MA-GY 7043: Linear Algebra II

Inner Product Spaces
Cauchy-Schwarz and Triangle Inequalities
Orthogonal Sets and Bases

Deane Yang

Courant Institute of Mathematical Sciences
New York University

February 13, 2025

Outline I

Dot Product on $\mathbb{R}^n$

- Recall that the **dot product** of

$$v = \begin{bmatrix} v^1 \\ \vdots \\ v^n \end{bmatrix}, \quad w = \begin{bmatrix} w^1 \\ \vdots \\ w^n \end{bmatrix} \in \mathbb{R}^n$$

  is defined to be

$$v \cdot w = v^1 w^1 + \cdots + v^n w^n = v^T w = w^T v$$

- The **norm** or **magnitude** of $v \in \mathbb{R}^n$ is defined to be

$$|v| = \|v\| = \sqrt{v \cdot v}$$

- If v and w are nonzero and the angle at 0 from v to w is θ, then

$$\cos\theta = \frac{v \cdot w}{|v||w|}$$

Properties of Dot Product

- The dot product is **bilinear** because for any $a, b \in \mathbb{R}$ and $u, v, w \in \mathbb{R}^n$,

$$(au + bv) \cdot w = a(u \cdot w) + b(v \cdot w)$$
$$u \cdot (av + bw) = a(u \cdot v) + b(u \cdot w)$$

- It is **symmetric**, because for any $v, w \in \mathbb{R}^n$,

$$v \cdot w = w \cdot v$$

- It is **positive definite**, because for any $v \in \mathbb{R}^n$,

$$v \cdot v \geq 0$$

and

$$v \cdot v > 0 \iff v \neq 0$$

Inner Product on Real Vector Space

- Let V be an n-dimensional real vector space
- Consider a function
$$\alpha : V \times V \to \mathbb{R}$$
- It is **bilinear** if for any $a, b \in \mathbb{R}$ and $u, v, w \in \mathbb{R}^n$,
$$\alpha(au + bv, w) = a\alpha(u, w) + b\alpha(v, w)$$
$$\alpha(u, av + bw) = a\alpha(u, v) + b\alpha(u, w)$$
- It is **symmetric** if for any $v, w \in \mathbb{R}^n$,
$$\alpha(v, w) = \alpha(w, v)$$
- It is **positive definite** if for any $v \in \mathbb{R}^n$,
$$\alpha(v, v) \geq 0$$
and
$$\alpha(v, v) > 0 \iff v \neq 0$$
- Any positive definite symmetric bilinear function on a **real** vector space V is called an **inner product**

Hermitian Inner Product on $\mathbb{C}^n$

- Recall that if $z = x + iy \in \mathbb{C}$, then
$$\bar{z} = x - iy \text{ and } z\bar{z} = \bar{z}z = x^2 + y^2$$
- If A is a complex matrix, its **Hermitian adjoint** is defined to be
$$A^* = \bar{A}^T$$
- The Hermitian inner product on $\mathbb{C}^n$ of
$$v = \begin{bmatrix} v^1 \\ \vdots \\ v^n \end{bmatrix},\ w = \begin{bmatrix} w^1 \\ \vdots \\ w^n \end{bmatrix} \in \mathbb{C}^n$$
is defined to be
$$(v, w) = v^1\bar{w}^1 + \cdots + v^n\bar{w}^n = v^T\bar{w} = \bar{w}^T v = w^* v \in \mathbb{C},$$
- The **norm** of $v \in \mathbb{C}^n$ is defined to be
$$|v| = \|v\| = \sqrt{(v, v)}$$
- **No** geometric interpretation of the Hermitian inner product

Not a Real Inner Product

- **Not** bilinear, because if $c \in \mathbb{C}$,

$$(v, cw) = \bar{c}(v, w)$$

- **Not** symmetric, because

$$(w, v) = \overline{(v, w)}$$

- It is positive definite, because for any $v \in \mathbb{C}^n$, $(v, v) \in \mathbb{R}$,

$$(v, v) = v^1\bar{v}^1 + \cdots + v^n\bar{v}^n = |v^1|^2 + \cdots + |v^n|^2 \geq 0,$$

  and

$$(v, v) \neq 0 \iff v \neq 0$$

Properties of Hermitian Inner Product on $\mathbb{C}^n$

- It is a linear function of the first argument, because for any $a, b \in \mathbb{C}$, $u, v, w \in \mathbb{C}^n$,

$$(au + bv, w) = a(u, w) + b(v, w)$$

- It is **Hermitian**, which means

$$(v, w) = \overline{(w, v)}$$

- Therefore, for any $a, b \in \mathbb{C}$ and $u, v, w \in \mathbb{C}^n$,

$$(u, av + bw) = \bar{a}(u, v) + \bar{b}(a, w)$$

Inner Product of a Vector Space Over $\mathbb{F}$

- Assume $\mathbb{F}$ is $\mathbb{R}$ or $\mathbb{C}$
- An **inner product** over a vector space V is a function

$$(\cdot,\cdot) : V \times V \to \mathbb{F}$$

with the following properties: For any $a, b \in F$ and $u, v, w \in V$,

$$\begin{aligned}(au + bv, w) &= a(u, w) + b(v, w)\\ (w, v) &= \overline{(v, w)}\\ (v, v) &\geq 0\\ (v, v) \neq 0 &\iff v \neq 0\end{aligned}$$

- If $\mathbb{F} = \mathbb{R}$, this is the same definition as before
- If $\mathbb{F} = \mathbb{C}$, this is the definition of a Hermitian inner product

Examples

- For each $v \in \mathbb{F}^n$, denote $v^* = \bar{v}^T$
- The standard inner product on $\mathbb{F}^n$ is

$$(v, w) = w^* v,$$

  which is the dot product on $\mathbb{R}^n$ and the standard Hermitian inner product on $\mathbb{C}^n$
- An inner product on the space of polynomials of degree n or less and with coefficients in $\mathbb{F}$ is

$$(f, g) = \int_{t=0}^{t=1} f(t)\overline{g(t)}\, dt$$

- An inner product on the space of matrices with n rows and m columns is

$$(A, B) = \operatorname{trace}(B^* A) = \sum_{1 \le k \le m} \sum_{1 \le j \le n} \bar{B}^j_k A^j_k,$$

  where $B^* = \bar{B}^T$

Nondegeneracy Property

- Fact: If a vector $v \in V$ satisfies the following property:
$$\forall w \in V,\ (v, w) = 0,$$
then $v = 0$
  - Proof: Setting $w = v$, it follows that
  $$(v, v) = 0 \text{ and therefore } v = 0$$
- Corollary: If $v_1, v_2 \in V$ satisfy the property that
$$\forall w \in V,\ (v_1, w) = (v_2, w),$$
then $v_1 = v_2$
- Corollary: If '$L_1, L_2 : V \to W$ are linear maps such that
$$\forall v \in V,\ w \in W,\ (L_1(v), w) = (L_2(v), w),$$
then $L_1 = L_2$
  - Proof: Given $v \in V$,
  $$\forall w \in W,\ (L_1(v), w) = (L_2(v), w),$$
  which implies $L_1(v) = L_2(v)$
  - Since this holds for all $v \in V$, it follows that $L_1 = L_2$

Fundamental Inequalities

- **Cauchy-Schwarz inequality:** For any $v, w \in V$,

  $$|(v, w)| \leq |v||w|$$

  and

  $$|(v, w)| = |v||w|$$

  if and only if there exists $s \in \mathbb{F}$ such that

  $$v = sw \text{ or } w = sv$$

- **Triangle inequality:** For any $v, w \in V$,

  $$|v + w| \leq |v| + |w|$$

  and

  $$|v + w| = |v| + |w|$$

  if and only if $v = \pm w$

Proof When $\mathbb{F} = \mathbb{R}$

- If $v = 0$ or $w = 0$, equality holds
- Let

$$\begin{aligned} f(t) &= |v - tw|^2 \\ &= (v - tw, v - tw) \\ &= |v|^2 - 2t(v, w) + t^2|w|^2 \\ &= \left(t|w| - \frac{(v, w)}{|w|} \right)^2 + |v|^2 - \frac{(v, w)^2}{|w|^2} \end{aligned}$$

- f has a unique minimum when $t = t_{\min}$, where

$$t_{\min} = \frac{(v, w)}{|w|} \text{ and } f(t_{\min}) = |v|^2 - \frac{(v, w)^2}{|w|^2}$$

Proof of Cauchy-Schwarz (Part 1)

- If $v = 0$ or $w = 0$, equality holds
- If $w \neq 0$, let $f : \mathbb{F} \to \mathbb{R}$ be the function

$$\begin{aligned} f(t) &= |v - tw|^2 \\ &= (v - tw, v - tw) \\ &= |v|^2 - t(w, v) - \bar{t}(v, w) + |t|^2|w|^2 \end{aligned}$$

- If f has a minimum at $t_0 \in \mathbb{F}$, then its directional derivative at t_0 is zero in any direction $\dot{t}$

$$\begin{aligned} 0 &= \left.\frac{d}{ds}\right|_{s=0} f(t_0 + s\dot{t}) \\ &= -\dot{t}(w, v) - \bar{\dot{t}}(v, w) + (t_0\bar{\dot{t}} + \bar{t}_0\dot{t})|w|^2 \\ &= \dot{t}(\bar{t}_0 - (w, v)) + \bar{\dot{t}}(t_0|w|^2 - (v, w)) \\ &= \dot{t}(\overline{t_0 - (v, w)}) + \bar{\dot{t}}(t_0|w|^2 - (v, w)) \end{aligned}$$

Proof of Cauchy-Schwarz (Part 2)

- In particular, if

$$\dot{t} = t_0|w|^2 - (v, w),$$

we get

$$|t_0|w|^2 - (v, w)|^2 = 0,$$

- Therefore, the only critical point of f is

$$t_0 = \frac{(v, w)}{|w|^2}$$

- Since f is always nonnegative, it follows that

$$0 \leq f(t_0) = |v|^2 - \frac{|(v, w)|^2}{|w|^2}$$

which implies the Cauchy-Schwarz inequality

Proof of Cauchy-Schwarz (Part 3)

- If $w \neq 0$ and $|(v, w)| = |v||w|$, then

$$0 = |v|^2 - \frac{|(v, w)|^2}{|w|^2} = f(t_0) = |v - t_0 w|^2,$$

which implies that

$$v = t_0 w$$

Proof of Triangle Inequality

- The triangle inequality follows easily from Cauchy-Schwarz inequality

$$\begin{aligned} |v+w|^2 &= (v+w, v+w) \\ &= |v|^2 + (v,w) + (w,v) + |w|^2 \\ &\leq |v|^2 + |(v,w)| + |(w,v)| + |w|^2 \\ &\leq |v|^2 + 2|v||w| + |w|^2 \\ &= (|v|+|w|)^2 \end{aligned}$$

- If $|v+w| = |v| + |w|$, then

$$|(v,w)| = |(v,w)| = |v||w|,$$

which implies $v = tw$ and therefore

$$|t+1|^2|w|^2 = |tw+w|^2 \quad = |tw|^2 + |w|^2 = (|t|^2+1)|w|^2,$$

which implies that $t = \bar{t}$, i.e., $t \in \mathbb{R}$

Polarization Identities

- On $\mathbb{R}^n$

$$(v, w) = \frac{1}{4}(|v + w|^2 - |v - w|^2)$$

- On $\mathbb{C}^n$

$$(v, w) = \frac{1}{4}(|v + w|^2 + i|v + iw|^2 - |v - w|^2 - i|v - iw|^2)$$

Norm Defined by Inner Product

- The norm of $v \in V$,
$$|v| = \sqrt{(v, v)}$$
satisfies the following properties for any $s \in \mathbb{F}$, $v, w \in V$
$$|sv| = |s||v| \qquad \text{(Homogeneity)}$$
$$|v| \geq 0 \qquad \text{(Nonnegativity)}$$
$$|v| = 0 \iff v = 0 \qquad \text{(Nondegeneracy)}$$
$$|v + w| \leq |v| + |w| \qquad \text{(Triangle inequality)}$$
- Homogeneity and the triangle inequality imply convexity: For any $0 \leq t \leq 1$ and $v, w \in V$,
$$|(1-t)v + tw| \leq (1-t)|v| + t|w|$$

Norm

- A norm on a vector space V over $\mathbb{F}$ is a function

$$g : V \to \mathbb{R},$$

that satisfies for any $s \in \mathbb{F}$ and $v, w \in V$,

$$|sv| = |s||v| \quad \text{(Homogeneity)}$$
$$|v| \geq 0 \quad \text{(Nonnegativity)}$$
$$|v| = 0 \iff v = 0 \quad \text{(Nondegeneracy)}$$
$$|v + w| \leq |v| + |w| \quad \text{(Triangle inequality)}$$

Examples of Norms

- Given $1 \leq p < \infty$, the ℓ_p norm of $v \in \mathbb{F}^n$ is defined to be

$$|v|_p = (|v^1|^p + \cdots + |v^n|^p)^{1/p}$$

- The ℓ_∞ norm of $v \in \mathbb{F}^n$ is defined to be

$$|v|_\infty = \max(|v^1|, \ldots, |v^n|) = \lim_{p \to \infty} |v|_p$$

- The L_p norm of a continuous function $f : [0, 1] \to \mathbb{C}$ is defined to be

$$\|f\|_p = \left(\int_{x=0}^{x=1} |f(x)|^p \, dx \right)^{1/p}$$

- The L_∞ norm of a continuous function $f : [0, 1] \to \mathbb{C}$ is defined to be

$$\|f\|_\infty = \sup\{|f(x)| \ : \ 0 \leq x \leq 1\} = \lim_{p \to \infty} \|f\|_p$$

Parallelogram Identity

- A norm $|\cdot|$ on a vector space V satisfies the parallelogram identity

$$|v+w|^2 + |v-w|^2 = 2(|v|^2 + |w|^2),\ \forall v, w \in V$$

if and only if there is an inner product on V such that

$$|v|^2 = (v, v)$$

Orthogonality For Standard Dot Product on $\mathbb{R}^n$

- The following are synonyms: orthogonal, perpendicular, normal
- On $\mathbb{R}^n$,
  - Two vectors v_1, v_2 are called **orthogonal** if

$$v_1 \cdot v_2 = 0$$

  - A basis $(v_1, \ldots, v_n)$ is called **orthonormal** if for any $1 \leq i, j \leq n$,

$$v_i \cdot v_j = \delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

Orthogonality on an Inner Product Space

- Let V be an n-dimensional vector space over $\mathbb{F}$ with inner product $(\cdot, \cdot)$
- Two vectors v_1, v_2 are **orthogonal** if

$$(v_1, v_2) = 0$$

- Vectors $v_1, \ldots, v_k$ are **mutually orthogonal** if for every $1 \leq i < j \leq k$,

$$(v_i, v_j) \neq 0$$

- Mutually orthogonal vectors must all be nonzero
- A set of muturally orthogonal vectors is called an **orthogonal set**

Linear Independence of Orthogonal Set

- An orthogonal set is linearly independent, because if

$$a^1 v_1 + \cdots + a^k v_k = 0,$$

then for any $1 \leq j \leq k$,

$$0 = (v_j, a^1 v_1 + \cdots + a^k v_k) = a^j (v_j, v_j)$$

Since $v_j \neq 0$, $(v_j, v_j) \neq 0$ and therefore $a^j = 0$

- If

$$v = a^1 v_1 + \cdots + a^k v_k,$$

then for each $1 \leq j \leq k$,

$$a^j = \frac{(v, v_j)}{|v_j|}$$

and

$$v = \frac{(v, v_1)}{|v_1|} + \cdots + \frac{(v, v_k)}{|v_k|}$$

- Any orthogonal set of n vectors is a basis

Orthonormal Set and Basis

- $\{v_1, \ldots, v_k\} \subset V$ is called an **orthonormal** set if for any $1 \leq i, j \leq k$,
$$(v_i, v_j) = \delta_{ij}$$
- If $\mathbb{F} = \mathbb{C}$, such a set is also called a **unitary** set
- An orthonormal set of n elements is called an **orthonormal** or **unitary** basis
- Any orthogonal set $\{v_1, \ldots, v_k\}$ can be turned into an orthonormal set,
$$\{\frac{v_1}{|v_1|}, \ldots, \frac{v_k}{|v_k|}$$
- An orthormal or unitary basis is an orthonormal set with n elements,
$$E = (e_1, \ldots, e_n) \subset V$$
- If $v = a^1 e_1 + \cdots + a^n e_n$, then
$$a_j = (v, e_j)$$
- I.e.,
$$v = (v, e_1)e_1 + \cdots + (v, e_n)e_n$$

Example: Finite Fourier Decomposition (Part 1)

- For each $-N \leq k \leq N$, consider

$$\begin{aligned} v_k : [0, 2\pi] &\to \mathbb{C} \\ \theta &\mapsto e^{ik\theta} \end{aligned}$$

- Let

$$V = \{a^{-N} v_N + \cdots + a^0 + \cdots + a^N v_N \ : \ (a^1, \ldots, a^N) \in \mathbb{C}^{2N+1}\}.$$

- V is a $(2N+1)$-dimensional complex vector space
- Consider the inner product

$$(f_1, f_2) = \int_{\theta=0}^{\theta=2\pi} f_1(\theta) \bar{f_2}(\theta)\, d\theta$$

Finite Fourier Decomposition (Part 2)

- If $j \neq k$, then

$$
\begin{aligned}
(v_j, v_k) &= \int_{\theta=0}^{\theta=2\pi} e^{i(j-k)\theta}\, d\theta \\
&= \left.\frac{e^{i(j-k)\theta}}{i(j-k)}\right|_{\theta=0}^{\theta=2\pi} \\
&= 0 \\
(v_k, v_k) &= \int_{\theta=0}^{\theta=2\pi} 1\, d\theta \\
&= 2\pi
\end{aligned}
$$

- Therefore, $(v_{-N}, \ldots, v_N)$ is an orthogonal basis, and $(u_{-N}, \ldots, u_N)$, where

$$u_k = \frac{v_k}{\sqrt{2\pi}},\ -N \leq k \leq N,$$

is an orthonormal basis

Finite Fourier Decomposition (Part 3)

- Given any $f : C^0([0, 2\pi])$, let

$$f_N(\theta) = a^{-N} u_{-N} + \cdots + a^N u_N,$$

where

$$a^k = (f, u_k) = \frac{1}{\sqrt{2\pi}} \int_{\theta=0}^{\theta_2\pi} f(\theta) e^{-ik\theta}\, d\theta$$

- When is f_N is a good approximation to f?
- When is

$$f = \sum_{k=-\infty}^{k=\infty} a^k u_k?$$

Orthogonal Complement

- Let V be a real vector space with inner product $(\cdot, \cdot)$
- Given a subspace $E \subset V$, define its **orthogonal complement** to be the subspace
$$E^{\perp} = \{v \in V \;:\; \forall e \in E, (v, e) = 0\}$$
- $E \cap E^{\perp} = \{0\}$, because if
$$v \in E \cap E^{\perp},$$
then
$$|v|^2 = (v, v) = 0,$$
- If $v_1, v_2 \in E$, $w_1, w_2 \in E^{\perp}$, and
$$v_1 + w_1 = v_2 + v_2,$$
then
$$v_1 - v_2 = w_2 - w_1 \in E \cap E^{\perp}$$
and therefore, $v_1 = v_2$ and $w_1 = w_2$
- If follows that $E \oplus E^{\perp}$ is a subspace of V

Orthogonal Decomposition

- For each $v \in E \oplus E$, there exist unique $v_1 \in E$ and $v_2 \in E^\perp$ such that
$$v = v_1 + v_2$$
- Define the orthogonal projection maps
$$\begin{aligned} P_E : E \oplus E^\perp &\to E \\ v &\mapsto v_1 \end{aligned}$$
and
$$\begin{aligned} P_E^\perp : E \oplus E^\perp &\to E^\perp \\ v &\mapsto v_2 \end{aligned}$$

Orthogonal Projection Maps

- $P_E, P_E^{\perp}$ are linear maps
- $P_E : E \oplus E^{\perp} \to E$ is projection onto E:

$$\forall v \in E, \ P_E(v) = v$$

- $P_E^{\perp} : E \oplus E^{\perp} \to E^{\perp}$ is projection onto $E^{\perp}$:

$$\forall v \in E^{\perp}, \ P_E^{\perp}(v) = v$$

- Orthogonal decomposition: For any $v \in E \oplus E^{\perp}$,

$$P_E(v) \in E$$
$$P_E^{\perp}(v) \in E^{\perp}$$
$$v = P_E(v) + P_E^{\perp}(v)$$

Orthogonal Projection Minimizes Distance to a Subspace

- Observe that $v - P_E(v) = P_E^{\perp}(v) \in E^{\perp}$
- **Fact:** For each $v \in E \oplus E^{\perp}$ and $w \in E$,

$$|v - P_E(v)| \le |v - w|$$

  and equality holds if and only if $w = P_E(v)$
- **Proof:** Let $v = v_1 + v_2$, where

$$v_1 = P_E(v) \in E \text{ and } v_2 = v - P_E(v) \in E^{\perp}$$

- Then for any $w \in E$,

$$\begin{aligned} |v - w|^2 &= |v - P_E(v) + P_E(v) - w|^2 \\ &= (v_2 + (v_1 - w), v_2 + (v_1 - w)) \\ &= (v_2, v_2) + 2(v_1 - w, v_2) + (v_1 - w, v_1 - w) \\ &\ge |v - P_E(v)|^2 \end{aligned}$$

  and equality holds if and only if

$$|v_1 - w, v_1 - w|^2 = (v_1 - w, v_1 - w) = 0$$