Section: Module 1

Review of Linear Algebra

Kentaro Nakamura

GOV 2003

January 30th, 2026

Today's agenda

- Review of linear algebra
  - Recommended resources: Matrix cookbook

- We will cover
  - Inner product, Norm
  - Column Space
  - Linear Independence
  - Rank, Inverse, Trace, Determinant

Norm

Definition (ℓ^p-norm)

The **ℓ^p-norm** of vector $\boldsymbol{x} = (x_1, \cdots, x_n)^\top$ is defined as

$$||\boldsymbol{x}||_p = \Big(\sum_{i=1}^{n} |x_i|^p\Big)^{\frac{1}{p}}$$

- **Quiz**: Calculate ℓ^1-norm and ℓ^2-norm of $\boldsymbol{x} = [1, 3, 5]^\top$.

Inner Product

Definition (Inner Product)

The **inner product** of two vectors, $\boldsymbol{x} = [x_1, \cdots, x_n]^\top$ and $\boldsymbol{y} = [y_1, \cdots, y_n]^\top$, is defined as

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \boldsymbol{x}^\top \boldsymbol{y} = \sum_{i=1}^{n} x_i y_i$$

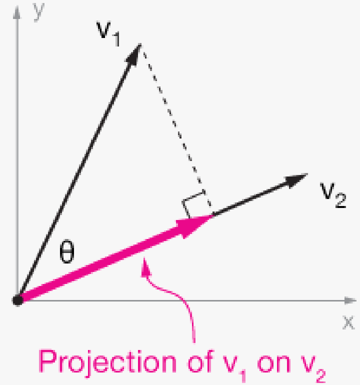
or equivalently,

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = ||\boldsymbol{x}||_2 ||\boldsymbol{y}||_2 \cos\theta$$

where θ is angle between $\boldsymbol{x}$ and $\boldsymbol{y}$.

- When inner product is zero, they are said to be **orthogonal**

Projection of Vectors



- Projection of $\boldsymbol{v}_1$ onto $\boldsymbol{v}_2$ is given by

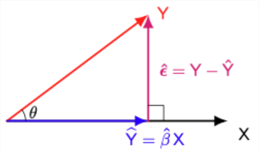
$$\underbrace{||\boldsymbol{v}_1||\cos\theta}_{\text{Length}} \cdot \underbrace{\frac{\boldsymbol{v}_2}{||\boldsymbol{v}_2||}}_{\text{Direction}} = \frac{\boldsymbol{v}_1^\top \boldsymbol{v}_2}{\boldsymbol{v}_2^\top \boldsymbol{v}_2}\boldsymbol{v}_2$$

OLS as a Projection

- Let $\boldsymbol{Y} \in \mathbb{R}^n$ be an outcome vector and $\boldsymbol{X} \in \mathbb{R}^n$ is a vector of the explanatory variable
  - n is the number of observations

- From the previous page's result, the projection is

$$\frac{\boldsymbol{Y}^\top \boldsymbol{X}}{\boldsymbol{X}^\top \boldsymbol{X}} \boldsymbol{X} = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2} \boldsymbol{X} = \boldsymbol{X} \hat{\beta}$$



- **Takeaway**: OLS is regarded as a projection of $\boldsymbol{Y}$ onto $\boldsymbol{X}$
  - Multiple regression is a projection of $\boldsymbol{Y}$ onto **column space** spanned by $\boldsymbol{X}$
  - I will explain what the column space is soon.

- **NOTE**: This derivation assumes that $\boldsymbol{Y}$ and $\boldsymbol{X}$ are centered

Column Space / Row Space

Definition (Column Space / Row Space)

Let $\boldsymbol{A} = (\boldsymbol{a}_1, \cdots, \boldsymbol{a}_n) = (\boldsymbol{b}_1, \cdots, \boldsymbol{b}_m)^\top$ be an $m \times n$ matrix. Then, the **column space** $\mathcal{S}(\boldsymbol{A})$ is defined by

$$\mathcal{S}(\boldsymbol{A}) = \left\{ \sum_{i=1}^{n} c_i \boldsymbol{a}_i \middle| c_1, \cdots, c_n \in \mathbb{R} \right\}$$

The **row space** $\mathcal{R}(\boldsymbol{A})$ is also defined by

$$\mathcal{R}(\boldsymbol{A}) = \left\{ \sum_{j=1}^{m} d_j \boldsymbol{b}_j \middle| d_1, \cdots, d_m \in \mathbb{R} \right\}$$

- By definition, any element in column space is written as $\boldsymbol{Xb}$ where $\boldsymbol{b} \in \mathbb{R}^n$

Multiple Regression as Projection (1)

- OLS minimizes

$$\hat{\beta} = \arg\min_{\beta} ||\boldsymbol{Y} - \boldsymbol{X}\beta||_2^2$$

  - Geometrically, this finds the vector $\boldsymbol{X}\beta$ that is the closest in $\boldsymbol{Y}$
  - As all the elements in column space $\mathcal{S}(\boldsymbol{X})$ is represented as $\boldsymbol{X}\beta$ (check definition), this means finding the closest vector to $\boldsymbol{Y}$ in the column space $\mathcal{S}(\boldsymbol{X})$
  - Thus, we can think OLS as projection onto column space

- Recall that the solution of this is given by

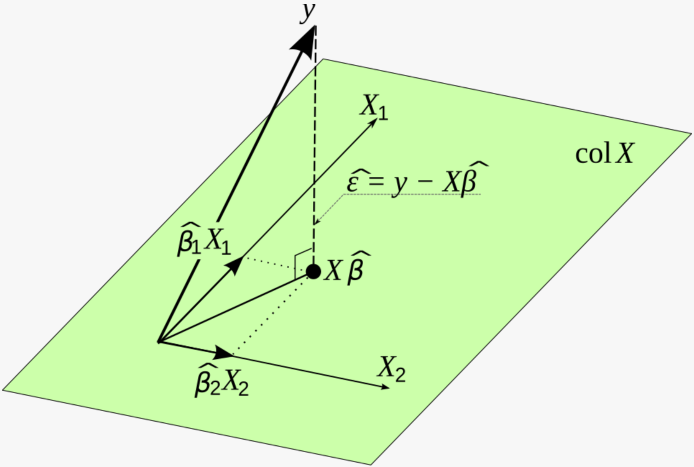
$$\hat{\beta} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{Y}$$

- This means that the predicted outcome is given by

$$\hat{\boldsymbol{Y}} = \boldsymbol{X}\hat{\beta} = \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{Y}$$

- OLS is an **orthogonal projection** onto column space in that $\boldsymbol{Y} - \hat{\boldsymbol{Y}}$ is orthogonal to any element in column space

Multiple Regression as Projection (2)



Multiple Regression as Projection (3)

- Let's prove the orthogonality.
  - Recall that to show orthogonality, we need to calculate the inner product.
  - Also, recall that the element of column space is written as $\boldsymbol{Xb}$ with some $\boldsymbol{b}$
  - Thus,

$$
\begin{aligned}
(\boldsymbol{Xb})^\top(\boldsymbol{Y} - \hat{\boldsymbol{Y}}) &= \boldsymbol{b}^\top \boldsymbol{X}^\top(\boldsymbol{Y} - \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{Y}) \\
&= \boldsymbol{b}^\top \boldsymbol{X}^\top(\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top)\boldsymbol{Y} \\
&= \boldsymbol{b}^\top(\boldsymbol{X}^\top - \boldsymbol{X}^\top)\boldsymbol{Y} = 0
\end{aligned}
$$

  - $\boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top$ is an example of **projection matrix**

- To compute $\hat{\beta}$, we need to calculate the inverse
  - What is inverse? When can we calculate it?
  - To understand it, we need to understand rank / determinant.