$\underset{\sim}{y} = X \underset{\sim}{b} + \underset{\sim}{e}$

STAT 714 fa 2025 Lec 01

Least squares estimation in linear models

Karl B. Gregory

University of South Carolina

These slides are an instructional aid; their sole purpose is to display, during the lecture, definitions, plots, results, etc. which take too much time to write by hand on the blackboard. They are not intended to explain or expound on any material.

$$\underset{n\times n}{A}\;\underset{n\times n}{A} = \underset{n\times n}{A}$$

Idempotent matrix

A $\boxed{\text{square}}$ matrix $\mathbf{A}$ is called *idempotent* if $\mathbf{A}^2 = \mathbf{A}$.

Exercise: [redacted] are idempotent matrices:

$$\begin{bmatrix} 3 & -2 \\ 3 & -2 \end{bmatrix} \qquad \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 3 & -2 \\ 3 & -2 \end{bmatrix}\begin{bmatrix} 3 & -2 \\ 3 & -2 \end{bmatrix} = \begin{bmatrix} 3 & -2 \\ 3 & -2 \end{bmatrix}$$

$V \subset \mathbb{R}^n$

Projection matrix

A $\boxed{\text{square}}$ matrix $\mathbf{P}$ is called a *projection matrix* onto the space V if

1. $\mathbf{P}$ is idempotent
2. for any $\mathbf{x}$, $\mathbf{Px} \in V$ ← brings all vectors inside V
3. for any $\mathbf{z} \in V$, $\mathbf{Pz} = \mathbf{z}$ ← leaves vectors in V unchanged.

Sometimes we call projection matrices simply "projections".

Theorem (Every idempotent matrix is a projection)

Every idempotent matrix is a projection onto its own column space.

Prove the result.

Let $\underset{n\times n}{P}$ be idempotent.

(i) ✓

(ii) Take any $\underset{\sim}{x} \in \mathbb{R}^n$.

Then $P\underset{\sim}{x} \in \text{Col } P$.

(iii) Take $\underset{\sim}{z} \in \underline{\text{Col } P}$. This means $\underset{\sim}{z} = P\underset{\sim}{c}$ for some $\underset{\sim}{c}$.

So $P\underset{\sim}{z} = P(P\underset{\sim}{c}) = P\underset{\sim}{c} = \underset{\sim}{z}$.

$\Rightarrow$ P is a proj. matrix onto Col P.

$$\tilde{P}\tilde{P} = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix} \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix} = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix}$$

$$\text{Col } P = \text{span}\left\{ \begin{bmatrix} 3 \\ 3 \end{bmatrix} \right\} = \text{Col } \tilde{P} = \text{span}\left\{ \begin{bmatrix} 1/2 \\ 1/2 \end{bmatrix} \right\}$$

Exercise: Let

$$\mathbf{P} = \begin{bmatrix} 3 & -2 \\ 3 & -2 \end{bmatrix}, \quad \tilde{\mathbf{P}} = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix}, \quad \text{and} \quad \mathbf{v} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

$$\sim \begin{bmatrix} 3 & -2 \\ 0 & 0 \end{bmatrix} \qquad \sim \begin{bmatrix} 1/2 & 1/2 \\ 0 & 0 \end{bmatrix}$$

1. Find $\mathbf{Pv}$, $\tilde{\mathbf{P}}\mathbf{v}$, $(\mathbf{I} - \mathbf{P})\mathbf{v}$, and $(\mathbf{I} - \tilde{\mathbf{P}})\mathbf{v}$.
2. Give the spaces onto which $\mathbf{P}$, $\tilde{\mathbf{P}}$, $(\mathbf{I} - \mathbf{P})$, and $(\mathbf{I} - \tilde{\mathbf{P}})$ project.
3. Draw pictures.

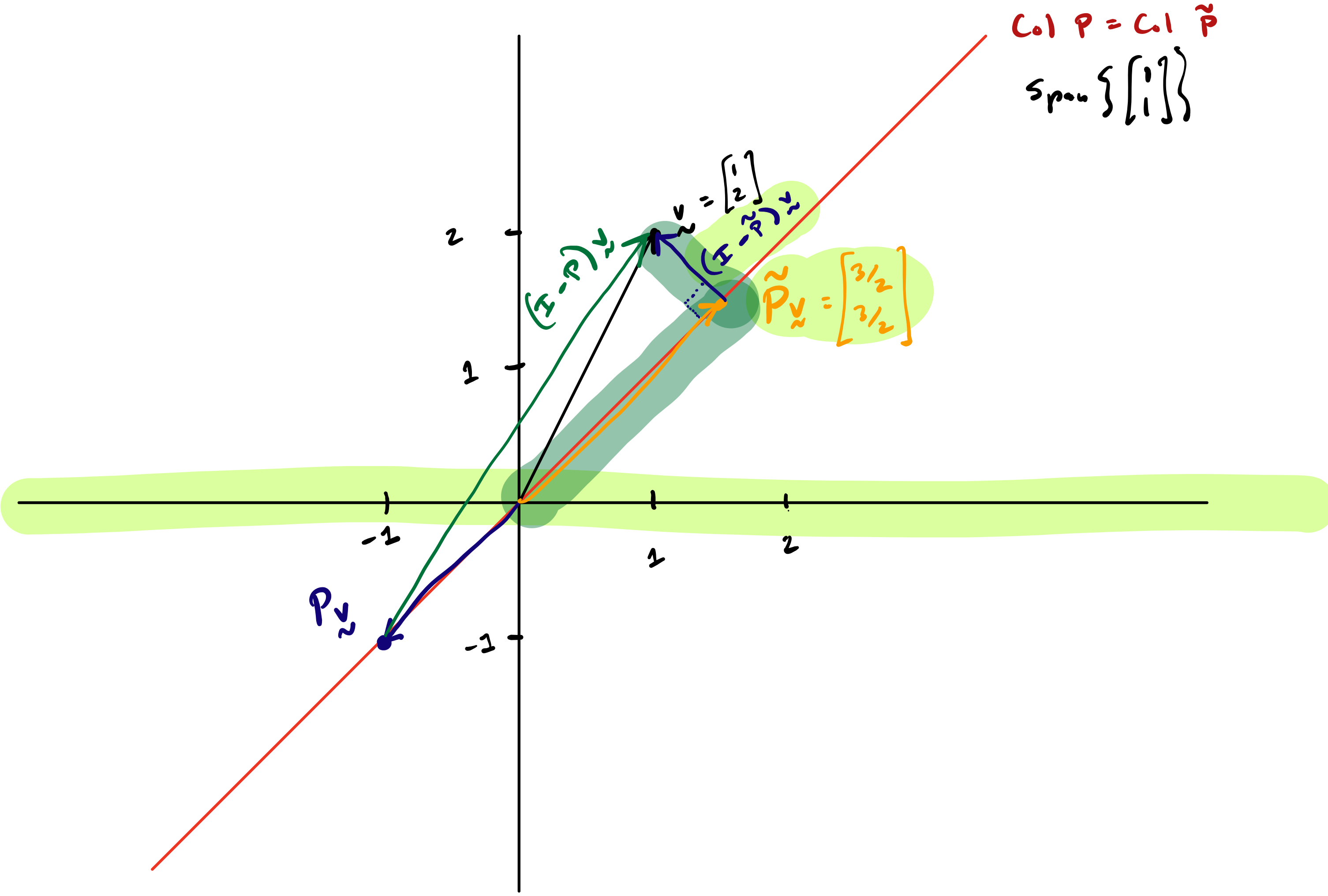
$$P\underset{\sim}{v} = \begin{bmatrix} 3 & -2 \\ 3 & -2 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} -1 \\ -1 \end{bmatrix}$$

$$(I - P)\underset{\sim}{v} = \underset{\sim}{v} - P\underset{\sim}{v} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} - \begin{bmatrix} -1 \\ -1 \end{bmatrix} = \boxed{\begin{bmatrix} 2 \\ 3 \end{bmatrix}}$$

$$\tilde{P}\underset{\sim}{v} = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 3/2 \\ 3/2 \end{bmatrix}$$

$$(I - \tilde{P})\underset{\sim}{v} = \underset{\sim}{v} - \tilde{P}\underset{\sim}{v} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} - \begin{bmatrix} 3/2 \\ 3/2 \end{bmatrix} = \begin{bmatrix} -1/2 \\ 1/2 \end{bmatrix}$$

$$\underset{\sim}{v} = P\underset{\sim}{v} + (I - P)\underset{\sim}{v} = \tilde{P}\underset{\sim}{v} + (I - \tilde{P})\underset{\sim}{v}$$

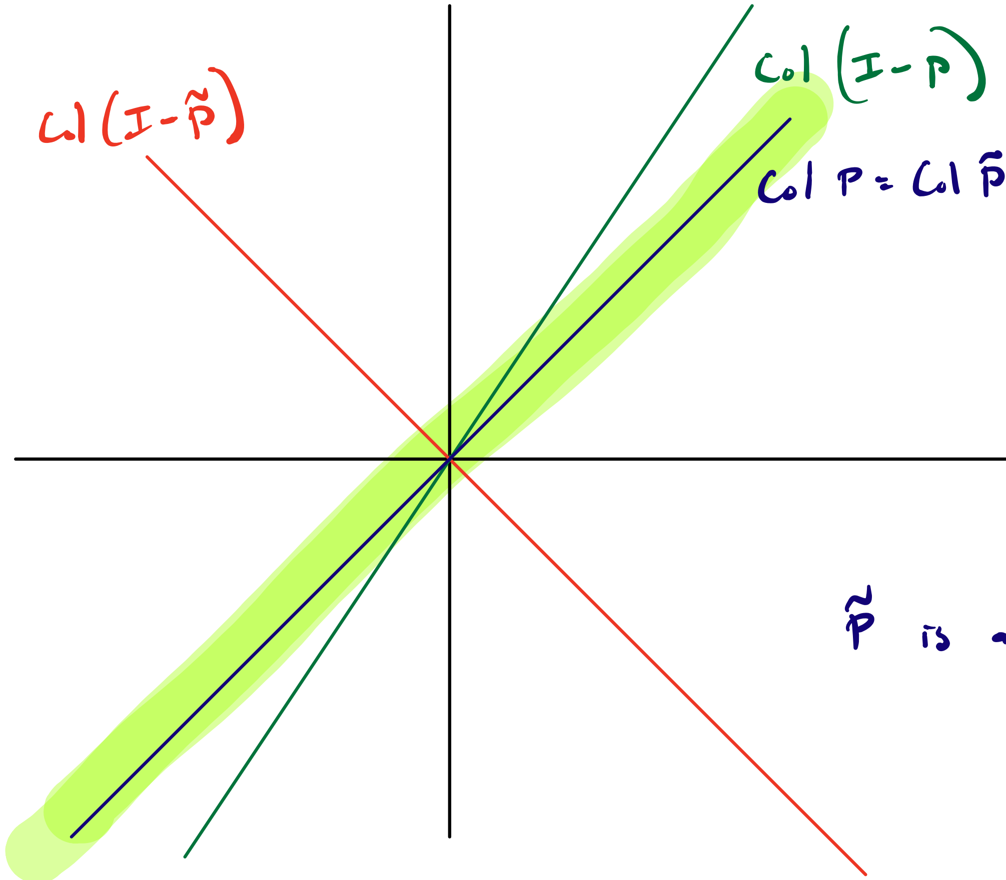


$$I - P = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 3 & -2 \\ 3 & -2 \end{bmatrix} = \begin{bmatrix} -2 & 2 \\ -3 & 3 \end{bmatrix} \sim \begin{bmatrix} -2 & 2 \\ 0 & 0 \end{bmatrix}$$

$$\text{Col}\,(I-P) = \text{Span}\left\{ \begin{bmatrix} -2 \\ -3 \end{bmatrix} \right\} = \text{Span}\left\{ \begin{bmatrix} 2 \\ 3 \end{bmatrix} \right\}$$

$$I - \tilde{P} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix} = \begin{bmatrix} 1/2 & -1/2 \\ -1/2 & 1/2 \end{bmatrix} \sim \begin{bmatrix} 1/2 & -1/2 \\ 0 & 0 \end{bmatrix}$$

$$\text{Col}\,(I-\tilde{P}) = \text{Span}\left\{ \begin{bmatrix} 1/2 \\ -1/2 \end{bmatrix} \right\} = \text{Span}\left\{ \begin{bmatrix} 1 \\ -1 \end{bmatrix} \right\}$$



$\tilde{P}$ is an orthogonal proj.

We like projections that let us orthogonally decompose any vector $\mathbf{x}$ as

$$\mathbf{x} = \underbrace{\mathbf{P}\mathbf{x}}_{\in \text{Col}\,\mathbf{P}} + \underbrace{(\mathbf{I} - \mathbf{P})\mathbf{x}}, \quad \text{where} \quad \mathbf{P}\mathbf{x} \cdot (\mathbf{I} - \mathbf{P})\mathbf{x} = 0.$$

Orthogonal projection

Let $\mathbf{P}$ be a projection matrix onto a subspace V. The projection is an *orthogonal projection* if $\boxed{(\mathbf{I} - \mathbf{P})}$ is the projection matrix onto $\boxed{V^{\perp}.}$

Discuss: Which projection matrix corresponds to an orthogonal projection?

$$\mathbf{P} = \begin{bmatrix} 3 & -2 \\ 3 & -2 \end{bmatrix}, \quad \tilde{\mathbf{P}} = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix}.$$

Not

orthogonal

Theorem (Symmetric, idempotent $\iff$ orthogonal projection)

A matrix $\boxed{\mathbf{P} \text{ is symmetric}}$ *and idempotent iff it is an* $\boxed{\text{orthogonal projection matrix.}}$

Prove the result.

"$\Rightarrow$" Assume P symm. idempotent. So P is a proj. onto $\text{Col } P$.

Want to show that $I - P$ is a projection matrix onto $(\text{Col } P)^\perp$.

Recall, $(\text{Col } P)^\perp = \text{Nul } P^T = \text{Nul } P$.

(i) $(I-P)(I-P) = I - P - P + PP \overset{\text{idemp}}{=} I - P$

(ii) Take any $\underset{\sim}{x}$.

Then $P(I-P)\underset{\sim}{x} = P\underset{\sim}{x} - \overbrace{PP}^{P}\underset{\sim}{x} = \underset{\sim}{0}$,

so $(I-P)\underset{\sim}{x} \in \text{Nul } P$.

(iii) Take $\underset{\sim}{z} \in \text{Nul } P$. Then

$$(I-P)\underset{\sim}{z} = \underset{\sim}{z} - \underbrace{P\underset{\sim}{z}}_{0} = \underset{\sim}{z}.$$

So $I-P$ is a proj. onto $(\text{Col } P)^\perp$.

"$\Leftarrow$" Suppose P is an orth. projection matrix.

Then $(I-P)$ is a projection onto $(\text{Col } P)^\perp$

We have $(\text{Col } P)^\perp = \text{Nul } P^T$.

$\boxed{\begin{array}{l} A\underset{\sim}{x} = \underset{\sim}{0} \quad \forall \underset{\sim}{x} \\ \Rightarrow A = 0. \end{array}}$

Then for any $\underset{\sim}{x}$

$$P^T(I-P)\underset{\sim}{x} = \underset{\sim}{0}.$$

Since above holds for all $\underset{\sim}{x}$, we have

$$P^T(I-P) = 0$$

$$\Leftrightarrow \quad P^T - P^T P = 0$$

$$\Leftrightarrow \quad P^T = P^T P,$$

where P^TP is symmetric, so P^T must also be symm.

$\Rightarrow$ P is symmetric.

So we write

$$P = PP$$

so that P is also idempotent.

$$A^T A = 0 \iff A = 0$$

Theorem (Uniqueness of orthogonal projection matrices)

If $\mathbf{P}_1$ *and* $\mathbf{P}_2$ *are orthogonal projections onto the same subspace then* $\mathbf{P}_1 = \mathbf{P}_2$.

Prove the result.

Suppose P_1, P_2, orthog. project onto $V \subset \mathbb{R}^n$.

Suff to show $(P_1 - P_2)^T (P_1 - P_2) = 0$.

We have

$$(P_1 - P_2)^T (P_1 - P_2) = P_1^T P_1 - P_1^T P_2 - P_2^T P_1 + P_2^T P_2$$

$$\overset{\text{symm}}{=} P_1 P_1 - P_1 P_2 - P_2 P_1 + P_2 P_2$$

$$\overset{\text{idemp}}{=} P_1 - \underbrace{P_1 P_2}_{= P_2} - \underbrace{P_2 P_1}_{= P_1} + P_2$$

We have

$$P_1 \underset{\sim}{x} \in V \qquad \forall \underset{\sim}{x} \in \mathbb{R}^n$$

$$\boxed{\begin{array}{l} A\underset{\sim}{x} = \underset{\sim}{0} \\ \quad \forall \underset{\sim}{x} \in \mathbb{R}^n \\ \Rightarrow A = 0 \end{array}}$$

$$\Rightarrow \quad P_2 \underbrace{P_1 \underset{\sim}{x}}_{\in V} = P_1 \underset{\sim}{x} \qquad \forall \underset{\sim}{x} \in \mathbb{R}^n$$

$$\Rightarrow \quad P_2 P_1 = P_1$$

Likewise $\quad P_2 \underset{\sim}{x} \in V \qquad \forall \underset{\sim}{x} \in \mathbb{R}^n$

$$\Rightarrow \quad P_1 \underbrace{P_2 \underset{\sim}{x}}_{\in V} = P_2 \underset{\sim}{x} \qquad \forall \underset{\sim}{x} \in \mathbb{R}^n$$

$$\Rightarrow \quad P_1 P_2 = P_2$$

$$\Rightarrow \quad (P_1 - P_2)^T (P_1 - P_2) = P_1 - P_2 - P_1 + P_2 = 0.$$

1. Projection and idempotent matrices

2. Generalized inverses ← Every matrix has one

For a matrix A a g-inv is any matrix G such that

$$AGA = A.$$

3. Least-squares geometry

Generalized inverse of a matrix

A matrix $\mathbf{G}$ which satisfies $\mathbf{AGA} = \mathbf{A}$ is called a *generalized inverse* of $\mathbf{A}$.

If A is invertible then $A\underset{\sim}{x} = \underset{\sim}{b}$ has unique solution $\underset{\sim}{x} = A^{-1}\underset{\sim}{b}$.

Theorem (Generalized inverses for solving systems of equations)

Suppose $\mathbf{Ax} = \mathbf{b}$ *is consistent and let* $\mathbf{G}$ *be a generalized inverse of* $\mathbf{A}$. *Then*

1. $\mathbf{Gb}$ *is a solution to* $\mathbf{Ax} = \mathbf{b}$.
2. $\hat{\mathbf{x}}$ *is a solution to* $\mathbf{Ax} = \mathbf{b}$ *iff there exists* $\mathbf{z}$ *such that* $\boxed{\hat{\mathbf{x}} = \mathbf{Gb} + (\mathbf{I} - \mathbf{GA})\mathbf{z}}$.

See Res A.12 and A.13 of Monahan (2008).

Prove the results.

(1) Assume $A\underset{\sim}{x} = \underset{\sim}{b}$ is consistent.

$\Rightarrow$ $\underset{\sim}{b} \in \text{Col } A$.

$\Rightarrow \quad \underset{\sim}{b} = A\underset{\sim}{c}$ for some $\underset{\sim}{c}$.

$\boxed{AGA = A}$

$$\Rightarrow \quad A(G\underset{\sim}{b}) = AGA\underset{\sim}{c}$$
$$= A\underset{\sim}{c}$$
$$= \underset{\sim}{b}$$

$\Rightarrow \quad \underset{\sim}{x} = G\underset{\sim}{b}$ is a solution to $A\underset{\sim}{x} = \underset{\sim}{b}$.

② "$\Rightarrow$" Let $\hat{\underset{\sim}{x}}$ be a solution to $A\underset{\sim}{x} = \underset{\sim}{b}$.

Then $A\hat{\underset{\sim}{x}} = \underset{\sim}{b}$

$$\Rightarrow \quad A(G\underset{\sim}{b} + \hat{x} - G\underset{\sim}{b}) = \underset{\sim}{b}$$

$A\hat{x} = \underset{\sim}{b}$

$$\Rightarrow \quad A(G\underset{\sim}{b} + \hat{x} - GA\hat{x}) = \underset{\sim}{b}$$

$$\Rightarrow \quad A(\underbrace{G\underset{\sim}{b} + (I - GA)\hat{x}}) = \underset{\sim}{b}.$$

a solution of the form

$G\underset{\sim}{b} + (I - GA)\underset{\sim}{z}$, with $\underset{\sim}{z} = \hat{x}$.

"$\Leftarrow$" Take any $\underset{\sim}{z}$. Then

$$A\left(G\underset{\sim}{b} + (I - GA)\underset{\sim}{z}\right)$$

$$= AG\underset{\sim}{b} + A(I - GA)\underset{\sim}{z}$$

$$= AG\underset{\sim}{b} + A\underset{\sim}{z} - \underbrace{AGA\underset{\sim}{z}}_{A\underset{\sim}{z}}$$

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

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

so $\hat{\underset{\sim}{x}} = G\underset{\sim}{b} + (I - GA)\underset{\sim}{z}$

is a solution to $A\underline{x} = \underline{b}$.

Theorem (Generalized inverse recipe using block structure)

Let $\mathbf{A}$ *be an* $m \times n$ *matrix with rank* r*. If we can partition* $\mathbf{A}$ *as*

$$\mathbf{A} = \begin{bmatrix} \mathbf{C} & \mathbf{D} \\ \mathbf{E} & \mathbf{F} \end{bmatrix}, \quad \text{with } \mathbf{C} \; r \times r \text{ invertible, then} \quad \mathbf{G} = \begin{bmatrix} \mathbf{C}^{-1} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{bmatrix}$$

is a generalized inverse of $\mathbf{A}$*, where the* $\mathbf{O}$ *matrices have the necessary dimensions.*

Similarly, if $\mathbf{F}$ *is* $r \times r$ *invertible, then* $\mathbf{G} = \begin{bmatrix} \mathbf{O} & \mathbf{O} \\ \mathbf{O} & \mathbf{F}^{-1} \end{bmatrix}$ *is a gen. inverse of* $\mathbf{A}$*.*

See Res A.10 and Cor A.3 of Monahan (2008).

Make gen. inv. of *any* matrix by permuting rows/columns to get such a partition. See Res A.11 of Monahan (2008).

Prove the first result.

$\underset{n\times n}{A}$ invertible $\Rightarrow$ $\underbrace{AA^{-1}}_{I}A = A$

$A^{\dagger}$ A^{+}

We sometimes denote the generalized inverse of a matrix $\mathbf{A}$ by $\mathbf{A}^-$.

Theorem (Projections constructed with a generalized inverse)

Let $\mathbf{A}^-$ be a generalized inverse of $\mathbf{A}$. Then

$AA^-A = A$

1. $\mathbf{A}\mathbf{A}^-$ *is a projection onto* $\operatorname{Col}\mathbf{A}$.
2. $(\mathbf{I} - \mathbf{A}^-\mathbf{A})$ *is a projection onto* $\operatorname{Nul}\mathbf{A}$.

See Res A.14 and A.15 of Monahan (2008).

Prove the result.

1 AA^- is a proj onto Col A.

(i) $\underbrace{AA^-A}_{A}A^- = AA^-$ Idempotent ✓

(ii) $\quad AA^-\underset{\sim}{x} \in \text{Col}\, A \quad \forall\, \underset{\sim}{x}$

(iii) Let $\underset{\sim}{z} \in \text{Col}\, A$. Then $\underset{\sim}{z} = A\underset{\sim}{c}$ for some $\underset{\sim}{c}$.

$$\Rightarrow \quad AA^-\underset{\sim}{z} = \underbrace{AA^-A}_{A}\underset{\sim}{c} = A\underset{\sim}{c} = \underset{\sim}{z}.$$

② $(I - A^-A)$ is proj onto $\text{Nul}\, A$

(i) $$(I - A^-A)(I - A^-A) = I - A^-A - A^-A + A^-\underbrace{AA^-A}_{A}$$
$$= I - A^-A$$

(ii) Take any $\underset{\sim}{x}$. Then

$$A(I - A^-A)\underset{\sim}{x} = A\underset{\sim}{x} - \underbrace{AA^-A}_{A}\underset{\sim}{x}$$
$$= A\underset{\sim}{x} - A\underset{\sim}{x}$$
$$= \underset{\sim}{0}.$$

$$\Rightarrow \quad (I - A^-A)\underset{\sim}{x} \in \text{Nul}\, A.$$

(iii) Take any $\underset{\sim}{z} \in \text{Nul}\, A$.

$$(I - A^-A)\underset{\sim}{z} = \underset{\sim}{z} - A^-A\underset{\sim}{z} = \underset{\sim}{z}.$$

For the remainder of the lecture, let $\mathbf{X}$ be an $n \times p$ matrix and $\mathbf{y}$ be a vector in $\mathbb{R}^n$.

We have in mind data coming from a model like

$$\mathbf{y} = \mathbf{X}\underset{\sim}{\mathbf{b}}_0 + \underset{\sim}{\mathbf{e}}$$

. . . but we are not thinking yet about the distribution of $\mathbf{e}$.

We consider least-squares "estimation" of $\mathbf{b}$, but no statistics yet—only geometry.

Equation:

$$X \underset{\sim}{b} = \underset{\sim}{y}$$

Least-squares solution

A *least-squares solution* to $\mathbf{X}\mathbf{b} = \mathbf{y}$ is a vector $\hat{\mathbf{b}} \in \mathbb{R}^p$ such that

$$\|\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}\| \leq \|\mathbf{y} - \mathbf{X}\mathbf{b}\| \quad \text{for all } \mathbf{b} \in \mathbb{R}^p.$$

Theorem (Least-squares solution iff solution to normal equations)

1. *The equation* $\mathbf{X}^T\mathbf{X}\mathbf{b} = \mathbf{X}^T\mathbf{y}$ *is consistent.*
2. $\|\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}\| \leq \|\mathbf{y} - \mathbf{X}\mathbf{b}\|$ *for all* $\mathbf{b} \in \mathbb{R}^p$ *if and only if* $\mathbf{X}^T\mathbf{X}\hat{\mathbf{b}} = \mathbf{X}^T\mathbf{y}$.

See Cor 2.1 and Res 2.3 of Monahan (2008).

We call the set of equations $\mathbf{X}^T\mathbf{X}\mathbf{b} = \mathbf{X}^T\mathbf{y}$ the *normal equations*.

Prove the result.

① $X^T X \underset{\sim}{b} = X^T \underset{\sim}{y}$ has a solution.

Recall: $\text{Col } X^T X = \text{Col } X^T$

Since $X^T \underset{\sim}{y} \in \text{Col } X^T$ we also have $X^T y \in \text{Col } X^T X$.

$\Rightarrow$ $X^T X \underset{\sim}{b} = X^T y$ is consistent.

② $\| \underset{\sim}{y} - X \hat{\underset{\sim}{b}} \| \le \| \underset{\sim}{y} - X \underset{\sim}{b} \| \quad \forall \ \underset{\sim}{b} \in \mathbb{R}^p$

$\underset{n \times p}{X}$

$\Leftrightarrow \quad X^T X \hat{\underset{\sim}{b}} = X^T \underset{\sim}{y}$.

"$\Rightarrow$"

$\| \underset{\sim}{y} - \underbrace{X \hat{\underset{\sim}{b}}}_{\in \text{Col } X} \| \le \| \underset{\sim}{y} - \underbrace{X \underset{\sim}{b}}_{\in \text{Col } X} \| \quad \forall \ \underset{\sim}{b} \in \mathbb{R}^p$

$\Rightarrow$ $X \hat{\underset{\sim}{b}}$ is $\text{Proj}_{\text{Col } X} \underset{\sim}{y}$.

$\Rightarrow$ $\underset{\sim}{y}$ is orthogonally and uniquely decomposed as

$$\underset{\sim}{y} = \underbrace{X \hat{\underset{\sim}{b}}}_{\text{Proj}_{\text{Col } X} \underset{\sim}{y}} + \underbrace{(\underset{\sim}{y} - X \hat{\underset{\sim}{b}})}_{\in (\text{Col } X)^\perp}$$

$\Rightarrow$ $y - X \hat{\underset{\sim}{b}}$ is orth to every column of X.

$$\Rightarrow \quad X^T (\underset{\sim}{y} - X\hat{\underset{\sim}{b}}) = \underset{\sim}{0}$$

$$\Rightarrow \quad X^T\underset{\sim}{y} - X^TX\hat{\underset{\sim}{b}} = \underset{\sim}{0}$$

$$\Rightarrow \quad X^TX\hat{\underset{\sim}{b}} = X^T\underset{\sim}{y}.$$

" $\Leftarrow$ " Suppose $X^TX\hat{\underset{\sim}{b}} = X^T\underset{\sim}{y}$.

$$\Rightarrow \quad X^TX\hat{\underset{\sim}{b}} - X^T\underset{\sim}{y} = \underset{\sim}{0}$$

$$\Rightarrow \quad X^T(X\hat{b} - y) = \underset{\sim}{0}$$

$$\Rightarrow \quad X^T(\underset{\sim}{y} - X\hat{\underset{\sim}{b}}) = \underset{\sim}{0}$$

$$\Rightarrow \quad \underset{\sim}{y} - X\hat{\underset{\sim}{b}} \in (\text{Col } X)^\perp$$

$$\Rightarrow \quad \underset{\sim}{y} = \underbrace{X\hat{\underset{\sim}{b}}}_{\in \text{Col } X} + \underbrace{(y - X\hat{\underset{\sim}{b}})}_{\in (\text{Col } X)^\perp}$$

$$\Rightarrow \quad X\hat{\underset{\sim}{b}} = \text{Proj}_{\text{Col } X}\, \underset{\sim}{y}$$

$$\Rightarrow \quad \|y - X\hat{\underset{\sim}{b}}\| \le \|y - X\underset{\sim}{b}\| \quad \forall \ \underset{\sim}{b} \in \mathbb{R}^p$$

Can also use calculus to obtain the normal equations...

For a real-valued function $Q(\mathbf{x})$ taking vectors in $\mathbb{R}^n$, define

$$\frac{\partial}{\partial \mathbf{x}} Q(\mathbf{x}) = \begin{bmatrix} \frac{\partial}{\partial x_1} Q(\mathbf{x}) \\ \vdots \\ \frac{\partial}{\partial x_n} Q(\mathbf{x}) \end{bmatrix}.$$

Theorem (Derivative of linear and quadratic forms)

For a vector $\mathbf{a}$ *and a matrix* $\mathbf{A}$, *we have*

$$\frac{\partial}{\partial \mathbf{x}} \mathbf{a}^T \mathbf{x} = \mathbf{a} \quad \text{and} \quad \frac{\partial}{\partial \mathbf{x}} \mathbf{x}^T \mathbf{A} \mathbf{x} = (\mathbf{A} + \mathbf{A}^T)\mathbf{x}.$$

A least-squares solution of $\mathbf{X}\mathbf{b} = \mathbf{y}$ is a minimizer of $Q(\mathbf{b}) = \|\mathbf{y} - \mathbf{X}\mathbf{b}\|^2$.

Exercise: Use fact that $\hat{\mathbf{b}}$ minimizes $Q(\mathbf{b})$ iff $\frac{\partial}{\partial \mathbf{b}} Q(\mathbf{b})\Big|_{\mathbf{b}=\hat{\mathbf{b}}} = \mathbf{0}$ to get normal eqs.

$$Q(\underset{\sim}{b}) = \|\underset{\sim}{y} - X\underset{\sim}{b}\|^2 = \underset{\sim}{y}^T\underset{\sim}{y} - 2\underset{\sim}{y}^T(X\underset{\sim}{b}) + \underset{\sim}{b}^T X^T X\underset{\sim}{b}$$

$$\frac{\partial}{\partial \underset{\sim}{b}} Q(\underset{\sim}{b}) = -2X^T\underset{\sim}{y} + 2X^TX\underset{\sim}{b} = 0$$

$$\iff X^TX\underset{\sim}{b} = X^T\underset{\sim}{y}$$

$A\underset{\sim}{x} = \underset{\sim}{b}$

$\hat{x} = Gb + (I - GA)\underset{\sim}{z}$

Theorem (Characterization of solutions to the normal equations)

The vector $\hat{\mathbf{b}}$ *is a solution to* $\boxed{\mathbf{X}^T\mathbf{X}\mathbf{b} = \mathbf{X}^T\mathbf{y}}$ *iff there exists a vector* $\mathbf{z}$ *such that*

$$\hat{\mathbf{b}} = (\mathbf{X}^T\mathbf{X})^{-}\mathbf{X}^T\mathbf{y} + (\mathbf{I} - (\mathbf{X}^T\mathbf{X})^{-}\mathbf{X}^T\mathbf{X})\mathbf{z}.$$

If $\mathbf{X}$ *has full-column rank, then* $\hat{\mathbf{b}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$ *is the unique solution.*

$\text{Col } X^T X = \text{Col } X^T$ $\quad$ $\dim \text{Col } X^T = \dim \text{Col } X$

$\Rightarrow$ X full rank $\Rightarrow$ $X^T X$ has full rank.

Prove the result.

These are helper results for constructing the orthogonal projection onto Col $\mathbf{X}$.

Theorem ("Cool result" and generalized inverse of $\mathbf{X}$)

1. $\boxed{\mathbf{X}^T\mathbf{X}\mathbf{A} = \mathbf{X}^T\mathbf{X}\mathbf{B} \iff \mathbf{X}\mathbf{A} = \mathbf{X}\mathbf{B}.}$
2. $(\mathbf{X}^T\mathbf{X})^-\mathbf{X}^T$ *is a generalized inverse of* $\mathbf{X}$.

See Res 2.4 and 2.5 of Monahan (2008).

Prove the results.

2

$$X^TX(X^TX)^-X^TX = X^TX$$

$$\iff X\underbrace{\boxed{(X^TX)^-X^T}}_{\text{g-inv. of } X}X = X \quad \text{(cool result)}$$

$$\Longleftrightarrow \quad X^T X \left[(X^T X)^-\right]^T X^T = X^T$$

(Assume $(X^T X)^-$ is symm)

$$\Longleftrightarrow \quad X^T \boxed{X (X^T X)^-} X^T = X^T$$

g-inv of X^T

Exercise: Characterize the set of solutions to the normal equations when

$$\mathbf{y} = \begin{bmatrix} 1 \\ -1 \\ 2 \\ 0 \end{bmatrix} \quad \text{and} \quad \mathbf{X} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}.$$

Theorem (Orthogonal projection onto Col $\mathbf{X}$)

The matrix $\mathbf{P_X} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^-\mathbf{X}^T$ *and the matrix* $\mathbf{I} - \mathbf{P_X}$ *are*

1. *projections onto* Col $\mathbf{X}$ *and* Nul $\mathbf{X}^T$*, respectively*
2. *invariant to the choice of generalized inverse*
3. *symmetric (therefore unique)*

See Thm 2.4 and Res 2.6 of Monahan (2008).

Prove the results.

1 (i) $X\boxed{(X^TX)^-X^T}X(X^TX)^-X^T = X(X^TX)^-X^T$

(ii) Take any $\underset{\sim}{v}$. Then $P_X\underset{\sim}{v} = X\boxed{(X^TX)^-X^T\underset{\sim}{v}} \in \text{Col}\, X$.

(iii) Take any $\underset{\sim}{z} \in \text{Col}\, X$. Then $\underset{\sim}{z} = X\underset{\sim}{c}$, for some $\underset{\sim}{c}$.

Then $\quad P_X \underset{\sim}{z} = X(X^TX)^- X^T \underset{\sim}{z}$

$$= X\boxed{(X^TX)^- X^T} X \underset{\sim}{c}$$

$$= X \underset{\sim}{c}$$

$$= \underset{\sim}{z}.$$

(i) $(I - P_X)(I - P_X) = I - P_X - P_X + P_X P_X$

$$= I - P_X$$

(ii) Take any $\underset{\sim}{v}$. Show that $(I - P_X)\underset{\sim}{v} \in \text{Nul } X^T$.

Assumed $(X^TX)^-$ was symmetric.

Write $\quad X^T(I - P_X)\underset{\sim}{v} = X^T\underset{\sim}{v} - X^T \underbrace{\boxed{X(X^TX)^-}}_{g \text{ inv. of } X^T} X^T \underset{\sim}{v}$

$$= X^T\underset{\sim}{v} - X^T\underset{\sim}{v}$$

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

(iii) Take $\underset{\sim}{z} \in \text{Nul } X^T$.

Then $\quad (I - P_X)\underset{\sim}{z} = \underset{\sim}{z} - X(X^TX)^- \underbrace{X^T\underset{\sim}{z}}_{\underset{\sim}{0}} = \underset{\sim}{z}.$

Result

We have $\mathbf{X}^T\mathbf{X}\mathbf{b} = \mathbf{X}^T\mathbf{y}$ *if and only if* $\mathbf{X}\mathbf{b} = \mathbf{P}_X\mathbf{y}$.

See Res 2.7 of Monahan (2008).

Prove the result.

"$\Rightarrow$" $\quad X^TX\underset{\sim}{b} = X^T\underset{\sim}{y} \quad \Rightarrow \quad X\boxed{(X^TX)^- X^T}X\underset{\sim}{b} = \underbrace{X(X^TX)^-X^T}_{P_X}\underset{\sim}{y}$

$\Rightarrow \quad X\underset{\sim}{b} = P_X\underset{\sim}{y}$

"$\Leftarrow$" $\quad X\underset{\sim}{b} = P_X\underset{\sim}{y} \quad \Rightarrow \quad X\underset{\sim}{b} = X(X^TX)^-X^T\underset{\sim}{y}$

g. inv of X^T

$\Rightarrow \quad X^TX\underset{\sim}{b} = X^T\boxed{X(X^TX)^-}X^T\underset{\sim}{y} = X^T\underset{\sim}{y}$

Is a least squares sol. to $Xb = y$.

Sums of squares

For $\hat{\mathbf{b}}$ satisfying $\mathbf{X}^T\mathbf{X}\hat{\mathbf{b}} = \mathbf{X}^T\mathbf{y}$ we define the

1. *fitted values* as $\hat{\mathbf{y}} = \mathbf{X}\hat{\mathbf{b}}$ $= P_X \underset{\sim}{y}$
2. *residuals* as $\hat{\mathbf{e}} = \mathbf{y} - \hat{\mathbf{y}}$
3. *total sum of squares (SST)* as $\|\mathbf{y}\|^2$
4. *regression sum of squares (SSR)* as $\|\hat{\mathbf{y}}\|^2$
5. *error sum of squares (SSE)* as $\|\hat{\mathbf{e}}\|^2$.

Theorem (Sum of squares decomposition)

We have SST = SSR + SSE, *or* $\|\mathbf{y}\|^2 = \|\hat{\mathbf{y}}\|^2 + \|\hat{\mathbf{e}}\|^2$.

Prove the result.

$$\|y\|^2 = \| P_X \underset{\sim}{y} + (I - P_X)\underset{\sim}{y} \|^2$$

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

$$= \hat{\underset{\sim}{y}}^T \underset{\sim}{y} + 2\underbrace{\hat{\underset{\sim}{y}}^T \hat{\underset{\sim}{e}}}_{=0} + \hat{\underset{\sim}{e}}^T \hat{\underset{\sim}{e}}$$

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

Monahan, J. F. (2008). *A primer on linear models*. CRC Press.