Linear Algebra II

陳信睿

January, 2023

Abstract

這篇筆記主要是因爲在預習線性代數二的時候，常常發現很多重要的定理都記不太起來，並且老師在下學期沒有選定指定的參考書，所以我就寫了這份筆記。主要是參考謝銘倫老師的影片 [2]，以及著名的線性代數教科書 [1] 所寫。

內容目前涵蓋了商空間、對偶空間以及內積空間的大部分內容，甚至比 “Linear Algebra”[1] 中還要多東西，像是 Hilbert space。不過我盡量把證明寫的精簡一點，同時我也省去了所有的範例。

I wrote this note because I often found that I could not remember many important theorems when I was studying Linear Algebra II, and my teacher did not choose a reference book for the next semester. The main reference is Professor Ming-Lun Hsieh's video [2], and the famous linear algebra textbook [1].

The content now covers most of the quotient space, dual space, and inner product space, even more than in “Linear Algebra”[1], like Hilbert space. I have tried to keep the proof as concise as possible, and I have also omitted all the examples.

Contents

1 Quotient and dual spaces

1.1 Quotient space

Definition 1 (Quotient space)**.** Let V be a vector space and let W be its subspace. Define an equivalence relation $\sim$ on V such that

$$v_1 \sim v_2 \text{ if } v_1 - v_2 \in W.$$

It is easy to verify that $\sim$ is indeed an equivalence relation on V. For each $v_0 \in V$, define $[v_0] = \{v \in V : v \sim v_0\}$ the equivalence class of v_0. Then, $\{[v] : v \in V\}$ is called the quotient space V/W.

Remark. The quotient space V/W is equipped with a natural vector (linear) structure, namely,

$$\begin{cases} [v_1] + [v_2] = [v_1 + v_2] & \text{, for all } v_1, v_2 \in V \\ c[v_1] = [cv_1] & \text{, for all } v_1 \in V \text{ and } c \in \mathbb{F} \end{cases}.$$

Although it is crucial that we shall check these natural addition and scalar multiplication are "well-defined", we omit here.

Definition 2 (Quotient maps)**.** There is a natural surjective map

$$\begin{aligned} \pi : V &\to V/W \\ v &\mapsto [v] \end{aligned},$$

which is called the quotient map. Moreover, it is a linear transformation.

Remark.

$$\begin{aligned} \ker \pi &= \{v \in V : \pi(v) = [0]\} \\ &= \{v \in V : [v] = [0]\} \\ &= \{v \in V : v - 0 \in W\} \\ &= W \ . \end{aligned}$$

Corollary. It follows from the dimension formula that $\dim_{\mathbb{F}} V/W = \dim_{\mathbb{F}} V - \dim_{\mathbb{F}} W$ whenever V is finite dimensional.

Here we give an alternative proof without using the dimensional formula. Since V has finite dimension, let $\mathcal{B} = \{w_1, w_2, \ldots, w_s\}$ be a basis of W and extend $\mathcal{B}$ to $\mathcal{A} = \{w_1, w_2, \ldots, w_r\}$ a basis of V. We claim that $\{[w_{s+1}], \ldots, [w_s]\}$ is a basis of V/W. To see this, we shall show that:

1. The set $\{[w_{s+1}], \ldots, [w_r]\}$ generates V/W.
   Suppose $[v] \in V/W$. Let $v = \sum_{i=1}^{r} \alpha_i w_i$, then

$$[v] = \left[\sum_{i=s+1}^{r} \alpha_i w_i\right] = \sum_{i=s+1}^{r} \alpha_i [w_i] \quad .$$

2. $\{[w_{s+1}], \ldots, [w_r]\}$ is a linear independent set over $\mathbb{F}$.
   Suppose $\sum_{i=s+1}^{r} \alpha_i \cdot [w_i] = [0]$, for some $\alpha_i \in \mathbb{F}$. Then,

$$\begin{aligned} &\left[\sum_{i=s+1}^{r} \alpha_i w_i\right] = [0] \\ \iff &\sum_{i=s+1}^{r} \alpha_i w_i \in W \\ \iff &\sum_{i=s+1}^{r} \alpha_i w_i = \sum_{j=1}^{s} \beta_j w_j, \text{ for some } \beta_j \in \mathbb{F}. \end{aligned}$$

   We conclude that α_i are all zeros, since $\mathcal{A}$ is a basis of V.

Discussions above show that $\dim_{\mathbb{F}} V/W = r - s = \dim_{\mathbb{F}} V - \dim_{\mathbb{F}} W$. Now, we shall study some properties about the quotient space V/W. The next theorem characterize the quotient space V/W by the following universal property.

Theorem 3. *Let T be a linear transformation from V to U, such that $\ker T$ contains W, namely $W \subset \ker T$. Then, T factors through π uniquely. That is, there exists a unique linear transformation $S : V/W \to U$ such that*

$$T = S \circ \pi.$$

Proof. Define $S : V/W \to U$ by

$$S([v]) = T(v).$$

We first show that S is a well-defined map, namely, if $[v] = [v']$, then $T(v) = T(v')$. Note that $[v] = [v'] \implies v - v' \in W \subset \ker T$, we conclude $T(v) = T(v')$. By definition, S is a linear transformation and $S \circ \pi = T$. The uniqueness of such S follows from the surjectivity of π. $\square$

Remark. The quotient space V/W with the quotient map π is the unique vector space satisfying the theorem. That is, if we are given $\pi' : V \to V'$ satisfying the property: for every linear transformation $T : V \to U$ with $W \subset \ker T$, there exists a unique $S' : V' \to U$ such that $S' \circ \pi' = T$. Then, $V' \simeq V/W$ uniquely.

Proof. From the assumptions, we have

$$\begin{cases} \exists!\ S : V/W \to V', \text{ such that } \pi' = S \circ \pi \\ \exists!\ S' : V' \to V/W, \text{ such that } \pi = S' \circ \pi' \end{cases}.$$

This shows $S \circ S' = \mathrm{Id}_{V'}$; $S' \circ S = \mathrm{Id}_{V/W}$ (using Theorem 3 again.) We conclude $V' \simeq V/W$ uniquely. $\square$

Corollary. Let $T : V \to W$ be a linear transformation. Then,

$$V/\ker T \simeq \mathrm{Im} T.$$

Hence, $\dim_{\mathbb{F}} V/\ker T = \dim_{\mathbb{F}} \mathrm{Im} T$.

Proof. From Theorem 3, we have: there exists a unique $S : V/\ker T \to W$, such that $T = S \circ \pi$. It follows from the surjectivity of π that $\mathrm{Im}S = \mathrm{Im}T$. We claim that S is injective. Note that

$$\begin{aligned}\ker S &= \{[v] \in V/\ker T : S([v]) = 0\} \\ &= \{[v] \in V/\ker T : T(v) = 0\} \\ &= \{[v] \in V/\ker T : v \in \ker T\} \\ &= \{[0]\}\,.\end{aligned}$$

Thus, S is a bijection. This completes the proof. $\square$

Now, let $T : V \to V$ be a linear transformation and let $W \subset V$ be a T-invariant subspace. Then, T induce a linear transformation $\widetilde{T}$ on V/W define by:

$$\begin{aligned}\widetilde{T} : V/W &\to V/W \\ [v] &\mapsto [T(v)]\end{aligned}.$$

This is a well-defined map since

$$\begin{aligned}[v] = [v'] &\implies v - v' \in W \\ &\implies T(v) - T(v') = T(v - v') \in W \\ &\implies [T(v)] = [T(v')].\end{aligned}$$

Now, let $\mathcal{B} = \{v_1, v_2, \ldots, v_s\}$ be a basis of W, and extend it to $\mathcal{A} = \mathcal{B} \sqcup \mathcal{B}'$, a basis of V. We have shown that $[\mathcal{B}'] = \{[v] : v \in \mathcal{B}'\}$ is a basis of V/W. Then, we have

$$[T]_{\mathcal{A}} = \left(\begin{array}{c|c} [T|_W]_{\mathcal{B}} & * \\ \hline 0 & \left[\widetilde{T}\right]_{[\mathcal{B}']} \end{array}\right).$$

We thus have

$$\begin{cases}\mathrm{ch}_T(x) = \mathrm{ch}_{T|_W}(x) \cdot \mathrm{ch}_{\widetilde{T}}(x) \\ \mathrm{m}_T(x) \text{ is divisible by } \mathrm{m}_{T|_W}(x)\end{cases}.$$

Corollary. If T is diagonalizable, then so is $\widetilde{T}$.

The corollary follows from the fact that $\mathrm{m}_T(x)$ is divisible by $\mathrm{m}_{\widetilde{T}}(x)$. We next shall discuss the concept of dual spaces.

1.2 Dual space

Definition 4 (Dual space)**.** Let V be a vector space over $\mathbb{F}$. It is well-known that $L(V, \mathbb{F})$ is a vector space over $\mathbb{F}$. It is called the dual space of V, and its elements are called linear functionals of V. We often write $V^\vee$ to denote the dual space of V.

Recall that:

Given two vector spaces V, W over $\mathbb{F}$. Then we have $L(V, W)$ is a vector space over $\mathbb{F}$ and

$$\dim_{\mathbb{F}} L(V, W) = \dim_{\mathbb{F}} V \cdot \dim_{\mathbb{F}} W.$$

Thus, we conclude that $\dim_{\mathbb{F}} V^\vee = \dim_{\mathbb{F}} V$ if $\dim_{\mathbb{F}} V < \infty$. Here we give an alternative proof.

Theorem 5. *Suppose V is a finite dimensional vector space over $\mathbb{F}$. Then, $\dim_{\mathbb{F}} V^\vee = \dim_{\mathbb{F}} V$.*

Proof. Let $\mathcal{B} = \{v_1, v_2, \ldots, v_n\}$ be a basis of V. Let us consider the following linear functionals:

$$\begin{aligned} v_i^\vee :\ & V \to \mathbb{F} \\ & \sum_{i=1}^n \alpha_i \cdot v_i \mapsto \alpha_i \end{aligned}$$

We claim that $\mathcal{B}^\vee = \{v_1^\vee, v_2^\vee, \ldots, v_n^\vee\}$ is a basis of $V^\vee$. We first show that $\mathcal{B}^\vee$ is linear independent. Suppose there exist $\beta_i \in \mathbb{F}$ such that

$$\sum_{i=1}^n \beta_i v_i^\vee = 0,$$

then

$$\sum_{i=1}^n \beta_i v_i^\vee (v_j) = 0.$$

This shows

$$\beta_i = 0, \text{ for all } i = 1, 2, \ldots, n.$$

Next we show that $\mathcal{B}^\vee$ generate $V^\vee$. Given $\ell \in V^\vee$. Then, from the linearity of ℓ, we have

$$\ell = \sum_{i=1}^n \ell(v_i) \cdot v_i^\vee.$$

We conclude that $\mathcal{B}^\vee$ is a basis of $V^\vee$. □

Remark. The basis $\mathcal{B}^\vee$ is called the dual basis of $\mathcal{B}$.

Given a linear transformation $T : V \to W$, it induces a linear transformation $T^\vee : W^\vee \to V^\vee$ between dual spaces defined by:

$$T^\vee(\ell)(v) := \ell(T(v)), \text{ for } \ell \in W^\vee \text{ and } v \in V.$$

It is easy to verify that $T^\vee$ is a linear transformation.

Theorem 6. *Let V, W be two finite dimensional vector spaces over $\mathbb{F}$. Let $\mathcal{A} = \{v_1, v_2, \ldots, v_n\}$ and $\mathcal{B} = \{w_1, w_2, \ldots, w_m\}$ be bases of V and W, respectively. Given $T : V \to W$. Then,*

$$[T]^{\mathrm{t}}_{\mathcal{A},\mathcal{B}} = [T^\vee]_{\mathcal{B}^\vee,\mathcal{A}^\vee}.$$

Proof. Let $A := [T]_{\mathcal{A},\mathcal{B}} = (a_{ij})_{n\times n}$ and $B := [T^\vee]_{\mathcal{B}^\vee,\mathcal{A}^\vee} = (b_{ij})_{n\times n}$. From the definition, we have

$$\begin{aligned} T(v_j) &= \sum_{i=1}^m a_{ij} w_i \\ T^\vee(w_i^\vee) &= \sum_{j=1}^n b_{ji} v_j^\vee \end{aligned}.$$

Then,

$$b_{ji} = T^\vee(w_i^\vee)(v_j) = w_i^\vee(T(v_j)) = w_i^\vee\left(\sum_{i=1}^{m} a_{ij}w_i\right) = a_{ij}.$$

This proves the theorem. $\square$

Theorem 7. *Let V be a vector space and let $W \subset V$ be a subspace. Then,*

$$(V/W)^\vee \simeq \{\ell \in V^\vee : W \subset \ker \ell\}.$$

Proof. We have known that there is a natural map $\pi : V \twoheadrightarrow V/W$. We claim that $\pi^\vee$ is the isomorphism that bijects $(V/W)^\vee$ and $\{\ell \in V^\vee : W \subset \ker \ell\}$. We first show that $\pi^\vee$ is injective. Suppose $\pi^\vee(\ell) = 0$, for some $\ell \in (V/W)^\vee$. Then,

$$\begin{aligned} & l(\pi(v)) = 0, \text{ for all } v \in V \\ \Longrightarrow\ & \ell([v]) = 0, \text{ for all } v \in V. \end{aligned}$$

This shows the injectivity of $\pi^\vee$. Hence, $(V/W)^\vee \simeq \mathrm{Im}\pi^\vee$. It suffices to show that $\mathrm{Im}\pi^\vee = \{\ell \in V^\vee : W \subset \ker \ell\}$.

1. $\mathrm{Im}\pi^\vee \subset \{\ell \in V^\vee : W \subset \ker \ell\}$.
   For each $S \in (V/W)^\vee$ and $w \in W$, we have

   $$\pi^\vee(S)(w) = S(\pi(w)) = S([w]) = S([0]) = 0.$$

2. $\{\ell \in V^\vee : W \subset \ker \ell\} \subset \mathrm{Im}\pi^\vee$.
   Let $\ell \in V^\vee$ such that $W \subset \ker \ell$. Theorem 3 asserts that there exists a unique $S : V/W \to \mathbb{F}$ such that $\ell = S \circ \pi$. This implies $\pi^\vee(S) = \ell$.

Discussions above complete the proof. $\square$

Corollary. Let $A \in M_{m\times n}(\mathbb{F})$. Then, $\mathrm{rank}A = \mathrm{rank}A^{\mathrm{t}}$.

Proof. Let $V = \mathbb{F}^n$, $W = \mathbb{F}^m$ and let $T : V \to W$ defined by

$$T(v) = A \cdot v.$$

Then it is equivalent to prove

$$\dim \mathrm{Im}T = \dim(\mathrm{Im}T^\vee).$$

By Theorem 7,

$$(W/\mathrm{Im}T)^\vee \simeq \{\ell \in W^\vee : \mathrm{Im}T \subset \ker \ell\} = \{\ell \in W^\vee : T^\vee(\ell) = 0\} = \ker(T^\vee). \tag{1}$$

Thus,

$$\dim W - \dim \mathrm{Im}T = \dim W/\mathrm{Im}T = \dim(W/\mathrm{Im}T)^\vee = \dim W^\vee - \dim \mathrm{Im}(T^\vee).$$

This completes the proof. $\square$

Theorem 8. *Let V and W are two finite vector spaces, and let $T : V \to W$ be a linear transformation. Then,*

1. *T is surjective if and only if $T^\vee$ is injective.*

2. *T is injective if and only if $T^\vee$ is surjective.*

Proof. In the proof of the previous corollary, we have shown in equation 1 that

$$(W/\mathrm{Im}T)^\vee \simeq \ker(T^\vee),$$

this proves the first assertion. Similarly, we have

$$(V/\ker T)^\vee \simeq \{\ell \in V^\vee : \ker T \subset \ker \ell\}. \tag{2}$$

We claim the set on the right hand side is $\mathrm{Im}(T^\vee)$.

1. $\{\ell \in V^\vee : \ker T \subset \ker \ell\} \subset \mathrm{Im}(T^\vee)$.
   Let $\ell \in V^\vee$ such that $\ker T \subset \ker \ell$. It is well-known that there exists a subspace $X \subset W$ such that $W = \mathrm{Im}T \oplus X$. Consider a transformation $s : W \to \mathbb{F}$ defined by:
   $$s(w) = \ell(v),$$
   where $w = T(v) + x$, for some $v \in V$ and $x \in X$. This is a well-defined map, since $\ker T \subset \ker \ell$. Note that s is a linear transformation and $\ell = s \circ T = T^\vee(s)$. This implies $\{\ell \in V^\vee : \ker T \subset \ker \ell\} \subset \mathrm{Im}(T^\vee)$.

2. $\mathrm{Im}(T^\vee) \subset \{\ell \in V^\vee : \ker T \subset \ker \ell\}$.
   Let $\ell \in \mathrm{Im}(T^\vee)$. Then, there exists $s \in W^\vee$ such that $\ell = T^\vee(s) = s \circ T$, thus $\ker T \subset \ker \ell$.

Discussions above with equation 2 show that

$$(V/\ker T)^\vee \simeq \mathrm{Im}(T^\vee),$$

which is equivalent to the second assertion. □

Remark. In the class, the teacher proved with another approach, which use the following property:

> Let V be a finite dimensional vector space, and let $V^{\vee\vee}$ be the dual space of $V^\vee$, then there is a natural identification, that is, there is an isomorphism $\phi : V \to V^{\vee\vee}$ defined by
>
> $$\phi : x \mapsto (\hat{x} : f \mapsto f(x)), \quad f \in V^\vee.$$

Next, we show that why we shall study dual spaces by the following theorem.

Theorem 9. *Let V be a finite dimensional vector space over $\mathbb{F}$. Let $\ell_1, \ell_2, \ldots, \ell_s \in V^\vee$ be linearly independent. Suppose $b_1, b_2, \ldots, b_s \in \mathbb{F}$ and put*

$$\Xi = \{v \in V : \ell_i(v) = b_i, \text{ for all } 1 \le i \le s\}.$$

Then, $\Xi \neq \varnothing$.

Proof. Consider the linear transformation $T : V \to \mathbb{F}^s$ defined by:

$$T : v \mapsto (\ell_1(v), \ell_2(v), \ldots, \ell_s(v)).$$

Then, $\dim(\ker T) = \dim V - s$. Here we omit the details of the proof. □

2 Inner product space

Definition 10 (Inner product)**.** Let V be a vector space over $\mathbb{F}$, where $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$. A function $\langle\cdot,\cdot\rangle : V \times V \to \mathbb{F}$ is called an inner product if the following conditions are satisfied:

1. $\langle x+y, z\rangle = \langle x, z\rangle + \langle y, z\rangle$, for all $x, y, z \in V$.
2. $\langle cx, y\rangle = c \cdot \langle x, y\rangle$, for all $x, y \in V$ and $c \in \mathbb{F}$.
3. $\langle x, y\rangle = \overline{\langle y, x\rangle}$, for all $x, y \in V$.
4. $\langle x, x\rangle \geq 0$, for all $x \in V$ and $\langle x, x\rangle = 0$ if and only if $x = 0$.

We write $(V, \langle\cdot,\cdot\rangle)$ for a vector space V together with an inner product structure $\langle\cdot,\cdot\rangle$. In the following text, $\mathbb{F}$ still stands for $\mathbb{R}$ or $\mathbb{C}$ unless otherwise stated.

We could also define the concept of norm (or length) of a vector $v \in V$.

Definition 11 (Norm)**.** For each $v \in V$, define the norm of v as $\|v\| = \langle v, v\rangle^{1/2}$.

Theorem 12 (Riesz representation Theorem on a finite dimensional space)**.** *Let* $(V, \langle\cdot,\cdot\rangle)$ *be an inner product space. Then,*

$$\begin{aligned} \Phi : V &\to V^\vee \\ v &\mapsto \Phi(v)(x) = \langle x, v\rangle \end{aligned}$$

is an isomorphism.

Proof. We first prove that Φ is injective. Note that

$$\ker \Phi = \{v \in V : \langle x, v\rangle = 0, \text{ for all } x \in V\} = \{0\}.$$

Since V is finite dimensional, we have $\dim_{\mathbb{F}} V = \dim_{\mathbb{F}} V^\vee$, thus Φ is an isomorphism. □

In other words, inner product $\langle\cdot,\cdot\rangle$ identifies V with its dual space $V^\vee$ when V is finite dimensional. We now start study how to represent an inner product structure with a matrix. Suppose V is a finite dimensional vector space, and let $\mathcal{A} = \{v_1, v_2, \ldots, v_n\}$ be a basis of V. For any $x, y \in V$, there exist α_i, β_i such that

$$x = \sum_{i=1}^n \alpha_i \cdot v_i\,; \quad y = \sum_{j=1}^n \beta_j \cdot v_j.$$

Then,

$$\langle x, y\rangle = \left\langle \sum_{i=1}^n \alpha_i \cdot v_i, \sum_{j=1}^n \beta_j \cdot v_j \right\rangle = \sum_{i=1}^n \sum_{j=1}^n \alpha_i \overline{\beta_j} \langle v_i, v_j\rangle.$$

Hence, if we let

$$\Omega = (\langle v_i, v_j\rangle) \in M_n(\mathbb{F}),$$

we have

$$\langle x, y\rangle = \begin{pmatrix} \alpha_1 & \alpha_2 & \ldots & \alpha_n \end{pmatrix} \cdot \Omega \cdot \begin{pmatrix} \overline{\beta_1} \\ \overline{\beta_2} \\ \vdots \\ \overline{\beta_n} \end{pmatrix}.$$

The matrix Ω is called the matrix of $\langle\ ,\ \rangle$ associated with $\mathcal{A}$.

Theorem 13 (change of basis)**.** *Let $\mathcal{B} = \{w_1, \dots, w_n\}$ be another basis of V. Assume that*

$$w_j = \sum_{i=1}^{n} a_{ij} v_i, \text{ for all } 1 \le j \le n.$$

Then,

$$\Omega' = A^{\mathrm{t}} \cdot \Omega \cdot \overline{A},$$

where Ω' is the matrix of $\langle\ ,\ \rangle$ associated with $\mathcal{B}$ and $A = (a_{ij})$.

Proof. Note that

$$\begin{aligned}
\langle w_i, w_j \rangle &= \left\langle \sum_{k=1}^{n} a_{ki} v_k, \sum_{l=1}^{n} a_{lj} v_l \right\rangle \\
&= \sum_{k=1}^{n} \sum_{l=1}^{n} a_{ki} \langle v_k, v_l \rangle \overline{a_{lj}} \\
&= \sum_{k=1}^{n} \sum_{l=1}^{n} {a_{ik}}^{\mathrm{t}} \langle v_k, v_l \rangle \overline{a_{lj}},
\end{aligned}$$

This proves the theorem. $\square$

Next, we shall ask whether we can define an inner product structure on V if we are given a matrix $\Omega \in M_n(\mathbb{F})$ and a basis $\mathcal{A}$ of V. The answer is no. In fact, the matrix can define an inner product structure on finite dimensional V if and only if it is positive definite. The next theorem gives the sufficient condition for a matrix being able to define an inner product.

Theorem 14. *If $\Omega = B \cdot B^*$ for some $B \in M_n(\mathbb{F})$ with $\det B \neq 0$, then $\langle\ ,\ \rangle_{\Omega,\mathcal{A}}$ is an inner product for any choice of $\mathcal{A}$.*

Proof. Let $\mathcal{A} = \{v_1, v_2, \dots, v_n\}$ be an arbitrary basis of V. It suffices to show the inner product defined by Ω satisfies the fourth axiom of Definition 10. If $x \in V$, then

$$x = \sum_{i=1}^{n} \alpha_i \cdot v_i, \text{ for some } \alpha_i \in \mathbb{F}.$$

We have

$$\begin{aligned}
\langle x, x \rangle_{\Omega,\mathcal{A}} &:= \begin{pmatrix} \alpha_1 & \alpha_2 & \dots & \alpha_n \end{pmatrix} \cdot \Omega \cdot \begin{pmatrix} \overline{\alpha_1} \\ \overline{\alpha_2} \\ \vdots \\ \overline{\alpha_n} \end{pmatrix} \\
&= \begin{pmatrix} \alpha_1 & \alpha_2 & \dots & \alpha_n \end{pmatrix} \cdot B \cdot B^* \cdot \begin{pmatrix} \overline{\alpha_1} \\ \overline{\alpha_2} \\ \vdots \\ \overline{\alpha_n} \end{pmatrix} \\
&= (yB) \cdot (yB)^*,
\end{aligned}$$

where $y = \begin{pmatrix}\alpha_1 & \alpha_2 & \dots & \alpha_n\end{pmatrix}$ is a row vector. Write $yB = \begin{pmatrix}\beta_1 & \beta_2 & \dots & \beta_n\end{pmatrix}$. We get

$$\langle x, x\rangle_{\Omega,\mathcal{A}} = \begin{pmatrix}\beta_1 & \beta_2 & \dots & \beta_n\end{pmatrix} \cdot \begin{pmatrix}\overline{\beta_1} \\ \overline{\beta_2} \\ \vdots \\ \overline{\beta_n}\end{pmatrix} = \sum_{i=1}^{n} |\beta_i|^2 \geq 0,$$

and $\langle x, x\rangle_{\Omega,\mathcal{A}} = 0$ if and only if $y = 0$. From the assumption that $\det B \neq 0$, it follows $x = 0$ if $\langle x, x\rangle = 0$. □

Definition 15 (Hermitian and positive definite matrix)**.** Let $\Omega \in M_n(\mathbb{F})$. Then,

1. Ω is said to be Hermitian if $\Omega^* = \Omega$.

2. Ω is said to be positive definite if Ω is Hermitian and

$$x \cdot \Omega \cdot x^* > 0, \text{ for all row vector } x \in \mathbb{F}^n \backslash \{0\}.$$

Remark. Let $\Omega \in M_n(\mathbb{F})$. Define a function $\langle \cdot, \cdot \rangle$ of two variables on the vector space $V = \mathbb{F}^n$ by

$$\langle x, y\rangle = x \cdot \Omega \cdot y^*, \text{ where } x \text{ and } y \text{ are row vectors.}$$

Then, $\langle\ ,\ \rangle$ is an inner product on V if and only if Ω is positive definite.

2.1 Orthogonal projection

Definition 16 (Perpendicular)**.** Let $(V, \langle\ ,\ \rangle)$ be an inner product space. Then, we say a vector v is perpendicular to w if

$$\langle v, w\rangle = 0.$$

We often write $v \perp w$ to indicate two vectors are perpendicular to each other.

Note that the Pythagorean theorem holds under this definition:

$$\boxed{\text{If } \langle v, w\rangle = 0, \text{ then } \|v + w\|^2 = \|v\|^2 + \|w\|^2.}$$

Now, we can define orthogonal projection of x to y.

Definition 17 (Orthogonal projection)**.** Given two vectors $x, y \in (V, \langle\ ,\ \rangle)$ $(y \neq 0)$. $\mathrm{Proj}_y(x)$ is the vector satisfying the following two conditions:

1. $\mathrm{Proj}_y(x)$ is parallel to y.

2. $x - \mathrm{Proj}_y(x) \perp y$.

From this definition, we can assume that $\mathrm{Proj}_y(x) = \alpha \cdot y$, for some $\alpha \in \mathbb{F}$. Since $x - \mathrm{Proj}_y(x) \perp y$, we have

$$\langle x - \alpha \cdot y, y\rangle = 0 \iff \alpha = \frac{\langle x, y\rangle}{\langle y, y\rangle}.$$

We conclude that

$$\boxed{\operatorname{Proj}_y(x) = \frac{\langle x, y\rangle}{\|y\|^2} \cdot y.}$$

Lemma 1. Let $x, y \in (V, \langle\ ,\ \rangle)$ $(y \neq 0)$. Then,

$$\left\|\operatorname{Proj}_y(x)\right\| \leq \|x\|.$$

Moreover, the equality holds if and only if x is parallel to y.

Proof. It follows from the Pythagorean theorem. □

Corollary. $|\langle x, y\rangle| \leq \|x\| \, \|y\|$, holds for all $x, y \in V$.

It immediate follows from Lemma 1. This inequality is known as "Cauchy's inequality".

Corollary. $\|x + y\| \leq \|x\| + \|y\|$, holds for all $x, y \in V$.

Proof. It is equivalent to prove $\|x + y\|^2 \leq (\|x\| + \|y\|)^2$.

$$\begin{aligned}
& \|x+y\|^2 \leq (\|x\| + \|y\|)^2 \\
\iff \quad & \langle x+y, x+y\rangle \leq \|x\|^2 + 2\|x\| \cdot \|y\| + \|y\|^2 \\
\iff \quad & \|x\|^2 + \langle x, y\rangle + \langle y, x\rangle + \|y\|^2 \leq \|x\|^2 + 2\|x\| \cdot \|y\| + \|y\|^2 \\
\iff \quad & \Re\langle x, y\rangle \leq \|x\| \cdot \|y\|.
\end{aligned}$$

Note that $\Re\langle x, y\rangle \leq |\langle x, y\rangle| \leq \|x\| \cdot \|y\|$. This proves the corollary. □

In general, if we were given a subspace $W \subset V$, we can discuss about $\operatorname{Proj}_W(x)$, the orthogonal projection of x to W.

Definition 18 (Generalization of orthogonal projection)**.** Let W be a subspace of V and let x be a vector in V. Then, $\operatorname{Proj}_W(x)$ is the vector satisfying the following two conditions:

1. $\operatorname{Proj}_W(x) \in W$.
2. $x - \operatorname{Proj}_W(x) \perp W$. That is, $x - \operatorname{Proj}_W(x)$ is perpendicular to any vectors in W.

The existence of $\operatorname{Proj}_W(x)$ in a finite dimensional vector space V follows from the following theorem.

Theorem 19. *Let V be a finite dimensional inner product space and let W be a subspace of V. Define $W^\perp$ as*

$$W^\perp := \{v \in V : \langle v, w\rangle = 0, \ \textit{for all } w \in W\}.$$

Then, $W^\perp$ is a subspace. Moreover, $V = W \oplus W^\perp$.

Proof. It is easy to see that $W^\perp$ is a subspace of V. Recall Theorem 12, we have an isomorphism:

$$\begin{aligned} V &\simeq V^\vee \\ v &\mapsto l_v(x) = \langle x, v \rangle . \end{aligned}$$

Note that the image of $W^\perp$ under this map is

$$\{l \in V^\vee : W \subset \ker l\} .$$

By Theorem 7, we have

$$W^\perp \simeq (V/W)^\vee .$$

Thus,

$$\begin{aligned} \dim_{\mathbb{F}} V &= \dim_{\mathbb{F}} W + (\dim_{\mathbb{F}} V - \dim_{\mathbb{F}} W) \\ &= \dim_{\mathbb{F}} W + \dim_{\mathbb{F}} V/W \\ &= \dim_{\mathbb{F}} W + \dim_{\mathbb{F}} W^\perp. \end{aligned}$$

We claim that $W \cap W^\perp = \{0\}$. Suppose $x \in W \cap W^\perp$, then $\langle x, x \rangle = 0$. This shows that x must be 0. We conclude that

$$V = W \oplus W^\perp.$$

□

If we are given a subspace $W \subset V$ and a vector x, then according to Theorem 19, there exist unique vectors $w_x \in W$, $w_x' \in W^\perp$ such that

$$x = w_x + w_x'.$$

We define $\mathrm{Proj}_w(x) := w_x$. We now discuss a new idea of (external) direct sum of vector spaces.

Definition 20 (Direct sum)**.** Let V_1, V_2 be two vector spaces. Define

$$V_1 \oplus V_2 := \{(v_1, v_2) \in V_1 \times V_2\} .$$

This space has a natural linear structure:

$$\begin{aligned} (v_1, v_2) + (v_1', v_2') &:= (v_1 + v_1', v_2 + v_2') \\ c(v_1, v_2) &:= (c \cdot v_1, c \cdot v_2) \end{aligned}$$

We shall say $V_1 \oplus V_2$ is the external direct sum of V_1 and V_2.

We can check that:

If W_1, W_2 are two subspaces of V, such that $W_1 \cap W_2 = \{0\}$. Then,

$$W_1 \oplus_{\text{in}} W_2 \simeq W_1 \oplus_{\text{out}} W_2,$$

where $\oplus_{\text{in}}$ is the original (internal) direct sum.

2.2 Orthonormal basis and Gram-Schimdt process

Definition 21 (Orthonormal basis)**.** A set of vectors $\{v_\alpha : \alpha \in \Lambda\}$ is an orthonormal set if $\langle v_\alpha, v_\beta \rangle = 0$ whenever $\alpha \neq \beta$, and $\|v_\alpha\| = 1$ for all $\alpha \in \Lambda$. An orthonormal basis is an orthonormal set which is a basis.

Lemma 2. If $\{v_1, v_2, \ldots, v_r\}$ is an orthonormal set, then it is linearly independent.

Proof. Suppose there exist $\alpha_i \in \mathbb{F}$ such that

$$\sum_{i=1}^{r} \alpha_i \cdot v_i = 0.$$

Then,

$$0 = \langle 0, v_i \rangle = \left\langle \sum_{i=1}^{r} \alpha_i \cdot v_i, v_i \right\rangle = \alpha_i.$$

This completes the proof. □

Remark.

1. If $\dim_{\mathbb{F}} V < \infty$, then any orthonormal set of cardinality equal to n is an orthonormal basis.

2. Let $\mathcal{A}$ be an orthonormal basis. Then, $\Omega = I_n$, where Ω is the matrix of $\langle\ ,\ \rangle$ associated with $\mathcal{A}$.

The existence of orthonormal bases in a finite dimensional inner product space follows from the next theorem. The technique to find such a basis is known as Gram-Schmidt process.

Theorem 22 (Gram-Schmidt process)**.** *Suppose $\{v_1, v_2, \ldots, v_r\}$ is linearly independent. Then, there exists an orthonormal set $\{w_1, w_2, \ldots, w_r\}$ such that*

$$\operatorname{span}_{\mathbb{F}}\{w_1, w_2, \ldots, w_r\} = \operatorname{span}_{\mathbb{F}}\{v_1, v_2, \ldots, v_r\}.$$

Proof. Define u_i and w_i recursively as:

$$\begin{aligned}
u_1 &= v_1 & w_1 &= \frac{u_1}{\|u_1\|} \\
u_2 &= v_2 - \langle v_2, w_1 \rangle \cdot w_1 & w_2 &= \frac{u_2}{\|u_2\|} \\
u_3 &= v_3 - \langle v_3, w_2 \rangle \cdot w_2 - \langle v_3, w_1 \rangle \cdot w_1 & w_3 &= \frac{u_3}{\|u_3\|} \\
&\vdots & &\vdots \\
u_k &= v_k - \sum_{i=1}^{k} \langle v_k, w_i \rangle \cdot w_i & w_k &= \frac{u_k}{\|u_k\|} \\
&\vdots & &\vdots
\end{aligned}$$

We claim that $\operatorname{span}_{\mathbb{F}}\{v_1, \ldots, v_k\} = \operatorname{span}_{\mathbb{F}}\{w_1, \ldots, w_k\}$ and $\{w_1, \ldots, w_k\}$ is an orthonormal set, for each $1 \leq k \leq r$. It is trivial when $k = 1$. Suppose this assertion is true for some $k = m < r$, then $\langle u_{m+1}, w_i \rangle = \langle v_{m+1}, w_i \rangle - \langle v_{m+1}, w_i \rangle = 0$ for $i \leq m$. Also, $v_{m+1} \notin \operatorname{span}_{\mathbb{F}}\{w_1, \ldots, w_m\} = \operatorname{span}_{\mathbb{F}}\{v_1, \ldots, v_m\}$, since $\{v_1, v_2, \ldots, v_r\}$ is linearly independent. We thus have $u_{k+1} \neq 0$, this completes the proof by mathematical induction on k. □

Corollary.

1. If $(V, \langle\ ,\ \rangle)$ is a finite dimensional inner product space over $\mathbb{F}$, then an orthonormal basis exists.

2. Let Ω be a positive definite matrix. From the remark of Definition 15, Ω defines an inner product on $V = \mathbb{F}^n$. Let P be an invertible matrix such that $Pe_i = w_i$, where $\{e_1, \ldots, e_n\}$ is the standard basis of V and $\{w_1, \ldots, x_n\}$ is one orthonormal basis of V with respect to the inner product defined by Ω. Then, Theorem 13 asserts
$$I_n = P^{\mathrm{t}} \cdot \Omega \cdot \overline{P} \implies \Omega = {P^{-1}}^{\mathrm{t}} \cdot \overline{P^{-1}}.$$
Let $Q = {P^{-1}}^{\mathrm{t}}$, then we conclude
$$\Omega = Q \cdot Q^*.$$

For each positive definite matrix $\Omega \in M_n(\mathbb{F})$, there is an invertible matrix $Q \in M_n(\mathbb{F})$ such that $\Omega = Q \cdot Q^*$.

Recall that in Theorem 19 we have shown the existence of $\mathrm{Proj}_W(x)$ when W is a subspace of finite dimensional vector space V. In fact, we can derive the same result but using a weaker condition.

Theorem 23 (orthogonal projection revisited)**.** *Let $(V, \langle\ ,\ \rangle)$ be an inner product space. (It could be infinite dimensional.) Let $W \subset V$ be a subspace with finite dimension. Then,* $\mathrm{Proj}_W(x)$ *exists uniquely. In fact,*
$$\mathrm{Proj}_W(x) = \sum_{i=1}^{n} \langle x, w_i \rangle \cdot w_i,$$
where $\{w_1, w_2, \ldots, w_n\}$ is an orthogonal basis of W.

Proof. We first show that $\langle x - \mathrm{Proj}_W(x), w \rangle = 0$, for all $w \in W$. Note that
$$\langle x - \mathrm{Proj}_W(x), w_i \rangle = \langle x, w_i \rangle - \langle x, w_i \rangle = 0,$$
for all $1 \le i \le n$. It remains to show $\mathrm{Proj}_W(x)$ is unique. Let $y \in W$ such that $x - y \in W^\perp$, then
$$\begin{aligned} \|\mathrm{Proj}_W(x) - y\|^2 &= \langle \mathrm{Proj}_W(x) - y, \mathrm{Proj}_W(x) - y \rangle \\ &= \langle \mathrm{Proj}_W(x) - x + x - y, \mathrm{Proj}_W(x) - y \rangle \\ &= \langle \mathrm{Proj}_W(x) - x, \mathrm{Proj}_W(x) - y \rangle + \langle x - y, \mathrm{Proj}_W(x) - y \rangle \\ &= 0 + 0 = 0\,. \end{aligned}$$
□

We now generalize the idea of orthogonal projection to the case when the subspace W is not given.

Definition 24 (Projection)**.** Let V be an inner product space over $\mathbb{F}$, and let $T : V \to V$ be a linear transformation.

1. We say T is a projection if $T^2 = T$.

2. We say T is an orthogonal projection if $T^2 = T$ and $(\mathrm{Im} T)^\perp = \ker T$.

Remark. Let $T : V \to V$ be an orthogonal projection defined as above. Then, $T(v) = \mathrm{Proj}_W(v)$, where $W := \mathrm{Im} T$.

2.3 Hilbert space

In the previous text, lots of properties of inner product spaces only hold when the space is finite dimensional. This subsection we shall introduce a kind of inner product space that act like a finite dimensional inner product space.

Definition 25 (Hilbert space)**.** Let $(V, \langle\ ,\ \rangle)$ be an inner product space. The norm $\|\cdot\|$ induces a metric d on V. V is said to be a Hilbert space, if (V, d) is a complete metric space in the sense that every Cauchy sequence converges. A subspace $W \subset V$ is closed if W is a Hilbert subspace.

Remark. In analysis, "closedness" of a subspace W means that every convergent sequence in W converges to a point in W. This definition coincides the above definition.

Theorem 26 (Existence of orthogonal projection)**.** *Let $(V, \langle\ ,\ \rangle)$ be a Hilbert space and let $W \subset V$ be a closed subset. Then,* $\mathrm{Proj}_W(x)$ *exists uniquely.*

Proof. Let $d := \inf_{w \in W} \|w - x\|$. We claim that there exist a vector $y_0 \in W$ such that $\|y_0 - x\| = d$. By the definition of infimum, there exist y_n such that

$$d \leq \|y_n - x\| < d + \frac{1}{n}.$$

We first show that (y_n) is a Cauchy sequence. Given $\epsilon > 0$. Let $N \in \mathbb{N}$ large enough so that

$$\frac{8d}{N} + \frac{4}{N^2} < \epsilon.$$

By the parallelogram law, we have

$$\begin{aligned}
\|y_n - y_m\|^2 &= 2(\|y_n - x\|^2 + \|y_m - x\|^2) - \|y_n + y_m - 2x\|^2 \\
&< 2\left(\left(d + \frac{1}{n}\right)^2 + \left(d + \frac{1}{m}\right)\right) - 4\left\|\frac{y_n + y_m}{2} - x\right\|^2 \\
&< 4\left(d + \frac{1}{N}\right)^2 - 4d^2 = \frac{8d}{N} + \frac{4}{N^2} < \epsilon,
\end{aligned}$$

where $n, m \geq N$. Hence, (y_n) is a Cauchy sequence. Suppose $y_n \to y_0$, then $\|y_0 - x\| = d$. We now show that $p = x - y_0 \in W^\perp$. Let us introduce two parameters $t \in \mathbb{F}$ and $w \in W$, then we have

$$\begin{aligned}
& \|p - t \cdot w\|^2 = \|x - y_0 - t \cdot w\|^2 \geq d^2 \\
\Longrightarrow\quad & \|p\|^2 + t^2 \cdot \|w\|^2 - 2\Re\big(\bar{t} \cdot \langle p, w \rangle\big) \geq d^2 \qquad (3) \\
\Longrightarrow\quad & t^2 \cdot \|w\|^2 - 2\Re\big(\bar{t} \cdot \langle p, w \rangle\big) \geq 0.
\end{aligned}$$

If $\langle p, w \rangle \neq 0$, then $\langle p, w \rangle = r \cdot \exp(i\theta)$ for some $r > 0$. We plug in $t = \epsilon \cdot \exp(i\theta)$ to (3), for small enough $\epsilon > 0$. Then,

$$\epsilon^2 \|w\|^2 \geq 2 \cdot \Re(\epsilon r),$$

which fail to be true when ϵ is small enough. Therefore, $y_0 = \lim y_n = \mathrm{Proj}_W(x)$. ☐

Next, we introduce the concept of bounded linear functional.

Definition 27 (Bounded linear functional)**.** Let $(V, \langle\ ,\ \rangle)$ be a Hilbert space over $\mathbb{F}$. A linear functional $l : V \to \mathbb{F}$ is said to be bounded if there exists $M > 0$ such that

$$|l(v)| \leq M \cdot \|v\|,$$

for all $v \in V$. The set of all bounded linear functional on V is denoted by $V^{\vee}_{\text{bdd}}$. In fact, we can similarly define the concept of bounded linear transformation.

Remark.

1. Any bounded linear functional is a continuous function, with respect to the norm of V and metric on $\mathbb{F}$.

2. Any finite dimensional inner product space V is a Hilbert space, moreover, $V^{\vee}_{\text{bdd}} = V^{\vee}$.

Theorem 28 (Riesz representation theorem)**.** *Let $(V, \langle\ ,\ \rangle)$ be a Hilbert space, and let $l \in V^{\vee}_{\text{bdd}}$ be a bounded linear functional, then there exist $y \in V$, such that*

$$l(x) = \langle x, y\rangle,$$

for all $x \in V$.

Proof. Let l be a bounded linear functional. Then, $N = \ker l$ is a closed subspace of V. (Recall that the preimage under a continuous function of a closed set is closed.) If N is V, then $l = 0$, and we can take $y = 0$. Now, we assume that $N \subsetneq V$, it follows from Theorem 26 that there exists $v \in N^{\perp}$. (Hence $l(v) \neq 0$.) Consider a function $\alpha(x) = l(x)/l(v)$, for all $x \in V$. Then,

$$\begin{aligned}
&l(x) = \alpha(x) \cdot l(v) \\
\Longrightarrow\ &l(x - \alpha \cdot v) = 0 \\
\Longrightarrow\ &x - \alpha \cdot v \in N \\
\Longrightarrow\ &\langle x - \alpha \cdot v, v\rangle = 0 \\
\Longrightarrow\ &\langle x, v\rangle = \alpha \cdot \langle v, v\rangle \\
\Longrightarrow\ &l(x) = \langle x, y\rangle, \text{ where } y = \frac{\overline{l(v)}}{\|v\|^2} \cdot v.
\end{aligned}$$

□

2.4 Adjoint linear transformation

Definition 29 (Adjoint linear transformation)**.** Let $(V, \langle\ ,\ \rangle)$ and $(W, \langle\ ,\ \rangle)$ be two inner product spaces over $\mathbb{F}$ and let $T : V \to W$ be a linear transformation. We define the adjoint of T is the transformation $T^* : W \to V$ such that:

$$\langle T^*(w), v\rangle = \langle v, T(w)\rangle,$$

for all $v \in V$ and $w \in W$.

We now show that T^* exists uniquely if both V and W are finite dimensional.

Theorem 30. *Let V and W be two finite dimensional inner product spaces and let $T : V \to W$ be a linear transformation. Then, T^* exists uniquely.*

Proof. By Theorem 22, there exist orthonormal bases of V and W, say $\mathcal{A} = \{v_1, \ldots, v_n\}$ and $\mathcal{B} = \{w_1, \ldots, w_m\}$, respectively. Let $[T]_{\mathcal{A},\mathcal{B}} = A = (a_{ij})_{m\times n}$. We now assume T^* exists, and let $[T^*]_{\mathcal{B},\mathcal{A}} = (b_{ij})_{n\times m}$. Then,

$$\begin{aligned} & \langle T^*(w_i), v_j\rangle = \langle w_i, T(v_j)\rangle \\ \Longrightarrow \quad & \left\langle \sum_{k=1}^{n} b_{ki}\cdot v_k, v_j \right\rangle = \left\langle w_i, \sum_{l=1}^{m} a_{lj}\cdot w_l \right\rangle \\ \Longrightarrow \quad & b_{ji} = \overline{a_{ij}}. \end{aligned}$$

This shows the uniqueness of T^*. In fact, this also shows the existence of T^*, since we can define:

$$\begin{aligned} T^* &: W \to V \\ [w]_{\mathcal{B}} &\mapsto A^*\cdot [w]_{\mathcal{B}}, \end{aligned}$$

where $[w]_{\mathcal{B}}$ denote the coordinate vector of w with respect to the basis $\mathcal{B}$. The calculations above implies T^* meets the condition of adjoint linear transformation. □

However, the adjoint of an operator is not always exist, especially in infinite dimensional inner product space. The next theorem asserts that some operators on Hilbert space has an adjoint. We shall first introduce the concept of bounded linear transformation.

Definition 31 (Bounded linear transformation)**.** Let $T : V \to W$ be a linear transformation between two normed space. Then T is said to be bounded if there exists $M > 0$ such that

$$\|T(v)\|_W \le M\cdot \|v\|_V\,, \text{ for all } v \in V.$$

Theorem 32 (existence of adjoint operators on Hilbert space)**.** *Let V be a Hilbert space. (Recall Definition 25.) Let $T : V \to V$ be a bounded linear operator. Then, T^*, the adjoint of T, exists.*

Proof. This is a corollary of the "Riesz representation theorem". For each $x \in V$, consider linear functionals:

$$l_{T,x}(y) = \langle T(y), x\rangle\,, \quad y \in V.$$

It is easy to check that $l_{T,x}$ is linear. We claim that if T is bounded then $l_{T,x}$ is bounded. Note that

$$|l_{T,x}(y)| = |\langle T(y), x\rangle| \le \|T(y)\|\,\|x\| \le M\,\|y\|\,\|x\|\,.$$

Thus, $l_{T,x}$ is bounded. It follows from Theorem 28 that there exists a unique $z \in V$ such that

$$l_{T,x}(y) = \langle y, z\rangle = \langle T(y), x\rangle\,, \text{ for all } y \in V$$

We define $T^*(x) := z$. It is easy to verify that T^* is a linear transformation. □

Theorem 33. *Let V, W be inner product spaces over $\mathbb{F}$, and let T_1, T_2 and T be linear transformations from V to W. Suppose T_1^*, T_2^* and T^* exist. Then, the following properties hold:*

1. $(T_1 + T_2)^* = T_1^* + T_2^*$.
2. $(\alpha\cdot T)^* = \overline{\alpha}\cdot T^*$, *for* $\alpha \in \mathbb{F}$.
3. *Let U be an inner product space and let $S : W \to U$ be a linear transformation with the adjoint exists. Then, $(S\circ T)^* = T^*\circ S^*$.*

4. $T^{**} = T$.

The proof is very straightforward, so we omit it.

Theorem 34. *Let $T : V \to W$ be a linear transformation between two "finite dimensional" inner product spaces. Then,*

1. $(\mathrm{Im}T)^{\perp} = \ker(T^*)$.
2. $(\ker T)^{\perp} = \mathrm{Im}(T^*)$.

Proof. To show the first assertion, suppose $w \in (\mathrm{Im}T)^{\perp}$, namely,

$$\begin{aligned} & \langle w, T(v)\rangle = 0, \text{ for all } v \in V. \\ \iff & \langle T^*(w), v\rangle = 0, \text{ for all } v \in V. \\ \iff & T^*(w) = 0. \\ \iff & w \in \ker(T^*). \end{aligned}$$

Similarly, for the second assertion. we assume that $v \in \mathrm{Im}(T^*)$, then $v = T^*(w)$ for some $w \in W$. Note that

$$\langle v, x\rangle = \langle T^*(w), x\rangle = \langle w, T(x)\rangle = 0, \text{ for all } x \in \ker T.$$

Thus, we conclude that $\mathrm{Im}(T^*) \subset (\ker T)^{\perp}$. By the dimensional formulas, we get $\mathrm{Im}(T^*) = (\ker T)^{\perp}$. □

Definition 35 (Unitary linear transformation (operator))**.** Let $T : V \to W$ be a linear transformation between two inner product spaces (probably infinite dimensional). T is called unitary if

$$\langle T(v_1), T(v_2)\rangle = \langle v_1, v_2\rangle,$$

for all $v_1, v_2 \in V$.

The next theorem gives a characterization of unitary operators.

Theorem 36. *Given a linear transformation $T : V \to W$ between two finite dimensional inner product spaces. Then the following statements are equivalent:*

1. *T is unitary.*
2. *$\|T(v)\| = \|v\|$, for all $v \in V$.*
3. *$T^* \circ T = \mathrm{Id}_V$.*
4. *T sends the orthonormal basis to an orthonormal set.*

Proof.
(1) $\Longrightarrow$ (2): Obvious.
(2) $\Longrightarrow$ (1): Consider $\|T(x+y)\|^2 = \|x+y\|^2$.

$$\begin{aligned} & \langle T(x), T(y)\rangle + \langle T(y), T(x)\rangle = \langle x, y\rangle + \langle y, x\rangle \\ \Longrightarrow\ & \Re(\langle T(x), T(y)\rangle) = \Re(\langle x, y\rangle). \end{aligned} \tag{4}$$

If $\mathbb{F} = \mathbb{R}$, then (4) shows that $\langle T(x), T(y)\rangle = \langle x, y\rangle$. If $\mathbb{F} = \mathbb{C}$, then plugging in $y \mapsto i \cdot y$ to equation (4) gives

$$\Re((-i) \cdot \langle T(x), T(y)\rangle) = \Re((-i) \cdot \langle x, y\rangle).$$

Together with equation 4 indicate that T is unitary.
(3) $\iff$ (1): T is unitary if and only if

$$\begin{aligned} & \langle T(x), T(y)\rangle = \langle x, y\rangle \text{ , for all } x, y \in V \\ \iff & \langle T^*T(x), y\rangle = \langle x, y\rangle \text{ , for all } x, y \in V \\ \iff & \langle (T^*T - \mathrm{Id}_V)(x), y\rangle = 0, \text{ for all } x, y \in V \\ \iff & (T^*T - \mathrm{Id}_V) \equiv 0. \end{aligned}$$

(1) $\iff$ (4): Let $\mathcal{A} = \{v_1, v_2, \ldots, v_n\}$ be an orthonormal basis of V. Then,

$$\langle T(v_i), T(v_j)\rangle = \langle v_i, v_j\rangle = \begin{cases} 1, & \text{if } i \neq j \\ 0, & \text{if } i = j \end{cases}.$$

Thus, $T(\mathcal{A}) = \{T(v_1), T(v_2), \ldots, T(v_n)\}$ is an orthonormal set.

(4) $\iff$ (1): Let $x, y \in V$ be two arbitrary vector in V. Let $\mathcal{A} = \{v_1, v_2, \ldots, v_n\}$ be an orthonormal basis of V. Assume

$$x = \sum_{i=1}^{n} \alpha_i \cdot v_i, \quad y = \sum_{i=1}^{n} \beta_i \cdot v_i.$$

Then,

$$\langle T(x), T(y)\rangle = \left\langle T(\sum_{i=1}^{n} \alpha_i \cdot v_i), T(\sum_{i=1}^{n} \beta_i \cdot v_i) \right\rangle = \sum_{i=1}^{n} \alpha_i \cdot \overline{\beta_i} = \langle x, y\rangle .$$

□

2.5 Spectral theory of normal operators

Definition 37 (Self-adjoint and normal operator)**.** Let $T : V \to V$ be a linear operator on an inner product space V.

1. We say T is self-adjoint, if $T = T*$.

2. We say T is normal, if $T \circ T^* = T^* \circ T$.

Remark. A linear operator $T : V \to V$ is unitary if and only if $T^* = T^{-1}$. (Assume that V is finite dimensional.) Thus, unitary operators and self-adjoint operators are normal.

In the rest of this section, if not specifically mentioned, V denotes the finite dimensional inner product space over $\mathbb{F}$ ($\mathbb{R}$ or $\mathbb{C}$.)

Theorem 38. *Given $T : V \to V$, a linear operator on finite dimensional space V. The the following statements are equivalent.*

1. T is normal.

2. $\|T(v)\| = \|T^*(v)\|$, for all $v \in V$.

Proof.
(1) $\implies$ (2): Note that

$$\langle T(v), T(v)\rangle = \langle T^*T(v), v\rangle = \langle TT^*(v), v\rangle = \langle T^*(v), T^*(v)\rangle .$$

(2) $\implies$ (1): Consider $\|T(x+y)\|^2 = \|T^*(x+y)\|^2$ (and $\|T(x+i\cdot y)\|^2 = \|T^*(x+i\cdot y)\|^2$ if $\mathbb{F} = \mathbb{C}$.) Expanding both equations gives

$$\langle T^*T(x), y\rangle = \langle TT^*(x), y\rangle , \text{ for all } x, y \in V.$$

Thus, $T \circ T^* \equiv T^* \circ T$. □

Corollary. Let $T : V \to V$ be a linear operator on a finite dimensional vector space V. Suppose T is normal, and v is an eigenvector of T with eigenvalue λ. Then, v is an eigenvector of T^* with eigenvalue $\overline{\lambda}$.

Proof. Since T is normal, $S = T - \lambda \cdot \mathrm{Id}_V$ is normal. (In fact, $p(T)$ is normal, for all $p(x) \in \mathbb{F}[x]$.) We have $Sv = 0$. From Theorem 38, we have $\|S^*v\| = \|Sv\| = 0$. Hence, v is in the kernel of $S^* = T^* - \overline{\lambda} \cdot \mathrm{Id}_V$. This completes the proof. □

We now prove an useful lemma.

Lemma 3. Let T be a linear operator on V, such that T^* exists. (We have assumed nothing about whether it is normal.) Then,

$$\ker T^*T = \ker T.$$

Proof. Obviously, $\ker T \subset \ker T^*T$. It suffices to show that $\ker T^*T \subset \ker T$. Let $v \in \ker T^*T$, then,

$$\begin{aligned} T^*T(v) = 0 &\implies & \langle T^*T(v), v\rangle = 0 \\ &\implies & \langle T(v), T(v)\rangle = 0 \\ &\implies & \|T(v)\| = 0 \\ &\implies & T(v) = 0. \end{aligned}$$

□

Theorem 39 (Semi-simplicity of normal operators)**.** *Suppose T is a normal operator on V. If $T^n \equiv 0$, for some $n \geq 1$. Then $T \equiv 0$.*

Proof. Let $S = T^*T$. By Lemma 3, it suffices to show $\ker S = V$. Since $T^n = 0$, we have $S^n = 0$. (T^* and T commute.) We may enlarge n so that $n = 2^k$ for some $k \in \mathbb{N}$. Note that

$$\left\| S^{2^{k-1}} v \right\|^2 = \left\langle S^{2^{k-1}} v, S^{2^{k-1}} v \right\rangle = \left\langle \left(S^{2^{k-1}} \right)^* S^{2^{k-1}} v, v \right\rangle = \left\langle S^{2^k} v, v \right\rangle = 0.$$

Repeating this process gives us $S = 0$. □

Before we introduce the next theorem (Theorem 40), we shall first prove another useful result.

Lemma 4. Let V be an inner product space over $\mathbb{F}$, and let $T : V \to V$ be a normal operator on V. Suppose $p(x)$ and $q(x)$ are polynomials in $\mathbb{F}$ with no common roots. Then,

$$\ker\left(p(T)\right) \perp \ker\left(q(T)\right),$$

that is, $\langle v, w\rangle = 0$, for all $v \in \ker\left(p(T)\right)$ and $w \in \ker\left(q(T)\right)$.

Proof. Since p, q have no common roots, there exist $A, B \in \mathbb{F}[x]$, such that

$$A(x)p(x) + B(x)q(x) = 1.$$

Let $v \in \ker(p(T))$ and $w \in \ker(q(T))$. We have $B(T)q(T)(v) = v$. Thus,

$$\langle v, w\rangle = \langle B(T)q(T)(v), w\rangle = \langle q(T)B(T)v, w\rangle = \langle B(T)v, q(T)^*(w)\rangle \overset{(\spadesuit)}{=} \langle B(T)v, 0\rangle = 0.$$

($\spadesuit$) is true since:

$$\begin{aligned} w \in \ker(q(T)) &\implies \|q(T)(w)\| = 0 \\ &\implies \|q(T)^*(w)\| = 0 \\ &\implies q(T)^*(w) = 0. \end{aligned}$$

$\square$

Theorem 40. *Let $(V, \langle\ ,\ \rangle)$ be an finite dimensional inner product space over $\mathbb{C}$. Let $T : V \to V$ be a normal operator on V. Then, T is diagonalizable. Moreover,*

$$V = \bigoplus_{i=1}^{s} E_{\lambda_i} = E_{\lambda_1} \oplus E_{\lambda_2} \oplus \cdots \oplus E_{\lambda_s}$$

is the orthogonal decomposition of eigenspaces of V. Recall that E_{λ_i} is the eigenspace that which has eigenvalue λ.

Here we give two proofs.

Proof. Let $\mathrm{ch}_T(x)$ be the characteristic polynomial of T. The fundamental theorem of algebra asserts that $\mathrm{ch}_T(x)$ splits completely, that is,

$$\mathrm{ch}_T(x) = \prod_{i=1}^{s}(x - \lambda_i)^{n_i}.$$

Then, we have learnt that $V = \bigoplus_{i=1}^{s} W_i$ in the theory of Jordan forms, where

$$W_i = \ker(T - \lambda_i \cdot \mathrm{Id}_V)^{n_i}.$$

Consider $T|_{W_i}$ on $(W_i, \langle\ ,\ \rangle|_{W_i \times W_i})$. Note that $T|_{W_i}$ is normal and that $(T|_{W_i} - \lambda_i \cdot \mathrm{Id}_{W_i})^{n_i} = 0$. By Theorem 39, we conclude $T|_{W_i} - \lambda_i \cdot \mathrm{Id}_{W_i} = 0$. This implies

$$W_i = \ker(T - \lambda_i \cdot \mathrm{Id}_V)^{n_i} = \ker(T - \lambda_i \cdot \mathrm{Id}_V) = E_{\lambda_i}.$$

It remains to show that each E_{λ_i} is orthogonal to each other. It follows by Lemma 4. $\square$

Here is an alternative proof using mathematical induction.

Proof. Let $\lambda \in \mathbb{C}$ be an eigenvalue of T. Then,

$$E_\lambda = \{v \in V : T(v) = \lambda \cdot v\} \neq \{0\}.$$

Decompose V into $E_\lambda \oplus E_\lambda^\perp$. ($V$ is finite dimensional.) We claim that $E_\lambda^\perp$ is a T-invariant subspace. Let $x \in E_\lambda^\perp$ and $v \in E_\lambda$. Then,

$$\langle T(x), v\rangle = \langle x, T^*(v)\rangle \overset{(\spadesuit)}{=} \langle x, \overline{\lambda} v\rangle = \lambda \langle x, v\rangle = 0.$$

The equality (♠) holds because of Corollary 2.5. On the other hand,

$$\dim E_\lambda^\perp < \dim V.$$

By induction, $T|_{E_\lambda^\perp}$ is diagonalizable and

$$E_\lambda^\perp = \bigoplus_i E_{\lambda_i}.$$

This completes the proof. □

However, Theorem 40 is not true for inner product space over $\mathbb{R}$. But we have the following theorem.

Theorem 41. *Let V be a finite dimensional inner product space over $\mathbb{R}$, and let $T : V \to V$ be a self-adjoint operator on V. Then, T is diagonalizable. Moreover,*

$$V = \bigoplus_{i=1}^{s} E_{\lambda_i},$$

and $E_{\lambda_i} \perp E_{\lambda_j}$ if $i \neq j$.

Proof. In view of the proofs of Theorem 40, it suffices to show that $\mathrm{ch}_T(x)$ splits completely in $\mathbb{R}$. Choose an orthonormal basis $\mathcal{A} = \{v_1, v_2, \ldots, v_n\}$ of V. Define a matrix

$$A := [T]_{\mathcal{A}} = (a_{ij})_{n\times n}.$$

Then, it is well-known that

$$[T^*]_{\mathcal{A}} = A^*.$$

Hence $A^* = A$ since T is self-adjoint. Now, assume $\lambda \in \mathbb{C}$ is an eigenvalue of T. Then, there exists $x \in \mathbb{C}^n \setminus \{0\}$ (column vector) such that

$$Ax = \lambda \cdot x.$$

Consider

$$\overline{\lambda}(x^* \cdot x) = (Ax)^* \cdot x = x^* \cdot A^* \cdot x = x^* \cdot A \cdot x = \lambda \cdot (x^* \cdot x).$$

This indicates

$$\lambda \cdot \|x\|^2 = \overline{\lambda} \cdot \|x\|^2 \implies \lambda \in \mathbb{R}.$$

□

Corollary. Let $A \in M_n(\mathbb{C})$ be a complex normal matrix, that is,

$$A^* \cdot A = A \cdot A^*.$$

Then, there exists an invertible matrix $P \in M_n(\mathbb{C})$ such that:

1. $P \cdot P^* = I_n$.
2. $P^{-1}AP$ is diagonal.

Proof. Let $V = \mathbb{C}^n$ be an inner product space equipped with the standard inner product structure. Let $T : V \to V$ be the operator defined by

$$v \mapsto A \cdot v.$$

Then, the standard basis is orthonormal and hence $A^* = A$ is equivalent to T is self-adjoint. It follows from Theorem 40 that

$$V = \bigoplus_{i=1}^{s} E_{\lambda_i}$$

is a orthogonal decomposition. For each E_{λ_i}, we choose an orthonormal basis

$$\mathcal{A}_i = \{v_{i1}, \ldots, v_{in_i}\}.$$

Then,

$$\mathcal{A} = \bigsqcup_{i=1}^{s} \mathcal{A}_i = \mathcal{A}_1 \sqcup \mathcal{A}_2 \sqcup \cdots \sqcup \mathcal{A}_s$$

is an orthonormal basis. (Because $E_{\lambda_i} \perp E_{\lambda_j}$.) Let P be the matrix sends the standard basis to $\mathcal{A}$. By Theorem 36, we conclude that $P \cdot P^* = P^* \cdot P = I_n$. Also, it is easy to see

$$P^{-1}AP = \begin{pmatrix} \lambda_1 I_{n_1} & 0 & \cdots & 0 \\ 0 & \lambda_2 I_{n_2} & \cdots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_s I_{n_s} \end{pmatrix}.$$

This completes the proof. □

Similarly, one can prove the following result:

Corollary. Let $A \in M_n(\mathbb{R})$ be a real matrix such that $A^{\mathrm{t}} = A$. Then, there exists an invertible matrix $P \in M_n(\mathbb{R})$ such that:

1. $P^{\mathrm{t}} \cdot P = P \cdot P^{\mathrm{t}} = I_n$.
2. $P^{-1}AP$ is diagonal.

Corollary. Let T be a self-adjoint operator on inner product space V over $\mathbb{F}$. Then, there exists $\lambda_i \in \mathbb{R}$ such that

$$T(v) = \lambda_1 \cdot \mathrm{Proj}_{E_{\lambda_1}}(v) + \lambda_2 \cdot \mathrm{Proj}_{E_{\lambda_2}}(v) + \cdots + \lambda_s \cdot \mathrm{Proj}_{E_{\lambda_s}}(v).$$

In Theorem 41, we show that every self-adjoint operator on vector space over $\mathbb{R}$ is diagonalizable. However, we do not deal with all normal operators. The next theorem is discussing operators over real inner product space.

Theorem 42. *Let $A \in M_n(\mathbb{R})$ be a real normal matrix, that is,*

$$A^{\mathrm{t}} \cdot A = A \cdot A^{\mathrm{t}}.$$

Then, there exists an invertible matrix $P \in M_n(\mathbb{R})$ such that:

1. $P \cdot P^{\mathrm{t}} = P^{\mathrm{t}} \cdot P = I_n$.

2. $P^{-1}AP = (\bigoplus_{i=1}^{s} \lambda_i I_{n_i}) \oplus (\bigoplus_{j=1}^{r} D_j^{\oplus m_j})$, *where all* $\lambda_i \in \mathbb{R}$, *and all* D_j *have the form:*

$$\begin{pmatrix} \alpha_j & \beta_j \\ -\beta_j & \alpha_j \end{pmatrix}.$$

Remark. Here, we have a little abuse of notation. We write $A \oplus B$ to represent

$$\begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix},$$

if both A and B are square matrices. Also, we write $P^{\oplus k}$ to mean $\bigoplus_{i=1}^{k} P = P \oplus P \oplus \cdots \oplus P$, for square matrix P.

Before we start proving this theorem, we shall first prove some useful lemmas.

Lemma 5. Let V be an inner product space over $\mathbb{R}$, and let $T : V \to V$ be a normal operator, such that

$$S^2 = -\mathrm{Id}_V.$$

Let $v_1 \in V \setminus \{0\}$ and $v_2 = S(v_1)$. Then,

$$S^*(v_1) = -v_2, \quad S^*(v_2) = v_1, \quad \langle v_1, v_2 \rangle = 0, \quad \|v_1\| = \|v_2\|.$$

Proof. Consider

$$\begin{aligned} & \|S^* v_1 + v_2\|^2 + \|S^* v_2 - v_1\|^2 \\ = & \langle S^* v_1, S^* v_1 \rangle + \langle S^* v_1, v_2 \rangle + \langle v_2, S^* v_1 \rangle + \langle v_2, v_2 \rangle \\ & \quad + \langle S^* v_2, S^* v_2 \rangle - \langle S^* v_2, v_1 \rangle - \langle v_1, S^* v_2 \rangle + \langle v_1, v_1 \rangle \\ = & \langle S v_1, S v_1 \rangle + 2 \cdot \langle S v_2, v_1 \rangle + \langle v_2, v_2 \rangle + \langle S v_2, S v_2 \rangle - 2 \cdot \langle S v_1, v_2 \rangle + \langle v_1, v_1 \rangle \\ = & \|S v_1 - v_2\|^2 + \|S v_2 + v_1\|^2 = 0. \end{aligned}$$

This prove the first two assertion. Note that

$$\langle v_1, v_2 \rangle = \langle v_1, S v_1 \rangle = \langle S^* v_1, v_1 \rangle = \langle -v_2, v_1 \rangle = -\langle v_1, v_2 \rangle$$

and that

$$\|v_2\|^2 = \langle v_2, v_2 \rangle = \langle S v_1, v_2 \rangle = \langle v_1, S^* v_2 \rangle = \langle v_1, v_1 \rangle = \|v_1\|^2.$$

□

From Lemma 5, we conclude that:

Continuing from the above definition, let

$$w_1 = \frac{v_1}{\|v_1\|}, \quad w_2 = \frac{v_2}{\|v_2\|}.$$

Then, $\{w_1, w_2\}$ is an orthonormal set. Moreover, $W := \mathrm{span}_{\mathbb{R}}\{w_1, w_2\}$ is S-invariant and S^*-invariant.

$$\left[S|_W \right]_{\{w_1, w_2\}} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

Lemma 6. Let $T : V \to V$ be a normal operator on a finite dimensional inner product space. Suppose

$$\mathrm{ch}_T(x) = \left((x-a)^2 + b^2\right)^m,$$

for some $b \neq 0$. Then, there exists an orthonormal basis $\mathcal{A}$ such that

$$[T]_{\mathcal{A}} = \begin{pmatrix} a & -b \\ b & a \end{pmatrix}^{\oplus m}.$$

Proof. Let $S = (T-a)/b$. Then, by Lemma 5, we have an orthonormal set $\mathcal{A}_1 = \{w_1, w_2\}$. Define $W_1 = \mathrm{span}_{\mathbb{R}}\{w_1, w_2\}$. Then,

$$\left[S|_{W_1}\right]_{\{w_1,w_2\}} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

That indicates that

$$\left[T|_{W_1}\right]_{\{w_1,w_2\}} = \begin{pmatrix} a & -b \\ b & a \end{pmatrix}.$$

We now claim that $W_1^\perp$ is a S-invariant subspace. Let $v \in W_1^\perp$ and $w \in W_1$, then

$$\langle Sv, w \rangle = \langle v, S^* w \rangle = 0,$$

since W_1 is also a S^*-invariant subspace. Similarly, we have an orthonormal set $\mathcal{A}_2 \subset W_1^\perp$ such that

$$\left[S|_{W_2}\right]_{\mathcal{A}_2} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix},$$

where W_2 is the subspace generated by $\mathcal{A}_2$. Also, $(W_1 \oplus W_2)^\perp$ is a S-invariant. Continuing this process give s

$$V = \bigoplus_{i=1}^{s} W_i,$$

each W_i is spanned by an orthonormal set $\mathcal{A}_i$ and

$$\left[S|_{W_i}\right]_{\mathcal{A}_i} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

Let $\mathcal{A} = \bigsqcup \mathcal{A}_i$, then

$$[T]_{\mathcal{A}} = \begin{pmatrix} a & -b \\ b & a \end{pmatrix}^{\oplus m}.$$

□

Lemma 7. Let $T : V \to V$ be a normal operator on a finite dimensional vector space V over $\mathbb{R}$. Suppose $\mathrm{m}_T(x) = \prod_{i=1}^{s} f_i$, where f_i are all irreducible. Then, f_i are all distinct.

Proof. Suppose not, then there exists an irreducible polynomial $f \in \mathbb{R}[x]$ such that $f = f_i$ for more than one i. Let us consider $W = \ker f^n(T)$, then W is a $f(T)$-invariant subspace. Note that $f(T)$ is normal on W and $f(T)^n \equiv 0$ on W. Thus, from Theorem 39, we conclude that

$$\ker f(T) = \ker f^n(T),$$

which leads to a contradiction. □

Proof of Theorem 42. From Lemma 7, we assume that

$$\mathrm{m}_T(x) = \left(\prod_{i=1}^{s}(x-\lambda_i)\right)\cdot\left(\prod_{j=1}^{r}\left((x-a_j)^2+b_j^2\right)\right).$$

From what we have learnt in the theory of Jordan forms,

$$V = \left(\bigoplus_{i=1}^{s}\ker\left(T-\lambda_i\right)\right)\oplus\left(\bigoplus_{j=1}^{r}\ker\left((T-a_j)^2+b_j^2\right)\right).$$

For simplicity, we define $W_i := \ker\left(T-\lambda_i\right)$ and $X_j := \ker\left((T-a_j)^2+b_j^2\right)$. It suffices to show that for each j, there exists a basis $\mathcal{A}_j$ such that

$$\left[T\big|_{X_j}\right]_{\mathcal{A}_j} = D_j^{\oplus m_j},$$

where

$$D_j := \begin{pmatrix} a_j & -b_j \\ b_j & a_j \end{pmatrix}.$$

This follows from Lemma 6. □

2.6 Applications of spectral theory of normal operators

This subsection is mainly deal with two topics:

1. Structure of orthogonal (unitary) operators.
2. Singular value decomposition (SVD).

In this subsection, we assume that V is a finite dimensional inner product space unless otherwise stated.

Definition 43 (Unitary groups and orthogonal groups)**.** Let V be a finite dimensional inner product space over $\mathbb{F}$.

1. If $\mathbb{F}=\mathbb{C}$, we define the unitary group
$$\mathrm{U}(V) = \left\{T : V \to V \mid T\cdot T^* = T^*\cdot T = \mathrm{Id}_V\right\}.$$
2. If $\mathbb{F}=\mathbb{R}$, we define the orthogonal group
$$\mathrm{O}(V) = \left\{T : V \to V \mid T\cdot T^* = T^*\cdot T = \mathrm{Id}_V\right\}.$$

We also define unitary groups and orthogonal groups by matrices. We write:

1. $\mathrm{O}_n(\mathbb{R}) = \{A \in M_n(\mathbb{R}) : A\cdot A^{\mathrm{t}} = I_n\}$ is the orthogonal group.
2. $\mathrm{U}_n(\mathbb{R}) = \{A \in M_n(\mathbb{C}) : A\cdot A^{*} = I_n\}$ is the unitary group.

Note that $\mathrm{U}_n(\mathbb{R})$ contains some complex matrices although it contains $\mathbb{R}$ in its “name”. We now focus on orthogonal groups.

Definition 44 (Reflection)**.** Let $T : V \to V$ be a linear operator. T is a reflection if there exists a $z \in V$ with $\|z\| = 1$ such that

$$T(x) = x - 2 \cdot \mathrm{Proj}_z(x) = x - 2 \cdot \langle x, z \rangle \cdot z, \text{ for all } x \in V.$$

We also say that T is the reflection over the hyperplane $\mathcal{H} = (\mathbb{R} \cdot z)^\perp$.

Remark. Suppose T is a reflection. Let $\mathcal{A}$ be an orthonormal basis of $\mathcal{H}$. Then, $\mathcal{A}' = \{z\} \sqcup \mathcal{A}$ is an orthonormal basis such that

$$[T]_{\mathcal{A}'} = \begin{pmatrix} -1 & 0 \\ 0 & I_{n-1} \end{pmatrix}.$$

This means that there exists a matrix $P \in \mathrm{O}_n(\mathbb{R})$, such that

$$P^{\mathrm{t}}[T]_{\mathcal{B}}P = \begin{pmatrix} -1 & 0 \\ 0 & I_{n-1} \end{pmatrix},$$

where $\mathcal{B}$ is the standard basis. Hence, we can define reflection on $M_n(\mathbb{R})$.

Definition 45. Let $A \in M_n(\mathbb{R})$ be a real matrix. A is a reflection if (and only if) there exist a $P \in \mathrm{O}_n(\mathbb{R})$ such that

$$P^{\mathrm{t}}AP = \begin{pmatrix} -1 & 0 \\ 0 & I_{n-1} \end{pmatrix}.$$

Lemma 8.

1. If $A \in \mathrm{O}_n(\mathbb{R})$ and $\lambda \in \mathbb{R}$ is an eigenvalue of A, then $\lambda = \pm 1$.
2. If $A \in \mathrm{U}_n(\mathbb{R})$ and $\lambda \in \mathbb{C}$ is an eigenvalue of A, then $|\lambda| = 1$.

Proof. Let $V = \mathbb{F}^n$. ($\mathbb{F}$ is $\mathbb{R}$ or $\mathbb{C}$.) Define the standard inner product $\langle \cdot, \cdot \rangle$ on V, namely,

$$\langle x, y \rangle = y^* \cdot x, \ x, y \text{ are column vectors.}$$

Then, $A \in \mathrm{O}_n(\mathbb{R})$ ($A \in \mathrm{U}_n(\mathbb{R})$) is a unitary operator on $(V, \langle \ , \ \rangle)$. If $\lambda \in \mathbb{F}$ is an eigenvalue of A, then there exists $v \in V \setminus \{0\}$ such that: $Av = \lambda v$

$$\langle v, v \rangle = \langle Av, Av \rangle = \langle \lambda \cdot v, \lambda \cdot v \rangle = \lambda \cdot \overline{\lambda} \cdot \langle v, v \rangle.$$

This implies $\lambda \cdot \overline{\lambda} = 1$. □

Theorem 46 (Cartan—Dieudonné Theorem)**.** *For every $A \in \mathrm{O}_n(\mathbb{R})$, then A is a product of reflections.*

Proof. From Theorem 42, we know that $A \in \mathrm{O}_n(\mathbb{R})$ can be written in the form

$$\left(\bigoplus_{i=1}^{s} (\lambda_i) \right) \oplus \left(\bigoplus_{j=1}^{r} D_j \right), \tag{5}$$

where D_j is

$$\begin{pmatrix} a_j & -b_j \\ b_j & a_j \end{pmatrix}, \text{ for some } a_j, b_j \in \mathbb{R}, \ b_j \neq 0.$$

Lemma 8 asserts that $\lambda_i = \pm 1$ in (5). It is easy to see that (by Definition 45) if $m < n$ and $X \in \mathrm{O}_m(\mathbb{R})$ is a reflection, then so is

$$\begin{pmatrix} X & 0 \\ 0 & I_{n-m} \end{pmatrix}.$$

Thus, it suffices to show that each D_j is a product of reflections on $\mathbb{R}^n$. Since each $D_j \in \mathrm{O}_2(\mathbb{R})$, we know that

$$D_j \cdot {D_j}^{\mathrm{t}} = I_2.$$

Therefore, $a_j^2 + b_j^2 = 1$, let $\theta \in [0, 2\pi)$ such that $a_j = \cos\theta$ and $b_j = \sin\theta$. Note that

$$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

We conclude that D_j is a product of two reflections. □

Next, we are going to discuss the singular value decomposition. We first define the singular decomposition of a matrix $A \in \mathrm{U}_n(\mathbb{R})$.

Definition 47 (Singular value decomposition (S.V.D.))**.** Let $A \in M_{m\times n}(\mathbb{C})$. If there exist $P \in \mathrm{U}_n(\mathbb{R})$ and $Q \in \mathrm{U}_m(\mathbb{R})$ such that

$$Q^* \cdot A \cdot P = \begin{pmatrix} \Sigma & O_{r\times(n-r)} \\ O_{(m-r)\times r} & O_{(m-r)\times(n-r)} \end{pmatrix} = \begin{pmatrix} \Sigma & 0 \\ 0 & 0 \end{pmatrix} \in M_{m\times n}(\mathbb{C}),$$

where O is the zero matrix in $M_{(m-r)\times(n-r)}(\mathbb{C})$ and $\Sigma \in M_r(\mathbb{C})$ is the diagonal matrix

$$\begin{pmatrix} \sigma_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_r \end{pmatrix},$$

with $\sigma_i \in \mathbb{R}$ and $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r$. Then,

$$A = Q \cdot \begin{pmatrix} \Sigma & 0 \\ 0 & 0 \end{pmatrix} \cdot P^*$$

is called the singular value decomposition.

Theorem 48 (Singular value decomposition)**.** *Let $A \in M_{m\times n}(\mathbb{C})$, then the singular value decomposition of A exists.*

It is equivalent to prove the following theorem. Although it is not quite trivial that the following theorem implies the singular value decomposition theorem, it is annoying to write it properly, so we omit the details here.

Theorem 49 (Linear transformation version)**.** *Let V, W be two finite dimensional inner product spaces over $\mathbb{F}$ ($\mathbb{R}$ or $\mathbb{C}$) and let $T : V \to W$ be a linear transformation. Then, there exist an orthonormal basis $\mathcal{B} = \{v_1, v_2, \ldots, v_n\}$ of V such that*

1. *$\{T(v_1), T(v_2), \ldots, T(v_r)\}$ is orthogonal.*
2. *$\{T(v_{r+1}), T(v_{r+2}), \ldots, T(v_n)\} = 0$.*

Proof. By Theorem 26, the adjoint T^* of T exists. Consider $S := T^* \circ T : V \to V$. Then S is self-adjoint. Applying the spectral theory for self-adjoint operators (Theorem 41 for $\mathbb{F} = \mathbb{C}$ or Theorem 40 for $\mathbb{F} = \mathbb{C}$), we can find an orthonormal basis $\mathcal{B} = \{v_1, v_2, \ldots, v_n\}$ consisting of eigevectors of S. Let λ_i be the eigenvalue of v_i (with respect to the transformation S), then

$$\langle T(v_i), T(v_j)\rangle = \langle T^*T(v_i), v_j\rangle = \langle S(v_i), v_j\rangle = \lambda_i \langle v_i, v_j\rangle .$$

This gives

$$\begin{cases} T(v_i) \perp T(v_j), & \text{if } i \neq j. \\ \|T(v_i)\|^2 = \lambda_i, & \text{for all } i. \end{cases}$$

This proves the theorem. $\square$

Singular value decomposition generalize the definition of "inverse matrix". We can define the pseudo inverse or the Moore-Penrose inverse.

Definition 50 (Pseudo inverse or Moore-Penrose inverse)**.** Let $A \in M_{m\times n}(\mathbb{C})$. Let the definition of P and Q be the same as in Definition 47. Then, the Moore-Penrose inverse is defined as

$$A^\dagger := P \cdot \begin{pmatrix} \Sigma^{-1} & 0 \\ 0 & 0 \end{pmatrix} \cdot Q^*.$$

We also can define Moore-Penrose inverse of linear transformation.

Definition 51 (Intrinsic definition of Moore-Penrose inverse)**.** Let V, W be two finite dimensional inner product spaces and let $T : V \to W$ be a linear transformation. Then, the Moore-Penrose inverse is the linear transformation $T^\dagger : W \to V$ defined by:

$$T^\dagger(w) = (T\big|_{(\ker T)^\perp})^{-1} \circ \mathrm{Proj}_{\mathrm{Im}T}(w).$$

Two definitions of the Moore-Penrose inverse aggee with each other. The Moore-Penrose inverse is invented to solve system linear equations.

Theorem 52. *Let V, W be two finite dimensioal inner product space and let $T : V \to W$. Given $b \in W$. Then, $T(x) = b$ has a solution in V if and only if*

$$b = T \cdot T^\dagger(b). \tag{6}$$

In addition, in this case, x is a solution if and only if

$$x = T^\dagger(b) + (\mathrm{Id}_V - T^\dagger \cdot T)(z), \text{ for some } z \in V.$$

Proof. $T(x) = b$ has a solution in V is equivalent to

$$\begin{aligned} b \in \mathrm{Im}T \iff& \quad \mathrm{Proj}_{\mathrm{Im}T}(b) = b \\ \iff& \quad T \circ T^\dagger(b) = b, \text{ by Definition 51.} \end{aligned}$$

To see the second assertion of the theorem, it suffices to show:

$$\ker T = \mathrm{Im}\big(\mathrm{Id}_V - T^\dagger \circ T\big).$$

However, it follows from the definition that $T^\dagger \circ T(v) = \mathrm{Proj}_{(\ker T)^\perp}(v)$. Thus,

$$\big(\mathrm{Id}_V - T^\dagger \circ T\big) = \mathrm{Proj}_{\ker T}.$$

This proves the theorem. $\square$

However, the equation is not always has a solution. In general, $T^{\dagger}(b)$ is the best approximation of solutions of $T(x) = b$ in the following sense:

$$\left\|T \circ T^{\dagger}(b) - b\right\| = \min_{x \in V} \|T(x) - b\|.$$

Theorem 53. *Let V, W be two finite dimensional inner product space and let $T : V \to W$ be a linear transformation. Given $b \in W$. Then, $T^{\dagger}(b)$ is the best approximation of solutions of $T(x) = b$.*

Proof. Since $T \cdot T^{\dagger} = \mathrm{Proj}_{\mathrm{Im}T}$, we have $\left(T \cdot T^{\dagger}(b) - b\right) \in (\mathrm{Im}T)^{\perp}$. Thus,

$$\begin{aligned}
\|T(x) - b\|^2 &= \left\|T(x) - T \cdot T^{\dagger}(b) + T \cdot T^{\dagger}(b) - b\right\|^2 \\
&= \left\|T(x) - T \cdot T^{\dagger}(b)\right\|^2 + \left\|T \cdot T^{\dagger}(b) - b\right\|^2 \\
&\geq \left\|T \cdot T^{\dagger}(b) - b\right\|^2.
\end{aligned}$$

The equality holds if $T(x) = T \cdot T^{\dagger}(b)$. $\square$

2.7 Bilinear forms

In the subsection, unless otherwise stated, we assume $\mathbb{F}$ is one of the following fields: $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$, or the finite field $\mathbb{Z}/p\mathbb{Z}$, and let V be a finite dimensional vector space over $\mathbb{F}$.

Definition 54. Let $\mathbb{F}$ be a field. Let

$$\Xi = \{n \in \mathbb{N} : n \cdot x = 0, \text{ for all } x \in \mathbb{F}\}.$$

Then, the characteristic of $\mathbb{F}$ is defined as:

$$\mathrm{char}\,(\mathbb{F}) = \begin{cases} \min \Xi & , \text{if } \Xi \neq \varnothing \\ 0 & , \text{otherwise} \end{cases}.$$

Definition 55 (Bilinear form)**.** Let V be a vector space over $\mathbb{F}$. Then, a bilinear form B is a function

$$B : V \times V \to \mathbb{F}$$

such that B is component-wise linear. That is, B is a linear function if we fix one variable.

Thus, the inner product on a vector space over $\mathbb{R}$ is a bilinear form. If B is a bilinear form on a finite dimensional vector space V, then it induce two linear maps from V to $V^{\vee}$.

$$\begin{aligned}
l_B &: V \to V^{\vee} \\
&\quad v \mapsto l_B(v)(w) = B(v, w), \text{ for all } w \in V \\
r_B &: V \to V^{\vee} \\
&\quad v \mapsto r_B(v)(w) = B(w, v), \text{ for all } w \in V
\end{aligned}$$

Conversely, given a linear transformation $f \colon V \to V^{\vee}$, f induces two bilinear forms:

$$\begin{aligned}
B_f^l(v, w) &:= f(v)(w) \\
B_f^r(v, w) &:= f(w)(v)
\end{aligned}$$

This explains there is a bijection between

$$\{\text{all bilinear forms } B : V \times V \to \mathbb{F}\} \simeq \mathrm{Hom}_{\mathbb{F}}(V, V^{\vee}).$$

We now fix a basis $\mathcal{B} = \{v_1, v_2, \ldots, v_n\}$ of V. We get an isomorphism

$$\begin{aligned}\{\text{all bilinear forms } B : V \times V \to \mathbb{F}\} &\longleftrightarrow M_n(\mathbb{F}) \\ B &\longleftrightarrow \Omega_{B,\mathcal{B}} = (B(v_i, v_j))\end{aligned}.$$

Similar to what we have shown in the theory of inner product space, if we change the basis to $\mathcal{A}$, then

$$\Omega_{B,\mathcal{A}} = P^{\mathrm{t}} \cdot \Omega_{B,\mathcal{B}} \cdot P,$$

where P is the matrix sends $\mathcal{B}$ to $\mathcal{A}$.

Recall that we have defined the (external) direct sum of two vector spaces in Definition 20. For two vector spaces V, W with bilinear forms B_v, B_w respectively, we can defined a bilinear form B on $V \oplus W$, defined by

$$B\left((v_1, w_1), (v_2, w_2)\right) := B_v(v_1, v_2) + B_w(w_1, w_2),$$

and we often write $B = B_v \oplus B_w$ This definition of the direct sum of bilinear forms agrees with the definition of internal direct sum in the following sense:

> Let (V, B) be a vector space with a bilinear form. W_1 and W_2 are subspaces of V such that $W_1 \oplus W_2 = V$. Then,
>
> $$B = B\big|_{W_1} \oplus B\big|_{W_2},$$
>
> if
>
> $$B(w_1, w_2) = 0, \text{ for all } w_1 \in W_1 \text{ and } w_2 \in W_2.$$

Next, we are going to define the concept of radical.

Definition 56 (Radical)**.** Let $B : V \times B \to V$ be a bilinear form. Define

$$\begin{aligned}\mathrm{rad}_L(V) &= \{v \in V : B(v, w) = 0, \text{ for all } w \in V\} \\ \mathrm{rad}_R(V) &= \{v \in V : B(w, v) = 0, \text{ for all } w \in V\}\end{aligned}$$

Definition 57 (Non-degenerate)**.** A bilinear form is non-degenerate if $\mathrm{rad}_R(V) = \{0\}$.

In fact, the following three statements are equivalent:

1. $\mathrm{rad}_R(V) = \{0\}$.
2. $\mathrm{rad}_L(V) = \{0\}$.
3. $\det \Omega_B \neq 0$.

Definition 58 (Alternating and symmetric bilinear forms)**.** Let $B : V \times V \to \mathbb{F}$ be a bilinear form.

1. B is alternating if $B(v, w) = -B(w, v)$, for all $v, w \in V$.
2. B is symmetric if $B(v, w) = B(w, v)$, for all $v, w \in V$.

We first discuss the alternating form. Now, suppose B is non-degenerate and alternating. Let $\mathcal{A} = \{v_1, v_2, \ldots, v_n\}$ be a basis of V and let $\Omega_B = (B(v_i, v_j))$ be the matrix attached to $\mathcal{A}$. Then,

$$\Omega_B{}^{\mathrm{t}} = -\Omega_B.$$

If a matrix $A \in M_n(\mathbb{F})$ satisfied $A^{\mathrm{t}} = -A$, then it is called skew-symmetric. Next, we want to find a basis $\mathcal{A}$ such that the matrix $\Omega_{B,\mathcal{A}}$ is as simple as possible.

Definition 59 (Symplectic basis)**.** A basis $\{e_1, e_2, \ldots, e_r, f_1, f_2, \ldots, f_r\}$ ($\dim V = 2r$) of V is called a symplectic basis for B if

1. $B(e_i, e_j) = B(f_i, f_j) = 0$, for all i, j.
2. $B(e_i, f_j) = 0$, if $i \neq j$.
3. $B(e_i, f_i) = 0$, for all i.

In other words, if $\mathcal{A}$ is a symplectic basis, then

$$\Omega_{B,\mathcal{A}} = \begin{pmatrix} O_r & I_r \\ -I_r & O_r \end{pmatrix},$$

where $O_r, I_r \in M_r(\mathbb{F})$ are the zero matrix and the identity matrix, respectively.

Theorem 60. *Assume* $\mathrm{char}(\mathbb{F}) \neq 2$*. If V is equipped with a non-degenerate and alternating form B, then* $\dim V$ *is even and V has a symplectic basis.*

Proof. B is alternating and $\mathrm{char}(\mathbb{F}) \neq 2$, so for any $v \in V$,

$$B(v, v) = -B(v, v) \implies B(v, v) = 0.$$

Let $e_1 \in V \setminus \{0\}$. Choose f_1 such that $B(e_1, f_1) = 1$. (This could be done because B is non-degenerate.) Let $W = \mathbb{F}e_1 \oplus \mathbb{F}f_1 = \mathrm{span}_{\mathbb{F}}\{e_1, f_1\}$. We define $W^{\perp}$ as

$$W^{\perp} := \{v \in V : B(v, w) = 0, \text{ for all } w \in W\}.$$

We claim $V = W \oplus W^{\perp}$ is an internal direct sum as vector space with bilinear form. To see this, it suffices to show that $V = W \oplus W^{\perp}$ is an internal direct sum as vector space. ($\because B(W, W^{\perp}) = 0$.)

1. W and $W^{\perp}$ is linearly independent.
   It is equivalent to prove $W \cap W^{\perp} = \{0\}$. Let $v \in W \cap W^{\perp}$. Then, $v = a \cdot e_1 + b \cdot f_1$ for some $a, b \in \mathbb{F}$.

   $$v \in W^{\perp} \implies B(v, e_1) = a = 0; \quad B(v, e_2) = b = 0.$$

2. W and $W^{\perp}$ generate V.
   It is equivalent to prove for each $v \in V$, there exist $a, b \in \mathbb{F}$ such that

   $$(v - a \cdot e_1 - b \cdot f_1) \in W^{\perp}.$$

   Some simple calculations show that

   $$a = B(v, f_1), \quad b = -B(v, e_1),$$

   satisfies the condition.

Thus, $(V,B) = (W, B|_W) \oplus (W^\perp, B|_{W^\perp})$. Note that $B|_{W^\perp}$ is a non-degenerate (why?) and alternating form. By induction, $\dim W^\perp = 2r-2$ for some $r \in \mathbb{N}$, and $W^\perp$ has a symplectic basis $\{e_2, e_3, \ldots, e_r, f_2, f_3, \ldots, f_r\}$ for $B|_{W^\perp}$. We conclude that $\dim V = 2r$ and $\{e_1, e_2, \ldots, e_r, f_1, f_2, \ldots, f_r\}$ is a symplectic basis for (V,B). □

Now, we discuss the non-degenerate symmetric form on V.

Theorem 61. *Assume* $\operatorname{char}(\mathbb{F}) \neq 2$. *If V is equipped with a non-degenerate and symmetric form B, then there exist a basis $\mathcal{A} = \{v_1, v_2, \ldots, v_n\}$ of V such that*

$$B(v_i, v_j), \text{ if } i \neq j.$$

In other words,

$$\Omega_{B,\mathcal{A}} = \begin{pmatrix} a_1 & & & \\ & a_2 & & \\ & & \ddots & \\ & & & a_n \end{pmatrix}$$

is a diagonal matrix where $a_i = B(v_i, v_i)$. Note that $a_i \neq 0$ since B is non-degenerate.

Proof. We claim that there exists $v \in V \setminus \{0\}$ such that $B(v,v) \neq 0$. If such v does not exist, then

$$2B(v,w) = B(v+w, v+w) - B(v,v) - B(w,w) = 0, \text{ for all } v, w \in V.$$

Since $\operatorname{char}(\mathbb{F}) \neq 2$, we have $B(v,w) = 0$ for all $v, w \in V$. Therefore, there exists $v_1 \in V \setminus \{0\}$ such that $B(v_1, v_1) \neq 0$. Let $W = \mathbb{F}v_1$ and let

$$W^\perp := \{v \in V : B(v, v_1) = 0\}.$$

Then, $(V,B) = (W, B|_W) \oplus (W^\perp, B|_{W^\perp})$ and we can proceed by induction. □

Next, we can classify all symmetric bilinear forms on finite dimensional vector space over $\mathbb{R}$. Let V be a real vector space with a symmetric bilinear form B. Suppose B is non-degenerate, then by Theorem 61, there is a basis $\mathcal{A} = \{v_1, v_2, \ldots, v_n\}$ such that

$$\Omega_{B,\mathcal{A}} = \begin{pmatrix} a_1 & & & \\ & a_2 & & \\ & & \ddots & \\ & & & a_n \end{pmatrix} \quad (a_i \neq 0.)$$

Replacing $\mathcal{A} = \{v_1, v_2, \ldots, v_n\}$ with $\mathcal{A}'$

$$\left\{ \frac{v_1}{\sqrt{|a_1|}}, \frac{v_2}{\sqrt{|a_2|}}, \ldots, \frac{v_n}{\sqrt{|a_n|}} \right\},$$

then

$$\Omega_{B,\mathcal{A}'} = \begin{pmatrix} \operatorname{sgn}(a_1) & & & \\ & \operatorname{sgn}(a_2) & & \\ & & \ddots & \\ & & & \operatorname{sgn}(a_n) \end{pmatrix}.$$

Thus, we can define the signature of a non-degenerate symmetric bilinear form by counting the positive and negative elements on the diagonal matrix $\Omega_{B,\mathcal{A}}$.

Definition 62 (Signature)**.** If B is a non-degenerate symmetric bilinear form on a vector space V over $\mathbb{R}$, then define the signature (r, s) of V so that

1. $r = \#\{i : \operatorname{sgn}(a_i) = 1\}$.
2. $s = \#\{i : \operatorname{sgn}(a_i) = -1\}$.

We have $r + s = \dim V$ since B is non-degenerate.

If B is degenerate (and symmetric), we also can define its signature. Note that $\operatorname{rad}_L(V) = \operatorname{rad}_R(V)$ (B is symmetric.) Then, $V/\operatorname{rad}(V)$ is a vector space, and induced a bilinear form $\widetilde{B}$ from B defined by

$$\widetilde{B}([v_1], [v_2]) = B(v_1, v_2).$$

It is easy to see that $\widetilde{B}$ is well-defined and one can check that $\widetilde{B}$ is a non-degenerate symmetric bilinear form. We define the signature of B to be the signature of $\widetilde{B}$.

Theorem 63. *Non-degenerate symmetric bilinear forms over finite dimensional real vector spaces are completely determined by their signature. That is, there exists a bijection preserving the bilinear form structure if two spaces have the same signature.*

Remark. The non-degenerate symmetric bilinear form B is positive definite (inner product) if the signature of B is $(\dim V, 0)$.

3 Applications of Linear Algebra

In this section, we will introduce some applications of linear algebra.

3.1 The number of common zeros of two polynomials

First, we look at the following question. Let $f(x), g(x)$ be two polynomials.

What is the size of $\#\{a \in \mathbb{C} : f(a) = g(a) = 0\}$. *(Counted with multiplicities)*

The solution to this question is answered by Étienne Bézout, a French mathematician. Actually, I am not pretty sure whether the result is discovered by him, but to prove the result, we have to introduce a matrix called Bézoutian. We first define $v(f, g)$ to simplify our notation.

Definition 64. Let $f, g \in \mathbb{C}[x]$. Define

$$\begin{aligned} v(f, g) &= \text{the common roots of } f(x) \text{ and } g(x) \text{ counted with multiplicities} \\ &= \deg\left(\gcd\left(f(x), g(x)\right)\right). \end{aligned}$$

Definition 65 (Bézoutian or Bézout matrix)**.** Let $f, g \in \mathbb{C}[x]$ be two polynomials and let $n = \max\{\deg f, \deg g\}$. Then the Bézoutian $B_{f,g} = (b_{ij}) \in M_n(\mathbb{C})$ is a matrix such that

$$\frac{f(x)g(y) - f(y)g(x)}{x - y} = \sum_{i=0}^{n-1}\sum_{j=0}^{n-1} x^i \cdot b_{ij} \cdot y^j = \begin{pmatrix} 1 & x & \cdots & x^{n-1} \end{pmatrix} \cdot B_{f,g} \cdot \begin{pmatrix} 1 \\ y \\ \vdots \\ y^{n-1} \end{pmatrix}.$$

If we define

$$V_n(x) = \begin{pmatrix} 1 & x & \cdots & x^{n-1} \end{pmatrix}^{\mathrm{t}},$$

then we have

$$\frac{f(x)g(y) - f(y)g(x)}{x - y} = V_n(x)^{\mathrm{t}} \cdot B_{f,g} \cdot V_n(y).$$

Theorem 66. *Let $f, g \in \mathbb{C}[x]$. We have*

$$v(f, g) = \operatorname{nullity}(B_{f,g}).$$

Before proving this theorem, we shall introduce some notations, some lemmas and an important theorem.

Definition 67. Given

$$f(x) = \sum_{i=0}^{n} a_i x^i \in \mathbb{C}[x] \text{ and } g(x) = \sum_{i=0}^{n} b_i x^i \in \mathbb{C}[x].$$

Define the following matrices associated with $f(x)$ (and possibly $g(x)$).

1. The Hankel matrix of $f(x)$.
   We usually write $H_f = (h_{ij}) \in M_n(\mathbb{C})$ to denote it. The entries are defined by:

$$h_{ij} = \begin{cases} a_{i+j-1} & \text{, if } i+j-1 < n \\ 0 & \text{, otherwise} \end{cases}.$$

   That is,

$$H_f = \begin{pmatrix} a_1 & a_2 & \cdots & \cdots & a_n \\ a_2 & \ddots & \ddots & a_n & \\ \vdots & \ddots & a_n & & \\ \vdots & a_n & & & \\ a_n & & & & \end{pmatrix} \in M_n(\mathbb{C}).$$

2. The Toeplitz matrix of $f(x)$.
   We usually write $T_f = (t_{ij}) \in M_n(\mathbb{C})$ to denote it. The entries are defined by:

$$t_{ij} = \begin{cases} a_{j-i} & \text{, if } i \leq j \\ 0 & \text{, otherwise} \end{cases}.$$

   That is

$$T_f = \begin{pmatrix} a_0 & a_1 & \cdots & a_{n-2} & a_{n-1} \\ & a_0 & a_1 & \cdots & a_{n-2} \\ & & \ddots & \ddots & \vdots \\ & & & a_0 & a_1 \\ & & & & a_0 \end{pmatrix} \in M_n \in \mathbb{C}.$$

3. The anti-diagonal matrix.
   We usually write $Z_n = (z_{ij}) \in M_n(\mathbb{C})$ to denote it. The entries are defined by

$$z_{ij} = \begin{cases} 1 & \text{, if } i+j = n+1 \\ 0 & \text{, otherwise} \end{cases}.$$

   That is,

$$Z_n = \begin{pmatrix} & & & 1 \\ & & \iddots & \\ & \iddots & & \\ 1 & & & \end{pmatrix} \in M_n\mathbb{C}.$$

4. The resultant of f and g.
   Let $n = \max\{\deg f, \deg g\}$. We write $R_{f,g} = (r_{ij}) \in M_{2n}(\mathbb{C})$ to denote the resultant. The entries are defined by

$$r_{ij} = \begin{cases} a_{j-i} & \text{, if } i \leq n \text{ and } 0 \leq j-i \leq n \\ b_{j-i} & \text{, if } i > n \text{ and } 0 \leq j-i \leq n \\ 0 & \text{, otherwise} \end{cases}.$$

That is,

$$R_{f,g} := \begin{pmatrix} a_0 & a_1 & \cdots & a_{n-1} & a_n & & & \\ & a_0 & a_1 & \cdots & a_{n-1} & a_n & & \\ & & \ddots & \ddots & \ddots & \ddots & \ddots & \\ & & & a_0 & a_1 & \cdots & a_{n-1} & a_n \\ b_0 & b_1 & \cdots & b_{n-1} & b_n & & & \\ & b_0 & b_1 & \cdots & b_{n-1} & b_n & & \\ & & \ddots & \ddots & \ddots & \ddots & \ddots & \\ & & & b_0 & b_1 & \cdots & b_{n-1} & b_n \end{pmatrix} \in M_{2n}(\mathbb{C}).$$

Remark. In all the above matrices, all "blank" entries represent 0.

From the above definitions, it is easy to see

$$Z \cdot H_f = \begin{pmatrix} a_n & & & \\ a_{n-1} & a_n & & \\ \vdots & \ddots & \ddots & \\ a_1 & \cdots & a_{n-1} & a_n \end{pmatrix}; \quad Z \cdot T_f = \begin{pmatrix} & & & a_0 \\ & & a_0 & a_1 \\ & \iddots & \iddots & \vdots \\ a_0 & a_1 & \cdots & a_{n-1} \end{pmatrix}.$$

Thus, we have

$$R_{f,g} = \left(\begin{array}{c|c} T_f & Z \cdot H_f \\ \hline T_g & Z \cdot H_g \cdot \end{array}\right).$$

Theorem 68. *Let $f, g \in \mathbb{C}[x]$. We have*

$$v(f, g) = \text{nullity}(R_{f,g}).$$

Proof. Let $n := \max\{\deg f, \deg g\}$ and let P_k be the set of all complex polynomials with degree less than k. In other words,

$$P_k := \{p \in \mathbb{C}[x] : \deg p < k\}.$$

Consider a linear transformation T defined by:

$$\begin{array}{rl} T : P_n \oplus P_n \to & P_{2n} \\ (u, v) \mapsto & u \cdot f + v \cdot g \end{array}.$$

Suppose $d(x) = \gcd(f(x), g(x))$, and we assume that $f(x) = h(x) \cdot d(x)$, $g(x) = k(x) \cdot d(x)$, and that $\gcd(h(x), k(x)) = 1$. Then,

$$\begin{aligned} \ker T &= \{(u, v) \in P_n \times P_n : u \cdot f + v \cdot g = 0\} \\ &= \{(u, v) \in P_n \times P_n : u \cdot h + v \cdot k = 0\} \\ &= \{(k \cdot \alpha, -h \cdot \alpha) : \alpha \in \mathbb{C}[x]\} \ (\because \gcd(k, h) = 1.) \end{aligned}$$

However, note the degree of α is less than $\deg d(x)$. This indicates that

$$\dim \ker T = \deg d(x) = v(f, g).$$

It suffices to show that $\dim \ker T = \text{nullity}(R_{f,g})$. It follows from the fact that

$$[T]_{\mathcal{B},\mathcal{A}} = {R_{f,g}}^{\mathrm{t}},$$

where $\mathcal{B}$ is the standard basis of $P_n \oplus P_n$ and $\mathcal{A}$ is the standard basis of P_{2n}. □

To prove Theorem 66, it remains to find the relation between $R_{f,g}$ and $B_{f,g}$.

Lemma 9. Let $f, g \in \mathbb{C}[x]$ and let $n = \max\{\deg f, \deg g\}$. H_f, T_f and Z are defined as above.

1. T_f and T_g commute, namely, $T_f \cdot T_g = T_g \cdot T_f$.
2. $X^{\mathrm{t}} = Z \cdot X \cdot Z$, for all $X \in M_n(\mathbb{C})$.
3. $H_f \cdot Z \cdot H_g = H_g \cdot Z \cdot H_f$.

The proof is omitted since it can be done by some simple calculations.

Lemma 10. Let $f, g \in \mathbb{C}[x]$. Then, $B_{f,g} = H_f \cdot T_g - H_g \cdot T_f$.

Proof. We write $R = R_{f,g}$ and $B = B_{f,g}$. It is easy to see that

$$\frac{x^n - y^n}{x - y} = V_n(x)^{\mathrm{t}} \cdot Z \cdot V_n(y).$$

Thus, we have

$$\begin{aligned}
&(x^n - y^n) \cdot \frac{f(x)g(y) - f(y)g(x)}{x - y} \\
=& V_n(x)^{\mathrm{t}} \cdot \Big(Z \cdot \big(f(x)g(y) - f(y)g(x)\big)\Big) \cdot V_n(y) \\
=& \begin{pmatrix} V_n(x) \cdot f(x) \\ V_n(x) \cdot g(x) \end{pmatrix}^{\mathrm{t}} \cdot \begin{pmatrix} 0 & Z \\ -Z & 0 \end{pmatrix} \cdot \begin{pmatrix} f(y) \cdot V_n(y) \\ g(y) \cdot V_n(y) \end{pmatrix} \\
=& V_{2n}(x)^{\mathrm{t}} \cdot R^{\mathrm{t}} \cdot \begin{pmatrix} 0 & Z \\ -Z & 0 \end{pmatrix} \cdot R \cdot V_{2n}(y) \text{ (by direct computation.)}
\end{aligned}$$

On the other hand, we have the left hand side is equal to

$$\begin{aligned}
&(x^n - y^n) \cdot V_n(x)^{\mathrm{t}} \cdot B_{f,g} \cdot V_n(y) \\
=& V_n(x)^{\mathrm{t}} \cdot \big(x^n \cdot B_{f,g}\big) \cdot V_n(y) - V_n(x)^{\mathrm{t}} \cdot \big(B_{f,g} \cdot y^n\big) \cdot V_n(y) \\
=& V_{2n}(x)^{\mathrm{t}} \cdot \begin{pmatrix} 0 & 0 \\ B & 0 \end{pmatrix} \cdot V_{2n}(y) - V_{2n}(x)^{\mathrm{t}} \cdot \begin{pmatrix} 0 & B \\ 0 & 0 \end{pmatrix} \cdot V_{2n}(y) \\
=& V_{2n}(x)^{\mathrm{t}} \cdot \begin{pmatrix} 0 & -B \\ B & 0 \end{pmatrix} \cdot V_{2n}(y).
\end{aligned}$$

Therefore, we conclude that

$$\begin{aligned}
& \begin{pmatrix} 0 & -B \\ B & 0 \end{pmatrix} = R^{\mathrm{t}} \cdot \begin{pmatrix} 0 & Z \\ -Z & 0 \end{pmatrix} \cdot R \\
\Longrightarrow & \begin{pmatrix} 0 & -B \\ B & 0 \end{pmatrix} = \begin{pmatrix} T_f{}^{\mathrm{t}} & T_g{}^{\mathrm{t}} \\ H_f{}^{\mathrm{t}} \cdot Z & H_g{}^{\mathrm{t}} \cdot Z \end{pmatrix} \cdot \begin{pmatrix} 0 & Z \\ -Z & 0 \end{pmatrix} \cdot \begin{pmatrix} T_f & Z \cdot H_f \\ T_g & Z \cdot H_g \end{pmatrix} \\
\Longrightarrow & \begin{pmatrix} 0 & -B \\ B & 0 \end{pmatrix} = \begin{pmatrix} Z \cdot T_f \cdot Z & Z \cdot T_g \cdot Z \\ H_f \cdot Z & H_g \cdot Z \end{pmatrix} \cdot \begin{pmatrix} Z \cdot T_g & H_g \\ -Z \cdot T_f & -H_f \end{pmatrix} \\
\Longrightarrow & \begin{pmatrix} 0 & -B \\ B & 0 \end{pmatrix} = \begin{pmatrix} 0 & * \\ X & 0 \end{pmatrix}, \text{ where } X = H_f \cdot T_g - H_g \cdot T_f \\
\Longrightarrow & B = X = H_f \cdot T_g - H_g \cdot T_f.
\end{aligned}$$

The above deductions have used Lemma 9. $\square$

Now, we can start proving Theorem 66.

Proof of Theorem 66. It suffices to show that $\text{nullity}(R_{f,g}) = \text{nullity}(B_{f,g})$. Without loss of generality, we assume that $\deg f \geq \deg g$. Hence, $a_n \neq 0$ if $n = \max\{\deg f, \deg g\}$. Consider

$$\begin{aligned}
\begin{pmatrix} I_n & O_n \\ T_f & ZH_f \end{pmatrix} \cdot R &= \begin{pmatrix} I_n & O_n \\ T_f & ZH_f \end{pmatrix} \cdot \begin{pmatrix} T_f & ZH_f \\ T_g & ZH_g \end{pmatrix} \\
&= \begin{pmatrix} T_f & ZH_f \\ T_f^2 + ZH_f T_g & T_f ZH_f + ZH_f ZH_g \end{pmatrix} \\
&= \begin{pmatrix} T_f & ZH_f \\ ZB + (ZH_g + T_f)T_f & (T_f + ZH_g)ZH_f \end{pmatrix} \\
&\qquad \text{(Recall that } B = H_f T_g - H_g T_f \text{ and Lemma 9)} \\
&= \begin{pmatrix} O_n & I_n \\ ZB & T_f + ZH_g \end{pmatrix} \cdot \begin{pmatrix} I_n & O_n \\ T_f & ZH_f \end{pmatrix} \\
&= \begin{pmatrix} O_n & I_n \\ Z & T_f + ZH_g \end{pmatrix} \cdot \begin{pmatrix} B & O_n \\ O_n & I_n \end{pmatrix} \cdot \begin{pmatrix} I_n & O_n \\ T_f & ZH_f \end{pmatrix}.
\end{aligned}$$

Note that

$$\begin{aligned}
\det \begin{pmatrix} I_n & O_n \\ T_f & ZH_f \end{pmatrix} &= \det(I_n) \cdot \det(Z \cdot H_f) \\
&= \det(Z) \cdot \det(H_f) \\
&= (-1)^n \cdot \det({H_f}^{\mathrm{t}}) \\
&= (-1)^n \cdot (a_n)^n \neq 0
\end{aligned}$$

and that

$$\begin{aligned}
\det \begin{pmatrix} O_n & I_n \\ Z & T_f + ZH_g \end{pmatrix} &= \det \begin{pmatrix} O_n & I_n \\ Z & T_f + ZH_g \end{pmatrix} \\
&= \det \begin{pmatrix} O_n & I_n \\ Z & T_f \end{pmatrix} \cdot \det \begin{pmatrix} I_n & H_g \\ O_n & I_n \end{pmatrix} \\
&= \det(T_f) \cdot \det(I_n T_f^{-1} Z) \cdot 1 \neq 0
\end{aligned}$$

Therefore, we conclude that

$$\text{nullity}(B) = \text{nullity} \begin{pmatrix} B & O_n \\ O_n & I_n \end{pmatrix} = \text{nullity}(R_{f,g}).$$

□

Theorem 66 together with Theorem 68 are called Jacobi-Darboux Theorem.

Remark. If we are given two polynomials $p, q \in \mathbb{C}[x, y]$ and we are asked to find all solutions to the equation

$$p(x, y) = 0, \qquad q(x, y) = 0.$$

We can use the following method. Fix y and we obtain two polynomials p_y and q_y with coefficients in $\mathbb{C}$. Then, (x_0, y_0) is a solution if $\det(B_{p_{y_0}, q_{y_0}}) = 0$.

3.2 Markov chain and the Perron-Frobenius Theorem

We first look at the following question:

> Suppose there are only two towns in the NTU COUNTRY, called the MATH town and the CSIE town. Suppose in every year, there are $s\%$ people from MATH moving to CSIE; and $t\%$ people from CSIE moving to MATH. Assume that there are no people died and born and no people moving out of the NTU COUNTRY. Then, we want to ask whether the population in these two towns will become steady.

Let S be the total population of the NTU COUNTRY, and let p_k and q_k be the percentage of the total population in two towns MATH and CSIE, respectively, at the k-th year. Write $v_k = (p_k, q_k)^{\mathrm{t}}$. Then, we have

$$v_{k+1} = \begin{pmatrix} 1 - s\% & t\% \\ s\% & 1 - t\% \end{pmatrix} v_k$$

Define

$$M := \begin{pmatrix} 1 - s\% & t\% \\ s\% & 1 - t\% \end{pmatrix},$$

then we wonder whether the limit

$$\lim_{k \to \infty} v_k = \lim_{k \to \infty} M v_k$$

exists? Above discussions give us the motivation to study Markov chain. The next two definitions are helpful for us rephrasing the problem.

Definition 69 (Steady state)**.** Given a matrix $M \in M_n(\mathbb{R})$, a steady state $v \in \mathbb{R}^n$ is an eigenvector of M with eigenvalue 1, namely, $M \cdot v = v$.

Definition 70 (Stochastic matrix)**.** Suppose $M \in M_n(\mathbb{R})$. $M = (m_{ij})$ is called a stochastic matrix if all its entries are nonegative and

$$\sum_{i=1}^{n} m_{ij} = 1,$$

for all $j \in [1, n]$.

We restate the problem as

> *Is the steady state of a stochastic matrix exists and is unique up to a scalar?*

In general, the answer is “no”. For instance, let $M = I_n$, then every state is a steady state. So, our goal is to find the sufficient condition when the steady state is unique.

Definition 71 (Positive matrix and non-negative matrix)**.** Given a matrix $M \in M_n(\mathbb{R})$.

1. M is positive (non-negative) if all its entries are positive (non-negative). We often write $M > 0$ or $M \geq 0$.

2. M is regular if M is non-negative and M^k is positive for some $k \in \mathbb{N}$. (The terminology “Regular” is sometimes confusing.)

Theorem 72. *Let $M \in M_n(\mathbb{R})$ be a stochastic matrix. If M is regular then a steady state of M is unique up to a scalar. In other words,* $\dim\ker(M - I_n) = 1$.

In fact, there is a more stronger result, however we shall introduce some other terminologies first.

Definition 73 (Spectral radius)**.** Let $A \in M_n(\mathbb{C})$ and let $\lambda_1, \lambda_2, \dots, \lambda_s$ be all the eigenvalues of A (roots of the characteristic polynomial). The spectral radius of A is defined as

$$\rho(A) := \max_{1 \le i \le s} |\lambda_i| .$$

Hence, we have $\rho(A) \ge 0$.

The stronger result mentioned above is the next theorem, which is proved by Oskar Perron (1907) and Georg Frobenius (1912).

Theorem 74 (Perron-Frobenius Theorem)**.** *Let $A \in M_n(\mathbb{R})$ be a regular matrix. Then, there exists a unique (up to a scalar) eigenvector $v \in \mathbb{R}^n$ with eigenvalue $\rho(A)$.*

Note that we does not assume $\rho(A)$ is an eigenvalue. Therefore, this theorem is pretty strong. Since it requires a lot of work to prove Theorem 74, we shall prove some theorems and lemmas first, instead.

Theorem 75 (Gelfond's formula)**.** *Let $A \in M_n(\mathbb{C})$. Then,*

$$\rho(A) = \lim_{k \to \infty} \left\| A^k \right\|^{1/k} .$$

Although this theorem is regard as a lemma of Theorem 74, we still need to decompose it into some small problems.

Lemma 11. Let A and B be two similar complex matrices. That is, there exists an invertible matrix $P \in M_n(\mathbb{C})$ such that

$$A = P^{-1} B P.$$

Then,

$$\lim_{k \to \infty} \left\| A^k \right\|^{1/k} = \lim_{n \to \infty} \left\| B^k \right\|^{1/k} ,$$

provided that $\lim \left\| A^k \right\|^{1/k}$ exists.

Proof. Let $t = \|P\| \cdot \|P^{-1}\| \ge \|P \cdot P^{-1}\| = 1$. Then,

$$\left\| A^k \right\| = \left\| P^{-1} \cdot B^k \cdot P \right\| \le \left\| P^{-1} \right\| \cdot \left\| B^k \right\| \cdot \|P\| = t \cdot \left\| B^k \right\| .$$

Similarly, we have

$$\left\| B^k \right\| \le t \cdot \left\| A^k \right\| .$$

We conclude

$$t^{-1/k} \left\| A^k \right\|^{1/k} \le \left\| B^k \right\|^{1/k} \le t^{1/k} \left\| A^k \right\|^{1/k} .$$

Taking the limit $k \to \infty$, we obtain $\lim \left\| A^k \right\|^{1/k} = \lim \left\| B^k \right\|^{1/k}$. $\square$

Proof of Theorem 75. If x is an eigenvector of eigenvalue λ, then

$$\begin{aligned} \left|A^k x\right| = |\lambda|^k \cdot |x| &\implies \left\|A^k\right\| \geq |\lambda|^k \\ &\implies \left\|A^k\right\|^{1/k} \geq |\lambda| . \end{aligned}$$

We find $\left\|A^k\right\|^{1/k} \geq \rho(A)$ for all $k \in \mathbb{N}$. It remains to prove that $\lim \left\|A^k\right\|^{1/k}$ exists and

$$\rho(A) \geq \lim_{k\to\infty} \left\|A^k\right\|^{1/k} .$$

From what we have learnt in the theory of Jordan forms and Lemma 11, we just need to consider the case when A is of Jordan form. We first consider the case when A is a Jordan block J_λ, that is,

$$A = \begin{pmatrix} \lambda & 1 & & & \\ & \lambda & 1 & & \\ & & \ddots & \ddots & \\ & & & \ddots & 1 \\ & & & & \lambda \end{pmatrix} = \lambda \cdot I_n + N \in M_n(\mathbb{C}),$$

where $N = J_0$. Then,

$$A^k = (\lambda \cdot I_n + N)^k = \sum_{i=0}^{k} \binom{k}{i} \lambda^{k-i} N^i = \sum_{i=0}^{n} \binom{k}{i} \lambda^{k-i} N^i \qquad (\text{if } k \geq n)$$

If we assume $k \geq n$, we then have

$$\left\|A^k\right\| = \left\| \sum_{i=0}^{n} \binom{k}{i} \lambda^{k-i} N^i \right\| \leq \sum_{i=0}^{n} \binom{k}{i} |\lambda|^{k-i} = |\lambda|^k \cdot p(k),$$

where

$$p(k) = \sum_{i=0}^{n} |\lambda|^{-i} \binom{k}{i}$$

is a polynomial in k. Thus,

$$\left\|A^k\right\|^{1/k} \leq |\lambda| \cdot p(k)^{1/k} \to |\lambda| .$$

Estimations above show that the theorem is true when A is a Jordan block. Now, we claim that if

$$A = B \oplus C = \begin{pmatrix} B & \\ & C \end{pmatrix},$$

then $\|A\| = \max\{\|B\|, \|C\|\}$. This claim proves the theorem, since $\left\|A^k\right\|^{1/k}$ converge to $\max_{1\leq i\leq s}\{|\lambda_i|\}$ when

$$A = \begin{pmatrix} J_1 & & & \\ & J_2 & & \\ & & \ddots & \\ & & & J_s \end{pmatrix}, \quad J_i \text{ are all Jordan blocks.}$$

We now start proving the claim. Let $B \in M_p(\mathbb{C})$ and $C \in M_q(\mathbb{C})$ and let $a = \max\{\|B\|, \|C\|\}$. Observe that for all $x \in \mathbb{C}^p$ and $y \in \mathbb{C}^q$, we have

$$\left|\begin{pmatrix} B & 0 \\ 0 & C \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}\right| = \sqrt{|Bx|^2 + |Cy|^2} \leq \sqrt{a^2 \cdot (|x| + |y|)}.$$

Hence, we conclude that

$$\left\|\begin{pmatrix} B & 0 \\ 0 & C \end{pmatrix}\right\| \leq a.$$

On the other hand, there exist $x_0 \in \mathbb{C}^p$ and $y_0 \in \mathbb{C}^q$ such that

$$|Bx_0| = \|B\| \cdot |x_0|, \qquad |Cy_0| = \|C\| \cdot |y_0|.$$

Then, we have

$$\left|\begin{pmatrix} B & 0 \\ 0 & C \end{pmatrix}\begin{pmatrix} x_0 \\ 0 \end{pmatrix}\right| \leq \|B\| \cdot \left|\begin{pmatrix} x_0 \\ 0 \end{pmatrix}\right|, \qquad \left|\begin{pmatrix} B & 0 \\ 0 & C \end{pmatrix}\begin{pmatrix} 0 \\ y_0 \end{pmatrix}\right| \leq \|C\| \cdot \left|\begin{pmatrix} 0 \\ y_0 \end{pmatrix}\right|,$$

this proves the theorem. □

Theorem 76. *Let $A \in M_n(\mathbb{R})$ be a positive real matrix. The following statements are true.*

1. *There is a positive vector $u \in \mathbb{R}^n_{>0}$ such that $A \cdot u = \rho(A)u$.*
2. *If $v \in \mathbb{C}^n$ is an eigenvector of A with eigenvalue λ satisfying $|\lambda| = \rho(A)$, then $\lambda = \rho(A)$.*
3. *The algebraic multiplicities of $\rho(A)$ is 1.*

Proof. Let $v \in \mathbb{C}^n$ be an eigenvector of A with eigenvalue λ satisfying $|\lambda| = \rho(A)$. Write $v = \begin{pmatrix} v_1 & v_2 & \dots & v_n \end{pmatrix}^{\mathrm{t}} \in \mathbb{C}^n$. `To Be Completed...` □

References

[1] S.H. Friedberg, A.J. Insel, and L.E. Spence. *Linear Algebra*. Featured Titles for Linear Algebra (Advanced) Series. Pearson Education, 2003. ISBN: 9780130084514.

[2] 謝銘倫. 線性代數二. https://www.youtube.com/playlist?list=PLVJXJebpO4PgW5WgeU_hIml7mSO8Eoggx. 2016.