\mathcal{N}(\mathbf{A}) = \{\mathbf{y} \mid \underset{m\times n}{\mathbf{A}}\underset{n\times 1}{\mathbf{y}} = \mathbf{0}\} \subset \mathbb{R}^n.$$

♣ Th: If $\mathbf{A}$ is $m \times n$ and $r(\mathbf{A}) = r$, then the rank of the null space of $\mathbf{A}$ is $n - r$, that is,

$$dim(C(\mathbf{A})) + dim(\mathcal{N}(\mathbf{A})) = n.$$

<u>ex</u> let $A = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \\ -1 & 1 & -3 \\ 1 & 2 & 0 \end{bmatrix}_{4\times 3}$ let's find $N(A)$

$Ay = 0 \Rightarrow$
$y_1 + y_2 + y_3 = 0$
$2y_1 + 2y_2 + 2y_3 = 0$
$-y_1 + y_2 + 3y_3 = 0$
$y_1 + 2y_2 + 0 = 0$

$\underset{3\times 1}{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_2 \end{bmatrix}$

$\Rightarrow -4y_2 + 2y_2 + 2y_2 = 0$
$\Rightarrow 2y_2 + y_2 - 3y_3 = 0 \Rightarrow y_3 = y_2$
$\Rightarrow y_1 = -2y_2$

$y = \begin{bmatrix} -2y_2 \\ y_2 \\ y_2 \end{bmatrix} = y_2 \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}, \quad y_2 \in \mathbb{R}$

$N(A) = \left\{ c\begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}, \ c \in \mathbb{R} \right\} \subset \mathbb{R}^3 \Rightarrow \overset{\text{dim}}{\overset{\|}{r}}(N(A)) = 1$

$\dim(C(A)) = 2$

★ Result: Recall the NEs:

$$\mathbf{X}^T\mathbf{X}\boldsymbol{\beta} = \mathbf{X}^T\mathbf{y}.$$

Suppose $\hat{\boldsymbol{\beta}}$ is a solution for the NEs.

- $\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}} \in C(\mathbf{X})$.
- $\underline{\mathbf{X}^T(\mathbf{y} - \underbrace{\mathbf{X}\hat{\boldsymbol{\beta}}}_{\hat{y}})} = \mathbf{X}^T\underbrace{\hat{\mathbf{e}}}_{y-\hat{y}} = 0 \Rightarrow \hat{\mathbf{e}} \in \mathcal{N}(\mathbf{X}^T)$

$X^Ty - X^TX\hat{\beta} = 0$

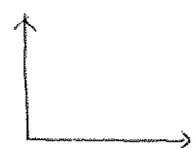
$\hat{y}$ $\hat{e}$

Q: What is the relationship between $C(\mathbf{X})$ and $\mathcal{N}(\mathbf{X}^T)$?

† Def **[Orthogonal]:**

- (Inner Product) The *inner product* between two vectors, $\mathbf{x}$ and $\mathbf{y}$ is $\mathbf{x}^T\mathbf{y}$
- (Orthogonal Vectors) Two vectors $\mathbf{x}$ and $\mathbf{y}$ are *orthogonal (or perpendicular)* (written $\mathbf{x} \perp \mathbf{y}$) if $\mathbf{x}^T\mathbf{y} = 0$.
- (Orthogonal Spaces) Two subspaces $\mathcal{M}_1$ and $\mathcal{M}_2$ are *orthogonal* if $\mathbf{x} \in \mathcal{M}_1$ and $\mathbf{y} \in \mathcal{M}_2$ implies $\mathbf{x}^T\mathbf{y} = 0$
- (Orthogonal Basis) $\{\mathbf{x}_1, \ldots, \mathbf{x}_r\}$ is an *orthogonal basis* for $\mathcal{M}$ if $\{\mathbf{x}_1, \ldots, \mathbf{x}_r\}$ is a basis of $\mathcal{M}$ and for $i \neq j$, $\mathbf{x}_i^T\mathbf{x}_j = 0$.

$\begin{bmatrix}1\\2\end{bmatrix}$ $\begin{bmatrix}3\\2\end{bmatrix}$ $\left[\begin{bmatrix}1\\0\end{bmatrix}\begin{bmatrix}0\\1\end{bmatrix}\right]$



† Def **[Orthogonal]:** — Contd

- (Orthonormal Basis) $\{\mathbf{x}_1, \ldots, \mathbf{x}_r\}$ is an *orthonormal basis* for $\mathcal{M}$ if $\{\mathbf{x}_1, \ldots, \mathbf{x}_r\}$ is an orthogonal basis and $\mathbf{x}_i^T \mathbf{x}_i = 1$ for $i = 1, \ldots, r$

- Note:

– $\mathbf{Q}$ has orthogonal columns $\Rightarrow$ $\mathbf{Q}^T\mathbf{Q} = \mathbf{D}$ (diagonal matrix)

– $\mathbf{Q}$ has orthonormal columns $\Rightarrow$ $\mathbf{Q}^T\mathbf{Q} = I$

||

$M_S^{\perp} \cap M = \{0\}$

† Def **[Orthogonal Complement]:** Let $\mathcal{S}$ be a vector space, and let $\mathcal{M}$ be a subspace of $\mathcal{S}$. Let $\mathcal{M}_{\mathcal{S}}^{\perp} = \{\mathbf{y} \in \mathcal{S} \mid \mathbf{y} \perp \mathcal{M}\}$. $\mathcal{M}_{\mathcal{S}}^{\perp}$ is called the *orthogonal complement* of $\mathcal{M}$ with such that respect to $\mathcal{S}$.

If $\mathcal{S}$ is taken as $\mathbb{R}^n$, then $\mathcal{M}_{\mathcal{S}}^{\perp} \equiv \mathcal{M}^{\perp}$ is simply referred to as the orthogonal complements of $\mathcal{M}$.

In layman's terms, every vector in $\mathcal{M} \subset \mathcal{S}$ is orthogonal to every vector in $\mathcal{M}_{\mathcal{S}}^{\perp}$.

♣ Th: Let $\mathcal{S}$ be a vector space, and let $\mathcal{M}$ be a subspace of $\mathcal{S}$.

- The orthogonal complement of $\mathcal{M}$ with respect to $\mathcal{S}$ is a subspace of $\mathcal{S}$;

$y = \hat{y} + \hat{e}$ $\quad$ $\mathcal{C}(X)$

- If $\mathbf{x} \in \overset{\mathbb{R}^n}{\mathcal{S}}$, $\mathbf{x}$ can be written <u>uniquely</u> as $\mathbf{x} = \mathbf{x}_0 + \mathbf{x}_1$ with $\underset{\hat{y}}{\mathbf{x}_0} \in \mathcal{M}$ and $\underset{\hat{e}}{\mathbf{x}_1} \in \mathcal{M}_{\mathcal{S}}^{\perp}$.

$N(X^T)$

- The ranks of these spaces satisfy the relation, $r(\mathcal{S}) = r(\mathcal{M}) + r(\mathcal{M}_{\mathcal{S}}^{\perp})$.

$S = \mathbb{R}^3 \qquad A = \begin{bmatrix} 1 & 1 & 2 \\ 1 & 0 & 3 \\ 1 & 0 & 3 \end{bmatrix} \qquad M = \mathcal{C}(A)$

$\Leftrightarrow \quad M = \{$ all the vectors with the form $\begin{bmatrix} a \\ b \\ b \end{bmatrix}$, $a, b \in \mathbb{R} \}$

What is $M^{\perp}$?

$M^{\perp} = \{$ all the vectors with the form $\begin{bmatrix} 0 \\ c \\ -c \end{bmatrix}$, $c \in \mathbb{R} \}$

$$\underbrace{\begin{bmatrix} x \\ y \\ z \end{bmatrix}}_{\in \mathbb{R}^3} = \underbrace{\begin{bmatrix} x \\ \frac{y+z}{2} \\ \frac{y+z}{2} \end{bmatrix}}_{\in M} + \underbrace{\begin{bmatrix} 0 \\ \frac{y-z}{2} \\ -\frac{y-z}{2} \end{bmatrix}}_{\in M^{\perp}}$$

$\mathcal{C}(X)$ $\quad N(X^T)$

$\hat{y} \in$ $\quad \hat{e} \in$

$\hat{y}^T \hat{e} = 0$

vectors spaces w/ vectors in $\mathbb{R}^m$

$x \in \mathcal{C}(A)$

$x = \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix}$

★ Result: For any matrix $\mathbf{A}$ ($m \times n$), $C(\underset{m\times n}{\mathbf{A}})$ and $\mathcal{N}(\underset{n \times m}{\mathbf{A}^T})$ are orthogonal complements in $\mathbb{R}^m$.

Let $r(A) = r(A^T) = r \quad \Rightarrow \quad \dim(\mathcal{C}(A)) = r$

$\dim(N(A^T)) = m - r$

Ⓐ Suppose $v \in \mathcal{C}(A)$ & $v \in N(A^T) \Rightarrow$ want to show $v = 0$

$v \in \mathcal{C}(A) \Rightarrow \exists\ c\ (\neq 0)$ s.t $v = Ac$

$v \in N(A^T) \Rightarrow A^T v = \underline{0 = A^T A c}$

$\Rightarrow$ consider $v^T v = (Ac)^T (Ac) = c^T \underbrace{A^T A c}_{=0} = 0$

$y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$

•**Bottom Line:** $C(\mathbf{X})$ and $\mathcal{N}(\mathbf{X}^T)$ are orthogonal complements in $\mathbb{R}^n$. So, $\hat{\mathbf{y}} \perp \hat{\mathbf{e}}$ and $\mathbf{y} = \hat{\mathbf{y}} + \hat{\mathbf{e}}$ is the unique decomposition.

Ⓑ Suppose $v \in \mathcal{C}(A)$ and $w \in N(A^T) \Rightarrow$ want to show $v \perp w$

$v \in \mathcal{C}(A) \Rightarrow \exists\ c\ (\neq 0)$ s.t $v = Ac$

$\Rightarrow v^T w = (Ac)^T w = c^T \underline{A^T w} = 0$

By Ⓐ & Ⓑ, $\mathcal{C}(A)$ and $N(A^T)$ are orthogonal compliments in $\mathbb{R}^m$

★ Summary! Recall the NEs:

$$\mathbf{X}^T\mathbf{X}\beta = \mathbf{X}^T\mathbf{y}.$$

Suppose $\hat{\beta}$ is a solution for the NEs.

- $\mathbf{y} \in \mathbb{R}^n$.
- $\hat{\mathbf{y}} = \mathbf{X}\hat{\beta} \in C(\mathbf{X})$.
- $\mathbf{X}^T(\mathbf{y} - \mathbf{X}\hat{\beta}) = \mathbf{X}^T\hat{\mathbf{e}} = 0 \Rightarrow \hat{\mathbf{e}} \in \mathcal{N}(\mathbf{X}^T)$ $\perp$
- $\hat{\mathbf{y}} \perp \hat{\mathbf{e}}$
- $\mathbf{y} = \hat{\mathbf{y}} + \hat{\mathbf{e}}$ is the unique decomposition.

- $\mathbf{y} = \hat{\mathbf{y}} + \hat{\mathbf{e}}$ is the unique decomposition.

$\Rightarrow$ The unique orthogonal decomposition of sums of squares from the Pythagorean Theorem,

$$||\mathbf{y}||^2 = ||\hat{\mathbf{y}} + \hat{\mathbf{e}}||^2 = ||\hat{\mathbf{y}}||^2 + ||\hat{\mathbf{e}}||^2$$

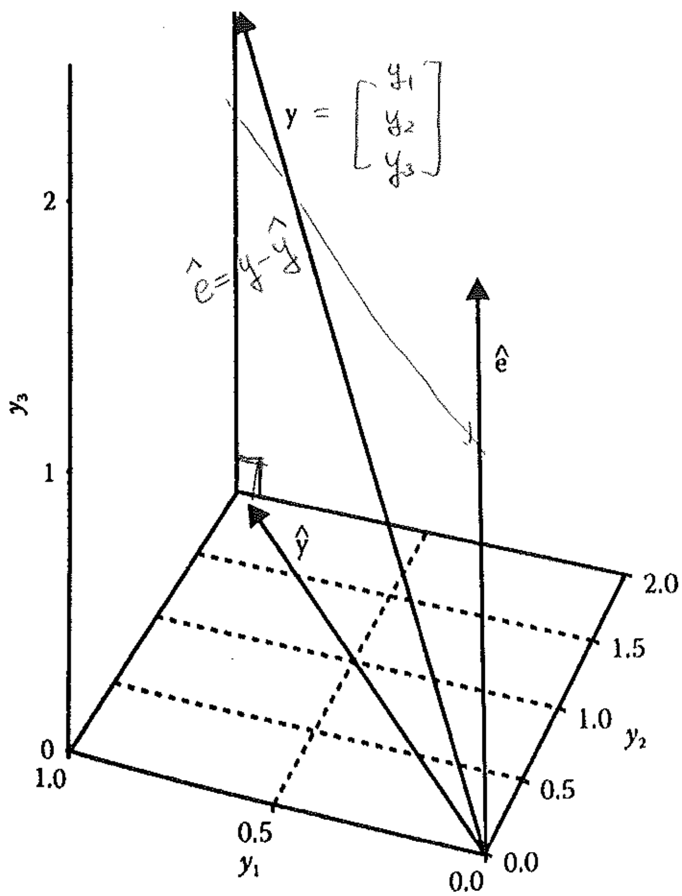
- $SST = ||\mathbf{y}||^2$: the total sum of squares
- $SSR = ||\hat{\mathbf{y}}||^2 = ||\mathbf{X}^T\hat{\beta}||^2$: the regression sum of squares
- $SSE = ||\hat{\mathbf{e}}||^2$: the error sum of squares

† Geometry of Least Squares:
$(\mathbf{y} - \mathbf{X}\hat{\beta})^T(\mathbf{y} - \mathbf{X}\hat{\beta}) = \min_{\beta}(\mathbf{y} - \mathbf{X}\beta)^T(\mathbf{y} - \mathbf{X}\beta)$

† Fact:

- Decomposition of $\mathbf{y} = \hat{\mathbf{y}} + \mathbf{y} - \hat{\mathbf{y}}$ into $\hat{\mathbf{y}} = \mathbf{X}\hat{\beta}$ (fitted value) and $\hat{\mathbf{e}} = \mathbf{y} - \hat{\mathbf{y}}$ (residuals)
- $\hat{\mathbf{y}} \in C(\mathbf{X})$
- $\hat{\mathbf{e}}$ is orthogonal to $\hat{\mathbf{y}}$ (that is, $\hat{\mathbf{e}} \in \mathcal{N}(\mathbf{X}^T)$)
- $\hat{\mathbf{y}}$ is the orthogonal projection of $\mathbf{y}$ onto $C(\mathbf{X})$
- $\hat{\mathbf{e}}$ is the orthogonal projection onto the orthogonal complement of $C(\mathbf{X})$. $= \mathcal{N}(X^T)$

†: The geometry provides good intuition for $n-$dimensional problems (but hard to visualize for $n > 3$).



† Geometry of Least Squares:

$$(\mathbf{y} - \mathbf{X}\hat{\beta})^T(\mathbf{y} - \mathbf{X}\hat{\beta}) = \min_{\beta}(\mathbf{y} - \mathbf{X}\beta)^T(\mathbf{y} - \mathbf{X}\beta)$$

$\Rightarrow$ LSE $\hat{\beta}$ is the vector that makes $\hat{\mathbf{y}} = \mathbf{X}\hat{\beta} \in C(\mathbf{X})$ closest to $\mathbf{y}$.

$\Rightarrow$ The vector in $C(\mathbf{X})$ that is closest to $\mathbf{y}$ is the perpendicular projection of $\mathbf{y}$ onto $C(\mathbf{X})$.

$\hat{y} = Py \in C(X)$

$\Rightarrow$ Then how to find orthogonal projection matrices?

★ Two ways to find the perpendicular projection matrix onto $C(\mathbf{X})$.

- △ A generalized inverse
- △ The Gram-Schmidt theorem

∗ Road map:

- Discuss the perpendicular projection matrix
- Discuss how to find perpendicular projection matrix

Consider $X = \begin{bmatrix} 2 \\ 1 \end{bmatrix} \in \mathbb{R}^2$, $r(\cancel{X}) = 1$

$$\mathcal{C}(X) = \left\{ \begin{bmatrix} 2a \\ a \end{bmatrix}, \quad a \in \mathbb{R} \right\}$$

$$\begin{bmatrix} b \\ -2b \end{bmatrix} \perp \mathcal{C}(X) \quad \text{for} \quad b \in \mathbb{R}$$

Now consider $M = \begin{bmatrix} 0.8 & 0.4 \\ 0.4 & 0.2 \end{bmatrix}$.

✠ <u>claim</u>: M is a perpendicular projection matrix onto $\mathcal{C}(X)$

Let's check (i) & (ii)

(i) $v = \begin{bmatrix} 2a \\ a \end{bmatrix} \in \mathcal{C}(X)$. $Mv = \begin{bmatrix} 0.8 & 0.4 \\ 0.4 & 0.2 \end{bmatrix} \begin{bmatrix} 2a \\ a \end{bmatrix}$

$$= \begin{bmatrix} 1.6a + 0.4a \\ 0.8a + 0.2a \end{bmatrix} = \begin{bmatrix} 2a \\ a \end{bmatrix} = v$$

(ii) $w = \begin{bmatrix} b \\ -2b \end{bmatrix} \perp \mathcal{C}(X)$, $Mw = \begin{bmatrix} 0.8 & 0.4 \\ 0.4 & 0.2 \end{bmatrix} \begin{bmatrix} b \\ -2b \end{bmatrix}$

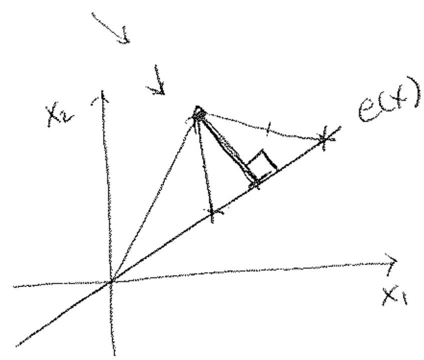
$$= \begin{bmatrix} 0.8b - 0.8b \\ 0.4b - 0.4b \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

✠ Check prop! $\mathcal{C}(X) = \mathcal{C}(M)$

✠ (i) $MM = \begin{bmatrix} 0.8 & 0.4 \\ 0.4 & 0.2 \end{bmatrix} \begin{bmatrix} 0.8 & 0.4 \\ 0.4 & 0.2 \end{bmatrix} = \begin{bmatrix} 0.64 + 0.16 & 0.32 + 0.08 \\ 0.32 + 0.08 & 0.16 + 0.04 \end{bmatrix}$

$$= \begin{bmatrix} 0.8 & 0.4 \\ 0.4 & 0.2 \end{bmatrix} = M \checkmark$$

(ii) M is symmetric!



† Def **[Projection]:** A square matrix $\mathbf{M}$ is a *perpendicular projection operator* (matrix) onto $C(\mathbf{X})$ if and only if

(i) $\mathbf{v} \in C(\mathbf{X})$ implies $\mathbf{Mv} = \mathbf{v}$ (projection)

(ii) $\mathbf{w} \perp C(\mathbf{X})$ implies $\mathbf{Mw} = 0$ (perpendicularity)

☆ Note: Any projection that is not a perpendicular projection is called an *oblique projection operator*.

★ Prop If $\mathbf{M}$ is a *perpendicular projection operator* onto $C(\mathbf{X})$, then $C(\mathbf{M}) = C(\mathbf{X})$.

♣ Th: $\mathbf{M}$ is a perpendicular projection operator on $C(\mathbf{M})$ if and only if (i) $\mathbf{MM} = \mathbf{M}$ (idempotent) and (ii) $\mathbf{M}^T = \mathbf{M}$ (symmetric).

♣ Th: Perpendicular projection operators are unique.

$y = \hat{y} + \hat{e}$

$I y = P_1 y + P_2 y$

$I = P_1 + P_2$

$P_2 = I - P_1$

♣ Th: Let $\mathbf{M}_1$ and $\mathbf{M}_2$ are perpendicular projection matrices on $\mathbb{R}^n$. $(\mathbf{M}_1 + \mathbf{M}_2)$ is the perpendicular projection matrix onto $C(\mathbf{M}_1, \mathbf{M}_2)$ if and only if $C(\mathbf{M}_1) \perp C(\mathbf{M}_2)$.

♣ Th: If $\mathbf{M}_1$ and $\mathbf{M}_2$ are symmetric, $C(\mathbf{M}_1) \perp C(\mathbf{M}_2)$, and $(\mathbf{M}_1 + \mathbf{M}_2)$ is the perpendicular projection matrix, then $\mathbf{M}_1$ and $\mathbf{M}_2$ are perpendicular projection matrices.

♣ Th: $\mathbf{M}$ and $\mathbf{M}_0$ are perpendicular projection matrices with $C(\mathbf{M}_0) \subset C(\mathbf{M})$. Then $\mathbf{M} - \mathbf{M}_0$ is a perpendicular projection matrix.

♣ Th:$\mathbf{M}$ and $\mathbf{M}_0$ are perpendicular projection matrices with $C(\mathbf{M}_0) \subset C(\mathbf{M})$. Then $C(\mathbf{M} - \mathbf{M}_0)$ is the orthogonal complement of $C(\mathbf{M}_0)$ with respect to $C(\mathbf{M})$. If $\mathbf{x} \in C(\mathbf{M})$ and $\mathbf{x} \perp C(\mathbf{M}_0)$, then $\mathbf{x} = \mathbf{M}\mathbf{x} = (\mathbf{M} - \mathbf{M}_0)\mathbf{x} + \mathbf{M}_0\mathbf{x} = (\mathbf{M} - \mathbf{M}_0)\mathbf{x}$. Thus, $\mathbf{x} \in C(\mathbf{M} - \mathbf{M}_0)$, so the orthogonal complement of $C(\mathbf{M}_0)$ with respect to $C(\mathbf{M})$ is contained in $C(\mathbf{M} - \mathbf{M}_0)$.

★ Cor: $r(\mathbf{M}) = r(\mathbf{M}_0) + r(\mathbf{M} - \mathbf{M}_0)$

★ Summary!

- Say $\mathbf{P}$ is the perpendicular projection matrix onto $C(\mathbf{X})$.
- I is the perpendicular projection operator onto $\mathbb{R}^n$.
- $(I - \mathbf{P})$: the perpendicular projection matrix onto the orthogonal complement of $C(\mathbf{X})$ with respect to $\mathbb{R}^n$. $= C(P) = N(X^T)$
- $r(I) = r(\mathbf{P}) + r(I - \mathbf{P}) = r + (n - r)$ where $r(C(\mathbf{X})) = r$.
- Decompose $\mathbf{y}$ into

$$\mathbf{y} = \mathbf{P}\mathbf{y} + (I - \mathbf{P})\mathbf{y} = \hat{\mathbf{y}} + \hat{\mathbf{e}}$$