Econ 703 Note 7

Chia-Min Wei*

August 24, 2025

1 The Inner Product on $\mathbb{R}^n$

Definition 1.1: The inner product between two vectors $u, v \in \mathbb{R}^n$, denoted by $u \cdot v$, is defined by

$$u \cdot v = \sum_{i=1}^{n} u_i v_i,$$

where $u = (u_1, ..., u_n)$ and $v = (v_1, ..., v_n)$.

There are many others notations for the inner product. The most common ones include $u^\intercal v$ and $\langle u, v\rangle$.

Theorem 1.1: The inner product on $\mathbb{R}^n$ satisfies the following properties:

(i) $\mathbf{0} \cdot u = 0$ for all $u \in \mathbb{R}^n$.

(ii) $u \cdot u \geq 0$ for all $u \in \mathbb{R}^n$. The equality holds if and only if $u = 0$.

(iii) $u \cdot v = v \cdot u$ for all $u, v \in \mathbb{R}^n$.

(iv) $(u + v) \cdot w = u \cdot w + v \cdot w$ for all $u, v, w \in \mathbb{R}^n$.

(v) $(au) \cdot v = a(u \cdot v)$ for all $a \in \mathbb{R}$ and $u, v \in \mathbb{R}^n$.

$\|u\| = \sqrt{u \cdot u}$ is called the **norm** of u. The Euclidean distance between two vectors u and v is then $d(u, v) = \|u - v\|$.

2 Cauchy Schwartz Inequality

Theorem 2.1 (Cauchy Schwartz Inequality): For all $u, v \in \mathbb{R}^n$,

$$(u \cdot v)^2 \leq \|u\|^2 \|v\|^2 .$$

The equality holds when $u = kv$ for some $k \in \mathbb{R}$.

*This TA Note is based on Prof. John Kennan's math camp lecture in 2025 at UW-Madison. All errors are mine.

Proof. Let $u, v \in \mathbb{R}^n$. For all $t \in \mathbb{R}$,

$$\begin{aligned}
(tu+v)\cdot(tu+v) &\geq 0 && \text{(Theorem 1.1 (ii))}\\
\implies tu\cdot(tu+v)+v\cdot(tu+v) &\geq 0 && \text{(Theorem 1.1 (iv))}\\
\implies (u\cdot u)t^2+(u\cdot v)t+(v\cdot u)t+v\cdot v &\geq 0 && \text{(Theorem 1.1 (iv))}\\
\implies \|u\|^2 t^2 + 2(u\cdot v)t + \|v\|^2 &\geq 0. && \text{(Theorem 1.1 (iii))}
\end{aligned}$$

Write $a = \|u\|^2$, $b = 2(u \cdot v)$ and $c = \|v\|^2$. Therefore, the quadratic function

$$f(t) = at^2 + bt + c \geq 0$$

for all $t \in \mathbb{R}$. So the discriminant $D \leq 0$:

$$\begin{aligned}
b^2 - 4ac \geq 0 &\implies 4(u\cdot v)^2 - 4\|u\|^2\|v\|^2 \leq 0\\
&\implies (u\cdot v)^2 \leq \|u\|^2\|v\|^2
\end{aligned}$$

When $u = kv$ for some $k \in \mathbb{R}$,

$$|u\cdot v|^2 = |k|^2|v\cdot v|^2 = |k|^2\|v\|^4 = \|kv\|^2\|v\|^2 = \|u\|^2\|v\|^2 .$$

□

Example 2.1: Given any data $\{(y_i, x_i)\}_{i=1}^n$ where $y_i, x_i \in \mathbb{R}^n$, the sample correlation is defined as

$$\hat{\rho}_n = \frac{\sum_{i=1}^n (y_i-\overline{y})(x_i-\overline{x})}{\sqrt{\sum_{i=1}^n (x_i-\overline{x})^2}\sqrt{\sum_{i=1}^n (y_i-\overline{y})^2}},$$

where $\overline{x} = \frac{1}{n}\sum_{i=1}^n x_i$ and $\overline{y} = \frac{1}{n}\sum_{i=1}^n y_i$. Let $u = (x_1 - \overline{x}, ..., x_n - \overline{x})$ and $v = (y_1 - \overline{y}, ..., y_n - \overline{y})$, By Cauchy-Schwartz Inequality,

$$|\hat{\rho}_n| = \frac{|u\cdot v|}{\|u\|\,\|v\|} \leq 1 \implies \hat{\rho}_n \in [-1, 1].$$

3 Projection

Definition 3.1: A nonempty subset W of $\mathbb{R}^n$ is called a **subspace** of $\mathbb{R}^n$ if the following two conditions hold:

(i) If $w \in W$ and $a \in \mathbb{R}$, then $aw \in W$.

(ii) If $u, v \in W$, then $u + v \in W$.

Theorem 3.1: Let $u \in \mathbb{R}^n$ and W be a subspace of $\mathbb{R}^n$. Then there exists a unique $\hat{u} = \arg\min_{w\in W} \|u - w\|^2$. Moreover, $\hat{u}$ satisfies $(u - \hat{u}) \cdot w = 0$ for all $w \in W$.

We will skip the proof now since it uses concepts in linear algebra.

Example 3.1: Consider a regression model $y_i \in \mathbb{R}$ and $x_i \in \mathbb{R}^k$,

$$y_i = x_i'\beta_0 + \epsilon_i.$$

Given data (y, X), we estimate the true parameter $\beta_0 \in \mathbb{R}^k$ by minimizing $\|y - \boldsymbol{X}\beta\|^2$, where

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \in \mathbb{R}^n, \quad \boldsymbol{X} = \begin{bmatrix} x_1' \\ x_2' \\ \vdots \\ v_n' \end{bmatrix} \in \mathbb{R}^{n\times k}.$$

One can check the set $W = \{w \in \mathbb{R}^n : w = \boldsymbol{X}\beta \text{ for some } \beta \in \mathbb{R}^k\}$ is a subspace of $\mathbb{R}^k$. Therefore, by the above theorem, there exists a unique $\hat{y} = \arg\min_{w\in W} \|y - w\|^2$. If $\boldsymbol{X}$ is of *full rank*, then $\hat{y}$ takes the form

$$\hat{y} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'y.$$

And the estimator of β_0,

$$\hat{\beta} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'y.$$

Observe that for any $w \in W$, say $w = \boldsymbol{X}\beta$,

$$\begin{aligned}
(y - \hat{y}) \cdot w &= (y - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'y) \cdot \boldsymbol{X}\beta \\
&= y \cdot (\boldsymbol{X}\beta) - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'y \cdot \boldsymbol{X}\beta \\
&= y'\boldsymbol{X}\beta - (\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'y)'(\boldsymbol{X}\beta) \\
&= y'\boldsymbol{X}\beta - y'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X}\beta \\
&= y'\boldsymbol{X}\beta - y'\boldsymbol{X}\beta = 0.
\end{aligned}$$

The fourth equation holds since $(\boldsymbol{X}'\boldsymbol{X})^{-1}$ is symmetric.