The Geometry of Vector Spaces

$\boldsymbol{x} \in E^N$: vector $\boldsymbol{x}$ belongs to an N-dimensional Euclidean space.

The **Inner product** of vectors $\boldsymbol{x}$ and $\boldsymbol{y} \in E^N$ is

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle \equiv \boldsymbol{x}^\top \boldsymbol{y}. \tag{1}$$

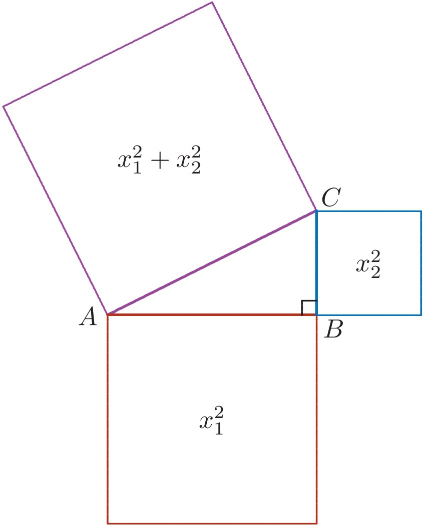
The **length** or **norm** of a vector $\boldsymbol{x}$ in E^N is

$$\|\boldsymbol{x}\| \equiv (\boldsymbol{x}^\top \boldsymbol{x})^{1/2} = \left(\sum_{i=1}^{N} x_i^2 \right)^{1/2}. \tag{2}$$

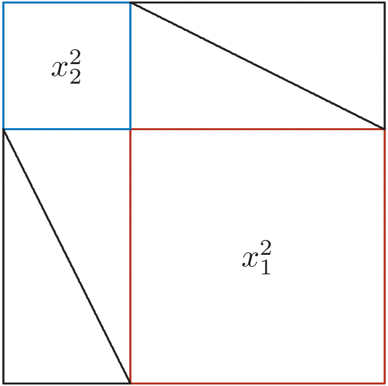
A vector $\boldsymbol{x}$ in E^2 has coordinates x_1 and x_2.

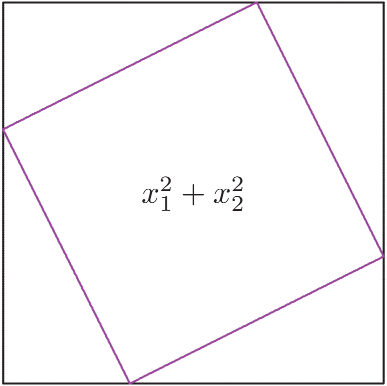
Pythagoras' Theorem tells us that the length of the vector $\boldsymbol{x}$, the hypotenuse of a right-angled triangle, is $(x_1^2 + x_2^2)^{1/2}$.

Pythagoras' Theorem

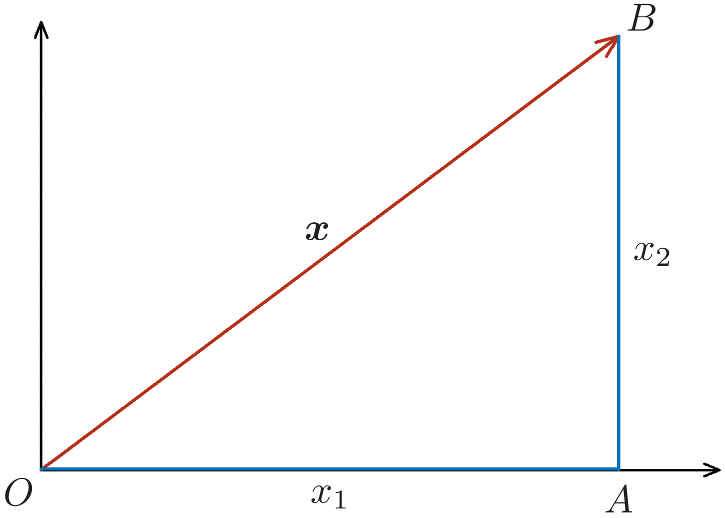


Proof of Pythagoras' Theorem

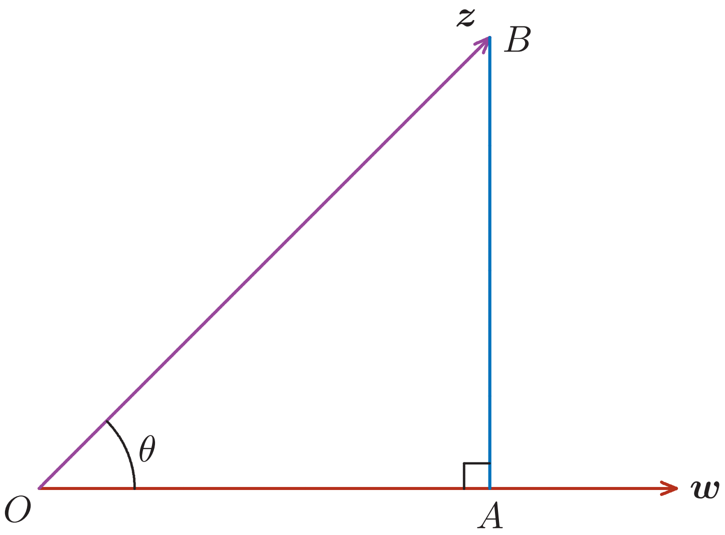




A Vector x in E^2



The Angle Between Two Vectors



Let $\boldsymbol{w}$ and $\boldsymbol{z}$ be vectors of length 1, and let θ denote the angle between them. This is illustrated in the figure just shown.

$\boldsymbol{w}$ has coordinates $(1, 0)$ and is represented by a horizontal line of length 1. Since $\boldsymbol{z}$ is also of length 1, $\|\boldsymbol{z}\| = 1$.

Because

$$\|\boldsymbol{z}\|^2 = \cos^2\theta + \sin^2\theta = 1, \tag{3}$$

the coordinates of $\boldsymbol{z}$ must be $(\cos\theta, \sin\theta)$.

The scalar product of $\boldsymbol{w}$ and $\boldsymbol{z}$ is

$$\langle \boldsymbol{w}, \boldsymbol{z} \rangle = \boldsymbol{w}^\top \boldsymbol{z} = w_1 z_1 + w_2 z_2 = z_1 = \cos\theta, \tag{4}$$

because $\boldsymbol{w}$ has coordinates $(1, 0)$.

More generally, if $\boldsymbol{x} = \alpha\boldsymbol{w}$ and $\boldsymbol{y} = \gamma\boldsymbol{z}$, for positive scalars α and γ, then $\|\boldsymbol{x}\| = \alpha$ and $\|\boldsymbol{y}\| = \gamma$. Thus

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \boldsymbol{x}^\top \boldsymbol{y} = \alpha\gamma\, \boldsymbol{w}^\top \boldsymbol{z} = \alpha\gamma \langle \boldsymbol{w}, \boldsymbol{z} \rangle. \tag{5}$$

Because $\boldsymbol{x}$ is parallel to $\boldsymbol{w}$, and $\boldsymbol{y}$ is parallel to $\boldsymbol{z}$, the angle between $\boldsymbol{x}$ and $\boldsymbol{y}$ is the same as that between $\boldsymbol{w}$ and $\boldsymbol{z}$, namely θ. Therefore,

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \|\boldsymbol{x}\| \, \|\boldsymbol{y}\| \cos\theta. \tag{6}$$

The inner product of any two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ is the product of their lengths times the cosine of the angle between them.

- $\cos\theta$ varies between -1 and 1.
- $\cos 0 = 1$, $\cos \pi/2 = 0$, and $\cos \pi = -1$.
- $\cos\theta$ is 1 for vectors that are parallel.
- $\cos\theta$ is 0 for vectors that are at right angles to each other.
- $\cos\theta$ is -1 for vectors that point in directly opposite directions.
- If $\boldsymbol{x}$ and $\boldsymbol{y}$ form a right angle, then $\cos\theta = 0$, and $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = 0$.
- If $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = 0$, then $\cos\theta = 0$ unless $\boldsymbol{x}$ or $\boldsymbol{y}$ is a zero vector. Except in that special case, $\theta = \pi/2$ when $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = 0$.

An implication of (6) is the **Cauchy-Schwartz inequality**,

$$|\boldsymbol{x}^\top \boldsymbol{y}| \leq \|\boldsymbol{x}\| \, \|\boldsymbol{y}\| . \tag{7}$$

This becomes an equality only when $\boldsymbol{x}$ and $\boldsymbol{y}$ are parallel, so that $|\cos\theta| = 1$.

Vectors that are at right angles are said to be **orthogonal** or **perpendicular**.

Now consider the regression model

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{u}. \tag{8}$$

Here $\boldsymbol{y}$ and each column of $\boldsymbol{X}$ can be thought of as vectors in E^N.

Denote the k columns of $\boldsymbol{X}$ (assumed to have full rank) as $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots \boldsymbol{x}_k$.

These k **basis vectors** are said to **span** a k-dimensional **subspace**.

The subspace they span may be denoted $\mathcal{S}(\boldsymbol{X})$ or $\mathcal{S}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k)$.

- $\mathcal{S}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k)$ consists of every vector that can be formed as a **linear combination** of the $\boldsymbol{x}_i$, $i = 1, \ldots, k$. Formally,

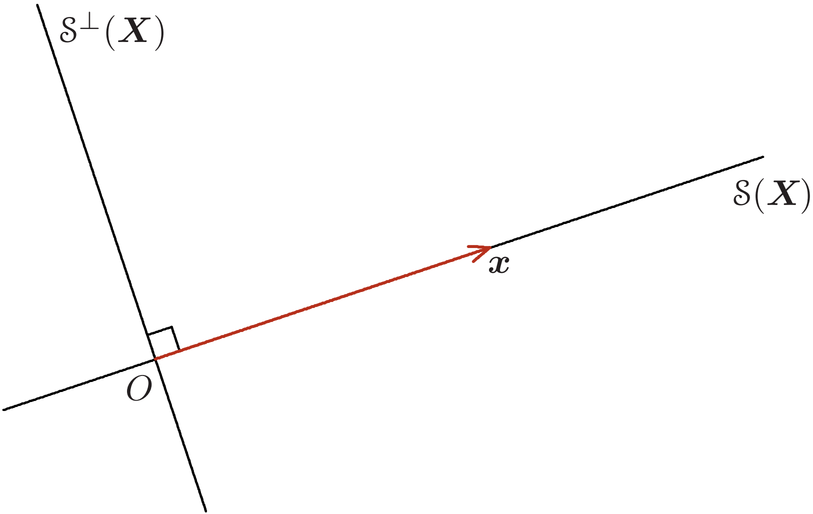
$$\mathcal{S}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k) \equiv \left\{ \boldsymbol{z} \in E^N \;\middle|\; \boldsymbol{z} = \sum_{i=1}^{k} b_i \boldsymbol{x}_i, \quad b_i \in \mathbb{R} \right\}. \tag{9}$$

- $\mathcal{S}(\boldsymbol{X})$ is called the subspace spanned by the $\boldsymbol{x}_i$, or the **column space** of $\boldsymbol{X}$, or the **span** of $\boldsymbol{X}$, or the span of the $\boldsymbol{x}_i$.
- For every dataset $[\boldsymbol{y}, \boldsymbol{X}]$, both $\boldsymbol{y}$ and every column of $\boldsymbol{X}$ lie in E^N, and $\boldsymbol{X}$ defines a k-dimensional subspace, $\mathcal{S}(\boldsymbol{X})$.
- The **orthogonal complement** of $\mathcal{S}(\boldsymbol{X})$ in E^N, denoted $\mathcal{S}^{\perp}(\boldsymbol{X})$, is the set of all vectors $\boldsymbol{w}$ in E^N that are orthogonal to everything in $\mathcal{S}(\boldsymbol{X})$:

$$\mathcal{S}^{\perp}(\boldsymbol{X}) \equiv \left\{ \boldsymbol{w} \in E^N \mid \boldsymbol{w}^{\top}\boldsymbol{z} = 0 \text{ for all } \boldsymbol{z} \in \mathcal{S}(\boldsymbol{X}) \right\}. \tag{10}$$

If dimension of $\mathcal{S}(\boldsymbol{X})$ is k, then dimension of $\mathcal{S}^{\perp}(\boldsymbol{X})$ is $N - k$.

The Spaces $\mathcal{S}(\boldsymbol{X})$ and $\mathcal{S}^{\perp}(\boldsymbol{X})$



- If $\boldsymbol{X}$ does not have full rank k, then the dimension of $\mathcal{S}(\boldsymbol{X})$ is less than k, and the dimension of $\mathcal{S}^{\perp}(\boldsymbol{X})$ is greater than $N-k$.
- When $\boldsymbol{X}$ has full rank, the columns of $\boldsymbol{X}$ are **linearly independent**.
- The vectors $\boldsymbol{x}_1$ through $\boldsymbol{x}_k$ are said to be **linearly dependent** if there is a vector $\boldsymbol{x}_j$, $1 \leq j \leq k$, and coefficients c_i such that

$$\boldsymbol{x}_j = \sum_{i \neq j} c_i \boldsymbol{x}_i. \tag{11}$$

- Equivalently, there exist coefficients b_i, at least one of which is nonzero, such that

$$\sum_{i=1}^{k} b_i \boldsymbol{x}_i = \mathbf{0}, \quad \text{or} \quad \boldsymbol{X}\boldsymbol{b} = \mathbf{0}. \tag{12}$$

- If $\boldsymbol{x}_j = \mathbf{0}$, then (12) is satisfied if we make b_j nonzero and set $b_i = 0$ for all $i \neq j$. So linear independence rules out $\boldsymbol{x}_j = \mathbf{0}$ for any j.

- If the columns of $\boldsymbol{X}$ are linearly dependent, there must exist a nonzero vector $\boldsymbol{b}$ such that $\boldsymbol{X}\boldsymbol{b} = \mathbf{0}$.
- Premultiplying the second equation in (12) by $\boldsymbol{X}^\top$ yields

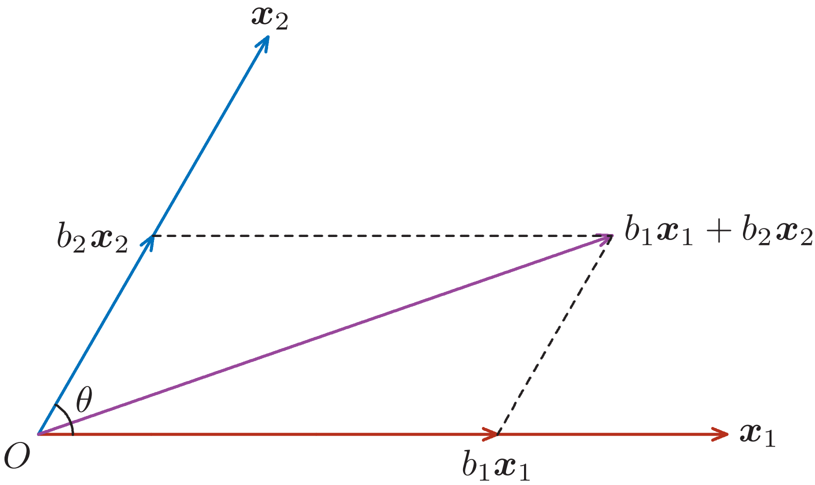
$$\boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{b} = \mathbf{0}. \tag{13}$$

  But if $\boldsymbol{X}^\top\boldsymbol{X}$ is invertible, there exists $(\boldsymbol{X}^\top\boldsymbol{X})^{-1}$ such that $(\boldsymbol{X}^\top\boldsymbol{X})^{-1}(\boldsymbol{X}^\top\boldsymbol{X}) = \mathbf{I}$, the $k \times k$ identity matrix.
- Thus (13) implies that

$$\boldsymbol{b} = \mathbf{I}\boldsymbol{b} = (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{b} = \mathbf{0}. \tag{14}$$

  But this is a contradiction, since we have assumed that $\boldsymbol{b} \neq \mathbf{0}$.
- We conclude that $\boldsymbol{X}^\top\boldsymbol{X}$ cannot have an inverse if the columns of $\boldsymbol{X}$ are linearly dependent.

A Two-Dimensional Subspace



When the dimension of $\mathcal{S}(\boldsymbol{X})$ is $k' < k$, $\mathcal{S}(\boldsymbol{X})$ must be identical to $\mathcal{S}(\boldsymbol{X}')$, where $\boldsymbol{X}'$ is an $N \times k'$ matrix consisting of any k' linearly independent columns of $\boldsymbol{X}$. Consider

$$\boldsymbol{X} = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 4 & 0 \\ 1 & 0 & 1 \\ 1 & 4 & 0 \\ 1 & 0 & 1 \end{bmatrix}. \tag{15}$$

The columns of this matrix are not linearly independent, since $\boldsymbol{x}_1 = \frac{1}{4}\boldsymbol{x}_2 + \boldsymbol{x}_3$. However, any two of the columns are linearly independent. Therefore,

$$\mathcal{S}(\boldsymbol{X}) = \mathcal{S}(\boldsymbol{x}_1, \boldsymbol{x}_2) = \mathcal{S}(\boldsymbol{x}_1, \boldsymbol{x}_3) = \mathcal{S}(\boldsymbol{x}_2, \boldsymbol{x}_3). \tag{16}$$

Henceforth, we will generally assume that the columns of any regressor matrix $\boldsymbol{X}$ are linearly independent.

The Geometry of OLS Estimation

When $\boldsymbol{X} = [\boldsymbol{x}_1 \;\; \boldsymbol{x}_2 \;\; \cdots \;\; \boldsymbol{x}_k]$, then

$$\boldsymbol{X}\boldsymbol{\beta} = [\boldsymbol{x}_1 \;\; \boldsymbol{x}_2 \;\; \cdots \;\; \boldsymbol{x}_k] \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} = \boldsymbol{x}_1\beta_1 + \boldsymbol{x}_2\beta_2 + \ldots + \boldsymbol{x}_k\beta_k = \sum_{i=1}^{k} \beta_i \boldsymbol{x}_i. \tag{17}$$

The moment conditions are

$$\boldsymbol{X}^\top(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}) = \mathbf{0}. \tag{18}$$

This tells us that the **residual vector** $\hat{\boldsymbol{u}} \equiv \boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}$ is orthogonal to every regressor.

Thus $\hat{\boldsymbol{u}}$ must lie in $\mathcal{S}^\perp(\boldsymbol{X})$.

(18) implies that $\hat{\boldsymbol{u}}$ is orthogonal to *every* vector in $\mathcal{S}(\boldsymbol{X})$. Since $\boldsymbol{X}^\top\hat{\boldsymbol{u}} = \mathbf{0}$, then for every $\boldsymbol{\beta}$

$$(\boldsymbol{X}\boldsymbol{\beta})^\top\hat{\boldsymbol{u}} = \boldsymbol{\beta}^\top\mathbf{X}^\top\hat{\boldsymbol{u}} = \mathbf{0}. \tag{19}$$

The vector $\boldsymbol{X}\hat{\boldsymbol{\beta}}$ is referred to as the vector of **fitted values**. Clearly, it lies in $\mathcal{S}(\boldsymbol{X})$, and, consequently, it must be orthogonal to $\hat{\boldsymbol{u}}$.

By Pythagoras' Theorem,

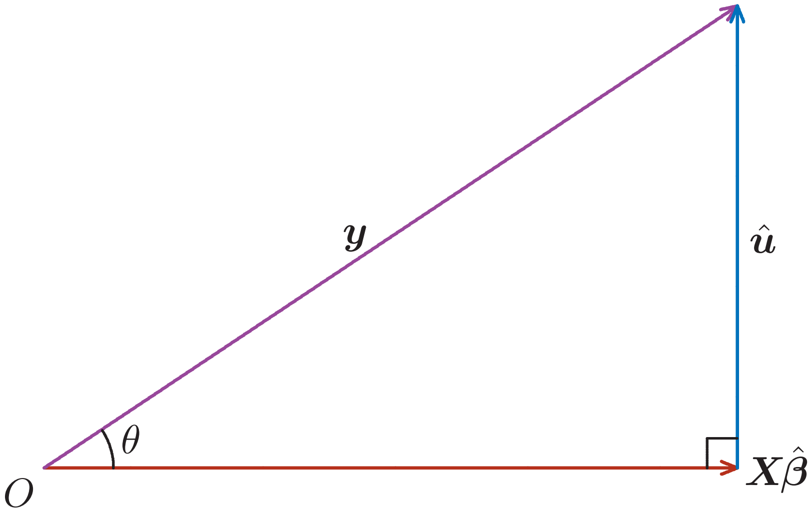
$$\|\boldsymbol{y}\|^2 = \|\boldsymbol{X}\hat{\boldsymbol{\beta}}\|^2 + \|\hat{\boldsymbol{u}}\|^2. \tag{20}$$

If we write out the squared norms as scalar products, this becomes

$$\boldsymbol{y}^\top\boldsymbol{y} = \hat{\boldsymbol{\beta}}^\top\mathbf{X}^\top\boldsymbol{X}\hat{\boldsymbol{\beta}} + (\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})^\top(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}). \tag{21}$$

The **total sum of squares**, or **TSS**, is equal to the **explained sum of squares**, or **ESS**, plus the **sum of squared residuals**, or **SSR**.

Residuals and Fitted Values



Projections and Projection Matrices

A **projection** maps each point of E^N into a point in a subspace of E^N.

An **orthogonal projection** maps any point into the point of the subspace that is closest to it.

An orthogonal projection of any vector on to a given subspace can be performed by premultiplying the vector by a suitable **projection matrix**. For an N-vector, it is an $N \times N$ matrix.

The projection matrices that yield fitted values and residuals are

$$\boldsymbol{P_X} = \boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top \quad \text{and} \tag{22}$$

$$\boldsymbol{M_X} = \mathbf{I} - \boldsymbol{P_X} = \mathbf{I} - \boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top, \tag{23}$$

where $\mathbf{I}$ is the $N \times N$ identity matrix.

Notice that $\boldsymbol{P_X}$ and $\boldsymbol{M_X}$ are both symmetric matrices.

The **invariant subspace** or **image** of $\boldsymbol{P_X}$ is $\mathcal{S}(\boldsymbol{X})$.

Since

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{y}, \tag{24}$$

then

$$\boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{y} = \boldsymbol{P_X y}, \tag{25}$$

and

$$\boldsymbol{P_X X b} = \boldsymbol{X b}. \tag{26}$$

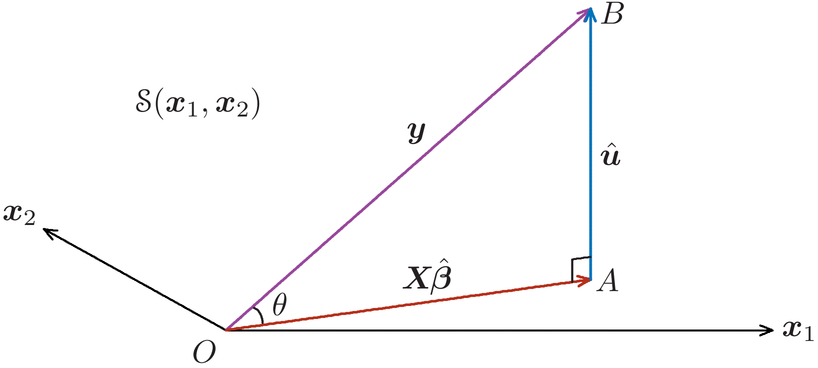
Thus $\boldsymbol{P_X}$ maps any vector in $\mathcal{S}(\boldsymbol{X})$ into itself.

Similarly, when $\boldsymbol{M_X}$ is applied to $\boldsymbol{y}$, it yields the vector of residuals

$$\boldsymbol{M_X y} = \big(\mathbf{I} - \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top\big)\boldsymbol{y} = \boldsymbol{y} - \boldsymbol{P_X y} = \boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{u}}. \tag{27}$$

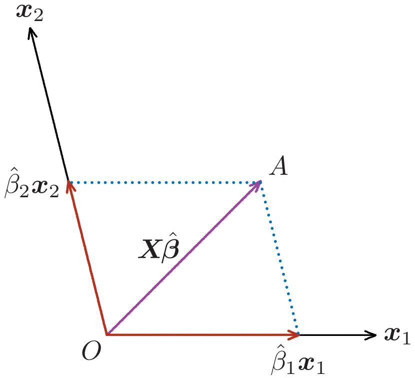
The image of $\boldsymbol{M_X}$ is $\mathcal{S}^\perp(\boldsymbol{X})$, which is the orthogonal complement of the image of $\boldsymbol{P_X}$.

Linear Regression in Three Dimensions

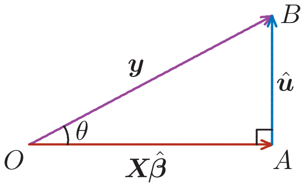


(a) $\boldsymbol{y}$ projected on two regressors

Linear Regression in Three Dimensions



(b) The span $\mathcal{S}(\boldsymbol{x}_1, \boldsymbol{x}_2)$ of the regressors



(c) The vertical plane through $\boldsymbol{y}$

The equality

$$(\boldsymbol{M_X y})^\top \boldsymbol{X} = \boldsymbol{y}^\top \boldsymbol{M_X X} \tag{28}$$

relies on the symmetry of $\boldsymbol{M_X}$. Then, from (23), we have

$$\boldsymbol{M_X X} = (\mathbf{I} - \boldsymbol{P_X})\boldsymbol{X} = \boldsymbol{X} - \boldsymbol{X} = \mathbf{O}, \tag{29}$$

For any matrix to represent a projection, it must be **idempotent**. When multiplied by itself, an idempotent matrix yields itself again. Thus,

$$\boldsymbol{P_X P_X} = \boldsymbol{P_X} \quad \text{and} \quad \boldsymbol{M_X M_X} = \boldsymbol{M_X}. \tag{30}$$

This must be the case, because $\boldsymbol{P_X a} = \boldsymbol{P_X P_X a}$ for any vector $\boldsymbol{a}$. Similarly, $\boldsymbol{M_X a} = \boldsymbol{M_X M_X a}$ for any vector $\boldsymbol{a}$.

Geometrically, $\boldsymbol{P_X}$ takes $\boldsymbol{a}$ to the closest point in $\mathcal{S}(\boldsymbol{X})$. Since the vector $\boldsymbol{P_X a}$ is already at that point, $\boldsymbol{P_X P_X a}$ has nowhere else to go.

Geometrically, $\boldsymbol{M_X}$ takes $\boldsymbol{a}$ to the closest point in $\mathcal{S}^\perp(\boldsymbol{X})$.

Since, from (23),

$$\boldsymbol{P_X} + \boldsymbol{M_X} = \mathbf{I}, \tag{31}$$

any vector $\boldsymbol{y} \in E^N$ is equal to $\boldsymbol{P_X y} + \boldsymbol{M_X y}$.

The pair of projections $\boldsymbol{P_X}$ and $\boldsymbol{M_X}$ are said to be **complementary projections**.

The projection matrices $\boldsymbol{P_X}$ and $\boldsymbol{M_X}$ define an **orthogonal decomposition** of E^N, because $\boldsymbol{M_X y}$ and $\boldsymbol{P_X y}$ lie in two orthogonal subspaces.

Evidently,

$$\boldsymbol{P_X M_X} = \mathbf{O}. \tag{32}$$

We may say that $\boldsymbol{P_X}$ and $\boldsymbol{M_X}$ **annihilate** each other.

Now consider any vectors $\boldsymbol{z} \in \mathcal{S}(\boldsymbol{X})$ and $\boldsymbol{w} \in \mathcal{S}^{\perp}(\boldsymbol{X})$.

Since $\boldsymbol{z} = \boldsymbol{P_X} \boldsymbol{z}$ and $\boldsymbol{w} = \boldsymbol{M_X} \boldsymbol{w}$,

$$\boldsymbol{z}^\top \boldsymbol{w} = (\boldsymbol{P_X}^\top \boldsymbol{z})^\top \boldsymbol{M_X} \boldsymbol{w} \tag{33}$$

$$= \boldsymbol{z}^\top \boldsymbol{P_X} \boldsymbol{M_X} \boldsymbol{w} = 0. \tag{34}$$

$\boldsymbol{M_X}$ annihilates all points that lie in $\mathcal{S}(\boldsymbol{X})$, and $\boldsymbol{P_X}$ likewise annihilates all points that lie in $\mathcal{S}^\perp(\boldsymbol{X})$.

An orthogonal decomposition $\boldsymbol{y} = \boldsymbol{P_X} \boldsymbol{y} + \boldsymbol{M_X} \boldsymbol{y}$ can be represented by a right-angled triangle, with $\boldsymbol{y}$ as the hypotenuse and $\boldsymbol{P_X} \boldsymbol{y}$ and $\boldsymbol{M_X} \boldsymbol{y}$ as the other two sides.

The relationship between TSS, ESS, and SSR given in (21) is just Pythagoras' Theorem:

$$\|\boldsymbol{y}\|^2 = \|\boldsymbol{P_X} \boldsymbol{y}\|^2 + \|\boldsymbol{M_X} \boldsymbol{y}\|^2. \tag{35}$$

Because the hypotenuse of a right-angled triangle is longer than either of the other two sides, ESS $\leq$ TSS, and SSR $\leq$ TSS.

Of course, projection matrices may be associated with any subspaces, not just $\mathcal{S}(\boldsymbol{X})$ and $\mathcal{S}^{\perp}(\boldsymbol{X})$. Here are some examples:

- $\boldsymbol{P_Z}$ is the matrix that projects orthogonally on to $\mathcal{S}(\boldsymbol{Z})$;
- $\boldsymbol{M_{X,W}}$ is the matrix that projects orthogonally off $\mathcal{S}(\boldsymbol{X}, \boldsymbol{W})$, or on to $\mathcal{S}^{\perp}(\boldsymbol{X}, \boldsymbol{W})$.
- $\boldsymbol{P}_{[\boldsymbol{X}_1\ \boldsymbol{X}_2]}$ is the matrix that projects orthogonally on to $\mathcal{S}(\boldsymbol{X}) = \mathcal{S}\big([\boldsymbol{X}_1\ \ \boldsymbol{X}_2]\big) = \mathcal{S}(\boldsymbol{X}_1, \boldsymbol{X}_2)$, where $\boldsymbol{X} = [\boldsymbol{X}_1\ \ \boldsymbol{X}_2]$.
- $\boldsymbol{P}_{\boldsymbol{X}_1} = \boldsymbol{P}_1$ is the matrix that projects orthogonally on to $\mathcal{S}(\boldsymbol{X}_1)$.
- $\boldsymbol{M}_{\boldsymbol{P_Z X}}$ is the matrix that projects orthogonally on to $\mathcal{S}^{\perp}(\boldsymbol{P_Z X})$.

Note that $\boldsymbol{P_X} = \boldsymbol{P}_{[\boldsymbol{X}_1\ \boldsymbol{X}_2]} = \boldsymbol{P}_{\boldsymbol{X}_1} + \boldsymbol{P}_{\boldsymbol{M}_1\boldsymbol{X}_2}$.

We can add the two projection matrices here to obtain another projection matrix, because everything in $\mathcal{S}(\boldsymbol{M}_1\boldsymbol{X}_2)$ is orthogonal to everything in $\mathcal{S}(\boldsymbol{X}_1)$.

In general, however, the sum of two projection matrices is not a projection matrix.

Let $\boldsymbol{A}$ be partitioned by its columns, which may be denoted $\boldsymbol{a}_i$, $i = 1, \ldots, k$:

$$\boldsymbol{X}\boldsymbol{A} = \boldsymbol{X}\begin{bmatrix}\boldsymbol{a}_1 & \boldsymbol{a}_2 & \cdots & \boldsymbol{a}_k\end{bmatrix} = \begin{bmatrix}\boldsymbol{X}\boldsymbol{a}_1 & \boldsymbol{X}\boldsymbol{a}_2 & \cdots & \boldsymbol{X}\boldsymbol{a}_k\end{bmatrix}. \tag{36}$$

Each block here takes the form $\boldsymbol{X}\boldsymbol{a}_i$, which is a linear combination of the columns of $\boldsymbol{X}$.

Thus any element of $\mathcal{S}(\boldsymbol{X}\boldsymbol{A})$ must also be an element of $\mathcal{S}(\boldsymbol{X})$. But any element of $\mathcal{S}(\boldsymbol{X})$ is also an element of $\mathcal{S}(\boldsymbol{X}\boldsymbol{A})$.

Any element of $\mathcal{S}(\boldsymbol{X})$ can be written as $\boldsymbol{X}\boldsymbol{\beta}$ for some $\boldsymbol{\beta} \in \mathbb{R}^k$. Since $\boldsymbol{A}$ is nonsingular, and thus invertible,

$$\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{X}\boldsymbol{A}\boldsymbol{A}^{-1}\boldsymbol{\beta} = (\boldsymbol{X}\boldsymbol{A})(\boldsymbol{A}^{-1}\boldsymbol{\beta}). \tag{37}$$

Since every element of $\mathcal{S}(\boldsymbol{X}\boldsymbol{A})$ belongs to $\mathcal{S}(\boldsymbol{X})$, and every element of $\mathcal{S}(\boldsymbol{X})$ belongs to $\mathcal{S}(\boldsymbol{X}\boldsymbol{A})$, these two subspaces must be identical.

The orthogonal projections $\boldsymbol{P_X}$ and $\boldsymbol{P_{XA}}$ are the same:

$$\begin{aligned} \boldsymbol{P_{XA}} &= \boldsymbol{XA}(\boldsymbol{A}^\top \boldsymbol{X}^\top \boldsymbol{XA})^{-1}\boldsymbol{A}^\top \boldsymbol{X}^\top && (38)\\ &= \boldsymbol{XAA}^{-1}(\boldsymbol{X}^\top \boldsymbol{X})^{-1}(\boldsymbol{A}^\top)^{-1}\boldsymbol{A}^\top \boldsymbol{X}^\top && (39)\\ &= \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top = \boldsymbol{P_X}. && (40) \end{aligned}$$

The fitted values and residuals depend on $\boldsymbol{X}$ only through $\boldsymbol{P_X}$ and $\boldsymbol{M_X}$, so they too must be invariant to any nonsingular linear transformation of the columns of $\boldsymbol{X}$.

Consider a temperature variable $\boldsymbol{T}$ in Celsius or $\boldsymbol{F}$ in Fahrenheit:

$$\boldsymbol{F} = 32\boldsymbol{\iota} + \tfrac{9}{5}\boldsymbol{T}. \tag{41}$$

If the constant is included in the transformation, then

$$\begin{bmatrix}\boldsymbol{\iota} & \boldsymbol{F}\end{bmatrix} = \begin{bmatrix}\boldsymbol{\iota} & \boldsymbol{T}\end{bmatrix} \begin{bmatrix}1 & 32\\ 0 & 9/5\end{bmatrix}. \tag{42}$$

Residuals and fitted values are unaffected.

Suppose the constant term and slope coefficient in the regression using Celsius are β_1 and β_2.

Similarly, the constant term and slope coefficient in the regression using Fahrenheit are α_1 and α_2.

Then

$$\beta_1 = \alpha_1 + 32\alpha_2 \quad \text{and} \quad \beta_2 = \tfrac{9}{5}\alpha_2. \tag{43}$$

It is often good to rescale regressors to ensure that:

- Coefficient estimates are of similar magnitude;
- The eigenvalues of $\boldsymbol{X}^\top\boldsymbol{X}$ do not differ so much that OLS estimates may be numerically unreliable.

We may redefine regressors so that the coefficients are meaningful:

$$\begin{aligned} \boldsymbol{y} &= \beta_1 + \beta_2\boldsymbol{x}_2 + \beta_3\boldsymbol{x}_3 + \boldsymbol{u} \\ &= \beta_1 + \beta_2(\boldsymbol{x}_2 + \boldsymbol{x}_3) + (\beta_3 - \beta_2)\boldsymbol{x}_3 + \boldsymbol{u}. \end{aligned} \tag{44}$$

The coefficient on $\boldsymbol{x}_3$ is now $\beta_3 - \beta_2$.