Linear Algebra Done Outrageous (Taiwanese Version)

Eric Liu

Contents

Chapter 1

Summary

1.1 Quick Recap

Abstract

In this section, V and W are two vector spaces over the same field $\mathbb{F}$, $\mathbb{F}$ can be interpreted as either the field of real numbers or the field of complex numbers, and S is subset of V.

We say S is

(a) **linearly independent** if to write 0 as finite linear combination of S, the coefficients must all be zero.

(b) S **spans** V if all $v \in V$ can be written as finite linear combination of S.

We say S is a **basis** if any of the following equivalent conditions hold true

(a) S is a maximum linearly independent subset of V.

(b) S is linearly independent and spans V.

(c) All elements of V can be uniquely written as some finite linear combination of S.

Suppose V has a basis of cardinality n. Using **Gauss elimination**, one see that all linear independent subset of V must have cardinality not greater than n, and all basis must have cardinality n. Therefore, for vector space with some basis of finite cardinality, it make sense to refer to its **dimension**. Let $\dim(V) = n$. Each linearly independent subset S of V of cardinality n must be a basis, otherwise we may construct a linearly independent set of cardinality greater than n. By an algorithm, one see that

(a) If S spans V and has cardinality greater than n, then there exists a subset of S that is a basis of V.

(b) If B is a basis of V and S is linearly independent, then there exists some subset U of B such that $S \cup U$ is a basis of V.

Let $F : V \to W$ be a linear map. It is clear that both image and kernel of F form some vector space. Therefore, it make sense to define the **rank** of F by

$$\text{rank}(F) \triangleq \dim(\text{Im}\, F)$$

By extending a basis of the kernel of F to a basis of the whole V, one can prove the **Rank-Nullity Theorem**

$$\dim(\text{Ker}\, F) + \text{rank}(F) = \dim(V)$$

We use V^* to denote the vector space of linear map from V to $\mathbb{F}$, the **dual space** of V. Let $F : V \to W$ be a linear map. Its **dual map** $F^* : W^* \to V^*$ is defined by

$$F^*(\varphi) \triangleq \varphi \circ F$$

If V is finite-dimensional with basis $\{v_1, \ldots, v_n\}$, then its **dual basis** $\{\varphi_1, \ldots, \varphi_n\}$ is defined by

$$\varphi_i(v_j) \triangleq \delta^i_j$$

And there exists a natural isomorphism between V and its double dual $(V^*)^*$ by identifying $v \in (V^*)^*$ as

$$v(\varphi) \triangleq \varphi(v)$$

Let $\{w_1, \ldots, w_m\}$ be a basis of W with dual basis $\{\xi_i, \ldots, \xi_m\}$. It is clear that

The matrices of F and F^* are transpose of each other.

This together with the observation that for finite dimensional V, W, we have

$$\text{rank}(F) + \dim(\ker F^*) = \dim(W)$$

gain us a quick proof that the dimensions of column space and row space of a fixed matrix are always equal.

Given some n-by-n square matrix A, its **determinant** is by definition

$$\det A \triangleq \sum_{\sigma \in S_n} \left(\operatorname{sgn} \sigma \prod_{k=1}^{n} A_{\sigma(k),k} \right)$$

If one identify the space of n-by-n square matrix over $\mathbb{F}$ with $(\mathbb{F}^n)^n$, determinant can be equivalently defined to be the unique alternating multilinear map from $(\mathbb{F}^n)^n$ to $\mathbb{F}$ such that

$$\det(I) = 1$$

By an algorithm, one see that, for each alternating multilinear map $F : (\mathbb{F}^n)^n \to \mathbb{F}$, we have

$$F(B) = (\det B) \cdot F(I) \text{ for all } B.$$

Therefore, if we observe for each n-by-n square matrix A over $\mathbb{F}$ that the map $B \mapsto \det AB$ is indeed alternating multilinear, then we immediately have the celebrated result $\det AB = \det A \det B$ as a corollary. This multiplicative property of determinant allow us to well define determinant for each linear epimorphism F over some finite-dimensional vector space by

$$\det F \triangleq \det([F])$$

where $[F]$ is the matrix representation of F with respect to some basis. Moreover, because the inverse of a linear map is linear if exists, we see that for linear map F and square matrix A,

$$F \text{ is invertible} \iff \det F \neq 0$$

And

$$A \text{ is invertible} \iff \det A \neq 0$$

Given some linear epimorphism F over V, we say $v \in V$ is an **eigenvector** with respect to the **eigenvalue** $\lambda \in \mathbb{F}$ if $v \neq 0$ and $F(v) = \lambda v$. If there exists some basis of V consisting eigenvectors of F, we say F is **diagonalizable**. Let A be some square matrix, if there exists some invertible square matrix P such that PAP^{-1} diagonal, we also say A is **diagonalizable**. Immediately, we see that if V is finite dimensional, then with respect to any basis of V

$$[F] \text{ is diagonalizable } \iff F \text{ is diagonalizable}$$

Let V be finite dimensional. We define the **characteristic polynomial** of F and A by

$$\det(tI - F) \text{ and } \det(tI - A)$$

Obviously, λ is an eigenvalue of F if and only if λ is a root of the characteristic polynomial of F.

1.2 Jordan-Chevalley Decomposition

Abstract

In this section, all vectors spaces are finite dimensional, and we say that V is a **direct sum** of some collection $\{U_i\}_{i\in I}$ of subspaces of V and writes $V = \bigoplus_{i\in I} U_i$ if for each $v \in V$ there exists some unique tuple $(u_i)_{i\in I}$ such that $u_i \neq 0$ for finite number of $i \in I$ and $v = \sum u_i$.

Given $F \in \text{End}(V)$, we know the kernels of its powers is increasing

$$\{0\} = \text{Ker } F^0 \subseteq \text{Ker } F^1 \subseteq \text{Ker } F^2 \subseteq \text{Ker } F^3 \subseteq \cdots$$

This sequence grows in good manner. If it stops growing at some points, say, $\text{Ker } F^n = \text{Ker } F^{n+1}$ for some $n \in \mathbb{Z}_0^+$, then it stops forever

$$\text{Ker } F^n = \text{Ker } F^{n+1} = \text{Ker } F^{n+2} = \text{Ker } F^{n+3} = \cdots$$

In particular, by counting dimensions, we know that this sequence must stops before reaching to the dimension of V in the sense that

$$\text{Ker}(F^{\dim V}) = \text{Ker}(F^{\dim V+1}) \tag{1.1}$$

Equation 1.1 allows us to elegantly unify distinct notions within a single framework. For instance, for each eigenvalue λ, we define the **generalized eigenspace** $G(\lambda, F)$ as

$$G(\lambda, F) \triangleq \text{Ker}(F - \lambda I)^{\dim V}$$

Similarly, we say that F is **nilpotent** if

$$\text{Ker } F^{\dim V} = V$$

Theorem 1.2.1. (Linear Independence of Generalized Eigenspaces) Let $F \in \text{End}(V)$. If $v_1, \dots, v_n$ are generalized eigenvectors with respect to distinct eigenvalues $\lambda_1, \dots, \lambda_n$, then they are linearly independent.

Proof. Suppose

$$a_1v_1 + \cdots + a_nv_n = 0 \tag{1.2}$$

Because v_1 is a generalized eigenvector, we may let k be the largest non-negative integer such that $(F - \lambda_1 I)^k v_1 \neq 0$, so that

$$w \triangleq (F - \lambda_1 I)^k v_1 \text{ is an eigenvector with eigenvalue } \lambda_1.$$

Now, noting that the set of endomorphism $\{(F-\lambda_1 I)^k, (F-\lambda_2 I)^{\dim V}, \dots, (F-\lambda_n I)^{\dim V}\}$ indeed commutes, we may apply the endomorphism $(F-\lambda_1 I)^k (F-\lambda_2 I)^{\dim V} \cdots (F-\lambda_n I)^{\dim V}$ onto both sides of Equation 1.2 and get

$$\begin{aligned} 0 &= a_1 (F-\lambda_2 I)^{\dim V} \cdots (F-\lambda_n I)^{\dim V} w \\ &= a_1 (\lambda_1-\lambda_2)^{\dim V} \cdots (\lambda_1-\lambda_n)^{\dim V} w \end{aligned}$$

Which implies $a_1 = 0$. WLOG, we have shown $a_1 = \cdots = a_n = 0$. ■

Equation 1.1 together with Rank-Nullity Theorem give us a theoretically crucial decomposition of V

$$V = \operatorname{Ker} F^{\dim V} \oplus \operatorname{Im} F^{\dim V} \tag{1.3}$$

As we will later see, Decomposition 1.3 plays a central role in the proof for Theorem 1.2.2.

Theorem 1.2.2. (Decomposition into Generalized Eigenspaces) Let V be a finite dimensional complex vector space, let $F \in \operatorname{End}(V)$, and let $\{\lambda_1, \dots, \lambda_m\}$ be the set of eigenvalues of F. We have

$$V = G(\lambda_1, F) \oplus \cdots \oplus G(\lambda_m, F)$$

Proof. This is proved by induction on the dimension of V. The two base cases correspond to the 0-dimensional and 1-dimensional cases, both of which are trivial. Now, suppose that such a decomposition always exists for complex vector spaces of strictly smaller dimension than V and their endomorphisms. By Equation 1.3, we may decompose V into

$$V = G(\lambda_1, F) \oplus U, \quad \text{ where } U = \operatorname{Im}(F-\lambda_1)^{\dim V}.$$

Noting that F and $F - \lambda_1 I$ commutes, we conclude that U is stable under F. Therefore, the restriction $F|_U$ defines an endomorphism $F|_U \in \operatorname{End}(U)$. By inductive hypothesis, we may decompose U into

$$U = G(\lambda_2, F|_U) \oplus \cdots \oplus G(\lambda_m, F|_U)$$

WLOG, it remains to show

$$G(\lambda_2, F|_U) = G(\lambda_2, F)$$

Arbitrary select $v \in G(\lambda_2, F)$. We may decompose $v = a_1 v_1 + \cdots + a_m v_m$, where $v_1 \in G(\lambda_1, F)$ and $v_n \in G(\lambda_n, F|_U)$ for $2 \le n \le m$. Because $(F-\lambda_2 I)^{\dim V} v_2 = 0$, we know

$$(F-\lambda_2 I)^{\dim V}(a_1 v_1 + a_3 v_3 + \cdots + a_m v_m) = 0$$

This implies $a_1 v_1 + a_3 v_3 + \cdots + a_m v_m \in G(\lambda_2, F)$. It now follows from Theorem 1.2.1 that $a_1 = a_3 = \cdots = a_m = 0$. We have shown $v \in G(\lambda_2, F|_U)$. ■

Given some finite dimensional vector space V and some $F \in \text{End}(V)$, we are particularly concerned with the existence and uniqueness of the **Jordan-Chevalley decomposition** of F, i.e, some diagonalizable $S \in \text{End}(V)$ such that

(a) $N \triangleq F - S$ is nilpotent.

(b) S and N commute.

If V is over $\mathbb{C}$, then Theorem 1.2.2 assert the existence of such decomposition by letting S maps $v \in G(\lambda_i, F)$ to $\lambda_i v$. To see such decomposition is unique, let $S \in \text{End}(V)$ be some Jordan-Chevalley decomposition of F and decompose V into

$$V = V_1 \oplus \cdots \oplus V_k$$

where V_i are the eigenspaces of S corresponding to distinct eigenvalues λ_i. Because F, S commute and $F - S$ is nilpotent, we may conclude that λ_i are indeed eigenvalues of F and

$$V_i \subseteq G(\lambda_i, F) \text{ for all } i$$

This together with Theorem 1.2.2 shows the uniqueness of Jordan-Chevalley decomposition of endomorphisms of finite dimensional vector spaces over $\mathbb{C}$.

Theorem 1.2.3. (Jordan forms of Nilpotent Operators) Let V be a finite-dimensional complex vector space, and let $N \in \text{End}(V)$ be nilpotent. There exists some $v_1, \dots, v_m \in V$ and nonnegative integer $n_1, \dots, n_m$ such that

(a) $N^{n_1} v_1, \dots, N v_1, v_1, \dots, N^{n_m} v_m, \dots, N v_m, v_m$ form a basis for V.

(b) $N^{n_1+1} v_1 = \cdots = N^{n_m+1} v_m = 0$.

Proof. This is also proved by induction on the dimension of V. The two base cases correspond to the 0-dimensional and 1-dimensional cases, both of which are trivial. We now prove the inductive case.

Because N is nilpotent, we know $\dim(\text{Im}\, N) < \dim V$. Noting that $\text{Im}\, N$ is stable under N, we see $N|_{\text{Im}\, N} \in \text{End}(\text{Im}\, N)$. From inductive hypothesis, we have $v_1, \dots, v_n \in \text{Im}\, N$ and nonnegative integer $k_1, \dots, k_n$ such that

$$\{N^{k_1} v_1, \dots, N v_1, v_1, \dots, N^{k_n} v_n, \dots, N v_n, v_n\} \text{ form a basis for } \text{Im}\, N.$$

and

$$N^{k_1+1} v_1 = \cdots = N^{k_n+1} v_n = 0.$$

Because $v_1,\dots,v_n \in \operatorname{Im} N$, we may let $u_1,\dots,u_n \in V$ satisfy $v_j = Nu_j$ for all j. To see

$$\{N^{k_1+1}u_1,\dots,Nu_1,u_1,\dots,N^{k_n}u_n,\dots,Nu_n,u_n\} \text{ is linearly independent,} \tag{1.4}$$

suppose some finite linear combination equals to 0. By applying N to this finite linear combination, we see the only possible nonzero coefficients are those of $N^{k_j+1}u_j$, otherwise

$$\{N^{k_1}v_1,\dots,Nv_1,v_1,\dots,N^{k_n}v_n,\dots,Nv_n,v_n\} \text{ would not be linearly independent.}$$

Knowing that the only possible nonzero coefficients are those of $N^{k_j+1}u_j = N^{k_j}v_j$, we may conclude that even these coefficients are zero, since $\{N^{k_1}v_1,\dots,N^{k_n}v_n\}$ is linearly independent in the first place. We may now expand Set 1.4 to a basis

$$\{N^{k_1+1}u_1,\dots,Nu_1,u_1,\dots,N^{k_n+1}u_n,\dots,Nu_n,u_n,w_1,\dots,w_p\} \tag{1.5}$$

for V. Because $\{N^{k_1+1}u_1,\dots,N^2u_1,Nu_1,\dots,N^{k_n+1}u_n,\dots,N^2u_n,Nu_n\}$ is a basis for $\operatorname{Im} N$, we may subtract each w_i with some element of

$$\operatorname{span}\{N^{k_1}u_1,\dots,Nu_1,u_1,\dots,N^{k_n+1}u_n,\dots,Nu_n,u_n\}$$

so that Set 1.5 form a desired basis for V. ■

Let V be some finite-dimensional complex vector space and $F \in \operatorname{End}(V)$. Let $S+N$ be the Jordan-Chevalley decomposition of F. The basis for V from Theorem 1.2.3 is called a **Jordan basis**. If we express F as a matrix with respect to this basis, we say that matrix is in its **Jordan form**. The Jordan form looks like

$$\begin{bmatrix} J_1 & & \\ & \ddots & \\ & & J_r \end{bmatrix}$$

where each block matrix looks like

$$J = \begin{bmatrix} \lambda_i & 1 & & \\ & \lambda_i & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_i \end{bmatrix}$$

Let $\lambda_1,\dots,\lambda_m$ be the distinct eigenvalues of F, and let r_i be the dimension of the largest Jordan block with eigenvalue λ_i. It is clear that the **minimal polynomial** of F, some polynomial m of smallest degree such that $m(F)=0$, take the form

$$m(x) = (x-\lambda_1)^{r_1}\cdots(x-\lambda_m)^{r_m}$$

It is also clear that even if V is over $\mathbb{R}$ instead of $\mathbb{C}$, we have the **Cayley-Hamilton**

Theorem, that is, $p(F) = 0$, where p is the characteristic polynomial of F.

Interestingly, we may also take a completely algebraic approach to construct the Jordan-Chevalley decomposition S, N and show that they are polynomials in F, without ever invoking Theorem 1.2.1, Theorem 1.2.2, nor any result that depends on them. The argument proceeds as follows:

Because $\mathrm{End}(V)$ is finite-dimensional, the minimal polynomial m exists. By the Fundamental Theorem of Algebra, we can factor m into linear terms:

$$m(x) = (x - \xi_1)^{r_1} \cdots (x - \xi_k)^{r_k}.$$

Consider the polynomials f_i defined by

$$f_i(x) \triangleq \prod_{j=1; j \neq i}^{k} (x - \xi_j)^{r_j}.$$

Since $\{f_1, \ldots, f_k\}$ are coprime, by Bézout's identity, we may find polynomials $q_i \in \mathbb{C}[x]$ such that

$$1 = \sum_{i=1}^{k} q_i f_i.$$

Define $\pi_i \in \mathrm{End}(V)$ by $\pi_i \triangleq (q_i f_i)(F)$. Because m divides $f_i f_j$ for $i \neq j$, we conclude that

$$\pi_i \pi_j = 0 \quad \text{for } i \neq j.$$

This, together with $\sum \pi_j = 1$, shows that π_i are **projections**:

$$\pi_i^2 = \pi_i \quad \text{for all } i.$$

It is clear that V is a direct sum of the images of these projections. Define

$$S \triangleq \sum_{i=1}^{k} \xi_i \pi_i.$$

Since S is a polynomial in F, we conclude that S and N commute. The fact that N is nilpotent follows from the definition of f_i.

1.3 Spectral

By a **norm**, we mean some **positive-definite** functional $\|\cdot\|: V \to \mathbb{R}$ that satisfies **absolute homogeneity** and **triangle inequality**. In this context, for $\|\cdot\|$ to be positive-definite, it must satisfy

$$\|v\| = 0 \implies v = 0$$

Observing that

$$\|0\| = \|0 + v\| \leq \|0\| + \|v\| \text{ for all } v \in V$$

we see norm must also be nonnegative.

By an **inner product**, we mean some **positive-definite** map $\langle \cdot, \cdot \rangle : V^2 \to \mathbb{F}$ that is linear in the first argument and satisfies **conjugate symmetry**:

$$\langle v, w \rangle = \overline{\langle w, v \rangle} \quad \text{for all } v, w \in V$$

In this context, for $\langle \cdot, \cdot \rangle : V^2 \to \mathbb{F}$ to be positive-definite, it has to satisfy

$$\langle v, v \rangle > 0 \quad \text{for all } v \neq 0$$

With an inner product equipped on V, we may discuss the **orthogonality** of vectors. Given a countable set $\{v_1, v_2, v_3, \dots\}$ of non-zero vectors, we use the term **Gram-Schmidt process** to refer to the process of defining

$$e_n \triangleq v_n - \frac{\langle v_n, e_{n-1} \rangle}{\|e_{n-1}\|^2} e_{n-1} - \cdots - \frac{\langle v_n, e_1 \rangle}{\|e_1\|^2} e_1$$

so that $\{e_1, e_2, e_3, \dots\}$ become orthogonal while

$$\operatorname{span}(v_1, \dots, v_n) = \operatorname{span}(e_1, \dots, e_n) \quad \text{for all } n.$$

Although implicit, Gram-Schmidt process may be considered one of the most important notions in inner product space. For example, the Gram-Schmidt process, together with **Pythagorean Theorem**, can be used to prove the **Cauchy-Schwarz Inequality**, which in term shows that the functional $f : V \to \mathbb{R}$ induced by

$$f(v) \triangleq \sqrt{\langle v, v \rangle}$$

indeed satisfies triangle inequality, thus forming a norm. In addition, given an endomorphism T of some finite-dimensional complex vector space V, by applying Gram-Schmidt process to a Jordan basis, we get an orthogonal basis under which T becomes an upper triangular matrix, called its **Schur's form**.

Let V, W be two inner product space over $\mathbb{F}$, and let $T : V \to W$ be some linear map. If some linear map $T^\dagger : W \to V$ satisfies

$$\langle Tv, w \rangle = \langle v, T^\dagger w \rangle \quad \text{for all } v \in V, w \in W,$$

we say $T^\dagger$ is an **adjoint** of T. If both V and W are finite-dimensional, then $T^\dagger$ exists uniquely, and its matrix representation is always the complex conjugate of that of T, regardless of the choices of bases. It is obvious that the double adjoint of a linear map between two finite-dimensional inner product spaces over $\mathbb{F}$ is itself.

If the underlying field is $\mathbb{C}$ and T is **normal**, then by direct computation, we see that its Schur's form must be diagonal. This argument is called **Spectral Theorem for complex finite-dimensional vector space**. If the underlying field is $\mathbb{R}$, then T being normal is not enough for T to be orthogonally diagonalized, since we do not have Schur's decomposition in the first place.[a]

[a] Actually, we do have Schur's decomposition in the sense that we may first decompose a real matrix into a complex Jordan matrix, and then applying the Gram-Schmidt process. Yet, because the matrix representation of adjoint of T when the underlying field is $\mathbb{R}$ is just the transpose of that of T instead of the conjugate transpose, direct computation yield nothing.

Theorem 1.3.1. (Spectral Theorem for real finite-dimensional vector space) Let V be a finite-dimensional real vector space, and let $T \in \mathrm{End}(V)$. If T is **self-adjoint**, then there exists an orthogonal eigenbasis for V with respect to T.

Proof. The is proved by induction on the dimension of V. The two base cases correspond to the 0-dimensional and 1-dimensional cases, both of which are trivial. We now prove the inductive case.

Let $v^\perp$ denote the space of vectors orthogonal to v. Because T is self-adjoint, if v is an eigenvector of T, then $v^\perp$ is stable under T. This reduce the problem into proving the existence of an eigenvector.

Let f be some real polynomial that sends T to zero. By Fundamental Theorem of Algebra, we may write

$$0 = f(T) = (T^2 + b_1 T + c_1 I) \cdots (T^2 + b_n T + c_n I)(T - \lambda_1 I) \cdots (T - \lambda_m I)$$

for some real b_i, c_i, λ_i such that $b_i^2 < 4c_i$ for all i. Because T is self-adjoint, for each i, by Cauchy-Schwarz inequality, we may compute for all $v \neq 0$ that

$$\begin{aligned}
\langle (T^2 + b_i T + c_i I)v, v\rangle &= \langle T^2 v, v\rangle + b_i\langle Tv, v\rangle + c_i\langle v, v\rangle \\
&= \langle Tv, Tv\rangle + b_i\langle Tv, v\rangle + c_i\|v\|^2 \\
&\geq \|Tv\|^2 - b_i\|Tv\| \cdot \|v\| + c_i\|v\|^2 \\
&= \left(\|Tv\| - \frac{b_i\|v\|}{2}\right)^2 + \left(c_i - \frac{b_i^2}{4}\right)\|v\|^2 > 0
\end{aligned}$$

This implies $T^2 + b_i T + c_i I$ are invertible, which implies for some j, the map $T - \lambda_j I$ is not invertible, i.e., for some j, the real number λ_j is an eigenvalue. ■

1.4 Matrix Exponential

Given some square matrix A over $\mathbb{F}$, we define its **matrix exponential** by

$$e^A \triangleq \sum_{n=0}^{\infty} \frac{A^n}{n!}$$

It

Chapter 2

NTU Math M.A. Program Entrance Exam

2.1 Year 113

Question 1

Let

$$A \triangleq \begin{bmatrix} -2 & -1 & 1 \\ 1 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

Find the Jordan-Chevalley decomposition of A and compute e^A.

Proof. Routine computation give us

$$J = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & -1 \end{bmatrix} \text{ and } P = \begin{bmatrix} 0 & 1 & -1 \\ 1 & -1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \text{ and } A = PJP^{-1}$$

Therefore, the Jordan-Chevalley decomposition of A is

$$A = D + N \text{ where } D = P \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix} P^{-1} \text{ and } N = P \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} P^{-1}$$

And

$$e^A = P \begin{bmatrix} e & 0 & 0 \\ 0 & e^{-1} & e^{-1} \\ 0 & 0 & e^{-1} \end{bmatrix} P^{-1}$$

■

Question 2

Let V be the space of polynomial in x over $\mathbb{R}$ of degree not higher than 2. Define an inner product for V by

$$\langle f, g\rangle \triangleq \int_{-1}^{1} fg$$

Find a polynomial $k(x,t)$ such that

$$f(x) = \int_{-1}^{1} k(x,t)f(t)dt \quad \text{for all } f \in V \tag{2.1}$$

Define $T \in \text{End}(V)$ by

$$T(a_2x^2 + a_1x + a_0) \triangleq 2a_2x + a_1$$

That is, $Tf \triangleq f'$. Find the adjoint of T.

Proof. Because of the linearity, for k to satisfy Equation 2.1 for all $f \in V$, it only has to satisfies

$$1 = \int_{-1}^{1} kdt \text{ and } x = \int_{-1}^{1} ktdt \text{ and } x^2 = \int_{-1}^{1} kt^2dt \tag{2.2}$$

If we write k in the form

$$k = f_0(x) + f_1(x)t + f_2(x)t^2 + \cdots$$

Then Equation 2.2 becomes

$$\begin{aligned} 1 &= \frac{2f_0(x)}{1} + \frac{2f_2(x)}{3} + \frac{2f_4(x)}{5} + \cdots \\ x &= \frac{2f_1(x)}{3} + \frac{2f_3(x)}{5} + \frac{2f_5(x)}{7} + \cdots \\ x^2 &= \frac{2f_0(x)}{3} + \frac{2f_2(x)}{5} + \frac{2f_4(x)}{7} + \cdots \end{aligned}$$

Thus, k can be

$$k(x,t) = \Big(\frac{9}{8} + \frac{-15}{8}x^2\Big) + \Big(\frac{3x}{2}\Big)t + \Big(\frac{-15}{8} + \frac{45}{8}x^2\Big)t^2$$

Routine computation give us an orthonormal basis

$$\left\{\sqrt{\frac{1}{2}}, \sqrt{\frac{3}{2}}x, \sqrt{\frac{45}{8}}(x^2-\frac{1}{3})\right\} \text{ for } V$$

With respect to this ordered basis, the matrix representation of T is

$$\begin{bmatrix} 0 & \sqrt{3} & 0 \\ 0 & 0 & \sqrt{15} \\ 0 & 0 & 0 \end{bmatrix}$$

Because the matrix representation of $T^\dagger$ with respect to the same basis is the conjugate transpose of that of T, we may now compute

$$T^\dagger(a_2x^2+a_1x+a_0)=\frac{15a_1}{2}(x^2-\frac{1}{3})+(3a_0+a_2)x$$

■

Question 3

Let V be the vector space of all n-by-n square matrices over $\mathbb{R}$, and let $F: V \to \mathbb{R}$ be linear. Suppose

$$F(AB)=F(BA) \text{ and } F(I)=n$$

for all $A, B \in V$. Show that F is the trace function.

Proof. Let $E_{i,j} \in V$ be the matrix whose only nonzero entry is 1, located in the i-th row and j-th column. If $i \neq j$, then

$$E_{i,i}E_{i,j}=E_{i,j} \text{ and } E_{i,j}E_{i,i}=0.$$

This, together with the linearity of F, shows that F depends only on the diagonal entries. For each permutation $\sigma \in S_n$, there exists a **permutation matrix** C such that

The $\sigma(i)$-th row of CA is identical to the i-th row of A.

And

The $\sigma(i)$-th column of AC^{-1} is identical to the i-th column of A.

Therefore,

$$A_{i,i}=(CAC^{-1})_{\sigma(i),\sigma(i)}.$$

Observing that

$$F(CAC^{-1}) = F(A(CC^{-1})) = F(A),$$

we have shown that F is stable under the permutation of diagonal entries. Define $\sigma \in S_n$ by

$$\sigma(i) \triangleq \begin{cases} i+1 & \text{if } 1 \leq i < n, \\ 1 & \text{if } i = n. \end{cases}$$

Let C_k be the permutation matrix corresponding to σ^k for all $1 \leq k \leq n$. The proof now follows from computing

$$nF(A) = \sum_{k=1}^{n} F(C_k A C_k^{-1}) = F\Big(\sum_{k=1}^{n} C_k A C_k^{-1}\Big) = F((\operatorname{tr} A)I) = n \operatorname{tr} A.$$

■

Question 4

Let U, V be two finite dimensional space over the same field. Let $T : U \to V$ be a linear map, and let $T^* : U^* \to V^*$ be its dual map. Prove

$$T \text{ is injective} \iff T^* \text{ is surjective}$$

And

$$T \text{ is surjective} \iff T^* \text{ is injective}$$

Proof. Let $\{T(u_1), \dots, T(u_n)\}$ be a basis for the image of T. Extend this to a basis $\{T(u_1), \dots, T(u_n), v_1, \dots, v_m\}$ for V. Let $\{\xi_1, \dots, \xi_{n+m}\}$ be its dual basis. It is clear that $\{\xi_{n+1}, \dots, \xi_{n+m}\}$ belongs to the kernel of T^*. Observe

$$T^* \xi_1 v_1 = 1$$

to conclude that $\xi_i \notin \operatorname{Ker} T^*$ for all $1 \leq i \leq n$. We have shown $\{\xi_{n+1}, \dots, \xi_{n+m}\}$ is a basis for the kernel of T^*. In other words,

$$\operatorname{rank} T + \operatorname{Dim}(\operatorname{Ker} T^*) = \dim V.$$

This, together with Rank-Nullity Theorem, proves both propositions. ■

Question 5

Let V be some finite-dimensional vector space over some field $\mathbb{F}$ and let $T \in \text{End}(V)$. Let f, g be two relatively prime polynomials. Prove

$$\text{Ker}(f(T)g(T)) = \text{Ker}\, f(T) \oplus \text{Ker}\, g(T).$$

Proof. To show that the two kernels form a direct sum, assume for contradiction that there exists a nonzero vector $v \in V$ such that v belongs to both kernels. Since f is a polynomial satisfying $f(T)v = 0$, there must exist a polynomial p of minimal degree such that $p(T)v = 0$. By polynomial division, we can write

$$f = pq + r, \quad \text{where } q, r \text{ are polynomials and } \deg r < \deg p.$$

Since r has a strictly smaller degree than p, the minimality of p implies that $r = 0$. Thus, f is divisible by p. WLOG, we may assume that g is also divisible by p. This contradicts the assumption that f and g are relatively prime polynomials. We have shown the two kernels indeed form a direct sum. It is clear that

$$\text{Ker}\, f(T) \oplus \text{Ker}\, g(T) \subseteq \text{Ker}(f(T)g(T))$$

We now prove the opposite. Let $v \in \text{Ker}(f(T)g(T))$. Because f, g are relatively prime, by **Bezout's identity**, there exists some polynomials a, b such that

$$af + bg = 1$$

The proof then follows from noting

$$(af)(T)v \in \text{Ker}\, g(T) \text{ and } (bg)(T)v \in \text{Ker}\, f(T)$$

∎

2.2 Year 112

Question 6

Let

$$\mathbf{v}_1 = (1,2,0,4), \mathbf{v}_2 = (-1,1,3,-3), \mathbf{v}_3 = (0,1,-5,-2), \mathbf{v}_4 = (-1,-9,-1,-4)$$

be vectors in $\mathbb{R}^4$. Let W_1 be the subspace spanned by $\mathbf{v}_1$ and $\mathbf{v}_2$, and let W_2 be the subspace spanned by $\mathbf{v}_3$ and $\mathbf{v}_4$. Find a basis for $W_1 \cap W_2$.

Proof. Use Gauss elimination to show

$$\begin{cases} \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\} \text{ is linearly independent.} \\ \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_4\} \text{ is linearly independent.} \end{cases}$$

Use Gauss elimination to show

$$3\mathbf{v}_1 + 2\mathbf{v}_2 + \mathbf{v}_3 + \mathbf{v}_4 = 0$$

And conclude

$$W_1 \cap W_2 = \text{span}\{\mathbf{v}_3 + \mathbf{v}_4\}$$

∎

Question 7

Let

$$A = \begin{bmatrix} 0 & 1 & -1 \\ 3 & -4 & 1 \\ 3 & -8 & 5 \end{bmatrix}$$

Find an invertible matrix $Q \in M_3(\mathbb{C})$ such that

$$QAQ^{-1} = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 12 \\ 0 & 1 & 1 \end{bmatrix}$$

Find an invertible matrix $P \in M_3(\mathbb{C})$ such that PAP^{-1} is diagonal. *Note that my wording differs slightly with the original.*

Proof. Noting that characteristic polynomial is defined uniquely up to change of basis,

$$\det(tI - QAQ^{-1}) = \det(Q(tI - A)Q^{-1}) = \det(tI - A)$$

We may quickly compute that the characteristic polynomial of A is $t(t-4)(t+3)$. Routine computation now give us

$$P = \begin{bmatrix} 1 & 7 & 0 \\ 1 & -1 & 1 \\ 1 & -29 & 1 \end{bmatrix} \text{ and } PAP^{-1} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & -3 \end{bmatrix}$$

Note that one may find Q by solving a 9-by-9 linear equation. Here we give a smarter approach. By looking at QAQ^{-1}, we may reduce the problem into finding $v \in \mathbb{C}^3$ such that

$$A^3 v = 12Av + A^2 v \text{ and } \{v, Av, A^2 v\} \text{ is linear independent}$$

So that

$$Q \triangleq \begin{bmatrix} v & Av & A^2 v \end{bmatrix} \text{ suffices.}$$

Because the characteristic polynomial of A is $t(t-4)(t+3)$, by Cayley-Hamilton Theorem, we know all $v \in \mathbb{C}^3$ satisfy the first condition. To satisfy the linear independence, we write

$$v = c_1 e_1 + c_2 e_2 + c_3 e_3$$

where e_1, e_2, e_3 are the eigenvectors

$$e_1 \triangleq \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \text{ and } e_2 \triangleq \begin{bmatrix} 7 \\ -1 \\ -29 \end{bmatrix} \text{ and } e_3 \triangleq \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$$

So that we may just looks for $c_1, c_2, c_3 \in \mathbb{C}$ that makes

$$\begin{bmatrix} c_1 & c_2 & c_3 \\ 0 & 4c_2 & -3c_3 \\ 0 & 16c_3 & 9c_3 \end{bmatrix} \text{ invertible.}$$

By computing the determinant, we see that an obvious choice is $c_1 = c_2 = c_3 = 1$. That is,

$$v = \begin{bmatrix} 8 \\ 1 \\ -27 \end{bmatrix} \text{ and } Q = \begin{bmatrix} 8 & 28 & 112 \\ 1 & -7 & -7 \\ -27 & -119 & -455 \end{bmatrix}$$

■

Question 8

Define **matrix trigonometry** by

$$\sin A \triangleq \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!} A^{2n+1}$$

Compute

$$\sin \begin{pmatrix} 1 & 3 \\ 0 & 1 \end{pmatrix}$$

Show that there exist no matrix $A \in M_2(\mathbb{R})$ such that

$$\sin A = \begin{pmatrix} 1 & 2022 \\ 0 & 1 \end{pmatrix}$$

Proof. Using the identity

$$\sin A = \frac{e^{iA} - e^{-iA}}{2i}$$

we may compute

$$\sin \begin{pmatrix} 1 & 3 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} \sin(1) & 3\cos(1) \\ 0 & \sin(1) \end{pmatrix}$$

Because $\sin(PBP^{-1}) = P(\sin B)P^{-1}$ for all B and all invertible P, and because

$$\sin \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} = \begin{pmatrix} \sin \lambda_1 & 0 \\ 0 & \sin \lambda_2 \end{pmatrix} \text{ and } \sin \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix} = \begin{pmatrix} \sin \lambda & \cos \lambda \\ 0 & \sin \lambda \end{pmatrix}$$

We may finish the proof by computing the Jordan form

$$\begin{pmatrix} 1 & 2022 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & \frac{1}{2022} \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & \frac{1}{2022} \end{pmatrix}^{-1}$$

that there exists no matrix $A \in M_2(\mathbb{R})$ such that

$$\sin A = \begin{pmatrix} 1 & 2022 \\ 0 & 1 \end{pmatrix}$$

∎

Question 9

Let $A = (a_{ij}) \in M_n(\mathbb{C})$, and let $\lambda_1, \ldots, \lambda_n$ be roots of characteristic polynomial of A counted with multiplicity. Show that

$$A \text{ is normal} \iff \|A\|_F = \sum_{k=1}^{n} |\lambda_k|^2$$

where $\|A\|_F$ is the Frobenius norm of A.

Proof. Because the underlying field is $\mathbb{C}$, we may apply **Schur's decomposition** to get an upper triangular matrix D and an unitary matrix Q such that

$$A = QDQ^{-1}$$

Because A and D has the same characteristic polynomial, we know the eigenvalues of A lie on the diagonal line of D counted with multiplicity. The proof now follows from computing

$$\|A\|_F = \operatorname{tr}(A^*A) = \operatorname{tr}(D^*D) = \|D\|_F$$

and **Spectral Theorem for complex finite-dimensional vector space**. ■

Question 10

Let $A, B \in M_n(\mathbb{C})$. Suppose that all of the eigenvalues of A and B are positive real numbers. Prove

$$A^4 = B^4 \implies A = B$$

Proof. ■

2.3 Year 111

Question 11

Let V be a finite-dimensional complex inner product space. Let $d \in \text{End}(V)$ satisfy

$$d^2 = 0$$

Let δ be the adjoint of d. Define $\Delta \in \text{End}(V)$ by

$$\Delta \triangleq d\delta + \delta d$$

Prove

(a) $\text{Ker}\, d\delta \subseteq \text{Ker}\, \delta$ and $\text{Ker}\, \delta d \subseteq \text{Ker}\, d$.

(b) $\text{Ker}\, \Delta = \text{Ker}\, d \cap \text{Ker}\, \delta$.

(c) $V = \text{Ker}\, \Delta \oplus \text{Im}\, d \oplus \text{Im}\, \delta$ is an orthogonal decomposition.

(d) $\text{Ker}\, d = \text{Ker}\, \Delta \oplus \text{Im}\, d$ is an orthogonal decomposition.

Proof. Because the underlying field is $\mathbb{C}$, we may express d in its **Jordan form**. Because $d^2 = 0$, we know its Jordan blocks all have eigenvalue 0 and have size no greater than 2. Knowing that the matrix representation of δ with respect to the same Jordan basis is just the conjugate transpose of d, we may easily prove all four propositions with computation, except

$$\text{Ker}\, \Delta \perp \text{Im}\, d \quad \text{and} \quad \text{Im}\, d \perp \text{Im}\, \delta \quad \text{and} \quad \text{Im}\, \delta \perp \text{Ker}\, \Delta$$

Direct computation with the hint $\delta = d^\dagger$ and $d^2 = 0$ shows $\text{Im}\, d \perp \text{Im}\, \delta$. Direct computation with hint $\text{Ker}\, \Delta = \text{Ker}\, d \cap \text{Ker}\, \delta$ shows the other two orthogonal relationships. ■

Question 12

Let $V = \mathbb{R}^n$ be the space of column vectors, and M a positive definite symmetric $n \times n$ real matrix. Suppose $A \in M_n(\mathbb{R})$ satisfies

$$MAM^{-1} = A^t$$

Show that there exists some $P \in M_n(\mathbb{R})$ such that

$$P^t MP = I \text{ and } P^{-1}AP \text{ is diagonal}$$

Chapter 3

Archived

3.1 Tensor Algebra

Abstract

In this section, by the term **ring**, we mean a ring with a multiplication identity, and by the term **real algebra**, we mean a real vector space equipped with a vector multiplication compatible with both scalar multiplication and addition. In this definition, for a real algebra A to be a ring, A must be associative. By the term **ideal**, we mean a 2-sided ideal. If we say a multi-linear map $M : V^k \to Z$ is **alternating**, we mean that M maps $(v_1, \dots, v_n)$ to 0 if two arguments coincide.

Given a finite collection $(V_1, \dots, V_n)$ of finite dimensional real vector space, by the term **tensor product of** $V_1, \dots, V_n$, we mean a real vector space usually denoted by $V_1 \otimes \cdots \otimes V_n$ and a multilinear map $\otimes : V_1 \times \cdots \times V_n \to V_1 \otimes \cdots \otimes V_n$ satisfying the universal property: If $B : V_1 \times \cdots \times V_n \to Z$ is a multilinear map, then there exists a unique linear map $\beta : V_1 \otimes \cdots \otimes V_n$ such that

$$B(v_1, \dots, v_n) = \beta(v_1 \otimes \cdots \otimes v_n)$$

In other words, we have the commutative diagram

$$\begin{tikzcd}
V_1 \times \cdots \times V_n \arrow[r, "B"] \arrow[d, "\otimes"'] & U \\
V_1 \otimes \cdots \otimes V_n \arrow[ur, dashed, "\exists! \beta"'] &
\end{tikzcd}$$

This approach indeed define a pair of vector space and multilinear map uniquely up to isomorphism, in the sense of Theorem 3.1.1, where we define the isomorphism between tensor product.

Theorem 3.1.1. (Uniqueness of Tensor product) Given a finite collection $(V_1, \dots, V_n)$ of finite dimensional real vector space, if $V_1 \otimes \cdots \otimes V_n, V_1 \otimes' \cdots \otimes' V_n$ both satisfy the universal property, then there exists an linear isomorphism $T : V_1 \otimes \cdots \otimes V_n \to V_1 \otimes' \cdots \otimes' V_n$ such that

$$T(v_1 \otimes \cdots \otimes v_n) = v_1 \otimes' \cdots \otimes' v_n$$

Proof. Because $V_1 \otimes \cdots \otimes V_n$ satisfies the universal property, there exists a linear map $T : V_1 \otimes \cdots \otimes V_n \to V_1 \otimes' \cdots \otimes' V_n$ such that

$$\otimes' = T \circ \otimes$$

It remains to show T is bijective. Similarly, because $V_1 \otimes' \cdots \otimes' V_n$ satisfies the universal property, there exists a linear map $T' : V_1 \otimes' \cdots \otimes' V_n \to V_1 \otimes \cdots \otimes V_n$ such that

$$\otimes = T' \circ \otimes'$$

Composing the two equations, we have

$$\otimes' = T \circ T' \circ \otimes'$$

It then follows from uniqueness of the induced linear map in universal property that $T \circ T' = \mathbf{id} : V_1 \otimes' \cdots \otimes' V_n \to V_1 \otimes' \cdots \otimes' V_n$. This implies T is indeed bijective. ■

We have shown that tensor products is unique up to isomorphism. A construction further shows that if B_i are bases for V_i, then

$$\{v_1 \otimes \cdots \otimes v_n : v_i \in B_i \text{ for all } 1 \leq i \leq n\} \text{ form a basis for } V_1 \otimes \cdots \otimes V_n$$

Theorem 3.1.2. (Associativity of the Tensor product) Given three finite-dimensional real vector spaces X, Y, Z, there exists a unique linear isomorphism $F : X \otimes Y \otimes Z \to (X \otimes Y) \otimes Z$ that satisfy

$$F(x \otimes y \otimes z) = (x \otimes y) \otimes z$$

Proof. Define $f : X \times Y \times Z \to (X \otimes Y) \otimes Z$ by

$$f(x, y, z) \triangleq (x \otimes y) \otimes z$$

It follows from the universal property that there exists a unique linear map $F : X \otimes Y \otimes Z \to (X \otimes Y) \otimes Z$ such that

$$F(x \otimes y \otimes z) = f(x, y, z) = (x \otimes y) \otimes z$$

It remains to show F is bijective. For all $z \in Z$, define $h_z : X \times Y \to X \otimes Y \otimes Z$ by

$$h_z(x, y) \triangleq x \otimes y \otimes z$$

If follows from the universal property that there exists a unique linear map $H_z : X \otimes Y \to X \otimes Y \otimes Z$ such that

$$H_z(x \otimes y) = h_z(x, y) = x \otimes y \otimes z$$

Define $h : (X \otimes Y) \times Z \to X \otimes Y \otimes Z$ by

$$h(v, z) \triangleq H_z(v)$$

It is clear that h in linear in $(X \otimes Y)$. We now show h is linear in Z, that is

$$H_{c_1 z_1 + z_2} = c_1 H_{z_1} + H_{z_2}$$

By definition,

$$\begin{aligned}(c_1 H_{z_1} + H_{z_2})(x \otimes y) &= c_1 x \otimes y \otimes z_1 + x \otimes y \otimes z_2 \\ &= x \otimes y \otimes (c_1 z_1 + z_2) = h_{c_1 z_1 + z_2}(x, y)\end{aligned}$$

It then follows from the uniqueness part of the universal property that $H_{c_1 z_1 + z_2} = c_1 H_{z_1} + H_{z_2}$. (done)

We have shown h is indeed bilinear. It follows from the universal property that there exists a unique linear map $H : (X \otimes Y) \otimes Z \to X \otimes Y \otimes Z$ such that

$$H((x \otimes y) \otimes z) = h(x \otimes y, z) = H_z(x \otimes y) = x \otimes y \otimes z$$

Let $\otimes : X \times Y \times Z \to X \otimes Y \otimes Z$ denotes the tensor product, we now have

$$\otimes = H \circ F \circ \otimes$$

It then follows from universal property that $H \circ F = \mathbf{id} : X \otimes Y \otimes Z \to X \otimes Y \otimes Z$. This implies F is indeed bijective. (done) ■

Let V be a finite-dimensional real vector space. By its **tensor algebra**, we mean any real associative algebra $T(V)$ with an injective linear map $i : V \to T(V)$ that satisfies the universal property: If A is a real associative algebra and $f : V \to A$ is a linear map, then there exists a unique algebra homomorphism $F : T(V) \to A$ such that the diagram

$$\begin{tikzcd}
V \arrow[r, "i"] \arrow[dr, "f"'] & T(V) \arrow[d, dashed, "F"] \\
& A
\end{tikzcd}$$

commutes. The proof that such definition is indeed unique up to isomorphism is similar to that of Theorem 3.1.1 and thus omitted. We now give the most useful construction.

Let V be finite-dimensional real vector space. We use the notation

$$T^n(V) \triangleq \overbrace{V \otimes \cdots \otimes V}^{n \text{ copies}}$$

and call $T^n(V)$ the **n-th tensor power of** V or the **n-fold tensor product of** V. Define

$$\begin{aligned} T(V) &\triangleq \bigoplus_{n=0}^{\infty} T^n(V) \\ &= \mathbb{R} \oplus V \oplus (V \otimes V) \oplus (V \otimes V \otimes V) \oplus \cdots \end{aligned}$$

and define for all $f, g \in T(V)$ the multiplication

$$(fg)(n) \triangleq \sum_{k=0}^{n} f(k)g(n-k)$$

where

$$\begin{aligned} &\Big(\sum_I a_I v_{I(1)} \otimes \cdots \otimes v_{I(k)} \Big) \Big(\sum_J b_J v_{J(1)} \otimes \cdots \otimes v_{J(l)} \Big) \\ &\triangleq \sum_{I,J} a_I b_J v_{I(1)} \otimes \cdots \otimes v_{I(k)} \otimes v_{J(1)} \otimes \cdots \otimes v_{J(l)} \end{aligned}$$

where $\{v_1, \dots, v_m\}$ is some basis for V, I run through the set of function that maps $\{1, \dots, k\}$ into $\{1, \dots, m\}$ and J run through the set of function that maps $\{1, \dots, l\}$ into $\{1, \dots, m\}$. For example, given two elements

$$(5, 0, v_1 \otimes v_2, 0, 0, \dots) \text{ and } (7, v_3, 0, 0, \dots)$$

of $T(V)$, their product is defined to be

$$(35, 5v_3, 7v_1 \otimes v_2, v_1 \otimes v_2 \otimes v_3, 0, 0, \dots)$$

Tedious effort shows that our multiplication is consistent with abuse of notation in the sense that if $f, g \in T(V)$ is defined by

$$f(k) \triangleq \begin{cases} w_1 \otimes \cdots \otimes w_n & \text{if } k = n \\ 0 & \text{if otherwise} \end{cases} \quad \text{and } g(k) \triangleq \begin{cases} w_{n+1} \otimes \cdots \otimes w_{n+l} & \text{if } k = l \\ 0 & \text{if otherwise} \end{cases}$$

then

$$(fg)(k) = \begin{cases} w_1 \otimes \cdots \otimes w_{n+l} & \text{if } n = k + l \\ 0 & \text{if otherwise} \end{cases}$$

does form an associative algebra with multiplication identity $1 \in \mathbb{R}$. Thus, $T(V)$ is in fact a ring. Let $I(V) \subseteq T(V)$ be the ideal generated by $\{v \otimes v : v \in V\}$. By definition, ideal $I(V)$ is a subgroup of $T(V)$. To see that $I(V)$ is closed under scalar multiplication, observe that for all $t \in \mathbb{R}$ and $x \in T(V)$, the scalar multiplication tx is identical to tx where t is treated as an element of $T(V)$, so it follows from definition of ideal that $I(V)$ is also a vector subspace of $T(V)$. Let $\{v_1, \dots, v_n\}$ be a basis for V, and let S be the set of function that maps $\{1, \dots, n\}$ into $\{1, \dots, k\}$. We know for a fact that

$$T^k(V) = \operatorname{span}\{v_{I(1)} \otimes \cdots \otimes v_{I(k)} : I \in S\}$$

If we define $I^k(V) \triangleq I(V) \cap T^k(V)$, one then have

$$I^k(V) = \operatorname{span}\{v_{I(1)} \otimes \cdots \otimes v_{I(k)} : I(j) = I(j+1) \text{ for some } j\} \tag{3.1}$$

This is proved by showing $I^0(V) \oplus I^1(V) \oplus I^2(V) \oplus \cdots$ is indeed the smallest ideal containing $\{v \otimes v : v \in V\}$. Define an equivalence class on $T(V)$ by

$$x \sim y \overset{\triangle}{\iff} x - y \in I(V)$$

Because ideal form a subgroup, we see that our definition indeed give an equivalence relation. We then can define on the set of equivalence class $T(V) \setminus I(V)$ addition, scalar multiplication and vector multiplication

$$[x] + [y] \triangleq [x + y] \text{ and } [x] \wedge [y] \triangleq [xy] \text{ and } c[x] \triangleq [cx]$$

which is well defined and form an algebra as one can check. We call this algebra $T(V) \setminus I(V)$ the **exterior algebra** $\wedge^*(V)$. Note that if we refer to $v, w \in T^k(V)$ as elements of $\wedge^*(V)$, we mean $[v], [w]$. Immediately, we see that the wedge product is **alternating** in the sense that if $v \in V$, then

$$v \wedge v = 0$$

and is **anti-symmetric** in the sense that if $v, w \in V$, then

$$v \wedge w = -w \wedge v$$

We use the notation

$$\wedge^k(V) \triangleq \left\{ [x] \in \wedge^*(V) : x \in \overbrace{V \otimes \cdots \otimes V}^{k \text{ copies}} \right\}$$

Immediately from Equation 3.1, we see that $\wedge^k(V)$ is the vector space

$$\text{span}\{v_{I(1)} \wedge \cdots \wedge v_{I(k)} : I \in S\}$$

where $\{v_1, \dots, v_n\}$ is a basis for V and S is the set of function that maps $\{1, \dots, k\}$ into $\{1, \dots, n\}$. If we define the vector subspace $I^k(V) \triangleq T^k(V) \cap I(V)$, there exists a natural vector space isomorphism

$$\wedge^k(V) \underset{\text{v.s.}}{\cong} T^k(V) \setminus I^k(V); [x] \leftrightarrow [x]$$

where $T^k(V) \setminus I^k(V)$ is the quotient vector space.

Theorem 3.1.3. (Universal mapping property for alternating k-linear map) For any vector space Z over $\mathbb{R}$ and any alternating k-linear map $f : V^k \to Z$, there is a unique linear map $F : \bigwedge^k V \to Z$ such that the diagram

$$\begin{tikzcd}
\bigwedge^k V \arrow[dr, "F"] & \\
V^k \arrow[u, "\wedge"] \arrow[r, "f"'] & Z
\end{tikzcd}$$

commute, i.e.,

$$F(v_1 \wedge \cdots \wedge v_k) = f(v_1, \dots, v_k) \text{ for all } v_1, \dots, v_k \in V$$

Proof. By universal property of tensor product, there exists unique linear map $h : T^k(V) \to Z$ such that

$$h(v_1 \otimes \cdots \otimes v_k) = f(v_1, \cdots, v_k)$$

Because f is alternating, we see from the characterization of $I^k(V)$ given in Equation 3.1 that h vanishes on $I^k(V)$. We then can induce a linear map

$$F : \wedge^k(V) \cong \frac{T^k(V)}{I^k(V)} \to Z$$

by $F([x]) \triangleq h(x)$. This then give us the desired

$$F(v_1 \wedge \cdots \wedge v_k) = h(v_1 \otimes \cdots \otimes v_k) = f(v_1, \ldots, v_k)$$

Note that F is unique because all such linear map take the same values on $\{v_{I(1)} \wedge \cdots \wedge v_{I(k)} : I \in S\}$ which spans $\wedge^k(V)$. ■

Let $\{w_1, \ldots, w_l\} \subseteq V$ be linear independent. An immediate consequence of the universal mapping property for alternating k-linear map is that one may define alternating multilinear $f : V^l \to \mathbb{R}$ by

$$B(v_1, \ldots, v_l) \triangleq \det(M) \text{ where } v_i = \sum_j M_{i,j} w_j$$

and see that $F : \wedge^l(V) \to \mathbb{R}$ take $w_1 \wedge \cdots \wedge w_l$ to 1. This implies that

$$w_1 \wedge \cdots \wedge w_l \neq 0$$

Theorem 3.1.4. (Anti-symmetry of wedge product) If $\alpha \in \wedge^k(V), \beta \in \wedge^l(V)$, then $\alpha \wedge \beta = (-1)^{kl}(\beta \wedge \alpha)$.

Proof. Let $v_1, \ldots, v_n$ be a basis of V. Let S_k be the space of function that maps $\{1, \ldots, k\}$into $\{1, \ldots, n\}$, S_l be the space of function that maps $\{1, \ldots, l\}$ into $\{1, \ldots, n\}$. We may then write

$$\alpha = \sum_{I \in S_k} a_I(v_{I(1)} \wedge \cdots \wedge v_{I(k)}) \text{ and } \beta = \sum_{J \in S_l} b_J(v_{J(1)} \wedge \cdots \wedge v_{J(l)})$$

and compute

$$
\begin{aligned}
\alpha \wedge \beta &= \sum_{I\in S_k, J\in S_l} a_I b_J (v_{I(1)} \wedge \cdots \wedge v_{I(k)} \wedge v_{J(1)} \wedge \cdots \wedge v_{J(l)}) \\
&= \sum_{I\in S_k, J\in S_l} (-1) a_I b_J (v_{I(1)} \wedge \cdots \wedge v_{J(1)} \wedge v_{I(k)} \cdots \wedge \cdots \wedge v_{J(l)}) \\
&= \sum_{I\in S_k, J\in S_l} (-1)^k a_I b_J (v_{J(1)} \wedge v_{I(1)} \wedge \cdots \wedge v_{I(k)} \wedge v_{J(2)} \wedge \cdots \wedge v_{J(l)}) \\
&= \sum_{I\in S_k, J\in S_l} (-1)^{kl} a_I b_J (v_{J(1)} \wedge \cdots \wedge v_{J(l)} \wedge v_{I(1)} \wedge \cdots \wedge v_{I(k)}) = (-1)^{kl} \beta \wedge \alpha
\end{aligned}
$$

■

Following from Theorem 3.1.4, Equation 3.1 and tedious effort, one can see that if $\{v_1, \ldots, v_n\}$ is a basis for V, then

$$\{v_{i_1} \wedge \cdots \wedge v_{i_k} : 1 \leq i_1 < \cdots < i_k \leq n\}$$

form a basis for $\wedge^k(V)$. If $A : V \to W$ is a linear map, we define linear map $\wedge^k A : \wedge^k(V) \to \wedge^k(W)$ by linear extension of

$$\wedge^k(A)(v_1 \wedge \cdots \wedge v_n) = Av_1 \wedge \cdots \wedge Av_n$$

Note that if $A : V \to V$ and $\dim(V) = n$, then $\wedge^n A : \wedge^n(V) \to \wedge^n(V)$ is given by the determinant since given basis $\{v_1, \ldots, v_n\}$, we have

$$
\begin{aligned}
\wedge^n A(v_1 \wedge \cdots \wedge v_n) &= \Big(\sum_j A_{j,1} v_j\Big) \wedge \cdots \wedge \Big(\sum_j A_{j,n} v_j\Big) \\
&= \sum_{\sigma \in S_n} A_{\sigma(1),1} \cdots A_{\sigma(n),n} v_{\sigma(1)} \wedge \cdots \wedge v_{\sigma(n)} \\
&= \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) A_{\sigma(1),1} \cdots A_{\sigma(n),n} v_1 \wedge \cdots \wedge v_n
\end{aligned}
$$

3.2 Operator Norm

Abstract

This section introduces the concept of the operator norm and proves some fundamental results related operator norm and finite-dimensional normed spaces. For example, we establish results such as a linear operator being bounded if and only if it is continuous and the equivalence of all norms on finite-dimensional vector spaces.

In this section, and particularly in functional analysis, we say a function T between two metric space is a **bounded operator** if T always map bounded set to bounded set. In particular, if T is a linear transformation between two normed space, we say T is a **bounded linear operator**. Now, suppose $\mathcal{X}, \mathcal{Y}$ are two normed space over $\mathbb{R}$ or $\mathbb{C}$. In space $L(\mathcal{X}, \mathcal{Y})$, alternatively, we can define

$$T \text{ is bounded} \quad \overset{\triangle}{\iff} \exists M \in \mathbb{R}, \forall x \in \mathcal{X}, \|Tx\| \leq M\|x\|$$

The proof of equivalency is simple. For $(\longrightarrow)$, observe

$$\|Tx\| = \|x\| \cdot \|T\frac{x}{\|x\|}\| \leq \Big(\sup\{\|Ty\| : \|y\| = 1\}\Big)\|x\|$$

For $(\longleftarrow)$, observe

$$\|Tx - Ty\| = \|T(x-y)\| \leq M\|x - y\|$$

We first show that a linear transformation is continuous if and only if it is bounded.

Theorem 3.2.1. (Liner Operator is Bounded if and only if it is Continuous) Given two normed space $\mathcal{X}, \mathcal{Y}$ over $\mathbb{R}$ or $\mathbb{C}$ and $T \in L(\mathcal{X}, \mathcal{Y})$, we have

$$T \text{ is a bounded operator} \iff T \text{ is continuous on } \mathcal{X}$$

Proof. If T is bounded, we see that T is Lipschitz.

$$\|Tx - Ty\| \leq M\|x - y\|$$

Now, suppose T is linear and continuous at 0. Let ϵ satisfy

$$\sup_{\|y\| \leq \epsilon} \|Ty\| \leq 1$$

Observe that for all $x \in \mathcal{X}$, we have

$$\|Tx\| = \frac{\|x\|}{\epsilon}\left\|T\frac{\epsilon x}{\|x\|}\right\| \leq \frac{\|x\|}{\epsilon}$$

■

Here, we introduce a new terminology, which shall later show its value. Given a set X, we say two metrics d_1, d_2 on X are **equivalent**, and write $d_1 \sim d_2$, if we have

$$\exists m, M \in \mathbb{R}^+, \forall x, y \in X, md_1(x, y) \leq d_2(x, y) \leq Md_1(x, y)$$

Now, given a fixed vector space V, naturally, we say two norms $\|\cdot\|_1, \|\cdot\|_2$ on V are **equivalent** if

$$\exists m, M \in \mathbb{R}^+, \forall x \in X, m\|x\|_1 \leq \|x\|_2 \leq M\|x\|_1$$

We say two metric d_1, d_2 on X are **topologically equivalent** if the topology they induce on X are identical.

A few properties can be immediately spotted.

(a) Our definition of $\sim$ between metrics of a fixed X is an equivalence relation.

(b) Our definition of $\sim$ between norms on a fixed V is an equivalence relation.

(c) Equivalent norms induce equivalent metrics.

(d) Equivalent metrics are topologically equivalent.

We now prove if V is finite-dimensional, then all norms on V are equivalent. This property will later show its value, as used to prove linear map of finite-dimensional domain is always continuous

Theorem 3.2.2. (All Norms on Finite-dimensional space are Equivalent) Suppose V is a finite-dimensional vector space over $\mathbb{R}$ or $\mathbb{C}$. Then

$$\text{all norms on } V \text{ are equivalent}$$

Proof. Let $\{e_1, \ldots, e_n\}$ be a basis of V. Define ∞-norm $\|\cdot\|_\infty$ on V by

$$\left\|\sum \alpha_i e_i\right\|_\infty \triangleq \max |\alpha_i|$$

It is easily checked that $\|\cdot\|_\infty$ is indeed a norm. Fix a norm $\|\cdot\|$ on V. We reduce the problem into

$$\text{finding } m, M \in \mathbb{R}^+ \text{ such that } m\|x\|_\infty \leq \|x\| \leq M\|x\|_\infty$$

We first claim

$$M = \sum \|e_i\| \text{ suffices}$$

Compute

$$\|x\| = \left\| \sum \alpha_i e_i \right\| \leq \sum |\alpha_i| \, \|e_i\| \leq \|x\|_\infty \sum \|e_i\| = M\|x\|_\infty \text{ (done)}$$

Note that reverse triangle inequality give us

$$\Big| \|x\| - \|y\| \Big| \leq \|x - y\| \leq M\|x - y\|_\infty \tag{3.2}$$

Then we can check that

(a) $\|\cdot\| : \Big(V, \|\cdot\|_\infty\Big) \to \mathbb{R}$ is Lipschitz continuous because of Equation 3.2.

(b) $S \triangleq \{y \in V : \|y\|_\infty = 1\}$ is sequentially compact in $\|\cdot\|$ and non-empty.

Now, by EVT, we know $\min_{y \in S} \|y\|$ exists. Note that $\min_{y \in S} \|y\| > 0$, since $0 \notin S$. We claim

$$m = \min_{y \in S} \|y\| \text{ suffices}$$

Fix $x \in V$ and compute

$$m\|x\|_\infty = \|x\|_\infty (\min_{y \in S} \|y\|) \leq \|x\|_\infty \cdot \left\| \frac{x}{\|x\|_\infty} \right\| = \|x\| \text{ (done) (done)}$$

■

Theorem 3.2.3. (Linear map of Finite-dimensional Domain is always Continuous) Given a finite-dimensional normed space $\mathcal{X}$ over $\mathbb{R}$ or $\mathbb{C}$, an arbitrary normed space $\mathcal{Y}$ over $\mathbb{R}$ or $\mathbb{C}$ and a linear transformation $T : \mathcal{X} \to \mathcal{Y}$, we have

$$T \text{ is continuous}$$

Proof. Fix $x \in \mathcal{X}, \epsilon$. We wish

$$\text{to find } \delta \text{ such that } \forall h \in \mathcal{X} : \|h\| \leq \delta, \|T(x+h) - Tx\| \leq \epsilon$$

Let $\{e_1, \dots, e_n\}$ be a basis of $\mathcal{X}$. Note that $\|\sum \alpha_i e_i\|_1 \triangleq \sum |\alpha_i|$ is a norm. Because $\mathcal{X}$ is finite-dimensional, we know $\|\cdot\|$ and $\|\cdot\|_1$ are equivalent. Then, we can fix $M \in \mathbb{R}^+$ such that

$$\|x\|_1 \leq M\|x\| \quad (x \in V)$$

We claim

$$\delta = \frac{\epsilon}{M(\max \|Te_i\|)} \text{ suffices}$$

Fix $\|h\| \leq \delta$ and express $h = \sum \alpha_i e_i$. Compute using linearity of T

$$\begin{aligned}
\|T(x+h) - Tx\| &= \|\sum \alpha_i T e_i\| \\
&\leq \sum |\alpha_i| \, \|Te_i\| \\
&\leq \|h\|_1 (\max \|Te_i\|) \\
&\leq M\|h\|(\max \|Te_i\|) = \epsilon \text{ (done)}
\end{aligned}$$

■

We now see that, because Linear transformation is bounded if and only if it is continuous and Linear map of finite-dimensional domain is always continuous, if $\mathcal{X}$ is finite-dimensional, then all linear map of domain $\mathcal{X}$ are bounded. A counter example to the generalization of this statement is followed.

Example 1 (Differentiation is an Unbounded Linear Operator)

$$\mathcal{X} = \Big(\mathbb{R}[x]|_{[0,1]}, \|\cdot\|_\infty\Big), D(P) \triangleq P'$$

Note that $\{x^n\}_{n\in\mathbb{N}}$ is bounded in $\mathcal{X}$ and $\{D(x^n)\}_{n\in\mathbb{N}}$ is not.

Now, suppose $\mathcal{X}, \mathcal{Y}$ are two fixed normed spaces over $\mathbb{R}$ or $\mathbb{C}$. We can easily check that the set $BL(\mathcal{X}, \mathcal{Y})$ of bounded linear operators from $\mathcal{X}$ to $\mathcal{Y}$ form a vector space over whichever field $\mathcal{Y}$ is over.

Naturally, our definition of boundedness of linear operator derive us a norm on $BL(\mathcal{X}, \mathcal{Y})$, as followed

$$\|T\|_{\text{op}} \triangleq \inf\{M \in \mathbb{R}^+ : \forall x \in \mathcal{X}, \|Tx\| \leq M\|x\|\} \tag{3.3}$$

Before we show that our definition is indeed a norm, we first give some equivalent definitions and prove their equivalency.

Theorem 3.2.4. (Equivalent Definitions of Operator Norm) Given two fixed normed space $\mathcal{X}, \mathcal{Y}$ over $\mathbb{R}$ or $\mathbb{C}$, a bounded linear operator $T : \mathcal{X} \to \mathcal{Y}$, and define $\|T\|_{\text{op}}$ as in Equation 3.3, we have

$$\|T\|_{\text{op}} = \sup_{x\in\mathcal{X}, x\neq 0} \frac{\|Tx\|}{\|x\|}$$

Proof. Define $J \triangleq \{M \in \mathbb{R}^+ : \forall x \in \mathcal{X}, \|Tx\| \leq M\|x\|\}$ and observe

$$J = \{M \in \mathbb{R}^+ : M \geq \frac{\|Tx\|}{\|x\|}, \forall x \neq 0 \in \mathcal{X}\}$$

This let us conclude

$$\sup_{x \in \mathcal{X}, x \neq 0} \frac{\|Tx\|}{\|x\|} = \min J = \|T\|_{\text{op}}$$

∎

It is now easy to see

$$\|T\|_{\text{op}} = \sup_{x \in \mathcal{X}, x \neq 0} \frac{\|Tx\|}{\|x\|} \tag{3.4}$$
$$= \sup_{x \in \mathcal{X}, \|x\|=1} \|Tx\| \tag{3.5}$$

It is not all in vain to introduce the equivalent definitions. See that the verification of $\|\cdot\|_{\text{op}}$ being a norm on $BL(\mathcal{X}, \mathcal{Y})$ become simple by utilizing the equivalent definitions.

(a) For positive-definiteness, fix non-trivial T and fix $x \in \mathcal{X} \setminus N(T)$. Use Equation 3.4 to show $\|T\|_{\text{op}} \geq \frac{\|Tx\|}{\|x\|} > 0$.

(b) For absolute homogenity, use Equation 3.5 and $\|Tcx\| = |c| \cdot \|Tx\|$.

(c) For triangle inequality, use Equation 3.5 and $\|(T_1 + T_2)x\| \leq \|T_1x\| + \|T_2x\|$.

Naturally, and very very importantly, Equation 3.4 give us

$$\|Tx\| \leq \|T\|_{\text{op}} \cdot \|x\| \qquad (x \in \mathcal{X})$$

This inequality will later be the best tool to help analyze the derivatives of functions between Euclidean spaces, and perhaps better, it immediately give us

$$\frac{\|T_1T_2x\|}{\|x\|} \leq \frac{\|T_1\|_{\text{op}} \cdot \|T_2\|_{\text{op}} \cdot \|x\|}{\|x\|} = \|T_1\|_{\text{op}} \cdot \|T_2\|_{\text{op}}$$

Then Equation 3.4 give us

$$\|T_1T_2\|_{\text{op}} \leq \|T_1\|_{\text{op}} \cdot \|T_2\|_{\text{op}}$$

3.3 Cauchy-Schwarz for Positive semi-definite Hermitian form

Sometimes, we do not require $\langle \cdot, \cdot \rangle : V^2 \to \mathbb{F}$ to be positive-definite, but only **positive semi-definite**, i.e. $\langle v, v \rangle \geq 0$ for all $v \in V$. In this case, we say $\langle \cdot, \cdot \rangle : V^2 \to \mathbb{F}$ is a **positive semi-definite Hermitian form**. If we again induce a functional $\|\cdot\| : V \to \mathbb{R}$ from some positive semi-definite Hermitian form using Equation **??**, then the functional $\|\cdot\| : V \to \mathbb{R}$ in general is not positive-definite, and is called a **semi norm**, since, as we later show, it still satisfies triangle inequality and absolute homogeneity.

Theorem 3.3.1. (Basic Property of Positive semi-definite Hermitian form) Given a positive semi-definite Hermitian form $\langle \cdot, \cdot \rangle : V^2 \to \mathbb{F}$ and $x, y \in V$, we have

$$\langle x, x \rangle = 0 \implies \langle x, y \rangle = 0$$

Proof. Assume $\langle x, y \rangle \neq 0$. Fix $t > \frac{\|y\|^2}{2|\langle x,y \rangle|^2}$. Compute

$$\begin{aligned}
\|y - t\langle y, x\rangle x\|^2 &= \|y\|^2 + \|(-t)\langle y, x\rangle x\|^2 + \langle -t\langle y, x\rangle x, y\rangle + \langle y, -t\langle y, x\rangle x\rangle \\
&= \|y\|^2 + t^2 |\langle x, y\rangle|^2 \|x\|^2 - t\langle y, x\rangle\langle x, y\rangle - t\langle x, y\rangle\langle y, x\rangle \\
&= \|y\|^2 - 2t |\langle x, y\rangle|^2 < 0 \text{ CaC}
\end{aligned}$$

∎

Theorem 3.3.2. (Cauchy-Schwarz Inequality) Given a positive semi-definite Hermitian form $\langle \cdot, \cdot \rangle : V^2 \to \mathbb{C}$ on vector space V over $\mathbb{C}$, we have

$$|\langle x, y\rangle| \leq \|x\| \cdot \|y\|$$

Moreover, the equality hold true if x, y are linearly independent. In addition, if $\langle \cdot, \cdot \rangle : V^2 \to \mathbb{F}$ is an inner product, then the equality hold true only if x, y are linearly independent.

Proof. We first prove

$$|\langle x, y\rangle| \leq \|x\| \cdot \|y\| \qquad (x, y \in V)$$

Fix $x, y \in V$. Theorem 3.3.1 tell us $\|x\| = 0 \implies \langle x, y \rangle = 0$. Then we can reduce the problem into proving

$$\frac{|\langle x, y\rangle|^2}{\|x\|^2} \leq \|y\|^2$$

Set $z \triangleq y - \frac{\langle y,x\rangle}{\|x\|^2}x$. We then have

$$\langle z, x\rangle = \langle y - \frac{\langle y,x\rangle}{\|x\|^2}x, x\rangle = \langle y, x\rangle - \frac{\langle y,x\rangle}{\|x\|^2}\langle x, x\rangle = 0$$

Then from $y = z + \frac{\langle y,x\rangle}{\|x\|^2}x$, we can now deduce

$$\begin{aligned}\langle y, y\rangle &= \langle z + \frac{\langle y,x\rangle}{\|x\|^2}x, z + \frac{\langle y,x\rangle}{\|x\|^2}x\rangle \\ &= \langle z, z\rangle + \left|\frac{\langle y,x\rangle}{\langle x,x\rangle}\right|^2 \langle x, x\rangle \\ &= \langle z, z\rangle + \frac{|\langle x,y\rangle|^2}{\langle x,x\rangle}\end{aligned}$$

Because $\langle z, z\rangle \geq 0$, we now have

$$\langle y, y\rangle = \langle z, z\rangle + \frac{|\langle x,y\rangle|^2}{\langle x,x\rangle} \geq \frac{|\langle x,y\rangle|^2}{\langle x,x\rangle} \text{ (done)}$$

The equality hold true if and only if $\langle z, z\rangle = 0$. This explains the other two statements regarding the equality. ■

Now, with Cauchy-Schwarz Inequality, we can check the triangle inequality

$$\begin{aligned}\|x+y\|^2 &= \langle x+y, x+y\rangle \\ &= \langle x,x\rangle + \langle x,y\rangle + \langle y,x\rangle + \langle y,y\rangle \\ &= \langle x,x\rangle + \langle y,y\rangle + 2 \text{ Re } \langle x,y\rangle \\ &\leq \|x\|^2 + \|y\|^2 + 2\,|\langle x,y\rangle| \\ &\leq \|x\|^2 + \|y\|^2 + 2\|x\| \cdot \|y\| = \Big(\|x\| + \|y\|\Big)^2\end{aligned}$$

indeed holds true for space equipped with only positive semi-definite Hermitian form.