Geometry of Least Squares

$$\underset{\sim}{y} = X\underset{\sim}{\beta} + \underset{\sim}{u}$$

$\underset{\sim}{y}_{n\times 1}$, $X_{n\times k}$, $\beta_{k\times 1}$, $u_{n\times 1}$

Least Squares, $\Rightarrow$ $\hat{\underset{\sim}{\beta}} = (X^T X)^{-1} X^T \underset{\sim}{y}$

Our interest lies in $\hat{\beta}$:
<u>Numerical</u> <u>properties</u>.

2.2. Geometry of Vector Spaces.

$\underset{\sim}{y}_{n\times 1}$ and it belongs to what is called a Euclidean Space, denoted E^n.

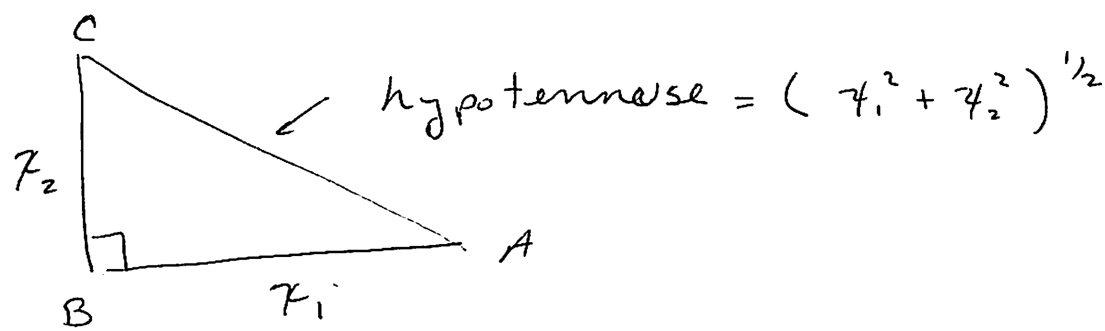
$$\underset{\sim}{y} \in E^n$$

The length of a vector (<u>NORM</u>)

is $\|\underset{\sim}{x}\| = (\underset{\sim}{x}^T \underset{\sim}{x})^{1/2}$

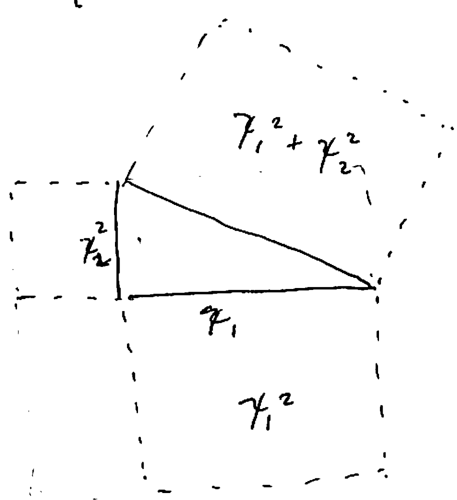
$$= (x_1^2 + x_2^2 + \cdots + x_n^2)^{1/2}$$

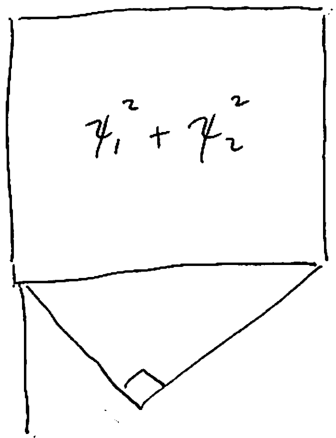
Pythagorean Theorem

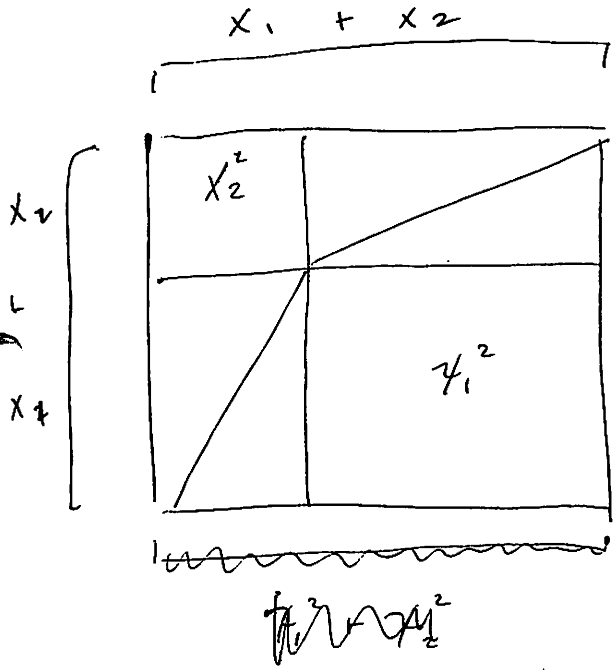


x_1 & x_2 are the lengths of the 2 sides of the triangle.

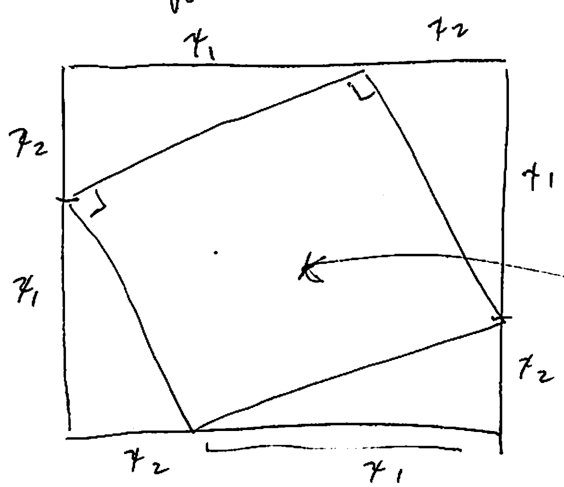
Proof.







4 copies of the right triangle contained in this Box



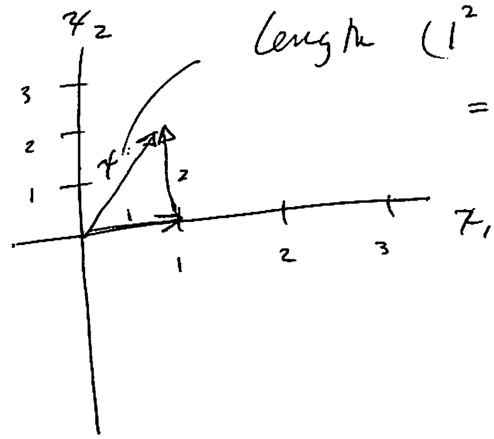
Same 4 copies, what is left must be equal to

$x_1^2 + x_2^2$

The elements of a vector can be thought of as coordinates in n-space.

Example

Let $\underset{\sim}{x} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$

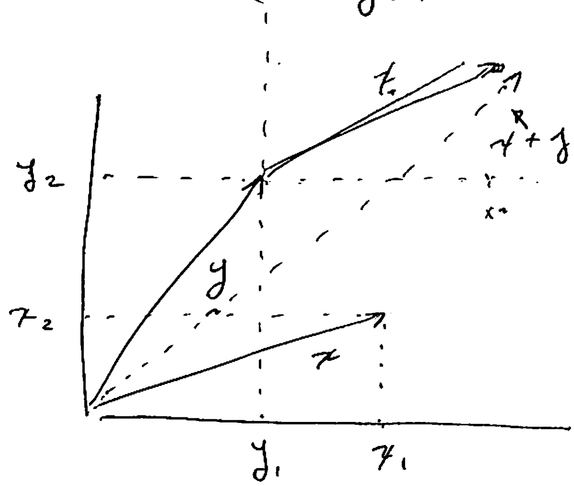


length $(1^2 + 2^2)^{1/2} = \sqrt{5}$

Geometry in 2-space.

$\underset{\sim}{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \qquad \underset{\sim}{y} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$

$\underset{\sim}{x} + \underset{\sim}{y} = \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \end{pmatrix}$ <u>Vector Addition</u>



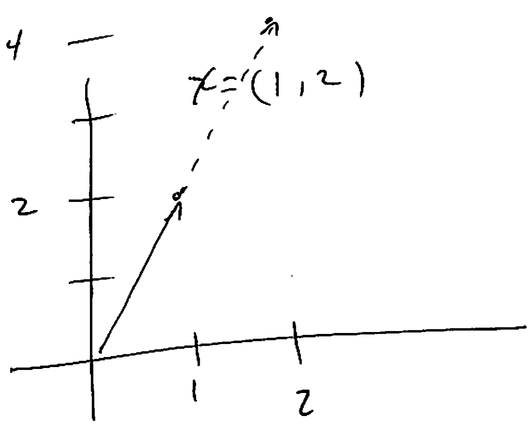
Multiply vector by scalar

$$\alpha \underset{\sim}{x} = \begin{pmatrix} \alpha x_1 \\ \alpha x_2 \end{pmatrix}$$

length $\|\alpha x\|$

$$(\alpha^2 x_1^2 + \alpha^2 x_2^2)^{1/2}$$
$$= |\alpha| \, \|x\|$$

So, you can shorten or lengthen any vector by multiplying it by a scalar. Its position is <u>not</u> changed much.



$2x = 2, 4$

$$|\alpha| = \sqrt{\alpha^2}$$

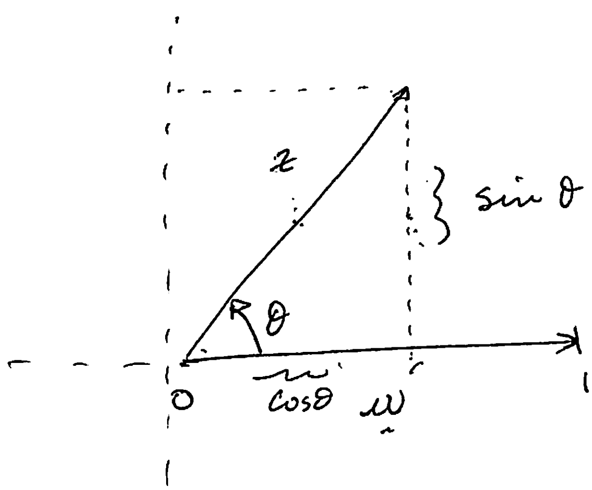
Geometry of Scalar Products

Take 2 vectors with length = 1

$\underset{\sim}{w}$, $\underset{\sim}{z}$

$\underset{\sim}{w} = (1, 0)$ note $(1^2 + 0^2)^{1/2} = 1$

The vector $\underset{\sim}{z}$ is also of length 1 and is [at] an angle $\underset{\sim}{\theta}$ relative to w



$\sin\theta = \frac{O}{H}$

$\cos\theta = \frac{A}{H}$

$z = (\cos\theta, \sin\theta)$

SIN ~~Cos~~ $\frac{O}{H}$

Cos ~~Sin~~ $\frac{A}{H}$

Then, By Pythagoras Theorem $a^2 + b^2 = c^2$

$\cos^2\theta + \sin^2(\theta) = 1$

$\underset{\sim}{w} = (1, 0) \qquad \underset{\sim}{z} = (\cos\theta, \sin\theta)$

$w_T \cancel{z}$

$w^T z = \cos^2\theta$

More generally, let $\underset{\sim}{x} = \alpha \underset{\sim}{w}$ $\qquad \alpha, \gamma > 0$

$\underset{\sim}{y} = \gamma \underset{\sim}{z}$

$$\underset{\sim}{x}^T \underset{\sim}{y} = \alpha\gamma\, w^T z = \alpha\gamma \cos\theta$$

x is parallel to w $\Rightarrow$ angle between

y " " to z $\qquad$ the 2 is same.

$\alpha = \|x\|, \; \gamma = \|y\|$

since $\|x\| = \|\alpha w\| = (\alpha^2 w_1^2 + \alpha_2^2 w_2^2)^{1/2}$

$= |\alpha| \, \|w\|$

and $\|w\| = 1$ so,

$\|x\| = |\alpha|$

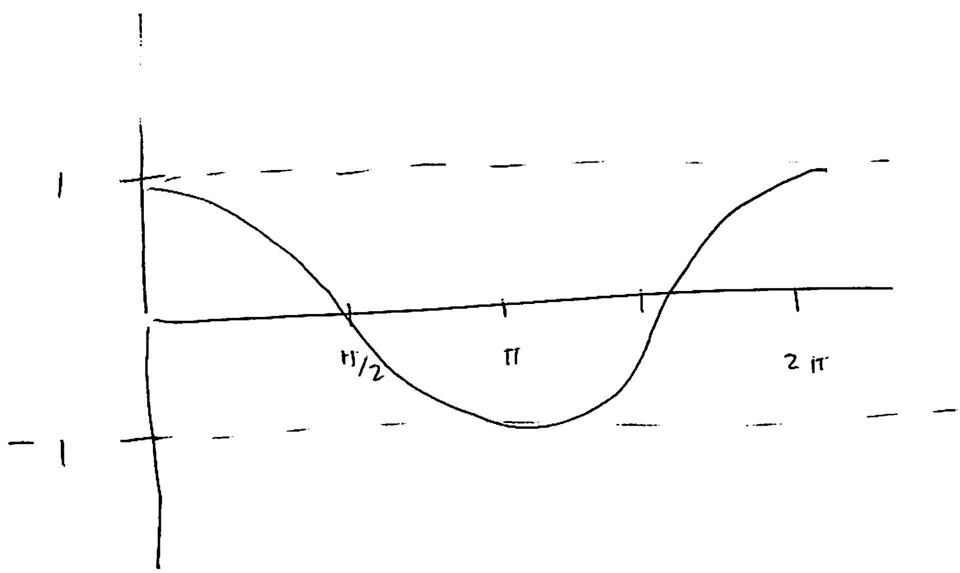
$$\Rightarrow \underset{\sim}{x}^T \underset{\sim}{y} = \|x\| \|y\| \cos\theta$$

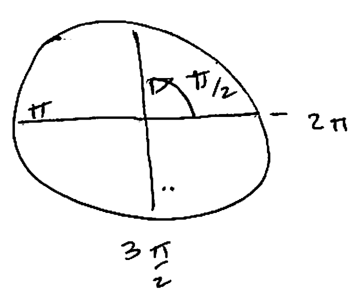
This is true of vectors of any length

$$\underset{\sim}{x}^T \underset{\sim}{y} = \|x\|\|y\| \cos\theta$$

θ tells us something about the directional relationship between 2 vectors.

$$-1 \le \cos\theta \le 1$$





$\pi/2$ - 90° $\cos\theta = \underline{\underline{0}}$

- $\cos\theta = \underline{\underline{1}}$

0, & 2π if parallel.

If at right angles $-$ $\cos\theta = 0$

If parallel $-$ $\cos\theta = 1$

If pointing in opposite directions $\cos\theta = -1$

$\Rightarrow$ 2 vectors at right angles

$$\underset{\sim}{x}^T \underset{\sim}{y} = \|x\|\,\|y\| \cos\theta = \|x\|\,\|y\| \cdot 0 = 0$$

These vectors are said to be
<u>orthogonal</u>.

$$\cancel{\underset{\sim}{x}^T \underset{\sim}{y}} =$$

Also, since $\cos(\theta)$ is always between -1 and 1 then

$$|x^T y| \leq \|x\|\,\|y\|$$

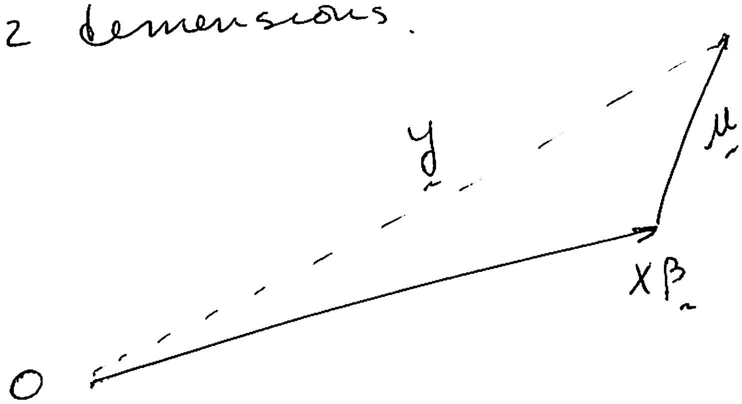
which is the Cauchy-~~Sch~~ Schwartz inequality.

Subspace of $\underset{\sim}{E}^n$

$$\underset{\sim}{y} = X\underset{\sim}{\beta} + \underset{\sim}{u}$$

Each of these are n-vectors.

On the R.H.S there are 2 vectors - and we can draw 2 vectors in 2 dimensions.



Since they are equal, $\underset{\sim}{y}$ is implied!

The matrix $X = \{ \underset{\sim}{x}_1 \ \underset{\sim}{x}_2 \cdots \underset{\sim}{x}_k \}$

has k columns, n rows with

$n \geq k$. The columns of X define

a subspace of X (K dimension). These

are called <u>Basis Vectors</u>

The subspace associated with these

k basis vectors is denoted

$S(X)$ or $S(\underset{\sim}{x}_1 \ \underset{\sim}{x}_2 \cdots \underset{\sim}{x}_n)$. The

Basis vectors are said to <u>Span</u>

this subspace.

~~[illegible] [illegible]~~ The subspace

$S(X)$ consists of every vector

that can be formed as a linear

combination of $\underset{\sim}{x}_i$ $i = 1, .., k$.

$\mathcal{S}(X)$ is called the subspace spanned by the columns of ~~X~~ X.

Sometimes called the <u>column space</u> of X.

or just the <u>span</u> of X.

The orthogonal complement of $\mathcal{S}(X)$ in E^n ($\mathcal{S}^{\perp}(X)$) is the set of all vectors in E^n that are orthogonal to everything in the column space of X. (or span of X).

$$\mathcal{S}(\underset{\sim}{x}_1 \; \underset{\sim}{x}_2 \; \cdots \; x_k) \equiv \Big\{ z \in E^n \;\Big|\; z = \sum_{i=1}^{k} b_i x_i \quad b_i \in \mathbb{R} \Big\}$$

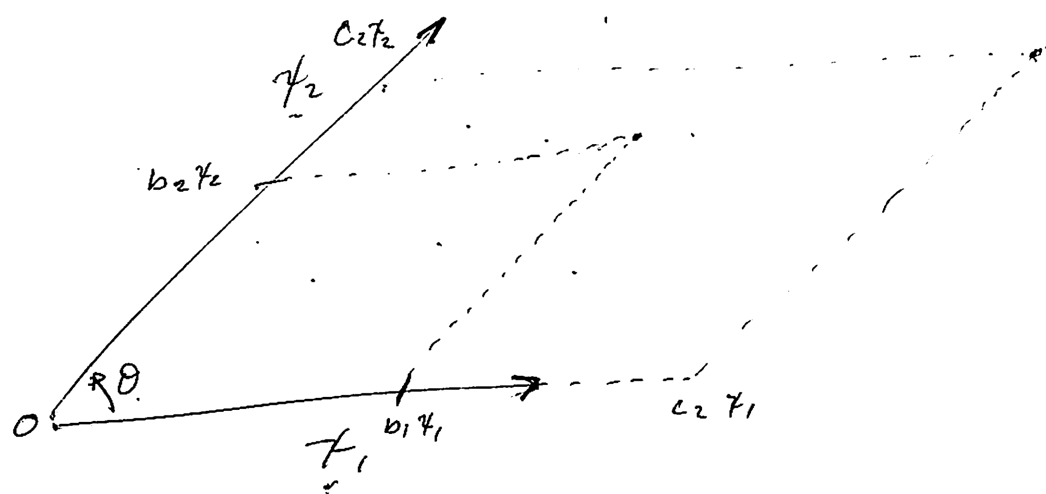
$\underset{\sim}{w}$ are the orthog complement, hence

$$z^T w = 0$$

since $\mathcal{S}^{\perp}(x_1, \ldots, x_k) \equiv \{ w \in E^n \mid w^T z = 0 \text{ for all } z \in \mathcal{S}(X) \}$

Consider the 2 dimensional subspace

$S(x_1, x_2)$



The column space is all linear combs of x_1, x_2

$$z_1 = b_1 x_1 + b_2 x_2$$
$$z_2 = c_1 x_1 + c_2 x_2$$

Basically this defines the entire <u>plane</u>.

Linear Independence

To get LS, X^TX must be invertible, and to be invertible X must have a rank of k.

If the Rank$(X) < k$ then there is at least 1 column in X that can be expressed as a linear comb of the others.

$$\underset{\sim}{x}_j = \sum_{i \neq j} c_i \underset{\sim}{x}_i$$

or

$$\sum_{i \neq j} c_i \underset{\sim}{x}_i - \underset{\sim}{x}_j = 0$$

$$\sum_{i=1}^{k} b_i \underset{\sim}{x}_i = 0$$

note $\quad c_i = \dfrac{b_i}{b_j}$

In Matrix Terms this can be written

$$X \underset{\sim}{b} = \underset{\sim}{0}.$$

If the columns of X are linearly independent, then $X^T X$ is invertible.

Proof:

If linearly dependent, then there is some <u>nonzero</u> ~~v~~ vector $\underset{\sim}{b}$ s.t.

$X\underset{\sim}{b} = 0.$

Premultiply by X^T

$$X^T X \underset{\sim}{b} = X^T \cdot 0 = \underset{\sim}{0}$$

$$(X^T X)^{-1} (X^T X)\underset{\sim}{b} = (X^T X)^{-1} \cdot 0$$

$$\underset{\sim}{b} = 0$$

This is a <u>contradiction</u> $\therefore$

~~the~~ $(X^T X)^{-1}$ cannot exist when columns of X are linearly dep. $\square$.

NOTE:

This is also sufficient

Geometry of OLS

Any point in a subspace of $S(X)$ can be represented as a linear comb. of the columns of X.

$X_{n \times k}$ Rank$(X) = k$

$\therefore$ $S(X)$ is k-dimensional subspace. The basic vectors span this space.

$$X = \{ \underset{\sim}{i} \;\; \underset{\sim}{x}_2 \;\; \underset{\sim}{x}_3 \;\; \cdots \;\; \underset{\sim}{x}_k \}$$

$X\underset{\sim}{\beta}$ is a linear comb of columns of X.

$$X\underset{\sim}{\beta} = \underset{\sim}{x}_1 \beta_1 + \underset{\sim}{x}_2 \beta_2 + \cdots + \underset{\sim}{x}_k \beta_k$$

$X\underset{\sim}{\beta}$ Belongs to $S(X)$

$X\hat{\beta}$ " " $S(X)$ $\hat{\beta}$ OLS

Any $X\tilde{\beta}$ " " $S(X)$!

<u>OLS</u> $\hat{\beta} = (X^T X)^{-1} X^T \underset{\sim}{y}$

satisfies

$$\underset{k \times n}{X^T}\underset{n \times 1}{(y - X\hat{\underset{\sim}{\beta}})} = \underset{k \times 1}{\underset{\sim}{0}}$$ k orthog <u>conditions</u>

$X^T \hat{e} = 0 \Rightarrow \hat{e}$ is orthogonal to X

$\Rightarrow \hat{e}$ is in orthog complement.

1) The OLS residuals $\hat{u}$ are orthogonal to <u>each</u> <u>regressor</u>

$$\underset{1 \times n}{\underset{\sim}{x}_k^T} (\underset{\sim}{y} - X \underset{n \times 1}{\underset{\sim}{\beta}}) = 0$$ here

2) The OLS residuals are orthog to every vector in $S(X)$ (with 1 above as special case)

if $X^T(y - X\hat{\beta}) = 0$ then $\beta^T X^T (y - X\hat{\beta}) = 0$

$\beta^T X^T \hat{u} = 0$

For any vector β. So, if $\beta = \hat{\beta}$ (OLS)

then

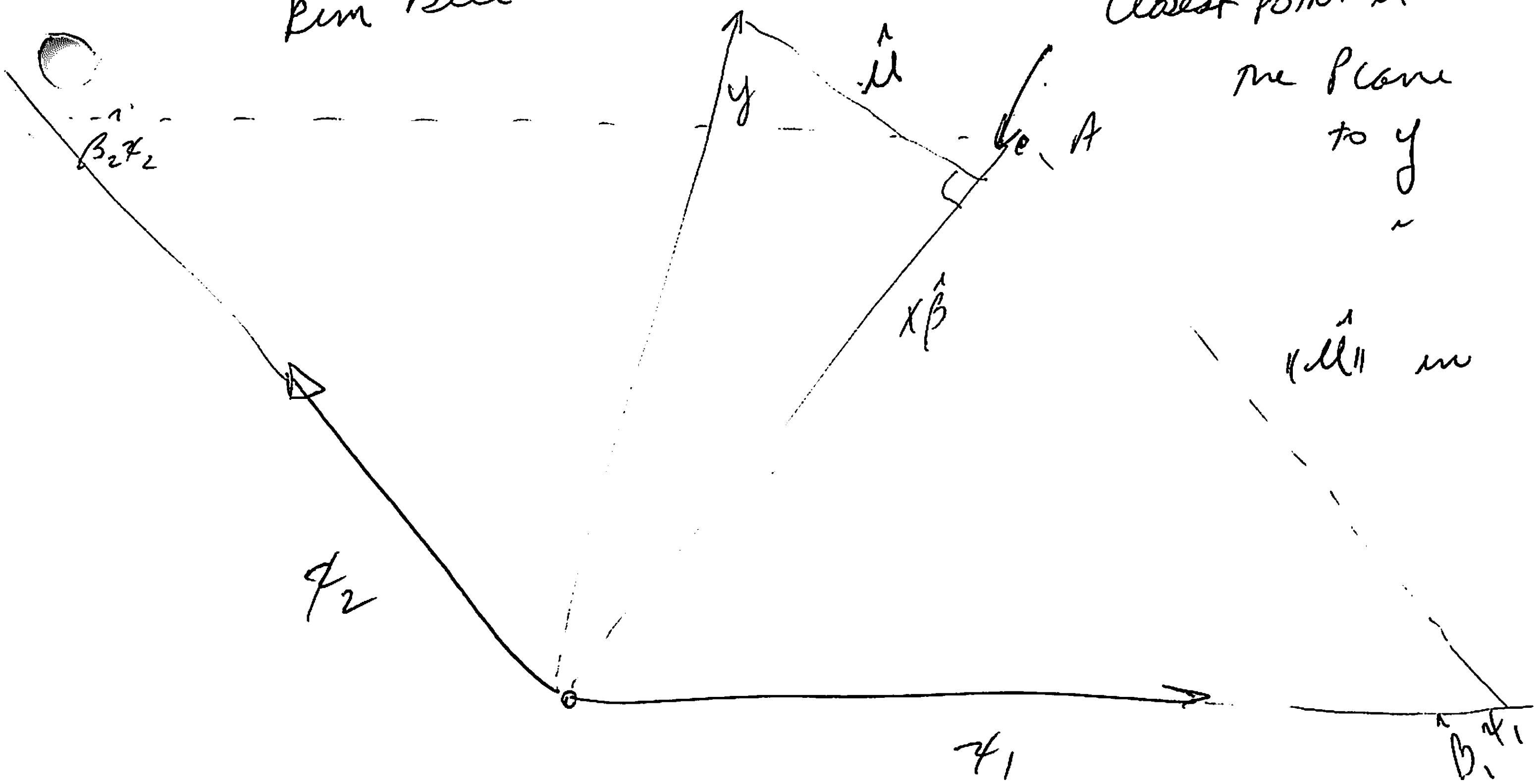
$$\hat{\beta}^T X^T \hat{u} = 0$$

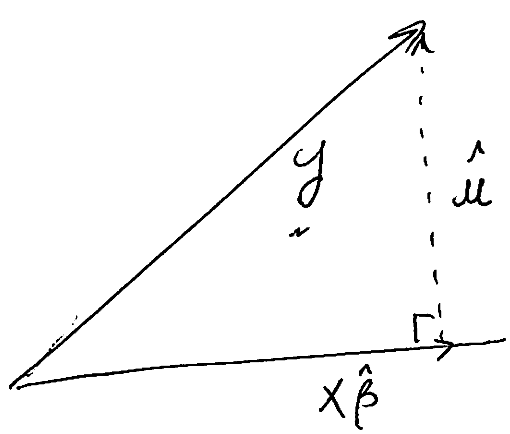
$$\hat{\beta}^T X^T = \hat{Y} - \text{fitted values.}$$

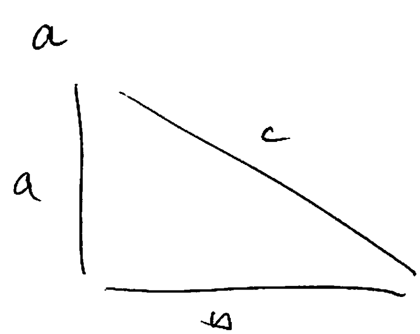
(3) Fitted values from OLS are orthog to the OLS residuals.

Kim Bell -

Closest point in the plane to y







$$a^2 + b^2 = c^2$$

$$\|\hat{u}\|^2 + \|X\hat{\beta}\|^2 = \|y\|^2$$

$$\hat{u}^T\hat{u} + \hat{\beta}^T X^T X^T \hat{\beta} = y^T y$$

$$SSR + ESS = TSS$$