Section: Module 2

Review of Linear Algebra / OLS

Kentaro Nakamura

GOV 2003

February 6th, 2026

Today's agenda

- Review of linear algebra (continued)
    - Linear Independence / Basis
    - Inverse / Rank / Determinant / Trace
    - Projection Matrix / Annihilator Matrix
- I uploaded linear algebra notebook on Canvas
- Logistics:
    - Problem Set 0 is due on next Monday 10am
    - Problem Set 1 is due on next Friday 1:30pm (before section)

Linear Independence

Definition (Linear Independence of Vectors)

A set of vectors $\{\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_k\}$ is said to be linearly independent if it satisfies the following condition: if $c_1, c_2, \cdots, c_k \in \mathbb{R}$ satisfies

$$c_1\boldsymbol{a}_1 + \cdots + c_k\boldsymbol{a}_k = 0,$$

then $c_1 = c_2 = \cdots = c_k = 0$.

Example (Linear Independent Example)

Consider $\boldsymbol{a}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \boldsymbol{a}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$. Then, $c_1\boldsymbol{a}_1 + c_2\boldsymbol{a}_2 = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}$ and thus if $c_1\boldsymbol{a}_1 + c_2\boldsymbol{a}_2 = \mathbf{0}$, then $c_1 = c_2 = 0$.

Basis and Dimensions

Definition (Basis of Vector Space)

Let $V \subset \mathbb{R}^n$ be a vector space. When any $\boldsymbol{v} \in V$ is represented as

$$\boldsymbol{v} = c_1 \boldsymbol{v}_1 + \cdots + c_k \boldsymbol{v}_k, \quad c_1, \cdots, c_k \in \mathbb{R}$$

with a linearly independent set $\{\boldsymbol{v}_1, \cdots, \boldsymbol{v}_k\}$, such set is called **basis** of V.

Definition (Dimensions)

The dimension of a vector space V, denoted by $\dim(V)$, is the number of vectors in a basis of V.

Rank

Definition (Rank)

Let $\boldsymbol{A}$ be an $m \times n$ matrix. Then, rank is defined as the dimension of column space (or equivalently the dimension of row space). I.e.,

$$\mathrm{rank}(\boldsymbol{A}) := \dim(\mathcal{S}(\boldsymbol{A}))(= \dim(\mathcal{R}(\boldsymbol{A})))$$

- I.e., rank is maximal number of linearly independent columns of A
- The definition above tells us that $\mathrm{rank}(\boldsymbol{A}) = \mathrm{rank}(\boldsymbol{A}^\top)$

Example (Full rank matrix / Rank-deficient matrix)

Consider $\boldsymbol{A} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \boldsymbol{B} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix}, \boldsymbol{C} = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$. Then,

$$\mathrm{rank}(\boldsymbol{A}) = 2 \Rightarrow \text{full-rank}$$
$$\mathrm{rank}(\boldsymbol{B}) = 2 \Rightarrow \text{full-rank}$$
$$\mathrm{rank}(\boldsymbol{C}) = 1 \Rightarrow \text{rank-deficient}$$

Inverse

Definition (Inverse Matrix)

Let $\boldsymbol{A}$ be an $m \times m$ matrix. If there exists an $m \times m$ matrix $\boldsymbol{B}$ such that $\boldsymbol{AB} = \boldsymbol{BA} = I$, then $\boldsymbol{B}$ is called the **inverse** of $\boldsymbol{A}$, which is denoted by $\boldsymbol{A}^{-1}$

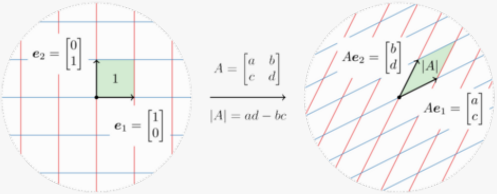
- If $\boldsymbol{A}$ has its inverse, then $\boldsymbol{A}$ is said to be **invertible** (or **non-singular**).
- A matrix $\boldsymbol{A} \in \mathbb{R}^{m \times m}$ is invertible if one of the following is true:
  - A is full rank (i.e., $\mathrm{Rank}(\boldsymbol{A}) = m$)
  - The determinant of $\boldsymbol{A}$ is non-zero
- Properties of Inverse
  - $(\boldsymbol{A}^{-1})^{-1} = \boldsymbol{A}$
  - $(\boldsymbol{AB})^{-1} = \boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$
  - $(\boldsymbol{A}^{\top})^{-1} = (\boldsymbol{A}^{-1})^{\top}$

Determinant (1)

- Determinant is one way to check invertibility
- In the case of 2×2 square matrix, determinant is calculated as follows:

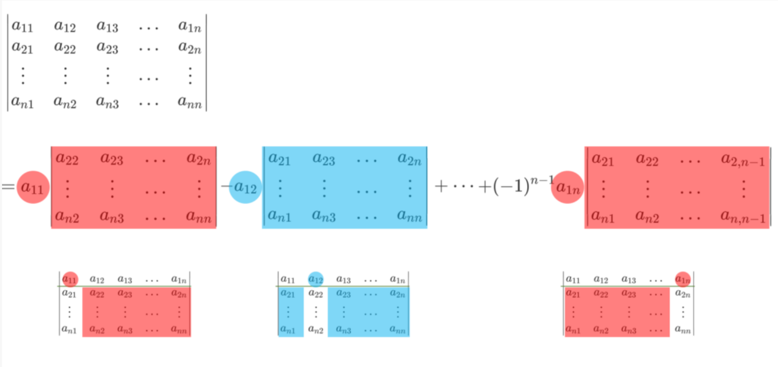
$$\det(\boldsymbol{A}) = |\boldsymbol{A}| = \left| \begin{bmatrix} a & b \\ c & d \end{bmatrix} \right| = ad - cb$$

- Intuitively, determinant represents the scaling factor for volume (or area in 2D) under a linear transformation
  - Thus, when determinant is 0, the volume (or area) under linear transformation is zero, which means that vectors are linearly dependent.



Determinant (2)

- **Cofactor Expansion**: general way to calculate the determinants



$$\begin{vmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn} \end{vmatrix} = a_{11} \begin{vmatrix} a_{22} & a_{23} & \cdots & a_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{n2} & a_{n3} & \cdots & a_{nn} \end{vmatrix} - a_{12} \begin{vmatrix} a_{21} & a_{23} & \cdots & a_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{n1} & a_{n3} & \cdots & a_{nn} \end{vmatrix} + \cdots + (-1)^{n-1} a_{1n} \begin{vmatrix} a_{21} & a_{22} & \cdots & a_{2,n-1} \\ \vdots & \vdots & \cdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{n,n-1} \end{vmatrix}$$

- Properties
  - $|\boldsymbol{I}_m| = 1$, where $\boldsymbol{I}_m$ is the $m \times m$ identity matrix
  - $|\boldsymbol{A}^\top| = |\boldsymbol{A}|$
  - If $\boldsymbol{A} = (a_{ij})$ is a triangular matrix, $|\boldsymbol{A}| = a_{11}a_{22}\cdots a_{mm}$
  - $|\boldsymbol{AB}| = |\boldsymbol{A}| \times |\boldsymbol{B}|$
  - The determinant of $\boldsymbol{A}$ is zero if and only if $\boldsymbol{A}$ is not full rank

Trace

Definition (Trace of a square matrix)

The trace of an $m \times m$ matrix $\boldsymbol{A} = (a_{ij})$, denoted by $\mathrm{tr}(\boldsymbol{A})$, is define to be the sum of the diagonal elements, that is

$$\mathrm{tr}(\boldsymbol{A}) = \sum_{i=1}^{m} a_{ii} = a_{11} + \cdots + a_{mm}$$

- Properties of Trace
  - $\mathrm{tr}(\boldsymbol{A}) = \mathrm{tr}(\boldsymbol{A}^\top)$ for $m \times m$ matrix A
  - $\mathrm{tr}(c\boldsymbol{A}) = c \times \mathrm{tr}(\boldsymbol{A})$ for $m \times m$ matrix A and constant $c \in \mathbb{R}$
  - $\mathrm{tr}(\boldsymbol{A} + \boldsymbol{B}) = \mathrm{tr}(\boldsymbol{A}) + \mathrm{tr}(\boldsymbol{B})$ for $m \times m$ matrices A and B
  - $\mathrm{tr}(\boldsymbol{ABC}) = \mathrm{tr}(\boldsymbol{BCA}) = \mathrm{tr}(\boldsymbol{CAB})$
  - If x is scalar, $x = \mathrm{tr}(x)$ (very useful in proof)

Quadratic Form

Definition (Quadratic Form)

Let $A = (a_{ij})$ be an $n \times n$ real symmetric matrix (i.e., $a_{ij} = a_{ji}$) and $\boldsymbol{x} = (x_1, \cdots, x_n)^\top$ be an $n \times 1$ vector. A **quadratic form** $Q_A(\boldsymbol{x})$ is a function from $\mathbb{R}^n$ to $\mathbb{R}$ defined by

$$Q_A(\boldsymbol{x}) = \boldsymbol{x}^\top A\boldsymbol{x} = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_i x_j$$

Example (Quadratic Form)

When $A = \begin{bmatrix} 2 & 3 \\ 3 & -1 \end{bmatrix}$ and $\boldsymbol{x} = \begin{bmatrix} x_1 & x_2 \end{bmatrix}^\top$,

$$Q_A(\boldsymbol{x}) = \boldsymbol{x} = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 2x_1 + 3x_2 \\ 3x_1 - 1x_2 \end{bmatrix} = 2x_1^2 - 6x_1x_2 - x_2^2$$

Sign of Quadratic Form

Definition (Signs of Quadratic Form)

Let $Q_A(\boldsymbol{x}) = \boldsymbol{x}^\top A\boldsymbol{x}$ be a quadratic form based on $n \times n$ real symmetric matrix A. Then,

- If $Q_A(\boldsymbol{x})$ is positive for all $\boldsymbol{x}(\neq \mathbf{0})$, then A is said to be **positive definite**
- If $Q_A(\boldsymbol{x})$ is non-negative for all $\boldsymbol{x}(\neq \mathbf{0})$, then A is said to be **semi-positive definite**
- If $Q_A(\boldsymbol{x})$ is negative for all $\boldsymbol{x}(\neq \mathbf{0})$, then A is said to be **negative definite**
- If $Q_A(\boldsymbol{x})$ is greater than 0 for some $\boldsymbol{x}$ and less than 0 for some $\boldsymbol{x}$, then A is said to be indefinite

- When $\boldsymbol{A} \geq \boldsymbol{B}$ for two matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, this means that $\boldsymbol{A} - \boldsymbol{B}$ is semi-positive definite.

- $\boldsymbol{AA}^\top$ is always semi-positive definite
  - **Proof**: $\boldsymbol{x}^\top(\boldsymbol{AA}^\top)\boldsymbol{x} = ||\boldsymbol{A}^\top\boldsymbol{x}||_2^2 \geq 0$ for any $\boldsymbol{x}$.

Projection Matrix

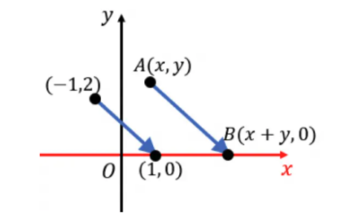
Definition (Projection Matrix)

A matrix $\boldsymbol{P} \in \mathbb{R}^{m \times m}$ is a **projection matrix** if $\boldsymbol{P}^2 = \boldsymbol{P}$.

- The property $\boldsymbol{P}^2 = \boldsymbol{P}$ is called **idempotent**

Example

Consider $\boldsymbol{P} = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$. As $\boldsymbol{P}^2 = \boldsymbol{P}$, $\boldsymbol{P}$ is an projection matrix.



Orthgonal Projection

Definition (Orthogonal Projection)

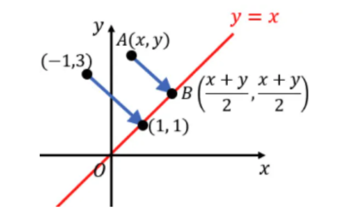
$\boldsymbol{P}$ is an **orthogonal projection matrix** if $\boldsymbol{P}^2 = \boldsymbol{P}$ and $\boldsymbol{P}^\top = \boldsymbol{P}$.

- **Key Property (orthogonality)**: $\boldsymbol{y} - \boldsymbol{P}\boldsymbol{y} \perp\!\!\!\perp \boldsymbol{P}\boldsymbol{z}$ for all $\boldsymbol{z}$

$$(\boldsymbol{y} - \boldsymbol{P}\boldsymbol{y})^\top \boldsymbol{P}\boldsymbol{z} = \boldsymbol{y}^\top \boldsymbol{P}\boldsymbol{z} - \boldsymbol{y}^\top \boldsymbol{P}^\top \boldsymbol{P}\boldsymbol{z} = \boldsymbol{y}^\top \boldsymbol{P}\boldsymbol{z} - \boldsymbol{y}^\top \boldsymbol{P}^2 \boldsymbol{z}$$
$$= \boldsymbol{y}^\top \boldsymbol{P}\boldsymbol{z} - \boldsymbol{y}^\top \boldsymbol{P}\boldsymbol{z} = 0.$$

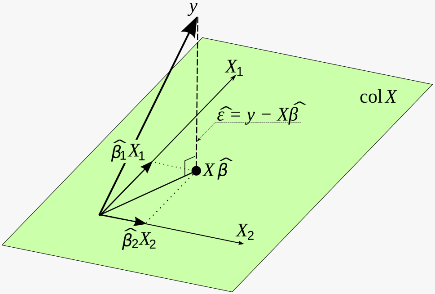
Example

$\boldsymbol{P} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$. As $\boldsymbol{P}^2 = \boldsymbol{P}$ and $\boldsymbol{P}^\top = \boldsymbol{P}$, $\boldsymbol{P}$ is an orthogonal projection.



OLS and Orthogonal Projection

- Consider the linear model $\boldsymbol{Y} = \boldsymbol{X}\beta + \epsilon$. Recall that OLS estimator of β is given by
$$\hat{\beta} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{Y}$$
- As $\boldsymbol{Y} = \boldsymbol{X}\beta + \epsilon$,
$$\hat{\boldsymbol{Y}} = \boldsymbol{X}\hat{\beta} = \underbrace{\boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top}_{:=\boldsymbol{P}_X} \boldsymbol{Y}$$
- Now, notice that $\boldsymbol{P}_X^2 = \boldsymbol{P}_X$ and $\boldsymbol{P}^\top = \boldsymbol{P}$ (Prove this!), and thus $\boldsymbol{P}_X$ is an orthogonal projection matrix



Orthogonal Complement of Vector Space

Definition (Orthogonal Complement)

Let $V \subset \mathbb{R}^n$ be a vector space. The **orthogonal complement** of V, denoted by $V^\perp$, consists of those vectors in $\mathbb{R}^n$ that are orthogonal to every vector in V; that is,

$$V^\perp = \{\boldsymbol{y} \in \mathbb{R}^n : \boldsymbol{x} \cdot \boldsymbol{y} = 0 \text{ for all } \boldsymbol{x} \in V\}$$

Theorem (Orthogonal Decomposition)

Let $V \subset \mathbb{R}^n$ be a vector space. Then, any vector $\boldsymbol{x}$ which is not in V can be uniquely expressed by the sum of two vectors in V and $V^\perp$, respectively: i.e.,

$$\boldsymbol{x} = \boldsymbol{y} + \boldsymbol{z} \quad (\boldsymbol{y} \in V, \boldsymbol{z} \in V^\perp)$$

Annihilator Matrix

Definition (Annihilator Matrix)

$\boldsymbol{M} = \boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top$ is called **annihilator matrix**.

- Notice that

$$\boldsymbol{M}\boldsymbol{X} = (\boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top)\boldsymbol{X} = \boldsymbol{X} - \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{X} = \boldsymbol{X} - \boldsymbol{X} = 0$$

- Annihilator matrix gives you residual since

$$\boldsymbol{M}\boldsymbol{Y} = (\boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top)\boldsymbol{Y} = \boldsymbol{Y} - \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{Y} = \boldsymbol{Y} - \hat{\boldsymbol{Y}}$$

  - Annihilator matrix is a projection onto orthogomal complement of the column space of $\boldsymbol{X}$
- We can also show that annihilator matrix is symmetric ($\boldsymbol{M}^\top = \boldsymbol{M}$) and idempotent $\boldsymbol{M}^2 = \boldsymbol{M}$
  - That is, annihilator matrix is a projection onto orthogonal complement of the column space

Residual and Annihilator Matrix

- Residual is a projection of $\boldsymbol{Y}$ onto $S^{\perp}(\boldsymbol{X})$ (i.e., orthogonal complement of the column space)

- Orthogonality 1:

$$\hat{\boldsymbol{Y}}^{\top}\hat{\boldsymbol{\epsilon}} = (\boldsymbol{P}_X\boldsymbol{Y})^{\top}\boldsymbol{M}\boldsymbol{Y} = \boldsymbol{Y}^{\top}\boldsymbol{P}_X^{\top}(\boldsymbol{I} - \boldsymbol{P}_X)\boldsymbol{Y} = 0$$

- Orthogonality 2:

$$\boldsymbol{X}_k^{\top}\hat{\boldsymbol{\epsilon}} = (\boldsymbol{P}_X\boldsymbol{X}_k)^{\top}\boldsymbol{M}\boldsymbol{Y} = \boldsymbol{X}_k^{\top}\boldsymbol{P}_X^{\top}(\boldsymbol{I} - \boldsymbol{P}_X)\boldsymbol{Y} = 0$$

  - $\boldsymbol{P}_X\boldsymbol{X}_k = \boldsymbol{X}_k$ because $\boldsymbol{P}_X\boldsymbol{X}_k = \boldsymbol{P}_X\boldsymbol{X}\boldsymbol{e}_k = \boldsymbol{X}(\boldsymbol{X}^{\top}\boldsymbol{X})^{-1}\boldsymbol{X}^{\top}\boldsymbol{X}\boldsymbol{e}_k = \boldsymbol{X}\boldsymbol{e}_k = \boldsymbol{X}_k$

- Orthogonality 3: $\boldsymbol{x}^{\top}\hat{\boldsymbol{\epsilon}} = 0$ for any $\boldsymbol{x} \in S(\boldsymbol{X})$
  - Proof: Recall the definition of column space: $\boldsymbol{x} = \boldsymbol{X}a$ for some a.

$$\boldsymbol{x}^{\top}\hat{\boldsymbol{\epsilon}} = (\boldsymbol{X}a)^{\top}\hat{\boldsymbol{\epsilon}} = a^{\top}\boldsymbol{X}^{\top}\boldsymbol{M}\boldsymbol{Y} = a^{\top}\boldsymbol{X}^{\top}(\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}^{\top}\boldsymbol{X})^{-1}\boldsymbol{X}^{\top}) = 0$$

- Zero mean: $\sum_{i=1}^{n}\hat{\epsilon}_i = 0$:
  - As $\boldsymbol{X}$ includes intercept, there exists a such that $\boldsymbol{1} = \boldsymbol{X}a$. As a result, by the previous orthogonality 3, we have zero mean.