Linear Algebra II

Minerva

2020-2nd

Contents

Chapter 1

Jordan form and rational form

1.1 What is a Jordan form?

1.1.1 Motivation

We assume that $\dim V \leq \infty, T : V \to V$ and $ch_T(x) = \prod_{i=1}^{k}(x - \lambda_i)^{m_i}$ splits over F. Previously, we have consider the case where $\dim E_{\lambda_i} = m_i$ for all i. In such a case, there exists a basis $\mathcal{B}$ for V such that $[T]_{\mathcal{B}}$ is diagonal. However, if $\dim E_{\lambda_i} < m_i$ for some i, can we find a "nice basis" such that $[T]_{\mathcal{B}}$ is "simple" and easy to do computation using it?

We give the result first.

1.1.2 Goal

If $ch_T(x)$ splits over F, then there exists a basis $\mathcal{B}$ such that $[T]_{\mathcal{B}}$ is of the form

$$[T]_{\mathcal{B}} = \begin{pmatrix} A_1 & & & O \\ & A_2 & & \\ & & \ddots & \\ O & & & A_k \end{pmatrix}$$

where

$$A_i = \begin{pmatrix} \lambda_i & 1 & & & O \\ 0 & \lambda_i & 1 & & \\ & & \ddots & \ddots & \\ & & & \lambda_i & 1 \\ O & & & & \lambda_i \end{pmatrix} \tag{$*$}$$

Such a matrix $[T]_{\mathcal{B}}$ is called a **Jordan normal form** (or **Jordan canonical form**) of T. Each matrix of the form $(*)$ is called a **Jordan block** and $\mathcal{B}$ is called a **Jordan canonical basis**. Notice that a $[T]_{\mathcal{B}}$ is easy to do computation.

- Since $[T]_{\mathcal{B}}$ is block diagonal matrix, we have

$$[T]_{\mathcal{B}}^n = \begin{pmatrix} A_1^n & & & O \\ & A_2^n & & \\ & & \ddots & \\ O & & & A_k^n \end{pmatrix}$$

- $A_i^k = (\lambda_i I + N)^k$ where N is matrix with entry on counter diagonal is 1 and others are 0.
  - $\lambda_i I$ and N commute
  - $N^{m_i} = 0$

  So we have

$$A_i^k = (\lambda_i + N)^k = \sum_{j=0}^{k} \binom{k}{j} \lambda_i^{k-j} N^j = \sum_{0 \le j < m_i} \binom{k}{j} \lambda_i^{k-j} N^j$$
$$= \begin{pmatrix} z_{0k} & z_{1k} & z_{2k} & \cdots & \\ 0 & z_{0k} & \ddots & & \\ & & \ddots & & \vdots \\ & & & z_{0k} & z_{1k} \\ O & & & 0 & z_{0k} \end{pmatrix} \quad \left(\text{where } z_{jk} = \binom{k}{j} \lambda_i^{k-j}\right)$$

1.2 Review

In this section, we will review the knowledge what we had learned and we will use it when we prove Jordan normal form.

1.2.1 T-invariant subspace

Definition 1.2.1 (T-invariant subspace)**.** If $T : V \to V$ is a linear operator. A subspace of W of V is a **T-invariant subspace** if $T(W) \subseteq W$

Example 1.2.1. For any $f(x) \in F[x]$, $\ker f(T)$ is a T-invariant subspace.

Definition 1.2.2. Let $v \in V$. The subspace $Z(v; T) := \operatorname{span}\{T^k(v) : k \in \mathbb{N}_0\}$ is called the **cyclic T-invariant subspace** generated by v

Theorem 1.2.1. If $k = \dim Z(v; T) < \infty$, then

- $\{v, T(v), ..., T^{k-1}(v)\}$ is a basis for $Z(v; T)$
- If $a_0 v + a_1 T(v) + \cdots + a_{k-1} T^{k-1}(v) + T^k(v) = 0$, then

$$ch_{T|_{Z(v;T)}} = x^k + a_{k-1} x^{k-1} + \cdots a_1 x + a_0$$

Theorem 1.2.2. Assume $\dim V \le \infty$. Let W be a T-invariant subspace of V. Then $ch_W | ch_T$

1.2.2 Direct sum

Definition 1.2.3. Let $W_1, ..., W_k$ be subspace of V. We say V is the **direct sum** of $W_1, ..., W_k$ if $V = W_1 + W_2 + \cdots W_k$ and $W_i \cap \sum_{j \ne i} W_j = \{0\}$
If V is direct sum of $W_1, ..., W_k$, we write $V = W_1 \oplus W_2 \oplus \cdots \oplus W_k$

Property 1.2.1. TFAE

- $V = W_1 \oplus \cdots \oplus W_k$
- $V = W_1 + \cdots + W_k$ and if $v_1 + \cdots + v_k = 0$ then $v_j = 0\ \forall j$.

- Each $v \in V$ can be written as $v = v_1 + \cdots + v_k$ for some $v_j \in W_j$ uniquely.
- If $\mathcal{B}_j$ is a basis for W_j, $j = 1, \ldots, k$, then $\mathcal{B} = \bigcup_{j=1}^{k} \mathcal{B}_j$ is a basis for V.
- $\exists$ basis $\mathcal{B}_j$ for W_j such that $\mathcal{B} = \bigcup_{j=1}^{k} \mathcal{B}_j$ is a basis for V.

Theorem 1.2.3. Assume that $\dim V < \infty$ and $V = W_1 \oplus \cdots \oplus W_k$. Then

$$ch_T(x) = \prod_{i=1}^{k} ch_{T|_W}(x)$$

Also, if $\mathcal{B}_i$ is a basis for W_i, then $\mathcal{B} = \bigcup\limits_{i=1}^{k} \mathcal{B}_i$ is a basis for V and

$$[T]_{\mathcal{B}} = \begin{pmatrix} [T|_{W_1}]_{\mathcal{B}_1} & & \\ & \ddots & \\ & & [T|_{W_k}]_{\mathcal{B}_k} \end{pmatrix}$$

Remark 1.2.1. Thus, to prove that a Jordan form exists for T. We will prove that $\exists T$-invariant subspaces $W_1, W_2, ..., W_k$ such that $V = W_1 \oplus \cdots W_k$ and each W_i has a basis $\mathcal{B}_i$ such that

$$[T|_{W_i}]_{\mathcal{B}_i} = \begin{pmatrix} \lambda_i & 1 & & & O \\ 0 & \lambda_i & 1 & & \\ & & \ddots & \ddots & \\ & & & \lambda_i & 1 \\ O & & & & \lambda_i \end{pmatrix}$$

1.2.3 Polynomial rings

Theorem 1.2.4. If $f(x), g(x) \in F[x]$ and $g(x) \neq 0$, then there exists unique polynomial $q(x)$ and $r(x)$ such that

$$f(x) = q(x)g(x) + r(x)$$

and $r(x) = 0$ or $\deg r(x) < \deg g(x)$ (which means $F[x]$ is ED)

Definition 1.2.4. A nonempty set I of $F[x]$ is said to be an **ideal** if

- $f(x), g(x) \in I \implies f(x) - g(x) \in I$
- If $g(x) \in I$, then $f(x)g(x) \in I \ \forall f(x) \in F[x]$

Example 1.2.2. Let $T : V \to V$ be a linear operator.

- $$I = \{f(x) \in F[x] : f(T) = 0\}$$
  - If $f(x), g \in I$ i.e. $f(T) = g(T) = 0$, then $f(T) - g(T) = 0 \implies f - g \in \ I$
  - If $g(x) \in I$ i.e. if $g(T) = 0$, then $f(T) \cdot g(T) = 0 \implies fg \in I$

  Hence, I is an ideal in $F[x]$.

- Given $v \in V$, the set
$$I_T(v) = \{f(x) \in F[x] : f(T)v = 0\}$$
is an ideal
- Let W be a T-invariant subspace, then
$$I_{v,W} \text{ or } I_T(v, W) := \{f(x) \in F[x] : f(T)v \in W\}$$
is an ideal.

Theorem 1.2.5. If I is an ideal of $F[x]$, then $\exists$ a polynomial $g(x) \in F[x]$ such that
$$I = \{f(x)g(x) : f(x) \in F[x]\} = (g(x))$$
(Which means $F[x]$ is PID.)

Remark 1.2.2. Note that $g(x)$ has the smallest degree among all nonzero elements of I.

Definition 1.2.5 (principal ideal)**.** We say I is the **principal ideal** generated by $g(x)$ if $I = (g(x))$

Remark 1.2.3. $T : V \to V$. Recall that $I = \{f(x) \in F[x] | f(T) = 0\}$ is a ideal. Then the minimal polynomial $m_T(x)$ is defined to be the monic polynomial that generates I.

Definition 1.2.6. Let $f(x), g(x) \in F[x]$. If $h(x)$ is a polynomial such that $(h(x)) = (f(x)) + (g(x))$, then we say $g(x)$ is a **greatest common divisor**(GCD) of $f(x)$ and $g(x)$. If $(f(x)) + (g(x)) = (1)$, then we say f, g are relatively prime.

Definition 1.2.7. A nonconstant polynomial $f(x) \in F[x]$ is **irreducible** over if "$f(x) = g(x)h(x)$ for $g(x), h(x) \in F[x]$, then one of $g(x), h(x)$ is constant"

Compare $\mathbb{Z}$ and $F[x]$

	$\mathbb{Z}$	$F[x]$
ideal	$n\mathbb{Z}$	$(g(x))$
GCD	$n\mathbb{Z} + m\mathbb{Z} = \gcd(m, n)\mathbb{Z}$	$(f) + (g) = (\gcd(f, g))$
irreducible	prime number	irreducible polynomial
prime	$p\|ab \rightsquigarrow p\|a$ or $p\|b$	f : irr, $f\|gh \rightsquigarrow f\|g$ or $f\|h$
UFD	Fundamental theorem of arithmetic	$f(x) = ap_1(x)^{n_1} \cdots p_k(x)^{n_k}$ with unique decomposition

1.2.4 Kernel decomposition theorem

Theorem 1.2.6. (kernel decomposition theorem) $t : V \to V$. If $f(x)$ and $g(x)$ are relatively prime, then
$$\ker f(T)g(T) = \ker f(T) \oplus \ker g(T)$$

Corollary 1.2.1. Assume $\dim V < \infty$. Then $T : V \to V$ is diagonalizable $\iff$ $m_T(x)$ splits into a product of distinct linear factors over F.

1.3 Generalized eigenspace

1.3.1 Motivation and definition

We continue to prove Jordan form. Assume that $\dim V < \infty$. $T : V \to V$ and $ch_T(x)$ splits over F, say

$$ch_T(x) = \prod_{i=1}^{k} (x - \lambda_i)^{m_i}$$

where λ_i are distinct.
By Cayley-Hamilton theorem ($ch_T(T) = 0$).

$$V = \ker ch_T(T) = \ker \prod_{i=1}^{k} (T - \lambda_i I)^{m_i}$$

Then by the kernel decomposition theorem

$$V = \ker(T - \lambda_1)^{m_1} \oplus \cdots \oplus \ker(T - \lambda_k I)^{m_k} = \bigoplus_{i=1}^{k} \ker(T - \lambda_i I)^{m_i}$$

So we want to research the property of $\ker(T - \lambda_i I)^{m_i}$
Claim: If $v \in \ker(T - \lambda_i I)^{m_i}$, then

$$I = \{f(x) \in F[x], f(T)v = 0\} = ((x - \lambda_i)^p)$$

for some $p \leq m_i$

Proof: Assume $g(x)$ is a polynomial such that $I = (g(x))$. Now, $v \in \ker(T - \lambda_i I)^{m_i} \implies (T - \lambda_i I)^{m_i} v = 0 \implies (x - \lambda_i)^{m_i} \in I \implies g(x) \big| (x - \ \lambda)^{m_i} \implies g(x) = (x - \lambda_i)^p$ for some $p \leq m_i$ □

Definition 1.3.1. Let λ be an eigenvalue of $T : V \to V$, ($\dim V$ may be ∞). The set

$$K_\lambda = \{v \in V : (T - \lambda I)^p v = 0 \text{ for some } p \geq 1\}$$

is called the **generalized eigenspace** corresponding to λ. A nonzero element v in K_λ is called a **generalized eigenvector**.

Example 1.3.1. $\begin{array}{cccc} T: & F[x] & \longrightarrow & F[x] \\ & f & \longmapsto & f' \end{array} \implies K_0 = F[x]$

Theorem 1.3.1. Let λ be an eigenvalue of $T : V \to V$ ($\dim V$ may be ∞). Then

(i) K_λ is a T-invariant subspace of V

(ii) For any $\mu \neq \lambda$, the restriction of $(T - \mu I)$ to K_λ is $1 - 1$

Remark 1.3.1. If V is finite dimensional, say λ has multiplicity m, then $K_\lambda = \ker(T - \lambda I)^m$ and K_λ is a T-invariant subspace since $\ker f(T)$ is T-invariant for any $f(x) \in F[x]$

Proof:

(i) We first prove that K_λ is a subspace.

Suppose that $v_1, v_2 \in K_\lambda$, say $(T-\lambda I)^{p_1} v_1 = (T-\lambda I)^{p_2} = 0$, let $p = \max(p_1, p_2)$, then $(T-\lambda I)^p(v_1 + cv_2) = 0$ and it clear $0 \in K_\lambda$

Hence, K_λ is a subspace and it is clear that K_λ is T-invariant subspace.

(ii) We need to show $\ker(T-\mu I) \cap K_\lambda = \{0\}$

Suppose thta $v \in \ker(T-\mu I) \cap K_\lambda$, say $(T-\lambda I)^p v = 0$

Since $(x-\mu)$ and $(x-\lambda)^p$ are relatively prime i.e. $(x-\mu) + ((x-\lambda)^p) = (1)$

$\implies \exists\, a(x), b(x) \in F[x]$ s.t. $1 = a(x)(x-\mu) + b(x)(x-\lambda)^p$

$\implies I = a(T)(T-\mu I) + b(T)(T-\lambda I)^p$

$\implies v = Iv = a(T)(T-\mu I)v + b(T)(T-\lambda I)^p v = 0$

Hence, $\ker(T-\mu I) \cap K_\lambda = \{0\}$

□

Theorem 1.3.2. Assume that $\dim V < \infty$ and $ch_T(x) = \prod\limits_{i=1}^{k}(x-\lambda_i)^{m_i}$. Then $\dim K_i = m_i$

Proof: We have $V = \bigoplus\limits_{i=1}^{k} K_{\lambda_i}$. Here each K_{λ_i} is a T-invariant subspace since $K_{\lambda_i} = \ker(T - \lambda_i I)^{m_i}$. Let $T_i : K_{\lambda_i} \to K_{\lambda_i}$ be the restriction of T to K_{λ_i}. By Theorem 1.2.3.

$$ch_T(x) = \prod_{i=1}^{k} ch_{T_i}(x)$$

Since λ_i is the only eigenvalue of T_i (by Theorem 1.3.1.(ii)). We have $ch_{T_i}(x) = (x-\lambda_i)^{n_i}$, where $n_i := \dim K_i$. Compare both side of

$$\prod_{i=1}^{k}(x-\lambda_i)^{m_i} = ch_T(x) = \prod_{i=1}^{k} ch_{T_i}(x) = \prod_{i=1}^{k}(x-\lambda_i)^{n_i}$$

then $m_i = n_i = \dim K_i$. □

Theorem 1.3.3. Assume $\dim V < \infty$. Let K_λ be the generalized eigenspace corresponding to an eigenvalue λ of T. Then $\exists v_1, ..., v_r \in K_\lambda$ such that

$$K_\lambda = Z(v_1;T) \oplus \cdots \oplus Z(v_r;T)$$

Moreover, let $s_i = \dim Z(v_i;T)$ and arrange the subscripts such that $s_1 \geq s_2 \geq \cdots \geq s_r$. Then the sequence $s_1, s_2, ..., s_r$ is unique.

We leave the proof in Appendix 2.1

Notation 1.3.1. If $I_T(v, W) = ((x-\lambda)^s)$ for some s , then denote $s(v, W)$ for this s.

Remark 1.3.2. Since $v_i \in K_\lambda$, we have $(T-\lambda I)^p(v_i) = 0 \implies I_T(v_i) = ((x-\lambda)^{s_i})$.

Proof: Since $\{v_i, T(v_i), ..., T^{s_i-1}(v_i)\}$ is a basis for $Z(v_i;T)$. Say $I_T(v_i) = ((x-\lambda)^s)$.

If $s \leq s_i$: Since $(T-\lambda I)^s(v_i) = 0 \implies v_i, T(v_i), ..., T^s(v_i)$ are linearly independent $(\rightarrow\leftarrow)$. Thus, $s \geq s_i$. On the other hand, there is a non-trivial relation

$$T^{s_i}(v_i) + a_{s_i-1}T^{s_i-1}(v_i) + \cdots a_1 T(v_i) + a_0 = 0$$

$\implies$ this polynomial in $I_T(v_i) \implies s \leq s_i$. Hence, $s = s_i$ □

Remark 1.3.3. Now we choose a basis $\mathcal{B}_i$ for $Z(v;T)$ to be

$$\mathcal{B} := \left\{(T-\lambda I)^{s_i-1}(v_i), ..., (T-\lambda I)v_i, v_i\right\}$$

We have

$$T\left((T-\lambda I)^j v_i\right) = (T-\lambda I)^{j+1}v_i + \lambda(T-\lambda I)^j v_i$$

$$\implies [T|_{Z(v_i;T)}] = \begin{pmatrix} \lambda & 1 & & & O \\ 0 & \lambda & 1 & & \\ & 0 & \lambda & & \\ & & & \ddots & 1 \\ O & & & & \lambda \end{pmatrix}$$

which is the form what we want.

1.3.2 Existence and uniqueness of Jordan normal form

Theorem 1.3.4 (Existence of Jordan normal form)**.**

Proof: If $ch_T(x) = \prod_{i=1}^{k}(x-\lambda)^{m_i}$ splits over F, then by Thm 1.3.2,

$$V = \bigoplus_{i=1}^{k} K_{\lambda_i} = \bigoplus_{i=1}^{k}\bigoplus_{j=1}^{r_i} Z(v_{ij};T)$$

for some $v_{ij} \in K_{\lambda_i}$. Choose a basis for V to be

$$\mathcal{B} = \bigsqcup_{i=1}^{k}\bigsqcup_{j=1}^{r_i}\mathcal{B}_{ij}$$

where $B_{ij} = \{T^\ell(v_{ij}) : 0 \le \ell \le s_{ij}-1\}$. Then

$$[T]_{\mathcal{B}} = \begin{pmatrix} J_1 & & & O \\ & J_2 & & \\ & & \ddots & \\ O & & & J_k \end{pmatrix}$$

where each J_m is of the form

$$J_i = \begin{pmatrix} \lambda_i & 1 & & & O \\ 0 & \lambda_i & 1 & & \\ & & \ddots & \ddots & \\ & & & \lambda_i & 1 \\ O & & & & \lambda_i \end{pmatrix}$$

Hence, $[T]_{\mathcal{B}}$ is the Jordan normal form of T. $\square$

Theorem 1.3.5 (Uniqueness of Jordan normal form)**.**

Proof: Let $\widetilde{T} := T - \lambda I$. Recall that

$$\mathcal{B} = \bigcup_{i=1}^{r_i} \underbrace{\left\{v_i, \widetilde{T}(v_i), ..., \widetilde{T}^{s_i-1}(v_i)\right\}}_{\text{basis for } Z(v_i;T)}$$

is a basis for K_λ.
Let's determine $\dim\ker\widetilde{T}^\ell$ for $\ell \geq 1$. Observe that for $1 \leq k \leq s_i$

$$\widetilde{T}^{s_i-k}(v_i) \in \ker\widetilde{T}^\ell \iff \widetilde{T}^\ell\left(\widetilde{T}^{s_i-k}(v_i)\right) = 0 \iff s_i - k + \ell \geq s_i \iff k \leq \ell$$

On other hand, the remaining vectors

$$B' := \bigcup_{i=1}^{r} \left\{v_i, ..., \widetilde{T}^{s_i-\ell-1}(v_i)\right\}$$

has linearly independent images in $\widetilde{T}^\ell$, since $\widetilde{T}^\ell(\mathcal{B}') \subseteq B'$
In summary

$$\text{basis for } Z(v_i;T) : \left\{\underbrace{v_i, \widetilde{v}_i, ..., \widetilde{T}^{s_i-\ell-1}(v_i)}_{\text{form a basis for } \operatorname{Im}\widetilde{T}^\ell} \middle| \underbrace{\widetilde{T}^{s_i-\ell}(v_i), ..., \widetilde{T}^{s_i-1}(v_i)}_{\in\ker\widetilde{T}^\ell}\right\}$$

Hence,

$$\dim\ker\widetilde{T}^\ell = \sum_{i=1}^{r}\begin{cases}\ell & \text{if } s_i \geq \ell\\ s_i & \text{if } s_i < \ell\end{cases} = \sum_{i=1}^{r}\min(\ell, s_i)$$

$$\begin{aligned}\dim\ker\widetilde{T}^\ell - \dim\ker\widetilde{T}^{\ell-1} &= \sum_{i=1}^{r}(\min(\ell, s_i) - \min(\ell-1, s_i))\\ &= \sum_{i=1}^{r}\begin{cases}s_i - s_i = 0 & \text{if } s_i \leq \ell - 1\\ \ell - (\ell-1) & \text{if } s_i \geq \ell\end{cases} = \#\{s_i \geq \ell\}\end{aligned}$$

$$\begin{aligned}\implies \#\{s_i = \ell\} &= \#\{s_i \geq \ell + 1\} - \#\{s_i \geq \ell\}\\ &= (\dim\ker\widetilde{T}^\ell - \dim\ker\widetilde{T}^{\ell-1}) - (\dim\ker\widetilde{T}^{\ell+1} - \dim\ker\widetilde{T}^\ell)\end{aligned}$$

RHS is a quantity intrinsic to T and is independent of choices of $v_1, v_2, ... \implies$ The sequence $s_1, ...,$ is unique. $\square$

Remark 1.3.4. The proof of uniqueness can be visualized using the dot diagram.
Rule:

- r columns, each column represents on s_i
- The i-thm column has s_i dots representing vectors $\widetilde{T}^{s_i-1}(v_i), \widetilde{T}^{s_i-2}(v_i), ..., v_i$

Then first i rows forms a basis for $\ker\widetilde{T}^i$. The numbers of dots in the first ℓ rows $= \dim\ker\widetilde{T}^\ell$.
Thus, # on the ℓ-th row $= \dim\ker\widetilde{T}^\ell = \dim\ker\widetilde{T}^{\ell-1} = \#\{i : s_i = \ell\}$

Remark 1.3.5. # of Jordan block for $K_\lambda = \dim E_\lambda$

Recall that if λ is a eigenvalue with multiplity m, then by theorem 1.3.3, we can

Definition 1.3.2. A **partition** of a positive integer m is a non-increasing sequence of positive integers $s_1, ..., s_r$ such $s_1 + \cdots + s_r = m$
The # of partitions will be denoted by $p(m)$ called the **partition function**.

It's clear that for a given eigenvalue λ with multiplicity m,

$$\{\text{possible Jordan-forms for } K_\lambda\} \xleftrightarrow{1-1} \{\text{partition of } m\}$$

$\implies$ # of possible Jordan forms for $K_\lambda = p(m)$

Property 1.3.1. Given $f(x) = \prod_{i=1}^{k} (x - \lambda_i)^{m_i}$. There are $\prod_{i=1}^{k} p(m_i)$ possible Jordan-forms with char. poly. $f(x)$.

Problem: How many possible Jordan Form are there for given $ch_T(x)$ and $m_T(x)$.
Ans: Say $ch_T(x) = \prod_{i=1}^{k} (x - \lambda_i)^{m_i}, m_T = \prod_{i=1}^{k} (x - \lambda_i)^{n_i}$. Then number of possible Jordan forms

$$\prod_{i=1}^{k} \# \text{ of partitions of } m_i \text{ with largest part is } m_i$$

1.4 Rational canonical forms

1.4.1 Motivation and Goal

Note that in order for Jordan forms to exists, a prerequisite is that $ch_T(x)$ splits over F. However, there are F and T whose $ch_T(x)$ does not split over F
eg. $\begin{array}{cccc} T: & \mathbb{R}^2 & \longrightarrow & \mathbb{R}^2 \\ & (a,b) & \longmapsto & (b,-a) \end{array} \implies [T]_e = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ and the char. poly is $x^2 + 1$, which has no root in $\mathbb{R}$.
Goal: Find a simple matrix representation

Definition 1.4.1. Let $T : V \to V$ (V possibly ∞-dimensional) and $p(x)$ be an irreducible polynomial over F. We let K_p denote the subspace

$$K_p = \{v \in V : p(T)^n v = 0 \text{ for some } n \geq 1\}$$

Note that if λ is an eigenvalue, then $K_{x-\lambda}$ is a simply the generalized eigenspace K_λ.

For now on, we assume $\dim V < \infty$
Let $ch_T(x) = \prod_{i=1}^{k} p_i(x)^{m_i}$ be the (unique) factorization of $ch_T(x)$ into a product of irreducible polynomial. We have

$$V = \ker ch_T(T) = \bigoplus_{i=1}^{k} \ker p_i(x)^{m_i} = \bigoplus_{i=1}^{k} K_{p_i^{m_i}} = \bigoplus_{i=1}^{k} K_{p_i}$$

Note that each K_{p_i} is a T-invariant subspace.

Theorem 1.4.1. $\dim K_{p_i} = m_i \deg p_i$

Proof: Let $T_i = T|_{K_{p_i}}$. Then $K_{p_i} = \ker p_i(T)^{m_i}$.

Claim: $ch_{T_i}(x) = p_i(x)^{n_i}$ for some n_i

pf. Assume that $ch_{T_i} = p_i(x)^{n_i} g(x)$ with $(g, p_i) = 1$. Then

$$K_{p_i} = \underbrace{\ker p_i(T)^{n_i}}_{:=U_1} \oplus \underbrace{\ker g(T)}_{:=U_2}$$

It's an easy exercise to show $m_{T_i} = \operatorname{lcm}(m_{T_{U_1}}, m_{T_{U_2}})$. Now, since $K_{p_i} = \ker p_i(T)^{m_i}$, we have $m_{T_i}(x)|p_i(x)^{m_i}$. Thus, the minimal polynomial of T_{U_2} is $p_i(x)^s$ for some $s \geq 0$. But $(g, p_i) = 1 \implies$ minimal polynomial of T_{U_2} is $1 \implies g(x) = 1$ i.e. $ch_{T_i}(x) = p_i(x)^{n_i}$ for some n_i □

By Claim, we have

$$\prod_{i=1}^{k} p_i(x)^{n_i} = ch_T(x) = \prod_{i=1}^{k} ch_{T_i}(x) = \prod_{i=1}^{k} p_i(x)^{m_i} \implies n_i = m_i$$

□

Remark 1.4.1. Another proof in Theorem 1.4.1 (Field extension)
By field theory $\exists F'/F$ such that $ch_T(x)$ splits over F'. Extend the scalar of V to F' and denoted the new vector space by $V \otimes_F F'$. Now assume $ch_{T_i}(x) = p_i(x)^{m_i} g(x)$ for $(p_j, g) = 1$. Let λ be a root of $g(x)$ in F', which is in fact an eigenvalue and there is an eigenvector $v \neq 0$ in $K_{p_i} \otimes_F F'$ corresponding to λ. However, $v \notin K_{p_i} \otimes F'(p_i(T)^n(v) \neq 0)$ $(\rightarrow\leftarrow)$

Property 1.4.1. $T : V \to V$ is linear with $\dim_F V = n < \infty$. Then $m_T(x)|ch_T(x)|m_T(x)^n$. In particular, the irreducible factors of $m_T(x)$ is the same as the irreducible factors of $ch_T(x)$.

Proof: $m_T(x)|ch_T(x)$ OK! We prove $ch_T(x)|m_T(x)^n$:

Suppose $\deg m_T = d$, by $x^k - y^k = (x-y)(x^{k-1} + x^{k-2}y + \cdots + y^{k-1})$ and xI_N commute with $[T]_\beta$, we have

$$m_T(xI_n) = m_T(xI_n) - m_T([T]_\beta) = (xI_n - [T]_\beta)P \text{ for some } P \in M_n(F[x])$$

Take determinant in both side,

$$m_T(x)^n = \det(m_T(xI_n)) = \det(xI_n - [T]_\beta)\det(P) = \mathrm{ch}_T(x)\det P \implies ch_T(x)|m_T(x)^n$$

□

1.4.2 Existence and uniqueness of Rational form

Theorem 1.4.2. $T : V \to V, \dim V < \infty$. $p(x)$ is an irreducible factor of $ch_T(x)$. Then $\exists v_1, ..., v_r \in K_p$ such that

$$K_p = Z(v_1; T) \oplus \cdots \oplus Z(v_r; T)$$

Moreover, let s_i be the smallest integer such that $p(T)^{s_i}(v_i) = 0$ and arrange the subscripts such that $s_1 \geq s_2 \geq \cdots \geq s_r$. Then the sequence $s_1, ..., s_r$ is unique.

Remark 1.4.2.

(1) For $v \in K_p$, let s be the smallest integer such that $p(T)^s(v) = 0$, Then $\dim Z(v; T) = s \deg p$ and $I_v = (p(x)^s)$.

pf. In general, the characteristic polynomial if $T_{Z(v;T)}$ is the same as the minimal polynomial of $T_{Z(v;T)}$. Here the assumption that s is the smallest integer such that $p(T)^s(v) = 0$ means the minimal poly. of $T_{Z(v;T)}$ is $p(x)^s$ i.e. $I_v = (p(x)^s)$

(2) More generally, let W be a T-invariant subspace of K_p and s be the smallest integer such that $p^s(T)(v) \in W$. Then $I_{v,W} = (p(x)^s)$.

Notation 1.4.1. For $v \in K_p$ and W is a T-invariant subspace of K_p. If $I_{v,W} = (p(x)^s)$, then define $s(v, W) = s$

Outline of proof Theorem 1.4.2

(i) Let $W_0 = \{0\}$

(ii) For $i \geq 1$, assume that W_{i-1} have been defined. Choose $u \in K_p$ such that

$$s(u, W_{i-1}) := \max_{v \in K_p} s(v, W_{i-1}) := s_i$$

Claim: $\exists w \in W_{i-1}$ such that $p(T)^{s_i}(w) = p(T)^{s_i}(u)$

(iii) Let $v_i = u - w$ and claim :

- $W_{i-1} \cap Z(v_i; T) = \{0\}$
- $I_{v_i} = (p(x)^{s_i})$

Let $W_i = W_{i-1} \oplus Z(v_i; T)$

(iv) Repeat (ii),(iii) until $W_i = K_p$

We leave the detail proof in Appendix 5.1.2
Recall that is $\dim Z(v; T) = k$, then

$$\mathcal{B} = \{v, T(v), ..., T^{k-1}(v)\} \text{ is a basis for } Z(v; T)$$

We have $T^k(v) = -a_{k-1}T^{k-1}(v) - \cdots - a_1 T(v) - a_0 v$ for some a_j, then

$$[T|_{Z(v,T)}]_{\mathcal{B}} = \begin{pmatrix} 0 & 0 & & & & -a_0 \\ 1 & 0 & & & & -a_1 \\ & 1 & 0 & & & \vdots \\ & & & \ddots & & \\ & & & & 0 & -a_{k-2} \\ 0 & & & & 1 & -a_{k-1} \end{pmatrix} \tag{*}$$

Definition 1.4.2. For a polynomial $f(x) = x^k + a_{k-1}x^{k-1} + \cdots + a_0$, define the **companion matrix** of f is a matrix in $(*)$

Corollary 1.4.1. Assume $\dim V < \infty$, $T : V \to T$. Then exists a basis $\mathcal{B}$ for T such that $[T]_{\mathcal{B}}$ is a block matrix of the form

$$\begin{pmatrix} C_1 & & O \\ & \ddots & \\ O & & C_m \end{pmatrix}$$

where each C_j is the companion matrix of $p(x)^s$ for some irreducible factor of $ch_T(x)$ and some $s \geq 1$. Moreover, C_i is unique up to permutation.

Definition 1.4.3. A matrix representation of T of the form is called a rational canonical form and $\mathcal{B}$ is called a rational canonical basis. The factors $p_i(x)^{s_i}$ are called the **elementary divisors of** T

Remark 1.4.3. There is another definition of a rational canonical form, where

$$T = \begin{pmatrix} C_1' & & O \\ & \ddots & \\ O & & C_k' \end{pmatrix}$$

and C_i; is the companion matrix of some f_i such that $f_i | f_{i+1}$ $\forall i = 1, ..., k-1$. The polynomials $f_i(x)$ are $f_i(x)$ are called the invariant factors of T.

1.5 Real Jordan Form

Let $A \in M_n(\mathbb{R})$, $ch_A(x) = (x-\lambda_1)^{m_1} \cdots (x-\lambda_\ell)^{m_\ell}(x^2+a_1x+b_1)^{n_1} \cdots (x^2+a_kx+b_k)^{n_k}$ with $\alpha_i \pm \beta_i\sqrt{-1}$ are roots of $x^2 + a_ix + b_i = 0$. Since $ch_T(x)$ is not splits over $\mathbb{R}$. How can we find a good basis for A ?

Theorem 1.5.1. Let $A \in M_n(\mathbb{R})$, $ch_A(x) = (x-\lambda_1)^{m_1} \cdots (x-\lambda_\ell)^{m_\ell}(x^2+a_1x+b_1)^{n_1} \cdots (x^2+a_kx+b_k)^{n_k}$ with $\alpha_i \pm \beta_i\sqrt{-1}$ are roots of $x^2 + a_ix + b_i = 0$. Then $\exists$ invertible $P \in M_n(\mathbb{R})$ s.t.

$$P^{-1}AP = \begin{pmatrix} I_{\lambda_1} & & & & & \\ & \ddots & & & & \\ & & I_{\lambda_\ell} & & & \\ & & & J_{\mu_1} & & \\ & & & & \ddots & \\ & & & & & J_{\mu_k} \end{pmatrix}$$

where $I_{\lambda_i} = J_1(\lambda_i) \oplus \cdots \oplus J_r(\lambda_i)$ with $J_k(\lambda_i)$ is Jordan blocks corresponding λ_i and $I_{\mu_j} = J_1(\mu_j) \oplus \cdots \oplus J_r(\mu_j)$ with $J_k(\mu_j)$ is form

$$J_k(\mu_j) = \begin{pmatrix} \alpha_j & \beta_j & 1 & 0 & & & & \\ -\beta_j & \alpha_j & 0 & 1 & & & & \\ & & \alpha_j & \beta_j & 1 & 0 & & \\ & & -\beta_j & \alpha_j & 0 & 1 & & \\ & & & & \ddots & \ddots & \ddots & \\ & & & & & \ddots & \ddots & 1 & 0 \\ & & & & & & \ddots & 0 & 1 \\ & & & & & & & \alpha_j & \beta_j \\ & & & & & & & -\beta_j & \alpha_j \end{pmatrix}$$

is called **real Jordan block**.

Before proving the theorem, we see some property on vector space over $\mathbb{C}$.

Property 1.5.1.

- For a vector $v = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{pmatrix} \in \mathbb{C}^n$, we write $\overline{v} = \begin{pmatrix} \overline{z}_1 \\ \overline{z}_2 \\ \vdots \\ \overline{z}_n \end{pmatrix} \in \mathbb{C}^n$
- For subspace $W \subseteq \mathbb{C}^n$, we write $\overline{W} = \{\overline{w} : w \in W\}$

- If $W = \text{span}_{\mathbb{C}}\{w_1, ..., w_k\} \implies \overline{W} = \text{span}_{\mathbb{C}}\{\overline{w}_1, ..., \overline{w}_k\}$
- $v_1, ..., v_k \in \mathbb{C}^n$ are linearly independent, then $\overline{v}_1, ..., \overline{v}_k$ are linearly independent.
- $A \in M_{n\times n}(\mathbb{R}),\ \lambda \in \mathbb{C} \setminus \mathbb{R}$, then
$$W = \ker(A - \lambda I)^r \implies \overline{W} = \ker(A - \overline{\lambda} I)^r$$
and $W \cap \overline{W} = \{0\}$

Back to proof of Theorem 1.5.1

Proof: For real part I_{λ_i} : OK! Now, we deal with I_{μ_j} part!
Consider A in $M_{n\times n}(\mathbb{C})$, then we focus on the eigenvalues $\lambda = \alpha + \beta\sqrt{-1}$ and $\overline{\lambda} = \alpha - \beta\sqrt{-1}$
By Jordan form theory, we can find $v_1, ..., v_r \in K_\lambda \subseteq \mathbb{C}^n$ s.t. $K_\lambda = \bigoplus_{i=1}^{r} Z(v_i; A)$
Looking one Jordan block, we have basis $\{(a - \lambda I)^{s-1}v, ..., (A - \lambda I)v, v\}$ in $\mathbb{C}^n$ and write $(A - \lambda I)^{i-1}v = w_i = x_i + y_i + \sqrt{-1}$. Since $Aw_i = \lambda w_i + w_{i+1}$ (assume $w_{s+1} = 0$) and compare real part and imaginary part, then

$$\begin{cases} Ax_i = \alpha x_i - \beta y_i + x_{i+1} \\ Ay_i = \beta x_i + \alpha y_i + y_{i+1} \end{cases} \text{for } i = 1, ..., s-1 \text{ and } \begin{cases} Ax_s = \alpha x_s - \beta y_s \\ Ay_i = \beta x_s + \alpha y_s \end{cases}$$

Since $\{w_s, \overline{w}_s, w_{s-1}, \overline{w}_{s-1}, ..., w_1, \overline{w}_1\}$ are linearly independent, then

$$\left\{ x_i = \frac{w_i + \overline{w}_i}{2}, y_i = \frac{w_i + \overline{w}_i}{2i} \middle| i = 1, ..., s \right\}$$

are linearly independent. Finally, we use basis $\{x_s, y_s, x_{s-1}, y_{s-1}, ..., x_1, y_1\}$ to write down matrix representation :

$$\begin{pmatrix} \alpha_j & \beta_j & 1 & 0 & & & & \\ -\beta_j & \alpha_j & 0 & 1 & & & & \\ & & \alpha_j & \beta_j & 1 & 0 & & \\ & & -\beta_j & \alpha_j & 0 & 1 & & \\ & & & & \ddots & \ddots & \ddots & \\ & & & & & \ddots & \ddots & 1 & 0 \\ & & & & & & \ddots & 0 & 1 \\ & & & & & & & \alpha_j & \beta_j \\ & & & & & & & -\beta_j & \alpha_j \end{pmatrix}$$

□

Chapter 2

Matrix exponential

2.1 Definition

Definition 2.1.1. Let $A \in M_{n\times n}(\mathbb{C})$. Then the **exponential** of A denoted by e^A or $\exp A$ is defined to

$$e^A = I_n + \sum_{k=1}^{\infty} \frac{A^k}{k!}$$

Example 2.1.1.

- $A = \begin{pmatrix} \lambda_1 & & O \\ & \ddots & \\ O & & \lambda_k \end{pmatrix}$, then $e^A = \begin{pmatrix} e^{\lambda_1} & & O \\ & \ddots & \\ O & & e^{\lambda_k} \end{pmatrix}$
- Nilpotent matrix is easy to calculate, since it only finite sum.

The example show that when A is diagonal or nilpotent, it is easy to computes e^A. Thus the theory of Jordan forms will be very useful in computing e^A. More specifically, let Q be an invertible matrix such that $J = Q^{-1}AQ$ is the Jordan form of A.

Example 2.1.2. If $A = QJQ^{-1}$, then $e^A = Qe^JQ^{-1}$. Note

$$J = \begin{pmatrix} J_1 & & \\ & \ddots & \\ & & J_k \end{pmatrix} \implies e^{Jz} = \begin{pmatrix} e^{J_1 z} & & \\ & \ddots & \\ & & e^{J_k z} \end{pmatrix}$$

Now, Write $J_i = \lambda_i I + N_i$ and notice that : if $AB = BA$, then $e^{A+B} = e^A e^B$, so

$$e^{J_i z} = e^{\lambda_i z I} e^{N_i z} = e^{\lambda_i z} \cdot e^{N_i z}$$

If N_i is $n \times n$ matrix, then $N_i^n = O$ and

$$e^{N_i z} = \sum_{k=0}^{\infty} \frac{1}{k!}(N_i z)^k = \sum_{k=0}^{n-1} \frac{1}{k!}(N_i z)^k = \begin{pmatrix} 1 & \frac{z}{1!} & \frac{z^2}{2!} & \cdots & \cdots & \frac{z^{n-1}}{(n-1)!} \\ 0 & 1 & \frac{z}{1!} & \cdots & \cdots & \frac{z^{n-2}}{(n-2)!} \\ 0 & 0 & 1 & \ddots & \cdots & \frac{z^{n-3}}{(n-3)!} \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \ddots & \frac{z}{1!} \\ 0 & 0 & 0 & \cdots & \cdots & 1 \end{pmatrix}$$

If we consider the real Jordan form of A, so it is sufficiently calculate the exponent of real Jordan block J. If

$$J = \begin{pmatrix} D & I_2 & & & \\ & D & I_2 & & \\ & & D & \ddots & \\ & & & \ddots & I_2 \\ & & & & D \end{pmatrix} \in M_{2d\times 2d}(\mathbb{R}) \text{ where } D = \begin{pmatrix} \alpha & \beta \\ -\beta & \alpha \end{pmatrix}$$

Similarly to calculate exponent of Jordan form, we calculate $\exp(D)$ first.

$$D = \alpha I_2 + \beta \underbrace{\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}}_{:=K} \implies \exp(D) = \exp(\alpha I_2)\exp(\beta K) = e^{\alpha}\exp(\beta K)$$

$$K = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix},\ K^2 = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix},\ K^3 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix},\ K^4 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = I_2$$

(Note: K have similar structure with $\sqrt{-1}$). So

$$\exp(\beta K) = \sum_{s=0}^{\infty} \frac{1}{s!}(\beta K)^s = \sum_{s=0}^{\infty} \frac{(-1)^s\beta^{2s+1}}{(2s+1)!}\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} + \sum_{s=0}^{\infty} \frac{(-1)^s\beta^{2s}}{(2s)!}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} \cos\beta & \sin\beta \\ -\sin\beta & \cos\beta \end{pmatrix}$$

$$\implies e^D = e^{\alpha}\begin{pmatrix} \cos\beta & \sin\beta \\ -\sin\beta & \cos\beta \end{pmatrix}. \text{ Similarly } e^{Jz} = e^{\alpha z}\begin{pmatrix} \cos(\beta z) & \sin(\beta z) \\ -\sin(\beta z) & \cos(\beta z) \end{pmatrix}$$

Hence,

$$e^{Mz} = \operatorname{diag}\{e^{Jz}, ..., e^{Jz}\} \begin{pmatrix} I_2 & \frac{z}{1!}I_2 & \frac{z^2}{2!}I_2 & \cdots & \frac{z^{d-1}}{(d-1)!}I_2 \\ & I_2 & \frac{z}{1!}I_2 & \cdots & \frac{z^{d-2}}{(d-2)!}I_2 \\ & & I_2 & \cdots & \frac{z^{d-3}}{(d-3)!}I_2 \\ & & & \ddots & \vdots \\ & & & & I_2 \end{pmatrix}$$

2.2 System of linear differential equation with constant coefficient

Theorem 2.2.1. Let $A \in M_{n\times n}(\mathbb{C})$. Then the unique solution of

$$y'(z) = \begin{pmatrix} y_1'(z) \\ \vdots \\ y_n'(z) \end{pmatrix} = A \begin{pmatrix} y_1(z) \\ \vdots \\ y_n(z) \end{pmatrix} =: Ay(z)$$

with the initial condition $y(0) = y_0$ is $y(z) = e^{Az}y_0$

Before prove this theorem, we need define a norm on matrix space to describe the limits of matrix.

Definition 2.2.1 (max norm). For $A \in M_n(\mathbb{C})$, define the **max norm** of A

$$\|A\| = \max_{1\leq i,j\leq n} |a_{ij}|$$

Property 2.2.1. For all $A, B \in M_n(\mathbb{C})$,

$$\|A \cdot B\| \leq n \cdot \|A\| \cdot \|B\|$$

Now, we can prove Theorem 2.2.1

Proof: We prove the case in $\mathbb{R}$ and the case in $\mathbb{C}$ is similar.

- **Existence:** $y(z) = e^{Az} \cdot y_0$ is a solution

  Claim: $\dfrac{d}{dz}(e^{Az}) = Ae^{Az}$

  pf. For $h \in \mathbb{R}$, we have

$$\frac{e^{A(z+h)} - e^{Az}}{h} - Ae^{Az} = e^{Az}\left(\frac{e^{Ah} - I_n - Ah}{h}\right)$$

  Observe for all $x \in \mathbb{R}$, we have

$$|e^x - 1 - x| = \left|\sum_{k=2}^{\infty} \frac{x^k}{k!}\right| \leq \sum_{k=2}^{\infty} \frac{|x|^k}{k!} \leq |x| \sum_{k=1}^{\infty} \frac{|x|^k}{k!} = |x|\left(e^{|x|} - 1\right)$$

  With similar ideal, in $M_n(\mathbb{R})$, we have

$$\begin{aligned}\|e^B - I_n - B\| &= \left\|\sum_{k=2}^{\infty} \frac{B^k}{k!}\right\| \leq \frac{1}{n} \sum_{k=2}^{\infty} \frac{(n\|B\|)^k}{k!} \qquad \text{(By Property 2.2.1)}\\ &= \frac{1}{n}\left(e^{n\|B\|} - 1 - n\|B\|\right) \leq \left(e^{n\|B\|} - 1\right)\|B\|\end{aligned}$$

  Apply $B = Ah$, then

$$\left\|\frac{e^{Ah} - I_n - Ah}{h}\right\| \leq \frac{1}{h}\left(e^{n\|Ah\|} - 1\right)\|Ah\| = \left(e^{nh\|A\|} - 1\right)\|A\| \longrightarrow 0 \text{ as } h \longrightarrow 0$$

$$\Longrightarrow \lim_{h \to 0}\left(\frac{e^{A(z+h)} - e^{Az}}{h} h - Ae^{Az}\right) = 0$$

- **Uniqueness of solution** :

  Suppose $x(z), y(z)$ are two solutions. Consider $u(z) = x(z) - y(z)$. Then $u(z)$ satisfies $u'(z) = Au(z)$ and $u(0) = 0$.

$$\begin{aligned}\Longrightarrow u(t) &= A \cdot \underbrace{\int_0^t u(s_1)\ ds_2}_{\text{entrywise integral}} = A \int_0^t A \int_0^{s_1} u(s_2)\ ds_2 ds_1 = \cdots \\ &= A^k \int_0^t \int_0^{s_1} \cdots \int_0^{s_{k-1}} u(s_k) ds_k ds_{k-1} \cdots ds_1\end{aligned}$$

  We will prove that $u(z) = 0$ on any closed interval $[a, b] \subset \mathbb{R}$. Suppose $t \in [a, b]$ and let $M = \max_{t \in [a,b]} \|u(t)\|$, where $\|u(t)\| = \max_{1 \leq i \leq n} |u_i(t)|$ (Since $u_i(x)$ are continuous on $[a, b]$, thus M is exists). Thus, for any $t \in [a, b]$, we have

$$\begin{aligned}\|u(t)\| &= \left\|A^k \int_0^t \int_0^{s_1} \cdots \int_0^{s_{k-1}} u(s_k) ds_k ds_{k-1} \cdots ds_1\right\| \\ &\leq n \cdot \|A^k\| \cdot \left\|\int_0^t \int_0^{s_1} \cdots \int_0^{s_{k-1}} u(s_k) ds_k ds_{k-1} \cdots ds_1\right\| \\ &\leq (n\|A\|)^k M \left|\int_0^t \int_0^{s_1} \cdots \int_0^{s_{k-1}} 1 \cdot ds_k ds_{k-1} \cdots ds_1\right| \leq (n\|A\|)^k M \frac{|b-a|^k}{k!} \longrightarrow 0\end{aligned}$$

as $k \longrightarrow \infty$. Thus $\|u(t)\| = 0$ for all $t \in [a, b] \implies u(t) = 0$ for all $t \in [a, b] \implies u(t) = 0$ on whole $\mathbb{R} \implies x(t) = y(t)\ \forall t \in \mathbb{R}$.

□

Example 2.2.1. Solve $y''(z) + 2y'(z) + y(z) = 0$.

Let $y_1(z) = y(z), y_2(z) = y'(z)$, then

$$\begin{pmatrix} y_1'(z) \\ y_2'(z) \end{pmatrix} = \underbrace{\begin{pmatrix} 0 & 1 \\ -1 & -2 \end{pmatrix}}_{:=A} \begin{pmatrix} y_1(z) \\ y_2(z) \end{pmatrix} \implies A = -I + \underbrace{\begin{pmatrix} 1 & 1 \\ -1 & -1 \end{pmatrix}}_{\text{nilpotent}}$$

$$e^{Az} = e^{-Iz} e^{\left(\begin{smallmatrix} 1 & 1 \\ -1 & -1 \end{smallmatrix}\right) z} = e^{z} \left(I + \begin{pmatrix} 1 & 1 \\ -1 & -1 \end{pmatrix} z \right) = \begin{pmatrix} (1+z)e^{z} & ze^{z} \\ -ze^{z} & (1-z)e^{z} \end{pmatrix}$$

Thus, solution are $y(z) = c_1(1+z)e^{-z} + c_2 ze^{-z} = c_1' e^{-z} + c_2' ze^{-z}$

If $y_1(z), y_2(z)$ are solutions of a linear differential equation, then $y_1(z) + cy_2(z)$ is also the solution. Thus, we usually write solutions of the DE as a linear combination of functions form a basis. For example, in example above, $c_1 e^{-z} + c_2 ze^{-z}$ is the general solution of $y''(z) + 2y'(z) + y(z) = 0$. If some additional conditions are given, e.g. $y(0) = a_1, y'(0) = a_2$. Then c_1, c_2 will be determined by additional condition.

Example 2.2.2. Solve $y''(z) + 9y(z) = 0$

Proof: We have $\begin{pmatrix} y(z) \\ y'(z) \end{pmatrix}' = \underbrace{\begin{pmatrix} 0 & 1 \\ -9 & 0 \end{pmatrix}}_{:=A} \begin{pmatrix} y(z) \\ y'(z) \end{pmatrix}$. The eigenvalues of A are $\pm 3i$. and $\begin{pmatrix} 1 \\ 3i \end{pmatrix}, \begin{pmatrix} 1 \\ -3i \end{pmatrix}$ are eigenvalue corresponding to $3i, -3i$, respectively. Thus, setting $Q = \begin{pmatrix} 1 & 1 \\ 3i & -3i \end{pmatrix}$, we have $Q^{-1}AQ = \begin{pmatrix} 3i & 0 \\ 0 & -3i \end{pmatrix} =: J$

$$\implies e^{Az} = Qe^{Jz}Q^{-1} = Q \begin{pmatrix} e^{3iz} & 0 \\ 0 & e^{-3iz} \end{pmatrix} Q^{-1} = \begin{pmatrix} \frac{e^{3iz}+e^{-3iz}}{2} & \frac{e^{3iz}-e^{-3iz}}{6i} \\ \frac{3i(e^{3iz}-e^{-3iz})}{2} & \frac{e^{3iz}+e^{-3iz}}{2} \end{pmatrix} = \begin{pmatrix} \cos 3z & \frac{1}{3}\sin 3z \\ -3\sin 3z & \cos 3z \end{pmatrix}$$

If the initial conditions are given as $y(0) = a_1, y'(0) = a_2$, then the solution is

$$a_1 \cos 3z + \frac{a_2}{3} \sin 3z$$

□

2.3 Matrix limits

Theorem 2.3.1. Let $A \in M_{n \times n}(\mathbb{C})$. Then $\lim_{k \to \infty} A^k$ exists if and only if

- All eigenvalue of A are in

$$\{z \in \mathbb{C} : |z| < 1\} \cup \{1\}$$

- If 1 is an eigenvalue, then $\dim E_1 =$ multiplicity of 1.

Proof: The proof is left as an exercise to the reader. □

<u>Question:</u> Suppose that each year 90% of city population stay in the city and 10% move to suburbs. 80% of suburbs population stay in suburbs and 20% move to city. Assume the number of all population will not change, will the populations of city and suburbs stabilize, oscillate or ?
<u>Solution:</u> Let a_n, b_n be the populations of city, suburbs, respectively in year n. We have

$$\begin{pmatrix} a_n \\ b_n \end{pmatrix} = \begin{pmatrix} 0.9 & 0.2 \\ 0.1 & 0.8 \end{pmatrix} \begin{pmatrix} a_{n-1} \\ b_{n-1} \end{pmatrix}$$

Since $\begin{pmatrix} 1 & 1 \end{pmatrix} A = \begin{pmatrix} 1 & 1 \end{pmatrix} \implies 1$ is an eigenvalue and $\operatorname{tr} A = 1.7 \implies 0.7$ is another eigenvalue. Let v_1, v_2 be an eigenvector corresponding to $1, 0.7$, respectively. Let $Q = \begin{pmatrix} v_1 & v_2 \end{pmatrix}$. Then $Q^{-1}AQ^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & 0.7 \end{pmatrix} \implies \lim_{k\to\infty} Q^{-1}A^kQ = Q^{-1} \lim_{k\to\infty} \begin{pmatrix} 1^n & 0 \\ 0 & 0.7^n \end{pmatrix} Q = \begin{pmatrix} 2/3 & 2/3 \\ 1/3 & 1/3 \end{pmatrix}$ (Notice that $\begin{pmatrix} 2/3 & 1/3 \end{pmatrix}$ is an eigenvalue corresponding to 1)

$$\implies \begin{pmatrix} a_n \\ b_n \end{pmatrix} = \begin{pmatrix} 2/3 & 2/3 \\ 1/3 & 1/3 \end{pmatrix} \begin{pmatrix} a_0 \\ b_0 \end{pmatrix} = \begin{pmatrix} 2(a_0 + b_0)/3 \\ (a_0 + b_0)/3 \end{pmatrix}$$

Chapter 3

Inner products

Through out the chapter, we assume that $F = \mathbb{R}$ or $\mathbb{C}$.

3.1 Definition

Definition 3.1.1. Let V be a vector space over F. An **inner product** $\langle \cdot, \cdot \rangle : V \times V \to F$ is a function such that $\forall\ x, y, z \in V, \forall\ c \in F$

- $\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$
- $\langle cx, y \rangle = c \langle x, y \rangle$
- $\langle x, y \rangle = \overline{\langle y, x \rangle}$
- $\langle x, x \rangle > 0$ if $x \neq 0$
  (Note that the condition $\langle x, x \rangle > 0$ implicitly say that $\langle x, x \rangle \in \mathbb{R}$)

Theorem 3.1.1. Let V be an inner product space. Then $\forall\ x, y, z \in V$, $\forall\ c \in F$

- $\langle x, y + z \rangle = \langle x, y \rangle + \langle x, z \rangle$
- $\langle x, cy \rangle = \overline{c} \langle x, y \rangle$
- $\langle x, 0 \rangle = \langle 0, x \rangle = 0$
- $\langle x, x \rangle = 0 \iff x = 0$
- If $\langle x, y \rangle = \langle x, z \rangle$ holds for all $x \in V$ then $y = z$.

Remark 3.1.1. A function $h : V \times V \to F$ is said to be

- **linear** in the first argument if

$$h(x + y, z) = h(x, z) + h(y, z),\ \ h(cx, y) = ch(x, y)$$

- **semilinear** in the second argument if

$$h(x, y + z) = h(x, y) + h(x, z),\ \ h(x, cy) = \overline{c}h(x, y)$$

- **sesquilinear** or **Hermitian** if it is linear in the first argument and semilinear in the second argument.

- **positive definite** if $h(x,x) > 0 \ \forall \ x \neq 0$ and $h(0,0) = 0$
- **semi-positive definite** if $h(x,x) \geq 0 \ \forall x \in V$
- **nondegenarate** if $\langle x, y\rangle = \langle x, z\rangle \ \forall \ x \in V \iff y = z$

Thus, an inner product can also be defined as a positive definite Hermition form.

Example 3.1.1. Define $\langle \cdot, \cdot \rangle$ on F^n by

$$\langle (a_1, ..., a_n), (b_1, ..., b_n)\rangle = \sum_{i=1}^{n} a_i \overline{b_i} \text{ or say } \langle x, y\rangle = x^T \overline{y}$$

which is called the **standard inner product** on F^n.

Example 3.1.2. Let $V = \mathbb{C}^2$. For $x = (a_1, a_2), y(b_1, b_2)$ define $\langle x, y\rangle$ by

$$\langle x, y\rangle := x \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \overline{y}^t$$

Check $\langle \cdot, \cdot \rangle$ is an inner product :

- linear conditions : Ok!
- $\langle y, x\rangle = \overline{\langle x, y\rangle}$:
$$\overline{\langle x, y\rangle} = \overline{x} A y^t = \left(\overline{x} A y^t\right)^t = y A \overline{x}^t = \langle y, x\rangle$$
- $\langle x, x\rangle \geq 0$
  Observe that
$$A = \begin{pmatrix} 1 & 1/2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 3/2 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1/2 & 1 \end{pmatrix}$$
  For $0 \neq x \in V$, let $(a, b) = x \begin{pmatrix} 1 & 1/2 \\ 0 & 1 \end{pmatrix} \neq 0$, then
$$\langle x, x\rangle = \begin{pmatrix} a & b \end{pmatrix} \begin{pmatrix} 3/2 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} \overline{a} \\ \overline{b} \end{pmatrix} = \frac{3}{2}|a|^2 + 2|b|^2 > 0$$

Example 3.1.3. $V = \{\text{complex value continuous function on } \mathbb{R}\}$. Define

$$\langle x, y\rangle = \int_0^1 f(t)\overline{g(t)} \ dt$$

is an inner product.

Example 3.1.4. Let $A \in M_{n\times n}(F)$. Then the **conjugate transpose** or **adjoint** of A is the matrix $A^* := \overline{A}^t$ i.e. $(A^*)_{ij} = \overline{A_{ji}}$

Remark 3.1.2. The standard inner product on $\mathbb{C}^n$ is often written xy^*

Example 3.1.5. $V = M_{n\times n}(F)$. Define $\langle A, B\rangle := \operatorname{tr}(AB^*)$
Check $\langle \cdot, \cdot \rangle$ is an inner product :

- linear conditions : OK!

- $\overline{\langle A, B\rangle} = \overline{\text{tr}(AB^*)} = \text{tr}(\overline{A}B^t) = \text{tr}((\overline{A}B^t)^t) = \text{tr}(A^*B) = \text{tr}(BA^*) = \langle B, A\rangle$
- $\langle A, A\rangle = \text{tr}(AA^*) = \sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}(A^*)_{ji} = \sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}\overline{A}_{ij} > 0$

This inner product is called the **Frobenius inner product** of $M_{n\times n}(F)$

Remark 3.1.3. $(AB)^* = B^*A^*$

Definition 3.1.2. Let V be an inner product space. For $v \in V$, the **norm** or the **length** of x is defined to be

$$\|x\| = \sqrt{\langle x, x\rangle}$$

Theorem 3.1.2. $\forall x \in V$, $c \in F$

- $\|cx\| = |c|\|x\|$
- $\|x\| \geq 0$ and $\|x\| = 0 \iff x = 0$
- **Cauchy-Schwarz inequality** : $|\langle x, y\rangle| \leq \|x\|\|y\|$
- **triangle inequality** : $\|x + y\| \leq \|x\| + \|y\|$

Proof:

- $\|cx\|^2 = \langle cx, cx\rangle = c\langle x, cx\rangle = c\overline{c}\langle x, x\rangle = \|c\|^2\|x\|^2$
- obvious
- If $y = 0$, then inequality holds trivially. Now, we assume $y \neq 0$. We have $\|x - cy\| \geq 0$ $\forall c \in F$ i.e.

$$\begin{aligned} 0 &\leq \langle x - cy, x - cy\rangle = \langle x, x\rangle - \langle cy, x\rangle - \langle x, cy\rangle + \langle cy, cy\rangle \\ &= \|x\|^2 - c\overline{\langle x, y\rangle} - \overline{c}\langle x, y\rangle + \|c\|^2\|y\|^2 \end{aligned}$$

  Now, we choose $c = \dfrac{\langle x, y\rangle}{\langle y, y\rangle}$, then we obtain

$$0 \leq \|x\|^2 - \frac{|\langle x, y\rangle|^2}{\|y\|^2} - \frac{|\langle x, y\rangle|^2}{\|y\|^2} + \frac{|\langle x, y\rangle|^2}{\|y\|^4}\|y^2\| \implies |\langle x, y\rangle|^2 \leq \|x\|^2\|y\|^2$$

- $\|x + y\|^2 = \langle x + y, x + y\rangle = \langle x, x\rangle + \langle x, y\rangle + \langle y, x\rangle + \langle y, y\rangle = \|x\|^2 + 2\text{Re}\langle x, y\rangle + \|y\|^2 \leq \|x\|^2 + 2|\langle x, y\rangle| + \|y\|^2 = (\|x\| + \|y\|)^2$

$\square$

Example 3.1.6. In the case of standard inner product on $\mathbb{C}^n$. Cauchy-Schwarz inequality

$$\left|\sum_{j=1}^{n} a_j\overline{b_j}\right| \leq \left(\sum_{j=1}^{n} |a_j|^2\right)^{1/2} \left(\sum_{j=1}^{n} |b_j|^2\right)^{1/2}$$

Recall that in $\mathbb{R}^2$ or $\mathbb{R}^3$, we have

$$\frac{\langle v_1, v_2\rangle}{\|v_1\|\|v_2\|} = \cos\theta,\ 0 \leq \theta \leq \pi$$

In particular, $\langle v_1, v_2\rangle = 0 \iff \theta = \pi/2$

Definition 3.1.3. Let V be an inner product space

- Two vector $x, y \in V$ are **orthogonal** or **perpendicular** if $\langle x, y\rangle = 0$
- A subset S of V is **orthogonal** if any 2 vectors in S are orthogonal.
- A vector space x in V is a unit vector if $\|x\| = 1$
- A subset S of V is **orthonormal** if S is orthogonal and every vector in S is a unit vector.

(Note that under our definition, an orthogonal subset may contain 0, but an orthonormal subset cannot have 0)

Example 3.1.7. $V = \{\text{continuous complex-valued functions on } \mathbb{R}\}$. Define

$$\langle f, g\rangle = \frac{1}{2\pi}\int_0^{2\pi} f(t)\overline{g(t)}\ dt$$

Then the set $\{f_n(t) := e^{int} : n \in \mathbb{Z}\}$ is orthonormal. Since

$$\langle f_n, f_m\rangle = \frac{1}{2\pi}\int_0^{2\pi} e^{i(n-m)t}\ dt = \begin{cases} 1 & \text{, if } n = m \\ \left.\dfrac{e^{i(n-m)t}}{2\pi i(n-m)}\right|_0^{2\pi} = 0 & \text{, if } n \neq m \end{cases}$$

Note that the symbol δ_{ij} defined by $\delta_{ij} = \begin{cases} 1 & \text{, if } i = j \\ 0 & \text{, if } i \neq j \end{cases}$ is called the **Kronecker delta symbol**.

3.2 The Gram-Schmidt orthogonalization process and orthogonal complement

Definition 3.2.1. Let V be an inner product. A subset of V is an **orthonormal basis** if it is a basis that is orthonormal.

Theorem 3.2.1. 3.2 V an inner product space $S = \{v_1, ..., v_n\}$ a finite orthogonal subset of V, $v_j \neq 0$ for all j. If $v \in \text{span}(S)$, then

$$v = \sum_{k=1}^{n} \frac{\langle v, v_k\rangle}{\langle v_k, v_k\rangle} v_k$$

In particular, if S is orthonormal, then

$$v = \sum_{k=1}^{n} \langle v, v_k\rangle v_k$$

Proof: Say $v = a_1 v_1 + \cdots + a_k v_k$. We have

$$\langle v, v_k \rangle = \sum_{i=1}^{n} a_j \langle v_i, v_k \rangle = a_k \langle v_k, v_k \rangle \implies a_k = \frac{\langle v, v_k \rangle}{\langle v_k, v_k \rangle}$$

Since $v_k \neq 0 \rightsquigarrow \langle v_k, v_k \rangle \neq 0$. $\square$

Corollary 3.2.1. If S is an orthogonal subset of V consisting of nonzero vector, then S is linear independent.

Proof: Note that, by Theorem , if $0 = a_1 v_1 + \cdots + a_n v_n \in S$, then

$$a_i = \frac{\langle 0, v_k \rangle}{\langle v_k, v_k \rangle} = 0$$

So S is linearly independent. $\square$

Now, we wonder to know how to find an orthonormal basis? For example in $\mathbb{R}^2$.
Problem: Given $v_1, v_2 \in \mathbb{R}^2$, find $c \in \mathbb{R}$ such that $(v_2 - cv_1) \perp v_1$. This constant c must satisfy

$$0 = \langle v_2 - cv_1, v_1 \rangle \implies x = \frac{\langle v_2, v_1 \rangle}{\langle v_1, v_1 \rangle}$$

In general, we have this theorem.

Theorem 3.2.2 (Gram—Schmidt orthogonalization process). Given $S = \{w_1, ..., w_n\}$ is linearly independent. Define $S' = \{w_1, ..., w_n\}$ by

$$\begin{aligned}
v_1 &= w_1 \\
v_2 &= w_2 - \frac{\langle w_1, v_1 \rangle}{\langle v_1, v_1 \rangle} v_1 \\
v_2 &= w_2 - \frac{\langle w_1, v_1 \rangle}{\langle v_1, v_1 \rangle} v_1 - \frac{\langle w_1, v_2 \rangle}{\langle v_2, v_2 \rangle} v_2 \\
&\vdots \\
v_k &= w_k - \sum_{i=1}^{k-1} \frac{\langle w_k, v_i \rangle}{\langle v_i, v_i \rangle} v_i
\end{aligned}$$

Then S' is orthogonal consisting of nonzero vectors and satisfy $\text{span}(S) = \text{span}(S')$ i.e. S' is an orthogonal basis for $\text{span}(S)$. If we wnat an orthonormal basis simply let

$$S' = \left\{ \frac{v_1}{\|v_1\|}, ..., \frac{v_k}{\|v_k\|} \right\}$$

Proof: We'll prove by induction that

- $\langle v_k, v_j \rangle = 0 \ \forall j = 1, ..., k-1$
- $v_k \neq 0$

The statements clearly hold for v_1. Assume that the statement hold up to v_k. Consider $v_{k+1} = w_{k+1} - \sum_{j=1}^{k} \frac{\langle w_{k+1}, v_j \rangle}{\langle v_j, v_j \rangle} v_j$. For $i = 1, ..., k$, we have

$$\langle v_{k+1}, v_i \rangle = \left\langle w_{k+1} - \sum_{j=1}^{k} \frac{\langle w_{k+1}, v_j \rangle}{\langle v_j, v_j \rangle} v_j, v_i \right\rangle = \langle w_{k+1}, v_i \rangle - \sum_{j=1}^{k} \frac{\langle w_{k+1}, v_j \rangle}{\langle v_j, v_j \rangle} \langle v_j, v_i \rangle$$

By induction hypothesis, $\langle v_j, v_i \rangle = 0$ for all $1 \leq i, j \leq k$ and $i \neq j$. Thus

$$\langle v_{k+1}, v_i \rangle = \langle w_{k+1}, v_i \rangle - \frac{\langle w_{k+1}, v_i \rangle}{\langle v_i, v_i \rangle} \langle v_i, v_i \rangle = 0$$

Also, because

$$v_{k+1} = w_{k+1} + (\text{linear combination of } w_1, ..., w_k)$$

and $\{w_1, ..., w_{k+1}\}$ is assumed to be linearly independent. Thus $v_{k+1} \neq 0$.

Now, we show that $\operatorname{span}(S') = \operatorname{span}(S)$. It is clear that $\operatorname{span}(S') \subseteq \operatorname{span}(S)$. Since $\#(S) = \#(S')$ and both of S, S' are linearly independent, we have $\operatorname{span}(S) = \operatorname{span}(S')$. □

Example 3.2.1. Let $V = \mathbb{R}[x]$ and $\langle f, g \rangle = \int_{-1}^{1} f(x)g(x)dx$. Let $S = \{1, x, x^2\}$, then

$$\begin{aligned} v_1 &= 1 \\ v_2 &= x - \frac{\langle x, v_1 \rangle}{\langle v_1, v_1 \rangle} v_1 = x \\ v_3 &= x^2 - \frac{\langle x^2, v_1 \rangle}{\langle v_1, v_1 \rangle} v_1 - \frac{\langle x^2, v_2 \rangle}{\langle v_2, v_2 \rangle} v_2 = x^2 - \frac{1}{3} \end{aligned}$$

After orthonormalizing, we get

$$w_1 = \frac{1}{\sqrt{2}}, w_2 = \frac{\sqrt{3}}{\sqrt{2}} x, w_3 = \frac{\sqrt{45}}{\sqrt{8}} (x^2 - \frac{1}{3})$$

is a orthonormal basis for $\operatorname{span}(S)$.

Remark 3.2.1. The polynomials $\frac{1}{\sqrt{2}}, \frac{\sqrt{3}}{\sqrt{2}} x, ...$ derived this way are called the **Legendre polynomial**

Theorem 3.2.3. *V be a finited-dimensional inner product space. Then V has an orthonormal basis $\mathcal{B}$. More general, if $\mathcal{B} = \{v_1, ..., v_n\}$ and $x \in V$, then*

$$x = \sum_{k=1}^{n} \langle x, v_k \rangle v_k$$

Corollary 3.2.2. Let V be a finite-dimensional inner product space with an orthonormal basis $\mathcal{B} = \{v_1, ..., v_n\}$. Let $T : V \to V$ be a linear operator. Let $A = [T]_{\mathcal{B}}$, then $A_{ij} = \langle T(v_j), v_i \rangle$

Proof: By Theorem 3.2, $T(v_j) = \sum_{i=1}^{n} \langle T(v_j), v_i \rangle v_i \implies A_{ij} = \langle T(v_j), v_i \rangle$. □

Definition 3.2.2. Let V be an inner product space, S be a nonempty subset of V. The set

$$S^{\perp} := \{v \in V : \langle v, x \rangle = 0 \ \forall x \in S\}$$

is called the **orthogonal complement** of S.

Remark 3.2.2. $S^{\perp}$ is a subspace of V. For example, $\{0\}^{\perp} = V, V^{\perp} = \{0\}$

Theorem 3.2.4. Let W be a finite-dimensional subspace of an inner product space V ($\dim V < \infty$). Then $V = W \oplus W^\perp$. More precisely, let $\{v_1, ..., v_k\}$ be an orthonormal basis for W. For $x \in V$, let $w = \sum_{i=1}^{k} \langle x, v_i \rangle v_i$ and $w^\perp = x - w$, then $w \in W, w^\perp \in W^\perp$.

Proof: For $x \in V$, let $w = \sum_{i=1}^{k} \langle x, v_i \rangle v_i$ and $w^\perp := x - w$. Then $w \in W$ and

$$\langle w^\perp, v_i \rangle = \langle x, v_i \rangle - \sum_{j=1}^{k} \langle x, v_j \rangle \langle v_j, v_i \rangle = \langle x, v_i \rangle - \langle x, v_i \rangle = 0 \rightsquigarrow w^\perp \in W^\perp$$

So $x = w + w^\perp \in W + W^\perp$. Now, if $v \in W \cap W^\perp$. Since $v \in W$ and $v \in W^\perp$, we have $\langle v, v \rangle = 0$ i.e. $v = 0$. So $V = W \oplus W^\perp$. □

Remark 3.2.3. The condition that $\dim W < \infty$ is necessary. We see the counter example in below.

Example 3.2.2. Let $V = \{f \in C^0([0, 1])\}$. Define $\langle \cdot, \cdot \rangle$ by

$$\langle f, g \rangle = \int_0^1 f(x)g(x)\ dx$$

Let $W = \{f \in W : f(0) = 0\}$
Claim: $W^\perp = \{0\}$ and hence $V \neq W + W^\perp$.
pf. Suppose $g(x) \in W^\perp$. Define $f \in V$ by $f : x \mapsto xg(x)$. Now, $g \in W^\perp$ and $f \in W$

$$\implies 0 = \langle f, g \rangle = \int_0^1 xg(x)^2 dx$$

Since $0 \geq xg(x)^2$ is continuous on $[0, 1]$, so $xg(x)^2 = 0\ \forall x \in [0, 1] \implies g = 0$

Definition 3.2.3. The vector w in the statement of Theorem 3.2.4 is called the **orthogonal projection** of x on W.

Corollary 3.2.3. In the notation of Theorem 3.2.4, the vector w is the unique vector in W closest to x. That is $\forall w' \in W$, we have

$$\|x - w'\| \geq \|x - w\|$$

and " $=$ " hold if and only if $w' = w$.

Proof: Recall that Pythagorean theorem, if $\langle x, y \rangle = 0$, then $\|x + y\|^2 = \|x\|^2 + \|y\|^2$.
According to Theorem 3.2.4, $x-w = w^\perp \in W^\perp$. Now, $w-w' \in W \implies \langle x-w, w-w' \rangle = 0$. By Pythagorean theorem,

$$\|x - w'\|^2 = \|x - w + w - w'\|^2 = \|x - w\|^2 + \|w - w'\|^2 \geq \|x - w\|^2$$

Also " $=$ " holds $\iff \|w - w'\| = 0 \iff w = w'$ □

Corollary 3.2.4. Assume that S is an orthonormal subset of a finite-dimensional inner product space V. Then S can extended to an orthonormal basis for V.

Proof: Let $W = \text{span}(S)$, then S is a linearly independent and hence a basis for W. By Theorem 3.2.4, $V = W \oplus W^{\perp}$. By Theorem 3.2.3, $W^{\perp}$ has a orthonormal basis S'. Then $S \sqcup S'$ is a basis for V. It;'s easy to see that $S \sqcup S'$ is orthonormal. $\square$

Example 3.2.3. $V = \{\text{continuous complex-valued function on } [-\pi, \pi]\}$. Define $\langle \cdot, \cdot \rangle$ by

$$\langle f, g \rangle = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t)\overline{g(t)dt}$$

We have seen that the set $\{f_n(t) = e^{int} : n \in \mathbb{Z}\}$ is orthonormal. Let

$$W = \text{span}\{f_{-n}, ..., f_{-1}, f_0, f_1, ..., f_n\}$$

Let $f(t) = t$. Then the w in Theorem 3.2.4 in this cases is

$$w = \sum_{k=-n}^{n} \langle f, f_k \rangle f_k$$

$$\langle f, f_k \rangle = \frac{1}{2\pi} \int_{-\pi}^{\pi} te^{-ikt}dt = \frac{-1}{2\pi t} te^{-ikt}\Big|_{-\pi}^{\pi} + \frac{1}{2\pi ik} \int_{-\pi}^{\pi} e^{-ikt}dt = \frac{(-1)^{k+1}}{ik} \quad k \neq 0$$

$$\implies w = -\sum_{k=-n}^{n} \frac{(-1)^k}{ik} f_k$$

By the Pythagorean theorem, $\|f\|^2 \geq \|w\|^2$ and

$$\|f\|^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} t^2 dt = \frac{\pi^2}{3}$$

Also, by generalization of Pythagorean theorem,

$$\|w\|^2 = 2 \sum_{k=1}^{n} \frac{1}{k^2} \implies \sum_{k=1}^{n} \frac{1}{k^2} \leq \frac{\pi^2}{6}$$

In fact, we have $\displaystyle\sum_{k=1}^{\infty} \frac{1}{k^2} = \frac{\pi^2}{6}$. This is an example of **Parseval's identity**.

3.3 Adjoint of a linear operator

Recall that the adjoint of $A \in M_{m \times n}(\mathbb{C})$ is defined to be $A^* := \overline{A}^t$. What's the meaning of A^* in terms of linear transformation?

Definition 3.3.1. A linear transformation form V to F (F arbitrary, not just $\mathbb{R}$ or $\mathbb{C}$) is called a linear functional.

Theorem 3.3.1. Let V be a finite-dimensional inner product space over F ($F = \mathbb{R}$ or $\mathbb{C}$) and let $g : V \to F$ be a linear functional. Then $\exists! y \in V$ such that $g(x) = \langle x, y \rangle \ \forall x$.

Proof: Let $\mathcal{B} = \{v_1, ..., v_n\}$ be an orthonormal basis.

First, we try to know that should y be?

Say $y = b_1 v_1 + \cdots + b_n v_n$. Then $g(v_i) = \langle v_i, y \rangle \iff g(v_i) = \overline{b}_i$ i.e. $b_i = \overline{g(v_i)}$. Which give us the uniqueness.

Let $y = \overline{g(v_1)}v_1 + \cdots + \overline{g(v_n)}v_n$. Then $\forall i, \langle v_i, y\rangle = g(v_i) \implies \forall x = a_1v_1 + \cdots + a_nv_n$.

$$\langle x, y\rangle = \sum_{i=1}^{n} a_i\langle v_i, \overline{g(v_i)}v_i\rangle = \sum_{i=1}^{n} a_i g(v_i) = g(x)$$

If y' is another vector such that $g(x) = \langle x, y'\rangle\ \forall x \in V \implies \langle x, y - y'\rangle = 0\ \forall x \in V \implies y - y' = 0$. Since $\langle\cdot,\cdot\rangle$ is nondegenerate. (a sesquilinear form $h : V \times V \to F$ is nondegenerate if $\langle x, y\rangle = 0\ \forall x \in V$, then $y = 0$) □

Remark 3.3.1. The proof uses only the properties that $\langle\cdot,\cdot\rangle$ is sesquilinear and nondegenerate. The property that $\langle\cdot,\cdot\rangle$ is positive definite is not needed.
(positive definite $\implies$ nondegenerate)

Let V be an inner product space $\dim V \leq \infty$ and $T : V \to V$ be a linear operator. Given $y \in V$, consider $g_y(x) := \langle Tx, y\rangle$. This is a linear functional. By Theroem 3.3.1, $\exists! y^* \in V$ such that $g_y(x) = \langle x, y^*\rangle$

Claim. The map $y \mapsto y^*$ is a linear transformation.

Proof: We need to check that

- $(y_1 + y_2)^* = y_1^* + y_2^* : \forall x$

$$\langle x, (y_1 + y_2)^*\rangle = \langle Tx, y_1 + y_2\rangle = \langle Tx, y_1\rangle + \langle Tx, y_2\rangle = \langle x, y_1^*\rangle + \langle x, y_2^*\rangle = \langle x, y_1^* + y_2^*\rangle$$

  Since $\langle\cdot,\cdot\rangle$ is nondegenerate, $(y_1 + y_2)^* = y_1^* + y_2^*$.

- $(cy)^* = cy^* : \forall x$

$$\langle x, (cy)^*\rangle = \langle Tx, cy\rangle = \overline{c}\langle Tx, y\rangle = \overline{c}\langle x, y^*\rangle = \langle x, cy^*\rangle$$

  Since $\langle\cdot,\cdot\rangle$ is nondegenerate, $(cy)^* = cy^*$.

□

Definition 3.3.2. The linear transformation $y \mapsto y^*$ is called the **adjoint** of T and is denoted by T^*.

Now let $\mathcal{B} = \{v_1, ..., v_n\}$ be an orthonormal basis for V. How are $[T]_{\mathcal{B}}$ and $[T^*]_{\mathcal{B}}$ related.
Let $A = [T]_{\mathcal{B}}$. By Corollary 3.2.2, $A_{ij} = \langle Tv_j, v_i\rangle$. Likewise, let $B = [T^*]_{\mathcal{B}}$, then $B_{ij} = \langle T^*v_j, v_i\rangle$. Now

$$B_{ij} = \langle T^*v_j, v_i\rangle = \overline{\langle v_i, T^*v_j\rangle} = \overline{\langle Tv_i, v_j\rangle} = \overline{A_{ji}} \implies B = A^*$$

Theorem 3.3.2. Let V be a finite-dimensional inner product space $T : V \to V$ a linear operator. Then $\exists! T^* : V \to V$ such that $\langle Tx, y\rangle = \langle x, T^*y\rangle\ \forall x, y \in V$. Moreover, if $\mathcal{B}$ is an orthonormal basis for V, then $[T^*]_{\mathcal{B}} = [T]_{\mathcal{B}}^*$.

Theorem 3.3.3. Let V be finite-dimensional inner product space, S, T : linear operators on V. Then

(a) $(S + T)^* = S^* + T^*$

(b) $(cT)^* = \overline{c}T^*$

(c) $(ST)^* = T^*S^*$

(d) $(T^*)^* = T$

(e) $I^* = I$

And the matrix version

(a) $(A+B)^* = A^* + B^*$

(b) $(cA)^* = \bar{c}B^*$

(c) $(AB)^* = B^*A^*$

(d) $(A^*)^* = A$

(e) $I^* = I$

Proof: $(a), (b), (e)$ is trivial, so we only proof $(c), (d)$.

(c) For all $x, y \in V$

$$\begin{aligned} &\langle x, (ST)^*y\rangle = \langle STx, y\rangle = \langle Tx, S^*y\rangle = \langle x, T^*S^*y\rangle \quad \forall x, y \in V \\ \implies &(ST)^*y = T^*S^*y \quad \forall y \in V \implies (ST)^* = T^*S^* \end{aligned}$$

(d) $\forall x, y \in V$

$$\langle x, Ty\rangle = \overline{\langle Ty, x\rangle} = \overline{\langle y, T^*x\rangle} = \langle T^*x, y\rangle = \langle x, (T^*)^*y\rangle$$

□

3.4 Application of data analysis

3.4.1 least square approximation

Problem: Given a set of data $\{(x_i, y_i) : i = 1, ..., m\}$ and draw it on $\mathbb{R}$. Find a line $y = ax + b$ that fits to the data "best". Here we say $y = ax + b$ is the best fit if

$$E = \sum_{i=1}^{m} (y_i - ax_i - b)^2$$

is minimized.
Ideal: Observe that E is square of the length of the vector

$$\begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} - \begin{pmatrix} ax_1 + b \\ \vdots \\ ax_m + b \end{pmatrix} = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} - \underbrace{\begin{pmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_m & 1 \end{pmatrix}}_{:=A} \begin{pmatrix} a \\ b \end{pmatrix}$$

This is a special case of the following problem :
Given $y \in F^m, A \in M_{m\times n}(F)$, where $F = \mathbb{R}$ or $\mathbb{C}$. Find $x \in F^n$ such that $\|y - Ax\|$ is minimized. Let $W = \{Ax : x \in F^n\}$ and w be the orthogonal projection of y on W. Then by Corollary 3.2.3

$$\min_{x\in F^n} \|y - Ax\| = \|w^\perp\|$$

Notation 3.4.1. We let $\langle\cdot,\cdot\rangle_n, \langle\cdot,\cdot\rangle_m$ denoted the standard inner product on F^n, F^m.

Lemma 3.4.1. Let $A \in M_{m\times n}(F), x \in F^n, y \in F^m$. Then

$$\langle Ax, y\rangle_m = \langle x, A^*y\rangle_n$$

Moreover, A^* is the unique matrix with this property.

Proof: Recall that if $v_1, v_2 \in F^n$ regarded as column vectors, then $\langle v_1, v_2\rangle_n = v_2^* v_1$. Now, regard x, y as column vectors

$$\langle x, A^*y\rangle_n = (A^*y)^*x = y^*(A^*)^*x = y^*Ax = \langle Ax, y\rangle_m$$

□

Lemma 3.4.2. For $A \in M_{m\times n}(F)$, we have

$$\operatorname{rank}(A^*A) = \operatorname{rank}(A)$$

Proof: By rank-nullity theorem, it is suffices to prove that $\operatorname{nullity}(A^*A) = \operatorname{nullity}(A)$. We claim that $\ker(A^*A) = \ker(A)$. Clearly, $\ker(A) \subseteq \ker(A^*A)$. Conversely, if $x \in \ker(A^*A)$ i.e. $A^*Ax = 0$, then $\langle x, A^*Ax\rangle = 0$. By Lemma 3.4.1, $\langle Ax, Ax\rangle = \langle x, A^*Ax\rangle = 0$. Since $\langle \cdot, \cdot\rangle$ is positive definite, we have $Ax = 0 \rightsquigarrow x \in \ker A$. □

Corollary 3.4.1. Let $A \in M_{m\times n}(F)$. If $\operatorname{rank}(A) = n$ then A^*A is invertible.

Theorem 3.4.1. Given $y \in F^m, A \in M_{m\times n}(F)$ $\exists x_0 \in F^n$ such that $A^*Ax_0 = A^*y$ and

$$\|y - Ax_0\| \le \|y - Ax\| \ \forall x \in F^n$$

Moreover, if $\operatorname{rank} A = n$, then x_0 is unique and is given by $x_0 = (A^*A)^{-1}A^*y$.

Proof: By Corollary 3.2.3, $\exists x_0 \in X$ such that $\|y - Ax_0\|$ is minimized and is a vector such that Ax_0 is the orthogonal projection of y on $W = \{Ax : x \in F^n\}$ i.e. x_0 satisfies

$$\langle Ax, y - Ax_0\rangle_m = 0 \ \forall x \in F^n$$

By Lemma 3.4.1, $\langle x, A^*(y - Ax_0)\rangle_n \ \forall x \in F^n \implies A^*(y - Ax_0) = 0$ i.e. $A^*Ax_0 = A^*y$.
If $\operatorname{rank}(A) = n$, by Corollary 3.4.1 A^*A is invertible. Hence, $x_0 = (A^*A)^{-1}A^*y$ is unique. □

Example 3.4.1. Given $(1.2), (2, 3), (3.5), (4, 7)$ in $\mathbb{R}^2$. Let $y = \begin{pmatrix}2\\3\\5\\7\end{pmatrix}$ and $A = \begin{pmatrix}1 & 1\\2 & 1\\3 & 1\\4 & 1\end{pmatrix}$. Then we want to find the minimized of $E = \left\| y - A\begin{pmatrix}a\\b\end{pmatrix}\right\|^2$. By Theorem 3.4.1, then best (a, b) is given by

$$\begin{pmatrix}a\\b\end{pmatrix} = (A^*A)^{-1}A^*y$$

Here $A^*A = \begin{pmatrix}30 & 10\\10 & 4\end{pmatrix}$ is invertible and $A^*y = \begin{pmatrix}51\\17\end{pmatrix} \implies \begin{pmatrix}a\\b\end{pmatrix} = \begin{pmatrix}30 & 10\\10 & 4\end{pmatrix}^{-1}\begin{pmatrix}51\\17\end{pmatrix} = \begin{pmatrix}1.7\\0\end{pmatrix}$. Hence, the best fit is $y = 1.7x$.

Remark 3.4.1. We can also consider the problem of finding the polynomial with $\deg n$. For example, if we want to find a parabola $y = ax^2 + bx + c$ that best fits the given data. In such a problem, we let

$$y = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix}, \quad A = \begin{pmatrix} x_1^2 & x_1 & 1 \\ \vdots & & \vdots \\ x_m^2 & x_m & 1 \end{pmatrix}$$

And find the minimal of $E = \|y - Av\|$ for $v \in F^3$.

3.4.2 Minimal solution to a system of linear equation

Problem: Suppose $Ax = b$ is consistent (i.e. having a solution). Find a solution s such that $\|s\|$ is minimal.
Ideal: Let $W = \ker A$. Let x_0 be the solution of $Ax = b$. Then every solution for $Ax = b$ can we write as $x_0 + v$ for some $v \in \ker A$. The minimal solution s is a vector such that $x_0 - s$ is the orthogonal projection of x_0 on W.

Lemma 3.4.3. $(\ker A)^\perp = \operatorname{Im} A^*$

Proof: If $x \in \operatorname{Im} A^*$, say $x = A^*v$ for some $v \in F^m$, then $\forall y \in \ker A$, we have

$$\langle x, y\rangle = \langle A^*x, y\rangle = \langle x, Ay\rangle = \langle x, 0\rangle = 0$$

This proves that $\operatorname{Im} A^* \subseteq (\ker A)^\perp$. Now

$$\begin{aligned} \dim(\ker A)^\perp &= n - \dim\ker A && \text{(Theorem 3.2.4)} \\ &= \operatorname{rank} A = \operatorname{rank} \overline{A} = \operatorname{rank} \overline{A}^t = \dim \operatorname{Im} A^* \end{aligned}$$

Combine with $\operatorname{Im} A^* \subseteq (\ker A)^\perp$, we have $\operatorname{Im} A^* = (\ker A)^\perp$ □

Claim 1. We have $s \in \operatorname{Im} A^*$ and is the unique solution of $Ax = b$ lying in $\operatorname{Im} A^*$.

Proof: We have proved that $(\ker A)^\perp = \operatorname{Im} A^*$. If s' is another solution of $Ax = b$ lying in $\operatorname{Im} A^*$, then $A(s - s') = 0 \implies s - s' \in \ker A \cap \operatorname{Im} A^* = \ker A \cap (\ker A)^\perp = \{0\}$ i.e. $s = s'$. □

Theorem 3.4.2. Let $A \in M_{m\times n}(F)$, $b \in F^m$. Assume that $Ax = b$ is consistent.

- There exists a unique minimal solution s of $Ax = b$.
- s is the unique solution of $Ax = b$ lying in $\operatorname{Im} A^*$

i.e. if $A(A^*u) = b$, then $s = A^*u$.

3.5 Normal and self-adjoint operators

Motivation: Let $T : V \to V$ be a linear operator. We have seen that there are advantages in using bases consisting of eigenvectors. We have also seen that there are advantages in using orthonormal bases.
Problem: Can we find bases that are orthonormal and consist of eigenvectors?
Observation: If V has an orthonormal basis $\mathcal{B}$ consisting of eigenvectors of T, then $[T]_\mathcal{B}$ is diagonal. Since $\mathcal{B}$ is orthonormal. We have $[T^*]_B = [T]_B^*$, which is also diagonal.

$$\implies [T]_\mathcal{B}[T^*]_\mathcal{B} = [T^*]_\mathcal{B}[T]_\mathcal{B} \implies TT^* = T^*T$$

Definition 3.5.1. Let T be a linear operator on an inner product space. We say T is **normal** if $TT^* = T^*T$. Also, a matrix $A \in M_{n\times n}(F)$ is said to be **normal** if $AA^* = A^*A$.

Example 3.5.1. If A is real skew-symmetric, then A is normal.

Question: Suppose that T is normal. Can we always find an orthonormal basis consisting of eigenvectors if T.
Answer: If $F = \mathbb{R}$, the answer is no. For example, $T_\theta : \mathbb{R}^2 \to \mathbb{R}^2$ which is rotation by θ with respect to the standard basis $\mathcal{B}$, we have

$$[T_\theta]_{\mathcal{B}} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

Then we have

$$[T]_{\mathcal{B}}[T]_{\mathcal{B}}^* = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}\begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = [T]_{\mathcal{B}}^*[T]_{\mathcal{B}}$$

However, we have seen that T has no eigenvalues in $\mathbb{R}$ unless $\theta = k\pi$ for some $k \in \mathbb{Z}$.
On the other hand, if $F = \mathbb{C}$, we'll show in Theorem **??** that the answer is yes.

Theorem 3.5.1. T : normal

(1) $\|Tx\| = \|T^*x\|\ \forall x \in V$

(2) $T - cI$ is normal for any $c \in F$

(3) If x is an eigenvalue of T, then x is an eigenvector of T^*. In fact, if $Tx = \lambda x$, then $T^*x = \overline{\lambda}x$.

(4) If λ_1, λ_2 are distinct eigenvalues of T with corresponding eigenvectors x_1, x_2, then x_1, x_2 are orthogonal.

Proof:

(1) We have $\langle Tx, Tx\rangle = \langle x, T^*Tx\rangle = \langle x, TT^*x\rangle = \langle x, (T^*)^*T^*x\rangle = \langle T^*x, T^*x\rangle$

(2) $(T - cI)^*(T - cI) = T^*T - c(T + T^*) + c^2I = TT^* - c(T + T^*) + c^2I = (T - cI)(T^* - cI)$

(3) x is an eigenvector of T with eigenvalue $\lambda \iff \underbrace{(T - \lambda I)}_{:=S}x = 0 \iff \langle Sx, Sx\rangle = 0 \iff \langle S^*x, S^*x\rangle = 0 \iff S^*x = 0 \iff T^*x = \overline{\lambda}x$.

Where we use the fact that S is normal (by (2)) and (1).

(4) Consider $\langle Tx_1, x_2\rangle$, we have

$$\langle Tx_1, x_2\rangle = \langle \lambda_1 x_1, x_2\rangle = \lambda_1\langle x_1, x_2\rangle$$

Also,

$$\langle Tx_1, x_2\rangle = \langle x_1, T^*x_2\rangle \overset{\text{by (3)}}{=\!=} \langle x_1, \lambda_2^* x_2\rangle = \lambda_2\langle x_1, x_2\rangle$$

Since $\lambda_1 \neq \lambda_2$, $\langle x_1, x_2\rangle = 0$.

$\square$

Theorem 3.5.2. Let V be a finite-dimensional complex inner product space. Assume $T : V \to V$ is normal. Then $\exists$ an orthonormal basis consisting of eigenvectors of T.

Before proof this theorem, we see want we have if we have not the condition of normal.

Theorem 3.5.3 (Schur)**.** Let V be a complex inner product space with $\dim V < \infty$ and $T : V \to V$. Then $\exists$ an orthonormal basis $\mathcal{B}$, such that $[T]_{\mathcal{B}}$ is upper triangular.

Remark 3.5.1. Given $T : \mathbb{C}^n \to \mathbb{C}^n$, by Theorem 3.2.3, $\exists$ orthonormal basis $\mathcal{B}$ such that $[T]_{\mathcal{B}}$ is upper triangular. Let $\mathcal{B}_0$ be the standard basis for $\mathbb{C}^n$. Then we have

$$[T]_{\mathcal{B}} = [\mathrm{id}]_{\mathcal{B}_0}^{\mathcal{B}}[T]_{\mathcal{B}_0}[\mathrm{id}]_{\mathcal{B}}^{\mathcal{B}_0}$$

Let $A = [T]_{\mathcal{B}_0}$ and $Q = [\mathrm{id}]_{\mathcal{B}}^{\mathcal{B}_0}$. If $\mathcal{B} = \{v_1, ..., v_n\}$, then

$$Q = \begin{pmatrix} v_1 & v_2 & \cdots & v_n \end{pmatrix}$$

Since $\mathcal{B}$ is orthonormal, we have

$$Q^*Q = \begin{pmatrix} v_1^* \\ v_2^* \\ \vdots \\ v_n^* \end{pmatrix} \begin{pmatrix} v_1 & v_2 & \cdots & v_n \end{pmatrix} = I_n \implies Q^{-1} = Q^*$$

i.e. Q is unitary. So we get the Matrix version of Theorem 3.5.3 :
If $A \in M_{n\times n}(\mathbb{C})$, then $\exists Q$: **unitary** $(def : Q^{-1} = Q)$ such that $Q^{-1}AQ$ is upper triangular.

Proof: We'll prove by induction on $n = \dim V$. For $n = 0, 1$: which is clear that will hold. For $n > 1$: Since V is over $\mathbb{C}$, T^* has an eigenvalue λ_n with a unit eigenvector v_n.

Notice that If W is a T^*-invariant subspace of V, then $W^\perp$ is T-invariant. Apply this property to $W = \mathbb{C}v_n$. We see that $W^\perp$ is a T-invariant subspace of dimension $n-1$. By induction hypothesis, $\exists$ orthogonal basis $\mathcal{B}' = \{v_1, ..., v_{n-1}\}$ for $W^\perp$ such that $[T|_{W^\perp}]_{\mathcal{B}'}$ is upper triangular. Let $\mathcal{B} = \{v_1, ..., v_n\}$, then

$$[T]_{\mathcal{B}} = \begin{pmatrix} & & & * \\ & [T|_{W^\perp}]_{\mathcal{B}'} & & * \\ & & & * \\ 0 & \cdots & 0 & * \end{pmatrix}$$

which is upper triangular since $[T|_{W^\perp}]_{\mathcal{B}'}$ is upper triangular. □

Back to the proof of Theorem 3.5.2 :

Proof: $(\Leftarrow)$: discussed earlier.

$(\Rightarrow)$: Let $\mathcal{B} = \{v_1, ..., v_n\}$ be an orthonormal basis such that $A = [T]_{\mathcal{B}}$ is upper triangular, as per Schur's theorem. Since A is upper triangular, v_1 is an eigenvector of T, say the corresponding to v_1 is λ_1. Consider A_{12}, we have $A_{12} = \langle Tv_2, v_1\rangle = \langle v_2, T^*v_1\rangle$. By Theorem 3.5.1, v_1 is also an eigenvector of T^* with eigenvalue $\overline{\lambda} \implies A_{12} = \langle v_2, \overline{\lambda}v_1\rangle = \lambda_1\langle v_2, v_1\rangle = 0$. Thus v_2 is an eigenvector of T, say with eigenvalue λ_2.

In general, suppose that we have proved that $\{v_1, ..., v_{n-1}\}$ are eigenvectors with eigenvalue $\lambda_1, ..., \lambda_{k-1}$. Then $\forall j \le k-1$, we have $A_{jk} = \langle Tv_k, v_j\rangle = \langle v_k, T^*v_j\rangle = \langle v_k, \overline{\lambda_j}v_j\rangle = \lambda_j\langle v_k, v_j\rangle = 0$ and hence v_k is an eigenvector. By induction, this proves that $A = [T]_{\mathcal{B}}$ is diagonal and thus $\mathcal{B}$ is orthonormal basis consist of eigenvector. □

Remark 3.5.2. Theorem 3.5.2 may not hold when $\dim V = \infty$. For example :
Consider the inner product space $\mathcal{H} = \mathbb{C}[0, 2\pi]$ with the orthonormal set S in Example 3.1.7.

Let $V = \text{span}(S)$, and let T and U be the linear operators on V defined by $T(f) = f_1 f$ and $U(f) = f_{-1}f$. Then

$$T(f_n) = f_{n+1} \text{ and } U(f_n) = f_{n-1}$$

for all integers n. Thus

$$\langle T(f_m), f_n\rangle = \langle f_{m_1}, n\rangle = \delta_{(m+1),n} = \delta_{m,(n-1)} = \langle f_m, f_{n-1}\rangle = \langle f_m, U(f_n)\rangle.$$

It follows that $U = T^*$.Furthermore, $TT^* = I = T^*T$; so T is normal.

We show that T has no eigenvectors. Suppose that f is an eigenvector of T, say $T(f) = \lambda f$ for some λ. Since V equals the span of S, we may write

$$f = \sum_{i=n}^{m} a_i f_i, \text{ where } a_m \neq 0.$$

Hence,

$$\sum_{i=n}^{m} a_i f_{i+1} = T(f) = \lambda f = \sum_{i=n}^{m} \lambda a_i f_i$$

By S is linearly independent, $a_m = 0$ ($\rightarrow\leftarrow$).

Definition 3.5.2. Let V be an inner product space. We say $T : V \to V$ is **self-adjoint** or **Hermitian** (usually used only when $F = \mathbb{C}$) is $T^* = T$.
Matrix version : If $A \in M_{m\times n}(F)$ satisfies $A^* = A$, then we say A is **self-adjoint** or **Hermitian**.

Lemma 3.5.1. Assume $\dim V < \infty$. $T : V \to V$ is self-adjoint. Then

(1) every eigenvalue of T is real

(2) $\text{ch}_T(x)$ splits completely over $\mathbb{R}$.

Proof:

(1) Say v is an eigenvector with eigenvalue λ. Then $\langle Tv, v\rangle = \langle \lambda v, v\rangle = \lambda\langle v, v\rangle$. On the other hand, $\langle Tv, v\rangle = \langle v, T^*v\rangle = \langle v, Tv\rangle = \langle v, \lambda v\rangle = \overline{\lambda}\langle v, v\rangle \implies \lambda = \overline{\lambda}$ i.e. $\lambda \in \mathbb{R}$.

(2) If $F = \mathbb{C}$: By Fundamental theorem of Algebra, $\text{ch}_T(x)$ splits over $\mathbb{C}$. Let λ be the root of $\text{ch}_T(x)$, then λ is eigenvalue. By (1), $\lambda \in \mathbb{R}$

If $F = \mathbb{R}$: let $\mathcal{B}$ be an orthonormal basis. Then $[T]_{\mathcal{B}} = [T^*]_{\mathcal{B}} = [T]^*_{\mathcal{B}} \implies A := [T]_{\mathcal{B}}$ is self-adjoint i.e. $A^t = A$. Now, define $L_A : \mathbb{C}^n \to \mathbb{C}^n$ by $L_A v = Av$. Since $A^* = A$ and the standard basis $\mathcal{B}_0$ is orthonormal, $L_A^* = L_A$ (since $[L_A^*]_{\mathcal{B}_0} = A^*$). Then $\text{ch}_{L_A}(x)$ splits over $\mathbb{R}$ completely. Then $\text{ch}_T(x) = \text{ch}_A(x) = \text{ch}_{L_A}(x)$ splits completely over $\mathbb{R}$.

□

Theorem 3.5.4. Suppose $F = \mathbb{R}$, $\dim V < \infty$, $T : V \to V$. Then $\exists$ an orthonormal basis for V consisting of eigenvectors of $T \iff T$ is self-adjoint.

Proof: ($\Rightarrow$) : If such a basis $\mathcal{B}$ exists then $[T]_{\mathcal{B}}$ is diagonal and $[T^*]_{\mathcal{B}} = [T]^*_{\mathcal{B}} = [T]_{\mathcal{B}}$, for first equation is by $\mathcal{B}$ is orthonormal, second equation is by $[T]_{\mathcal{B}}$ is diagonal and all eigenvalue are in $\mathbb{R}$. Thus $T^* = T$.

($\Leftarrow$) : Observe that the proof of Schur's theorem require only that $\text{ch}_T(x)$ splits completely. Since here T is a self-adjoint, lemma above says that $\text{ch}_T(x)$ splits completely over $\mathbb{R}$. Then the statement of Schur's theorem also holds here. □

3.6 Unitary and orthogonal operators

Definition 3.6.1. Let V be an inner product space and $T : V \to V$ be a linear operator. If $\|Tx\| = \|x\|\ \forall x \in V$, then we say T is **unitary** when $F = \mathbb{C}$, **orthogonal** when $F = \mathbb{R}$.
Matrix version : We say $A \in M_{n\times n}(\mathbb{C})$ is **unitary** if $A^*A = I$, $A \in M_{n\times n}(\mathbb{R})$ is **orthogonal** if $A^*A = I$.

Remark 3.6.1. If $\dim V < \infty$ and $\langle Tx, Tx\rangle = \langle x, x\rangle\ \forall x \in V$, then $\langle x, T^*Tx\rangle = \langle x, x\rangle$. According to the lemma below, we have $T^*T = I$. Thus, if $\mathcal{B}$ is an orthonormal basis for V, then $[T]_{\mathcal{B}}$ is a unitary/orthogonal.

Remark 3.6.2. Sometimes we also say T is an **isometry** if $\|Tx\| = \|x\|\ \forall x \in V$

Lemma 3.6.1. If $S : V \to V$ is self-adjoint and $\dim V < \infty$. Suppose that $\langle x, Sx\rangle = 0\ \forall x \in V$, then $S = 0$.

Proof: By Theorem 3.5.2+3.5.4, $\exists$ an orthonormal basis $\mathcal{B} = \{v_1, ..., v_n\}$ such that $[T]_{\mathcal{B}}$ is diagonal. Then $\forall i\ 0 = \langle v_i, Sv_i\rangle = \langle v_i, \lambda_i v_i\rangle = \overline{\lambda_i} \implies \lambda_i = 0 \forall i \implies S = 0.$ □

Theorem 3.6.1. $\dim V < \infty$, $T : V \to V$. The following statements are equivalent.

(a) $TT^* = T^*T = I$

(b) $\langle Tx, Ty\rangle = \langle x, y\rangle\ \forall x, y \in V$

(c) If $\mathcal{B}$ is an orthonormal basis for V, then $T(\mathcal{B})$ is also an orthonormal basis.

(Thus, in Euclidean space $\mathbb{R}^n$ or $\mathbb{C}^n$, an orthogonal/unitary operator can we thought of an operator that changes the choice of Cartesian coordinate system.)

(d) $\exists$ an orthonormal basis $\mathcal{B}$ for V such that $T(\mathcal{B})$ is an orthonormal basis.

(e) $\|Tx\| = \|x\|\ \forall x \in V$.

Proof:

- $(a) \Rightarrow (b)$: OK!
- $(b) \Rightarrow (c)$: Note that if $S = \{v_1, ..., v_k\}$ is orthonormal, then by assumption $\langle Tv_i, Tv_j\rangle = \langle v_i, v_j\rangle = \delta_{ij} \implies T(S)$ is orthonormal. Hence, if $\mathcal{B}$ is an orthonormal basis, then so is $T(\mathcal{B})$.
- $(c) \Rightarrow (d)$: Obvious.
- $(d) \Rightarrow (e)$: Let $\mathcal{B} = \{v_1, ..., v_k\}$ be an orthonormal basis such that $T(\mathcal{B})$ is also an orthonormal basis. For $x \in V$, we have $x = a_1v_1 + \cdots a_nv_n$ for some a_i. We have $\langle x, x\rangle = \sum_{i=1}^{n} |a_i|^2$ and
$$\langle Tx, Tx\rangle = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i\overline{a_j}\langle Tx_i, Tx_j\rangle = \sum_{1\le i,j\le n} a_i\overline{a_j}\delta_{ij} = \sum_{i=1}^{n} |a_i|^2 = \langle x, x\rangle$$
- $(e) \Rightarrow (a)$: By the lemma above.

□

Definition 3.6.2. $A, B \in M_{n\times n}(\mathbb{C})$, we say A, B are **unitarily/orthogonally** equivalent if exists a unitary/orthogonal matrix Q such that $A = Q^*BQ$.

Theorem 3.6.2 (Matrix version of Theorem 3.6.1)**.**
$A \in M_{n\times n}(\mathbb{C})$ is normal $\iff$ A is unitary equivalent to a diagonal matrix.
$A \in M_{n\times n}(\mathbb{R})$ is self-adjoint $\iff$ A is diagonally equivalent to a diagonal matrix.

Corollary 3.6.1. Let T be a linear operator on a finite-dimensional real inner product space V . Then V has an orthonormal basis of eigenvectors of T with corresponding eigenvalues of absolute value 1 if and only if T is both self-adjoint and orthogonal.

Proof: $(\Rightarrow)$: Suppose that V has an orthonormal basis $\{v_1, ..., v_n\}$ such that $T(v_i) = \lambda_i v_i$ and $|\lambda_i| = 1$ for all i. By Theorem 3.5.4, T is self adjoint. Thus, $(TT^*)(v_i) = T(\lambda_i v_i) = \lambda_i^2 v_i = v_i$ for each i. So $TT^* = I$ i.e. T is diagonal.

$(\Leftarrow)$: If T is self-adjoint, then by Theorem 3.5.4, we have that V possesses an orthonormal basis $\{v_1, ..., v_n\}$ such that $Tv_i = \lambda_i v_i$ for all i. If T is also orthogonal, we have

$$|\lambda_i| \cdot \|v_i\| = \|\lambda_i v_i\| = \|Tv_i\| = \|v_i\| \rightsquigarrow |\lambda_i| = 1 \ \forall i$$

□

Rigid motions in Euclidean space

Definition 3.6.3. We say $f : \mathbb{R}^n \to \mathbb{R}^n$ is a **rigid motion** if

$$\|f(x) - f(y)\| = \|x - y\| \ \forall x, y \in \mathbb{R}^n.$$

Theorem 3.6.3. If $f : \mathbb{R}^n \to \mathbb{R}^n$ is a rigid motion, then $\exists!\ T : \mathbb{R}^n \to \mathbb{R}^n$ is orthogonal operator and $\exists!$ a **translation** g (i.e. $g(x) = x - x_0$ for some fixed x_0) such that $f = g \circ T$.

Proof: Let $T : V \to V$ be defined by

$$T(x) = f(x) - f(0) \ \forall x \in V$$

We will show that T is an orthogonal operator, from which it follow that $f = g \circ T$, where g is the translation by $f(0)$. Observe that T us the composite of f and translation by $-f(0)$; hence T is a rigid motion. Furthermore, for any $x \in V$

$$\|T(x)\|^2 = \|f(x) - f(0)\|^2 = \|x - 0\|^2 = \|x\|^2$$

Now, we show that T is linear translation : $\forall x, y \in V$

$$\begin{aligned}\|T(x) - T(y)\|^2 &= \|T(x)\|^2 - 2\langle T(x), T(y)\rangle + \|T(y)\|^2 \\ &= \|x\|^2 - 2\langle T(x), T(y)\rangle + \|y\|^2\end{aligned}$$

Combine with $\|x - y\|^2 = \|x\|^2 - 2\langle x, y\rangle + \|y\|^2$ and $\|T(x) - T(y)\|^2 = \|x - y\|^2$, we have

$$\langle T(x), T(y)\rangle = \langle x, y\rangle.$$

Let $x, y \in V, a \in \mathbb{R}$, then

$$\begin{aligned}\|T(x + ay) - T(x) - aT(y)\|^2 &= \|(T(x + ay) - T(x)) - aT(y)\|^2 \\ &= \|T(x + ay) - T(x)\|^2 + a^2\|T(y)\|^2 - 2a\langle T(x + ay) - T(x), T(y)\rangle \\ &= \|(x + ay) - x\|^2 + a^2\|y\|^2 - 2a\left(\langle T(x + ay), T(y)\rangle - \langle T(x), T(y)\rangle\right) \\ &= 2a^2\|y\|^2 - 2a(\langle x + ay, y\rangle - \langle x, y\rangle = 0\end{aligned}$$

Thus $T(x+ay) = T(x) + aT(y)$ i.e. T is linear. Since T also preserves inner product, T is an orthogonal operator.
Uniqueness: Suppose that u_0, v_0 are in V and T, U are orthogonal operators on V such that

$$f(x) = T(x) + u_0 = U(x) + v_0 \ \forall x \in V \tag{*}$$

Substituting $x = 0$ in $(*)$ $\rightsquigarrow u_0 = v_0$ and hence the translation is unique. Therefore $T(x) = U(x) \ \forall x \in V$ i.e. $T = U$. $\square$

Theorem 3.6.4. $T : \mathbb{R}^2 \to \mathbb{R}^2$ is orthogonal operator, then $\det T = \pm 1$ and

$$\begin{array}{rcl} \det T = 1 & \iff T & \text{is rotation} \\ \det T = -1 & \iff T & \text{is reflection} \end{array}$$

Remark 3.6.3. Note that $\det T^* = \overline{\det(T)}$. Then if $TT^* = I$, then $(\det T)(\overline{\det T}) = 1 \implies |\det T| = 1$. In particular, if T is orthogonal $\rightsquigarrow \det T = \pm 1$.

3.7 Orthogonal projections and spectral theorem

- Recall that a linear operator $T : V \to V$ is said to be a projection if $T^2 = T$. If T is a projection, then $V = \operatorname{Im} T \oplus \ker T$ $(v = Tv + (v - Tv))$.
- Conversely, if $V = W_1 \oplus W_2$, we can define $T_i : V \to V$ by $T_i(v_1 + v_2) = v_i$ if $v_1 \in W_1, v_2 \in W_2$, then each T_i is a projection, called the projection on W_i along W_i', where $W_1' = W_2, W_2' = W_1$. Note that we have $T_1 + T_2 = O$.
- Now, if T is diagonalizable, say $V = \bigoplus_{i=1}^{k} E_{\lambda_i}$, where λ_i are the distinct eigenvalues of T, then we can define $T_i : V \to V$ by $T_i(v_1 + \cdots + v_k) = v_i$ if $v_j \in E_{\lambda_j}$. Then $T = \lambda_1 T_1 + \cdots + \lambda_k T_k$.
- Now, assume that V is an inner product space and T is normal/self-adjoint when $F = \mathbb{C}/F = \mathbb{R}$. By theorem 3.5.2+3.5.4, $\exists$ an orthonormal basis $\mathcal{B}$ such that $[T]_{\mathcal{B}}$ is diagonal and say $V = \bigoplus_{i=1}^{k} E_{\lambda_i}$ i.e.

$$[T]_{\mathcal{B}} = \begin{pmatrix} \lambda_1 I & & & O \\ & \lambda_2 I & & \\ & & \ddots & \\ O & & & \lambda_k I \end{pmatrix}$$

  Let T_i be defined as above, we have

$$[T_i]_{\mathcal{B}} = \begin{pmatrix} O & & & & \\ & \ddots & & & \\ & & I & & \\ & & & \ddots & \\ & & & & O \end{pmatrix}, \text{ only } (i,i)\text{-block is identity}$$

  Since $\mathcal{B}$ is a orthonormal basis having matrices of this form means that $E_{\lambda_i} \perp E_{\lambda_j}$ if $i \neq j$ and $E_{\lambda_i}^{\perp} = \bigoplus_{j \neq i} E_{\lambda_j}$, since $\bigoplus_{j \neq i} E_{\lambda_j} \subseteq E_{\lambda_i}^{\perp}$ and

$$V = E_{\lambda_i} \oplus E_{\lambda_i}^{\perp} \implies \dim E_{\lambda_i}^{\perp} = \dim V - \dim E_{\lambda_i} = \dim \bigoplus_{j \neq i} E_j$$

- If $V = \bigoplus_{i=1}^{n} W_i$, set $W_i' = \bigoplus_{j \neq i} W_i$, we have $W_i' = W_i^{\perp}$. Let T_i be the projection on W_i along W_i'. Then the discussion above says $(\operatorname{Im} T_i)^{\perp} = W_i' = \ker T_i$. Since $\dim V < \infty$, this also implies $(\ker T_i)^{\perp} = \operatorname{Im} T_i$

Definition 3.7.1. A projection T on an inner product space is said to be an **orthogonal projection** if $(\operatorname{Im} T)^{\perp} = \ker T$ and $(\ker T)^{\perp} = \operatorname{Im} T$ and $T^2 = T$.

Remark 3.7.1. When $\dim V = \infty$, $(W^{\perp})^{\perp}$ may not equal to W.

Theorem 3.7.1. Let V be an inner product space. Then T is an orthogonal projection $\iff$ T has an adjoint T^* and $T^2 = T = T^*$

Remark 3.7.2. $\dim V$ is not assumed to be finite, so the existence of T^* is not guaranteed. Not that when $\dim V < \infty$. It's easy to see the theorem holds. Say T is a orthogonal projection on a finite-dimensional space i.e. $V = \operatorname{Im} T \oplus \ker T$ and $\operatorname{Im} T \perp \ker T$. Choose an orthonormal basis $\mathcal{B}_1, \mathcal{B}_2$ for $\operatorname{Im} T, \ker T$ respectively. Then $\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2$ is an orthonormal basis for V and

$$[T]_{\mathcal{B}} = \begin{pmatrix} I & O \\ O & O \end{pmatrix}$$

Since $\mathcal{B}$ is orthonormal basis, $[T^*]_{\mathcal{B}} = [T]_{\mathcal{B}}^* = [T]_{\mathcal{B}} \implies T = T^*$.

Proof: $(\Rightarrow)$: Suppose that T is an orthogonal projection. For $x, y \in V$, write $x = x_1 + x_2,\ y = y_1 + y_2$ with $x_1, y_1 \in \operatorname{Im} T,\ x_2, y_2 \in \ker T$. Then $\langle x, Ty \rangle = \langle x_1 + x_2, y_1 \rangle = \langle x_1, y_1 \rangle$. Similarly, $\langle Tx, y \rangle = \langle x_1, y_1 + y_2 \rangle = \langle x_1, y_1 \rangle \implies T^*$ exists and is equal to T.

$(\Leftarrow)$: We need to prove $(\operatorname{Im} T)^{\perp} = \ker T,\ (\ker T)^{\perp} = \operatorname{Im} T$.

- $\ker T \subseteq (\operatorname{Im} T)^{\perp},\ \operatorname{Im} T \subseteq (\ker T)^{\perp}$:

  If $x \in \operatorname{Im} T, y \in \ker T$, then $x = Tx$ and hence $\langle x, y \rangle = \langle Tx, y \rangle = \langle x, Ty \rangle = 0$

- $(\operatorname{Im} T)^{\perp} \subseteq \ker T$: Suppose that y is a vector such that $\langle Tx, y \rangle = 0\ \forall x \in V \implies \langle x, Ty \rangle = 0\ \forall x \in V \implies Ty = 0 \implies y \in \ker T$

- $(\ker T)^{\perp} \subseteq \operatorname{Im} T$: For all $x \in (\ker T)^{\perp}$, $\forall y \in \ker T$, $\langle x, y \rangle = 0$. Notice that $\langle Tx, y \rangle = \langle T^* x, y \rangle = \langle x, Ty \rangle = 0$, then

  $$\langle Tx - x, y \rangle = 0\ \forall y \in \ker T \text{ and } T(Tx - x) = T^2 x - Tx = 0 \rightsquigarrow Tx - x \in \ker T$$

  In particular, $\langle Tx - x, Tx - x \rangle = 0 \rightsquigarrow x = Tx \in \operatorname{Im} T$.

Theorem 3.7.2 (Spectral theorem). V is a inner product space, $\dim V < \infty$. Let $T : V \to V$ be a linear operator and $\lambda_1, ..., \lambda_k$ be the distinct eigenvalues of T. Assume T is normal when $F = \mathbb{C}$/self adjoint when $F = \mathbb{R}$. Let $W_i = E_{\lambda_i}$ and T_i be the projection of V on W_i. Then

(a) $V = W_1 \oplus \cdots \oplus W_k$.

(b) Let $W_i' = \bigoplus_{j \neq i} W_i$. Then $W_i^{\perp} = W_i'$

(c) $T_i T_j = \delta_{ij} T_i$ for all $1 \leq i, j \leq k$

(d) $I = T_1 + \cdots + T_k$ is called **resolution of the identity**

(e) $T = \lambda_1 T_1 + \cdots + \lambda_k T_k$ is called **spectral decomposition**

Terminology, $\{\lambda_1, ..., \lambda_k\}$ is called the **spectrum** of T.

$\square$

Corollary 3.7.1. If $\mathbb{F} = \mathbb{C}$, then T is normal $\iff$ $T^* = g(T)$ for some $g(x) \in \mathbb{C}[x]$

Proof: $(\Rightarrow)$: Let $T = \lambda_1 T_1 + \cdots + \lambda_k T_k$ be spectral decomposition of T. Taking the adjoint of both sides, we have $T^* = \overline{\lambda_1} T_1 + \cdots + \overline{\lambda_k} T_k$, since each T_i is self-adjoint. Using the Lagrange interpolation formula, we may choose a polynomial g such that $g(\lambda_i) = \overline{\lambda_i}$ for $1 \leq i \leq k$. Then

$$g(T) = g(\lambda_1)T_1 + \cdots + g(\lambda_k)T_k = \overline{\lambda_1}T_1 + \cdots + \overline{\lambda_k}T_k = T^*$$

$(\Leftarrow)$: If $T^* = g(T)$ for some $g \in \mathbb{C}[x]$, then T commutes with $T^* \rightsquigarrow TT^* = T^*T$ i.e. T is normal. $\square$

3.8 Singular Value Decomposition and Pseudoinverse

Singular Value Decomposition is an important technique in computing pseudo inverse and matrix approximation. It is a powerful tool in engineering, statistic, and numerical analysis. Pseudoinverse is a generalization of inverse matrix that is used to solve the least square approximate solution to the system of linear equations.

3.8.1 Singular value Decomposition

Theorem 3.8.1. (SVD-matrix version) Let $A \in M_{m\times n}(F)$. ($F = \mathbb{R}$ or $\mathbb{C}$). Then there exist unitari (orthogonal) matrices $P \in M_n(F)$, $Q \in M_m(F)$ such that

$$Q^*AP = \left(\begin{array}{cccc|c} \sigma_1 & & & & \\ & \sigma_2 & & & O \\ & & \ddots & & \\ & & & \sigma_r & \\ \hline & & O & & O \end{array}\right)$$

for some $\sigma_i \in \mathbb{R}$, $\sigma_1 \geq \sigma_2 \geq \cdots \sigma_r > 0$, and $r = \operatorname{rank} A$.

Theorem 3.8.2 (SVD-linear map version). Let $T : V \to W$ be a linear map between inner product spaces over F ($F = \mathbb{R}$ or $\mathbb{C}$), $\dim_F V = n, \dim_F W = m$. Then, there exists an orthonormal basis $\{e_1, ..., e_n\}$ of V and orthonormal basis $\{f_1, ..., f_m\}$ of W such that

$$T(e_i) = \begin{cases} \sigma_i f_i & \text{for } i = 1, ...r \\ 0 & \text{for } i > r \end{cases}$$

for some $\sigma_i \in \mathbb{R}, \sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r$ and $r = \operatorname{rank} T$.

We proof the version of linear transformation and we will explain how to change to the matrix version in Remark 3.8.1.

Proof: $T : V \to W$, finite dimensional $\implies$ $T^* : W \to V$ exists. Consider $S = T^*T : V \to V$ which is self-adjoint. By spectral theorem, $\exists\{e_1, ..., e_n\}$: orthonormal basis of V s.t. e_i is eigenvector of S with eigenvalue λ_i.

Claim: $\{T(e_i)\}$ are orthogonal. ($T(e_i)$ may be zero!)
pf. $\langle Te_i, Te_j\rangle = \langle T^*Te_i, e_j\rangle = \langle \lambda_i e_i, e_j\rangle = \lambda_i\langle e_i, e_j\rangle$. □

$$\implies \begin{cases} T(e_i) \perp T(e_j) & \text{if } i \neq j \\ \lambda_i = \|T(e_i)\|^2 \geq 0 & \text{for all } i \end{cases}$$

Let $\sigma_i = \sqrt{\lambda_i}$. Reorder $\{e_i\}$ s.t. $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$, $r \leq n = \dim_F V$. Let $f_i = \frac{1}{\sigma_i}T(e_i)$ for $i = 1, 2, ..., r$. Then $\{f_1, ..., f_r\}$ is orthonormal set in W and $f_1, ..., f_r$ are eigenvectors of TT^* with eigenvalue λ_i, since

$$TT^*f_i = \frac{1}{\sigma_i}TT^*Te_i = \frac{1}{\sigma_i}T(\lambda_i e_i) = \lambda_i \cdot \frac{1}{\sigma_i}T(e_i) = \lambda_i f_i$$
$$\implies W = \underbrace{Ff_1 \oplus Ff_2 \oplus \cdots \oplus Ff_r}_{=\bigoplus_{i=1}^r E_{\lambda_i}} \oplus \underbrace{\ker(TT^*)}_{=E_0}$$

where E_λ is the eigenspaces of TT^* corresponding to λ. Let $\{f_{r+1}, ..., f_m\}$ be orthonormal basis of $\ker(TT^*)$. Then $\{f_1, ..., f_m\}$ is orthonormal basis of W.
Conclusion: V has basis $\{e_1, ..., e_n\}$: orthonormal eigenvectors of $S = T^*T$ with eigenvalues

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > 0 = \lambda_{r+1} = \cdots = \lambda_n$$

W has basis $\{f_1, ..., f_m\}$: orthonormal eigenvectors of $S = TT^*$ with eigenvalues

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > 0 = \lambda_{r+1} = \cdots = \lambda_m$$

Also, $T(e_i) = \sigma_i f_i$ for all $i = 1, 2, ..., n$ with $\sigma_i = \sqrt{\lambda_i} \geq 0$ □

Definition 3.8.1. σ_i are called the **singular values** of $T : V \to W$.

Remark 3.8.1. Let $T : F^n \to F^m$ with standard inner product s.t. $[T]_{\text{std}} = A \in M_{m\times n}(F)$. Let

$$P = (e_1\ e_2\ \cdots\ e_n) \in M_n(F) \text{ and } Q = (f_1\ f_2\ \cdots\ f_m) \in M_m(F).$$

Then P, Q : unitary (or orthogonal) s.t.

$$Q^*AP = \left(\begin{array}{cccc|c} \sigma_1 & & & & \\ & \sigma_2 & & & \\ & & \ddots & & O \\ & & & \sigma_r & \\ \hline & & O & & O \end{array}\right) =: \Sigma$$

Definition 3.8.2. Let $A \in M_{m\times n}(F)$. By SVD,

$$A = Q \cdot \left(\begin{array}{cccc|c} \sigma_1 & & & & \\ & \sigma_2 & & & \\ & & \ddots & & O \\ & & & \sigma_r & \\ \hline & & O & & O \end{array}\right)_{m\times n} \cdot P^* \in M_{m\times n}(F)$$

Define the **pseudoinverse** or **Moore-Penrose inverse** $A^\dagger$ by

$$A^\dagger = P \cdot \left(\begin{array}{cccc|c} \sigma_1^{-1} & & & & \\ & \sigma_2^{-1} & & & \\ & & \ddots & & O \\ & & & \sigma_r^{-1} & \\ \hline & & O & & O \end{array}\right)_{n\times m} \cdot Q^* \in M_{n\times m}(F)$$

Example 3.8.1. $A = \begin{pmatrix} 1 & 1 & -1 \\ 1 & 1 & -1 \end{pmatrix}$, then

$$A^*A = \begin{pmatrix} 2 & 2 & -2 \\ 2 & 2 & -2 \\ -2 & -2 & 2 \end{pmatrix} \text{ having eigenvalue } \lambda_1, \lambda_2, \lambda_3 = 6, 0, 0$$

and $e_1 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}$, $e_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}$, $e_3 = \frac{1}{\sqrt{6}} \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}$ be the orthonormal eigenvectors of A^*A. Then $f_1 = \frac{1}{\sigma_1} T(e_1) = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and choose $f_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix} \rightsquigarrow \{f_1, f_2\}$: orthonormal eigenvectors of AA^*. Hence,

$$Q^*AP = \begin{pmatrix} \sqrt{6} & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

where

$$A = \begin{pmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{-1}{\sqrt{3}} & 0 & \frac{2}{\sqrt{6}} \end{pmatrix}, \quad Q = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{pmatrix}$$

$$\implies A = Q \begin{pmatrix} \sqrt{6} & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} P^* : \text{singular value decomposition of } A$$

We compute the pseudoinverse of A :

$$A^\dagger = P \begin{pmatrix} \sqrt{6}^{-1} & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} Q^* = \frac{1}{6} \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ -1 & -1 \end{pmatrix}$$

3.8.2 Polar decomposition

Theorem 3.8.3. Let $A \in M_n(\mathbb{C})$. Then, there exists a unitary $W \in M_n(\mathbb{C})$ and a positive semi-definite P such that $A = WP$. Moreover, if A is invertible, then W, P are unique.
Note: For $n = 1 : \forall z \in \mathbb{C}$, we can write $z = e^{i\theta} \cdot r$, where $e^{i\theta}$ can be regard as a unitary and $r \geq 0$ is semi-definite.

Proof: Given $A \in M_n(\mathbb{C})$, by SVD, $\exists$ unitary U, V such that $A = U\Sigma V^* \implies A = UV^*V\Sigma V^*$. Let $W = UV^*$ and $P = V\Sigma V^*$, then W is unitary and P is self-adjoint with eigenvalues $\sigma_1, ..., \sigma_r, 0, ..., 0 \geq 0 \implies P$ is positive semidefinite.
For uniqueness, say $A = W_1P_1 = W_2P_2$, since A is invertible, P_1, P_2 are positive definite and invertible. So $P_1P_2^{-1} = W_2^*W_1$ is unitary i.e.

$$I = (P_1P_2^{-1})^*(P_1P_2^{-1}) = P_2^{-1}P_1^2P_2^{-1} \implies P_1^2 = P_2^2$$

Then P_1, P_2 has same eigenvector corresponding to same eigenvalue, then $P_1 = P_2$. □

3.8.3 Pseudoinverse and system of linear equation

Setting : Let $T : V \to W$ be a linear map between finite dimensional inner product space. Then we have $T^* : W \to V$. By orthogonal decomposition, we have

$$V = \ker(T^*T) \oplus (\ker(T^*T))^\perp = \ker T \oplus (\ker T)^\perp$$

$$W = \operatorname{Im} T \oplus (\operatorname{Im} T)^{\perp}$$

Now, we introduce two common linear transformations in below. :

$$T|_{(\ker T)^{\perp}} : \ker T^{\perp} \longrightarrow \operatorname{Im} T \text{ is an isomorphism}$$

Since if $x \in \ker T|_{\ker T^{\perp}}$, then $x \in \ker T \cap (\ker T)^{\perp} = \{0\}$. So $T|_{(\ker T)^{\perp}}$ is $1-1$. Combine with $\dim(\ker T)^{\perp} = \operatorname{Im} T$, we have $T|_{(\ker T)^{\perp}}$ is an isomorphism.

$$\operatorname{Proj}_{\operatorname{Im} T} : W \longrightarrow \operatorname{Im} T \text{ is orthogonal projection}$$

Now, we define the linear map version of pseudoinverse or Moore-Penrose inverse.

Definition 3.8.3. Let $T : V \to W$ be linear map between finite dimensional inner product spaces. The **pseudoinverse** or **Moore-Penrose inverse** of T, $T^{\dagger}$ is defined by

$$T^{\dagger} = \left(T|_{(\ker T)^{\perp}}\right)^{-1} \circ \operatorname{Proj}_{\operatorname{Im} T} : W \longrightarrow V$$

Precisely, $T^{\dagger}$ is the following composition :

$$W \xrightarrow[\text{projection}]{\text{orthogonal}} \operatorname{Im} T \xrightarrow{\left(T|_{(\ker T)^{\perp}}\right)^{-1}} (\ker T)^{\perp} \xhookrightarrow{\text{inclusion}} V$$

Consider SVD of T and write the map above by element, we have

$$\sum_{i=1}^{m} a_i f_i \longrightarrow \sum_{i=1}^{r} a_i f_i \longrightarrow \sum_{i=1}^{r} a_i \cdot \frac{1}{\sigma_i} f_i \hookrightarrow \sum_{i=1}^{r} a_i \cdot \frac{1}{\sigma_i} f_i$$

For matrix version, $\forall v \in R^m$, $Q^* v$ is v corresponding to $\{f_i\}$. $\Sigma^{\dagger} Q^* v$ is the vector after second translation (corresponding to $\{e_i\}$). $P \sum^{\dagger} Q^* v$ is $\Sigma^{\dagger} Q^* v$ as the standard basis. So the definition of pseudoinverse in matrix version is same as the special case of linear map version.

Property 3.8.1.

- $TT^{\dagger} = \operatorname{Proj}_{\operatorname{Im} T}$:

  pf. $T \circ \left(T|_{(\ker T)^{\perp}}\right)^{-1} \circ \operatorname{Proj}_{\operatorname{Im} T} = \operatorname{Proj}_{\operatorname{Im} T}$.

- $\operatorname{Im} T = \ker(TT^{\dagger} - 1)$ and $(\operatorname{Im} T)^{\perp} = \operatorname{Im}(1 - TT^{\dagger})$:

  pf. $W = \operatorname{Im} T \oplus (\operatorname{Im} T)^{\perp}$. $\operatorname{id}_W = \operatorname{proj}_{\operatorname{Im} T} + \operatorname{proj}_{(\operatorname{Im} T)^{\perp}} \implies 1 - TT^{\dagger} = \operatorname{Proj}_{(\operatorname{Im} T)^{\perp}}$. So

  $$\operatorname{Im}(1 - TT^{\dagger}) = (\operatorname{Im} T)^{\perp} \text{ and } \ker(1 - TT^{\dagger}) = \left((\operatorname{Im} T)^{\perp}\right)^{\perp} = \operatorname{Im} T$$

- $\ker T = \operatorname{Im}(T^{\dagger} T - 1)$:

  pf. $(\subseteq) : v \in \ker T \implies v = T^{\dagger} T(-v) - (-v) = (T^{\dagger} T - 1)(-v) \in \operatorname{Im}(T^{\dagger} T - 1)$.

  $(\supseteq) : v \in \operatorname{Im}(T^{\dagger} T - 1) \implies v = (T^{\dagger} T - 1)(w) \implies T(v) = TT^{\dagger} T(w) - T(w) = \operatorname{Proj}_{\operatorname{Im} T}(T(w)) - T(w) = T(w) - T(w) = 0 \implies v \in \ker T$.

Theorem 3.8.4. Let $T : V \to W$ be linear map between finite dimensional inner product spaces. Let $b \in W$.

(1) The equation $Tx = b$ has a solution in V if and only if $TT^{\dagger} b = b$ i.e. $T^{\dagger} b$ is a solution.

(2) If the solution $Tx = b$ is consistent, then all the solution is of the form $T^\dagger b + (T^\dagger T - 1)z$ for some $z \in V$.

This form are also solutions since $\mathrm{Im}(T^\dagger T - 1) = \ker T$.

(3) In general, $T^\dagger b$ is the best solution to $Tx = b$ in the following sence :

$$\|TT^\dagger b - b\| = \min_{x \in V} \|Tx - b\|$$

Proof:

(1) $Tx = b$ has a solution in $V \iff b \in \mathrm{Im}\, T \iff b \in \ker(TT^\dagger - 1) \iff TT^\dagger b = b$.

(2) Suppose x is a solution to $Tx = b$. Then $Tx = b = TT^\dagger b$. Then

$$T(T^\dagger b - x) = 0 \implies T^\dagger b - x \in \ker T = \mathrm{Im}(T^\dagger T - 1)$$

i.e. $T^\dagger b - x = (T^\dagger T - 1)z$ for some $z \in V$, i.e. $x = T^\dagger b - (T^\dagger T - 1)z$.

(3)

$$\begin{aligned}\|Tx - b\|^2 &= \| \underbrace{Tx - TT^\dagger b}_{\in \mathrm{Im}\, T} + \underbrace{TT^\dagger b - b}_{\in \mathrm{Im}(1 - TT^\dagger) = (\mathrm{Im}\, T)^\perp} \| \\ &= \|Tx - TT^\dagger b\|^2 + \|TT^\dagger b - b\|^2 \geq \|TT^\dagger b - b\|^2\end{aligned}$$

and the " = " holds when $Tx = TT^\dagger b \implies x - T^\dagger b \in \ker T$ i.e. $x = T^\dagger b + (T^\dagger T - 1)z$ for some $z \in V$.

□

Example 3.8.2. (1) : Consider the linear system

$$\begin{cases} x_1 + x_2 - x_3 = 1 \\ x_1 + x_2 - x_3 = 1 \end{cases} \implies \begin{pmatrix} 1 & 1 & -1 \\ 1 & 1 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

$$\implies A^\dagger = \frac{1}{6} \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ -1 & -1 \end{pmatrix} \text{ and } A^\dagger A = \frac{1}{3} \begin{pmatrix} 1 & 1 & -1 \\ 1 & 1 & -1 \\ -1 & -1 & 1 \end{pmatrix}$$

So all solution are $A^\dagger \begin{pmatrix} 1 \\ 1 \end{pmatrix} + (A^\dagger A - 1) \begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix} = \frac{1}{3} \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix} + \begin{pmatrix} -2 \\ 1 \\ -1 \end{pmatrix} z_1 + \begin{pmatrix} 1 \\ -2 \\ -1 \end{pmatrix} z_2 + \begin{pmatrix} -1 \\ -1 \\ -2 \end{pmatrix} z_3$

(2) : $\begin{cases} x_1 + x_2 - x_3 = 1 \\ x_1 + x_2 - x_3 = 2 \end{cases} \implies \begin{pmatrix} 1 & 1 & -1 \\ 1 & 1 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ has no solution! And the best solution is $A^\dagger \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}$

Chapter 4

Bilinear forms

4.1 Bilinear forms

Definition 4.1.1. Let V be a vector space over a field F. A function $H : V \times V \to F$ is called a **bilinear form** on V if H is linear in each variable. That is

- $H(cx_1 + x_2, y) = cH(x_1, y) + H(x_2, y)$ for all $x_1, x_2, y \in V$ and all $c \in F$
- $H(x, cy_1 + y_2) = cH(x, y_1) + H(x, y_2)$ for all $x, y_1, y_2 \in V$ and all $c \in F$

Example 4.1.1.

- If $F = \mathbb{R}$, then an inner product $\langle \cdot, \cdot \rangle$ on V is a bilinear form. However, if $F = \mathbb{C}$, then an inner product on V is not a bilinear form.
- Let $V = F^n$ and $A \in M_n(F)$. Define $H : V \times V \to F$ by $H(x, y) = x^t A y$. Then H is a bilinear form.
- If f and g are two linear functionals on V, define $H : V \times V \to F$ by $H(x, y) = f(x)g(y)$. Then H is a bilinear form.
- Let $V = M_n(F)$. Given $A \in M_m(F)$, we define $H : V \times V \to F$ by $H(X, Y) = \operatorname{tr}(X^t A Y)$. Then H is a bilinear form.

Property 4.1.1. Let H, H_1 and H_2 be bilinear forms on V.

- For a fixed $x \in V$, the function $L_x : V \to F$ defined by $L_x(y) = H(x, y)$ is a linear functional of V. Likewise, for a fixed $y \in V$, let $R_y : V \to F$ be defined by $R_y(x) = H(x, y)$ is also a linear functional of V.
- $H(x, 0) = H(0, x) = 0 \ \forall x \in V$.
- The function $J : V \times V \to F$ defined by $J(x, y) = H(y, x)$ is also a bilinear form.
- The function $H_1 + H_2 : V \times V \to F$ defined by $(H_1 + H_2)(x, y) = H_1(x, y) + H_2(x, y)$ is a bilinear form on V, called the **sum** of H_1 and H_2.
- For $c \in F$, the function $cH : V \times V$ defined by $(cH)(x, y) = cH(x, y)$ is a bilinear form on V.

Theorem 4.1.1. Let $\mathfrak{B}(V)$ denote the set of bilinear forms on V. Then $\mathfrak{B}(V)$ is a vector space over F.

Proof: Trivial. $\square$

Definition 4.1.2. Let $H \in \mathfrak{B}(V)$ and $\mathcal{B} = \{v_1, ..., v_n\}$ be a basis for V. Then the matrix $A \in M_n(F)$ defined by

$$A_{ij} = H(v_i, v_j)$$

is called the **matrix representation** (or the **Gram matrix**) of H with respect to the basis $\mathcal{B}$. We let $[H]_{\mathcal{B}}$ denoted this matrix.

Property 4.1.2. Let $\mathcal{B} = \{v_1, ..., v_n\}$ be a basis for V and H be a bilinear form on V. Then for $x, y \in V$, we have

$$H(x, y) = [x]_{\mathcal{B}}^t [H]_{\mathcal{B}} [y]_{\mathcal{B}}$$

where $[x]_{\mathcal{B}}$ and $[y]_{\mathcal{B}}$ are the coordinate vectors of x and y relative to $\mathcal{B}$ (as column vectors), respectively.

Proof: Suppose that $x = \sum\limits_{i=1}^{n} a_i v_i$ and $y = x = \sum\limits_{i=1}^{n} b_i v_i$. Let $A_{ij} = H(v_i, v_j)$. Then

$$[x]_{\mathcal{B}}^t [H]_{\mathcal{B}} [y]_{\mathcal{B}} = \sum_{1 \le i,j \le n} a_i A_{ij} b_j = \sum_{1 \le i,j \le n} a_i b_j H(v_i, v_j) = H(x, y)$$

$\square$

Theorem 4.1.2. Assume that $\dim V = n$. For any basis $\mathcal{B} = \{v_1, ..., v_n\}$ for V, the map $\psi_B : H \mapsto [H]_{\mathcal{B}}$ is an isomorphism of vector spaces from $\mathfrak{B}(V)$ to $M_n(F)$. In particular, $\dim \mathfrak{B}(V) = n^2$.

Proof: It's clear that ψ is linear. So we focus on $1-1$ and onto.

- $1-1$: If $[H]_{\mathcal{B}} = 0$. By Property 4.1.2, $\forall x, y \in V$

  $$H(x, y) = [x]_{\mathcal{B}}^t [H]_{\mathcal{B}} [y]_{\mathcal{B}} = 0$$

  Thus, H is identically 0.

- onto : Given $A \in M_n(F)$, define $H : V \times V \to F$ by

  $$H(x, y) = [x]_{\mathcal{B}}^t A [y]_{\mathcal{B}}^t$$

  It is clear that H is a bilinear form on V. Also, we have for all i, j,

  $$H(v_i, v_j) = [v_i]_{\mathcal{B}}^t A [v_j]_{\mathcal{B}}^t = e_i^t A e_j = A_{ij}$$

  where $\{e_1, ..., e_n\}$ denotes the standard basis for F^n. Hence, $[H]_{\mathcal{B}} = A$.

So $\psi : \mathfrak{B}(V) \xrightarrow{\sim} M_n(F)$ is isomorphism and thus $\dim \mathfrak{B}(V) = \dim M_n(F) = n^2$. $\square$

Theorem 4.1.3. Assume that $\dim V < \infty$ and $\mathcal{B}$ and $\mathcal{B}'$ be teo bases for V. Let $H \in \mathfrak{B}(V)$ and $Q = [\mathrm{id}_V]_{\mathcal{B}'}^{\mathcal{B}}$ be the matrix that changes the $\mathcal{B}'$-coordinates to $\mathcal{B}$-coordinates. Then

$$[H]_{\mathcal{B}'} = Q^t [H]_{\mathcal{B}} Q$$

Proof: By Property 4.1.2

$$H(x, y) = [x]_{\mathcal{B}'}^t [H]_{\mathcal{B}'} [y]_{\mathcal{B}'}$$

On the other hand, by well-known, for $v \in V$, we have $[v]_{\mathcal{B}} = Q[v]'_{\mathcal{B}}$. Hence,

$$H(x, y) = [x]_{\mathcal{B}}^t [H]_{\mathcal{B}} [y]_{\mathcal{B}} = [x]_{\mathcal{B}'}^t Q^t [H]_{\mathcal{B}} Q [y]_{\mathcal{B}'}$$

By the isomorphism between $\mathfrak{B}(V)$ and $M_n(F)$ established in Theorem 4.1.2, this implies that $Q^t [H]_{\mathcal{B}} Q$ is the matrix representation of H with respect to the basis $\mathcal{B}'$ i.e. $[H]_{\mathcal{B}'} = Q^t [H]_{\mathcal{B}} Q$ $\square$

Definition 4.1.3. We say two matrices $A, B \in M_n(F)$ are **congruent** if there exists an invertible matrix Q such that $B = Q^t AQ$.

4.2 Matrix representations of bilinear forms and dual space

Questions: What do "rank", "invertible", etc. means in the context of matrix representations $[H]_{\mathcal{B}}$ of bilinear forms? Also, is there a linear transformation that $[H]_{\mathcal{B}}$ represents?

Definition 4.2.1. Let V be a vector space over a field F. A linear transformation from V to F is called a **linear functional** on V. The space of all linear functionals on V is called the **dual space** of V and will be denoted by V^*

Remark 4.2.1. Note that if $\dim V = n < \infty$, then $\dim V^* = n$. (In general, the space of all linear transformations from an m-dimensional vector space to an n-dimensional vector space has dimension mn)

Let H be a bilinear form on a vector space V over a field F. Recall that for $x \in V$, the function $L_{H,x} : V \to F$ defined by

$$L_{H,x}(y) = H(x, y)$$

is a linear functional, i.e. an element of V^*. Likewise, for $y \in V$, the function $R_{H,y} : V \to F$ defined by

$$R_{H,y}(x) = H(x, y)$$

is also an element to V^*.
Now, define $\mathscr{L}_H : V \to V^*$ and $\mathscr{R}_H : V \to V^*$ by $\mathscr{L}_H(x) = L_{H,x}$ and $\mathscr{R}_H(y) = R_{H,y}$, respectively.

Property 4.2.1. The maps $\mathscr{L}_H$ and $\mathscr{R}_H$ defined above are both linear transformations form V to V^*.
Conversely, if $\mathscr{L}$ and $\mathscr{R}$ are linear transformations from V to V^*, then $H_{\mathscr{L}}, H_{\mathscr{R}} : V \times V \to F$ defined by $H_{\mathscr{L}}(x, y) = \mathscr{L}(x)(y)$ and $H_{\mathscr{R}}(x, y) = \mathscr{R}(y)(x)$ are both bilinear forms on V.
This give us the corresponding between V and V^*. If $\dim V < \infty$, then $\dim V = \dim V^*$.

Proof: For the first statement, we need to shoe that $\mathscr{L}_H(cx_1 + x_2) = c\mathscr{L}_H(x_1) + \mathscr{L}_H(x_2)$ in V^* for all $x_1, x_2 \in V$ and $c \in F$ i.e.

$$\mathscr{L}_H(cx_1 + x_2)(y) = c\mathscr{L}_H(x_1)(y) + \mathfrak{L}_H(x_2)(y)$$

The LHS is equal to

$$L_{H,cx_1+x_2}(y) = H(cx_1 + x_2, y)$$

while the RHS is

$$cL_{H,x_1}(y) + L_{H,x_2}(y) = cH(x_1, y) + H(x_2, y)$$

Since H is bilinear form, the two sides are equal. Hence, $\mathscr{L}_H(cx_1 + x_2) = c\mathscr{L}_H(x_1) + \mathscr{L}_H(x_2)$.
Similarly, $\mathscr{R}_H$ is also a linear transformation.
Conversely, we have

$$H_{\mathscr{L}}(cx_1 + x_2, y) = \mathscr{L}(cx_1 + x_2)(y)$$

Since $\mathscr{L}$ is a linear transformation, the RHS is equal to

$$c\mathscr{L}(x_1)(y) + \mathscr{L}(x_2)(y) = cH_{\mathscr{L}}(x_1, y) + H_{\mathscr{L}}(x_2, y)$$

This proves that $H_{\mathscr{L}}$ is linear in the first variable. Also,

$$H_{\mathscr{L}}(x, cy_1 + y_2) = \mathscr{L}(x)(cy_1 + y_2)$$

Since $\mathscr{L}(x) \in V^*$, the RHS is equal to

$$c\mathscr{L}(x)(y_1) + \mathscr{L}(x)(y_2) = cH_{\mathscr{L}}(x, y_1) + H_{\mathscr{L}}(x, y_2)$$

This proves that $H_{\mathscr{L}}$ is linear in the second variable. Hence $H_{\mathscr{L}}$ is the bilinear form. Similarly, $H_{\mathscr{R}}$ is also the bilinear form. □

Property 4.2.2. Assume that $\dim V < \infty$ and H is a bilinear form on V. Let $A = [H]_{\mathcal{B}}$ be the matrix representation of V with respect to some basis $\mathcal{B}$. Then

$$\operatorname{rank} \mathscr{L}_H = \operatorname{rank} A = \operatorname{rank} \mathscr{R}_H$$

If $\mathcal{B}'$ is another basis, then $[H]_{\mathcal{B}'} = Q^t[H]_{\mathcal{B}}Q$ for some invertible matrix Q, so $\operatorname{rank}[H]_{\mathcal{B}'} = \operatorname{rank}[H]_{\mathcal{B}}$ which is independent on the choice of basis.

Proof: To prove $\operatorname{rank} \mathscr{R}_H = \operatorname{rank} A$, it suffices to prove that

$$\operatorname{nullity} \mathscr{R}_H = \operatorname{nullity} A$$

By Property 4.1.2,

$$\mathscr{R}_H(y)(x) = R_{H,y}(x) = H(x, y) = [x]_{\mathcal{B}}^t A[y]_{\mathcal{B}}$$

Thus, $y \in \ker \mathscr{R}_H$ i.e. $R_{H,y} = 0 \iff [x]_{\mathcal{B}}^t A[y]_{\mathcal{B}} = 0\ \forall x \in V \iff A[y]_{\mathcal{B}} = 0$ i.e. $[y]_{\mathcal{B}} \in \ker A$. Therefore, $\operatorname{nullity} \mathscr{R}_H = \operatorname{nullity} A$ and thus $\operatorname{rank} \mathscr{R}_H = \operatorname{rank} A$. Similarly, $\operatorname{rank} \mathscr{L}_H = \operatorname{rank} A^t = \operatorname{rank} A$. □

Definition 4.2.2. Assume that $\dim V = n < \infty$. We define the **rank** of a bilinear form H of V to be $\operatorname{rank} \mathscr{L}_H (= \operatorname{rank} \mathscr{R}_H)$.

Corollary 4.2.1. Assume that H is bilinear form on an n-dimensional vector space V. Then the following are equivalent :

(1) $\operatorname{rank} H = n$.

(2) If $x \in V$ is a vector such that $H(x, y) = 0$ for all $y \in V$, then $x = 0$.

(3) If $y \in V$ is a vector such that $H(x, y) = 0$ for all $x \in V$, then $y = 0$.

Proof: It's cleat that (1) $\iff$ (2) and (1) $\iff$ (3) by Property 4.2.2 □

Definition 4.2.3. A bilinear form on a vector space V is said to be **nondegenerate** (or **nonsingular**) if (2) and (3) in the Corollary 4.2.1 hold. (If $\dim V < \infty$, then we only need to assume that either (2) or (3) holds.)

Definition 4.2.4 (dual basis)**.** Let $\mathcal{B} = \{v_i\}_{i \in I}$ be a basis for V, where I is an index set. For $i \in I$, define $\varphi_i \in V^*$ by setting

$$\varphi_i(v_i) = \delta_{ij}$$

and extending it linearly to the whole V. We let $\mathcal{B}^*$ denote the set $\{\varphi_i\}_{i \in I}$

Theorem 4.2.1. If $\dim V < \infty$, then $\mathcal{B}^*$ is a basis for V^*, called the **dual basis** of $\mathcal{B}$.

Proof: Assume that $\mathcal{B} = \{v_1, ..., v_n\}$ and let $\mathcal{B}^*$ be the dual basis of $\mathcal{B}$. For $f \in V^*$ and $i = 1, ..., n$, let $a_i = f(v_i)$. Set

$$\varphi = \sum_{i=1}^{n} a_i \varphi_i.$$

Since

$$\varphi(v_j) = \sum_{i=1}^{n} a_i \varphi_i v_j = \sum_{i=1}^{n} a_i \delta_{ij} = a_j = f(v_j) \implies f = \varphi$$

i.e. $f = \varphi \in \operatorname{span} \mathcal{B}^*$. Since $|\mathcal{B}^*| = |\mathcal{B}| = \dim V = \dim V^*$, this proves the theorem. $\square$

Remark 4.2.2. If $\dim V = \infty$, $\mathcal{B}^*$ is never a basis. For example let $f \in V^*$ be the linear functional with $f(v_i) \in I$ for all $i \in I$. Then $f \notin \operatorname{span} \mathcal{B}^*$. If $f = \sum_{i \in J} a_i \varphi_i$ for some $J \subseteq I$ with $|J| < \infty$, choose $j \in I \setminus J$, the $1 = f(v_j) = \sum_{i \in J} a_i \varphi_i(v_j) = 0$ $(\to\leftarrow)$.

Let $\mathcal{B} = \{v_1, ..., v_n\}$ be a basis for V and $\mathcal{B}^* = \{\varphi_1, ..., \varphi_n\}$ be its dual basis. Let H be a bilinear form and $\mathscr{R}_H : V \to V^*$ be defined by $R_H(y) = R_{H,y}$, where $R_{H,y}(x) = H(x, y)$. Let us compute $[\mathscr{R}_H]_{\mathcal{B}}^{\mathcal{B}^*}$. We have for all i, j,

$$\mathbb{R}_H(v_j)(v_i) = R_{H,v_j}(v_i) = H(v_i, v_j)$$

Thus,

$$\mathscr{R}_H(v_j) = \sum_{i=1}^{n} H(v_i, v_j)\varphi_i.$$

Therefore,

$$[\mathscr{R}_H]_{\mathcal{B}}^{\mathcal{B}^*} = (H(v_i, v_j))_{n \times n} = [H]_{\mathcal{B}}$$

Hence, $[H]_{\mathcal{B}}$ is the matrix $[\mathscr{R}_H]_{\mathcal{B}}^{\mathcal{B}^*}$ for the linear transformation $\mathscr{R}_H : V \to V^*$ with respective to $\mathcal{B}$ and $\mathcal{B}^*$.

Chapter 5

Appendix

5.1 Appendix of proof in Chapter 1

5.1.1 Proof of Theorem 1.3.3

First we introduce a notation for a T-invariant subspace W of K_λ and $v \in K_\lambda$. Observe that $(T-\lambda I)^p(v) = 0$ for some p and hence $I_T(v, W) = ((x-\lambda)^s)$ for some s. We let $s(v, W)$ denote this s.

Outline of proof of existence:

Set $W_0 = \{0\}$. Let v_1 be a vector in K_λ such that

$$s(v_1, W_0) = \max_{v \in K_\lambda} s(v, W_0) =: s_1$$

Let $W_1 = Z(v_1; T)$. Note that $\dim Z(v_1; T) = s_1$, since $I_T(v_1, W_0)$ is simply $I_T(v_1)$ which by assumption is $((x-\lambda)^{s_1})$. We have seen earlier that the integer s, such that $I_T(v) = ((x-\lambda)^s)$ is preceisely the dimension of $Z(v_1, T)$. To find v_2 a natural ideal is to choose v_2 such that

$$s(v_2; W_1) = \max_{v \in K_\lambda} s(v, W_1) =: s_2$$

However, there is one problem here. That is $Z(v_1; T) \cap Z(v_2; T)$ may not be $\{0\}$ i.e. $Z(v_1; T) + Z(u; T)$ may not be a direct sum of $Z(v_1; T)$ and $Z(u; T)$. In order for v_1, v_2 to satisfy $Z(v_1; T) \cap Z(u; T) = \{0\}$, we need to modify it. We claim that there exists w in W_1 such that $(T-\lambda)^{s_2}(u) = (T-\lambda)^{s_2}(w)$ and replace v_2 by $v_2 - w$. We claim that :

- $Z(v_1; T) \cap Z(v_2; T) = \{0\}$
- $\dim Z(v_2, T) = s_2$

In general for $i \geq 3$, choose u such that

$$s(u, W_{i-1}) = \max_{v \in K_\lambda} s(v; W_{i-1}) =: s_i$$

We claim that there exists w in W_{i-1} such that $(T-\lambda)^{s_i}(u) = (T-\lambda)^{s_i}(w)$ and let $v_i = u - w$. We claim that :

- $W_{i-1} \cap Z(v_i; T) = \{0\}$
- $\dim Z(v_i, T) = s_i$

Let $W_i = W_{i-1} \oplus Z(w_i; T)$ and continuous until $W_i = K_\lambda$.

Outline of outline of proof of existence:

(1) Let $W_0 = \{0\}$

(2) Choose $u \in K_\lambda$ s.t.
$$s(u, W_{i-1}) = \max_{v \in K_\lambda} s(v, W_{i-1}) =: s_i$$
and prove that $\exists w \in W_{i-1}$ s.t. $(T-\lambda)^{s_i}(u) = (T-\lambda)^{s_i}(w)$

(3) Let $v_i = u - w$, prove that

- $W_{i-1} \cap Z(v_i; T) = \{0\}$
- $\dim Z(v_i; T) = s_i$

Let $W_i = W_{i-1} \oplus Z(v_i; T)$

(4) Repeat (2),(3) until $W_i = K_\lambda$ (Since $\dim W_i > \dim W_{i-1}$)

Detailed proof: Let $W_0 = \{0\}$, we induction on i :
$i = 1$: $W_0 = \{0\}$, so $w = 0$. We indeed have $0 = (T-\lambda)^{s_1}(0) = (T-\lambda)^{s_1}(v_1)$.
Now, assume that $(2), (3)$ holds for $W_1, ..., W_{i-1}$.

Lemma: Let u be a vector in K_λ such that
$$s(u, W_{i-1}) = \max_{v \in K_\lambda} s(v, W_{i-1}) =: s_i$$
Then $\exists w \in W_{i-1}$ such that $(T-\lambda)^{s_i}(w) = (T-\lambda)^{s_i}(u)$

Proof: For convenience, let $\widetilde{T} = T - \lambda I$. For $j \leq i-1$, $\mathcal{B}_j := \{v_j, \widetilde{T}(v_j), ..., \widetilde{T}^{s_j - 1}(v_j)\}$ is a basis for $Z(v_j; T)$. Hence, $\mathcal{B} := \bigsqcup_{j=1}^{i-1} \mathcal{B}_j$ is a basis for $W_{j-1} = \bigoplus_{j=1}^{i-1} Z(v_j; T)$. Thus
$$\widetilde{T}^{s_i}(u) = \sum_{j=1}^{i-1} \sum_{k=0}^{s_j - 1} a_{jk} \widetilde{T}^k(v_j) \tag{*}$$
for some unique a_{jk}. Note that since $W_0 \subset W_1 \subset \cdots \implies s(v, W_1) \geq s(v, W_2) \geq \cdots$ for any $v \in K_\lambda$, thus $s_1 \geq s_2 \geq \cdots \geq s_{i-1} \geq s_i$. Our goal is to show that $a_{jk} = 0$ for any (j,k) with $k \leq s_i - 1$. Then
$$\widetilde{T}^{s_i}(u) = \sum_{j=1}^{i-1} \sum_{k=0}^{s_j - 1} a_{jk} \widetilde{T}^k(v_j) = \sum_{j=1}^{i-1} \sum_{k=s_i}^{s_j - 1} a_{jk} \widetilde{T}^k(v_j) = \widetilde{T}^{s_i}\left(\sum_{j=1}^{i-1} \sum_{k=s_i}^{s_j - 1} a_{jk}.\widetilde{T}^{k - s_i}(v_j) \right)$$
which complete our lemma.

For $m \leq i-1$, we may apply $\widetilde{T}^{s_m - s_i}$ to $(*)$, we have
$$\widetilde{T}^{s_m}(u) = \sum_{j=1}^{i-1} \sum_{k=0}^{s_j - 1} a_{jk} \widetilde{T}^{s_m - s_i + k}(v_j)$$
Note that LHS$\in W_{m-1}$ according to the definition of $S_m := \max_{v \in K_\lambda} s(v, W_{m-1})$. Thus
$$\sum_{j=m}^{i-1} \underbrace{\sum_{k=0}^{s_j - 1} a_{jk} \widetilde{T}^{s_m - s_i + k}(v_j)}_{\in Z(v_j; T)} = 0 \implies \sum_{k=0}^{s_j - 1} a_{jk} \widetilde{T}^{s_m - s_i + k}(v_j) = 0 \ \forall j$$
In particular when $j = m$, if $k \geq s_i \rightsquigarrow s_m - s_i + k \geq s_m$ and thus $\widetilde{T}^{s_m - s_i + k}(v_m) = 0$. Since $\mathcal{B}_j$ is linearly independent over F, $a_{mk} = 0 \ \forall k \leq s_i - 1$ □

Let $w \in W_{i-1}$ be a vector such that $(T-\lambda)^{s_i}(w) = (T-\lambda)^{s_i}(u)$ as in the previous lemma. Let $v_i = u - w$. Then

- $W_{i-1} \cap Z(v_i;T) = \{0\}$ ($\implies W_{i-1} + Z(v_i;T)$ is a direct sum of $Z(v_1;T), ..., Z(v_i;T)$) :

  pf. Recall that if W is a T-invariant and $v - v' \in W$, then $I_{v,W} = I_{v',W}$.

  Thus, $I_{v_i,W_{i-1}} = I_{u,W_{i-1}} = ((x-\lambda)^{s_i})$ i.e. $s(v_i, W_{i-1}) = s(u, W_{i-1}) = s_i$. Suppose that $v \in W_{i-1} \cap Z(v_i;T)$. Say $v = a_0 v_i + a_1 T(v_i) + \cdots + a_n T^n(v_i)$. Let $f(x) := a_0 + a_1 x + \cdots + a_n x^n$. Since $v \in W_{i-1}$ and $v = f(T)(v_i) \implies f(x) \in I_{v_i,W_{i-1}} = ((x-\lambda)^{s_i}) \implies f(x) = g(x)(x-\lambda)^{s_i}$. Then $v = f(T)(v_i) = g(T)(T-\lambda I)^{s_i}(v_i) = g(T)(T-\lambda I)^{s_i}(u-w) = 0$ □

- $\dim Z(v_i;T) = s_i$:

  pf. Recall that for $v \in K_\lambda$, $\dim Z(v_i;T) =$ the smallest integer s such that $(T-\lambda)^s(v) = 0$.

  Here we have found that $(T-\lambda)^{s_i}(v_i) = 0 \implies s \leq s_i$

  On the other hand. Since $I_{v_i,W_{i-1}} = ((x-\lambda)^{s_i}) \implies s \geq s_i$ and thus $s = s_i$

By induction, (2),(3) holds for all i. Since $\dim W_{i-1} < \dim W_i \leq \dim K_\lambda$. We can repeat (2),(3) until $W_i = K_\lambda$. □

5.1.2 Proof of Theorem 1.4.2

Claim in (ii): Let $u \in K_p$ be such that

$$s(u, W_{i-1}) = \max_{v \in K_p} s(v, W_{i-1}) = s_i$$

Then $\exists w \in W_{i-1}$ such that $p(T)^{s_i}(w) = p(T)^{s_i}(u) = 0$

Proof: Since $s(u, W_{i-1}) = s_i$, we have $p(T)^{s_i}(u) \in W_{i-1}$

$$\implies p(T)^{s_i}(u) = \sum_{j=1}^{i-1} f_j(T)(v_j) \tag{*}$$

for some polynomials $f_j(x)$. Observe that $s_1 \geq s_2 \geq \cdots \geq s_i$. For $m \leq i-1$, we have $s_m \geq s_i$ and we can apply $p(T)^{s_m - s_i}$ to $(*)$ and obtain

$$p(T)^{s_m}(u) = \sum_{j=1}^{i-1} p(T)^{s_m - s_i} f_j(T)(v_j) \tag{**}$$

Recall that s_m is defined to be $\max\limits_{v \in K_p} s(v, W_{m-1})$. Thus, the LHS of $(**)$ belongs to W_{m-1} i.e. LHS of (**) equal to $\sum\limits_{j=1}^{m-1} g_j(T)(v_j)$ for some $g_j(x)$

$$\implies \sum_{j=1}^{m-1} g_j(T)(v_j) = \sum_{j=1}^{i-1} p(T)^{s_m - s_i} f_j(T)(v_j) \implies \underbrace{p(T)^{s_m} f_m(T)(v_m)}_{\in Z(v_m;T)} + \sum_{j \neq m} \underbrace{h_j(T)(v_j)}_{\in Z(v_j;T)} = 0$$

for some $h_j(x) \in F[x]$. By Property 1.2.1, we have $p(T)^{s_m} f_m(T)(v_m) = 0$
By induction hypothesis of (iii), $I_{v_m} = (p(x)^{s_m}) \implies p(x)^{s_m} \big| p(x)^{s_m - s_i} f_m(x) \implies p(x)^{s_i} \big| f_m(x)$, say $f_m(x) = p(x)^{s_i} \widetilde{f}_m(x)$ for some $\widetilde{f}_m$. Recall $(*)$,

$$p(T)^{s_i}(u) = \sum_{j=1}^{i-1} f_j(T)(v_j) = \sum_{j=1}^{i-1} p(T)^{s_i} \widetilde{f}_j(T)(v_j) = p(T)^{s_i} \underbrace{\left(\sum_{j=1}^{i-1} \widetilde{f}_j(T)(v_j) \right)}_{:= w \in W_{j-1}}$$

Hence, we find $w \in W_{i-1}$ s.t. $p(T)^{s_i}(u) = p(T)^{s_i}(w)$. □

Claim in (iii): Let u, w be as in (ii). Let $v_i := u - w$. Then

- $W_{i-1} \cap Z(v_i; T) = \{0\}$
- $I_{v_i} = (p(x)^{s_i})$

Proof:

- Assume $v \in W_{i-1} \cap Z(v_i; T)$. We have $v = f(T)(v_i)$ for some $f(x) \in F[x]$. This vector is in $W_{i-1} \implies f(x) \in I_{v,W_{i-1}}$. Recall that in Remark 1.4.2, we have seen that $I_{v,W_{i-1}} = (p(x)^{s_i})$. So we have $p(x)^{s_i} \big| f(x)$, say $f(x) = p(x)^{s_i}\widetilde{f}(x) \implies v = f(T)(v_i) = \widetilde{f}(T)p(T)^{s_i}(v_i) = 0$
- Since $v_i \in K_p := \{v \in V : p(T)^n v = 0 \text{ for some } n\} \implies I_{v_i} = (p(x)^s)$. Now recall that $s(v_i, W_{i-1}) = s_i \implies p(T)^{s_i} \in W_{i-1}$ and s_i is the smallest integer with the property $\implies s \geq s_i$

  On the other hand, $p(T)^{s_i}(v_i) = p(T)^{s_i}(u - w) = 0 \implies s_i \geq s \implies s = s_i$

  (Note that $p(x)$ is irreducible is necessarily)

□

5.2 Hilbert space (I)

Let $(V, \langle\cdot,\cdot\rangle)$ be an inner product space.

Definition 5.2.1.

- A sequence of vectors $v_1, v_2, v_3, ... \in V$ is **Cauchy sequence** if

$$\forall\varepsilon > 0 \exists N \in \mathbb{N} \text{ such that } \forall m, n \geq N,\ \|v_m - v_n\| < \varepsilon$$

- We say a sequence of vectors $\{v_n\}_{n=1}^{\infty}$ **converges to** $v \in V$ if

$$\forall\varepsilon > 0, \exists N \text{ such that } \forall n \geq N, \|v_n - v\| < \varepsilon$$

  If the limits exists, then the limits is unique and denoted by $\lim\limits_{n\to\infty} v_n = v$

- An inner product space $(V, \langle\cdot,\cdot\rangle)$ is a **Hilbert space** if any Cauchy sequence in V converges to a vector in V.

Example 5.2.1. $V = \mathbb{R}, \mathbb{C}$

$$V = \ell^2(\mathbb{R}) = \left\{ (a_1, a_2, a_3, ...) \middle| a_i \in \mathbb{R}, \sum_{i=1}^{\infty} |a_i|^2 < \infty \right\},\ \langle (a_n), (b_n) \rangle = \sum_{i=1}^{\infty} a_i b_i$$

$$V = \ell^2(\mathbb{C}) = \left\{ (a_1, a_2, a_3, ...) \middle| a_i \in \mathbb{C}, \sum_{i=1}^{\infty} |a_i|^2 < \infty \right\},\ \langle (a_n), (b_n) \rangle = \sum_{i=1}^{\infty} a_i \overline{b_i}$$

are all Hilbert space. But $V = \mathbb{Q}$ is not Hilbert space.

Definition 5.2.2. A subspace $W \subseteq V$ is a **closed subspace** if any Cauchy sequence $\{w_n\}_{n=1}^{\infty} \subseteq W$ converges to some vector $w \in W$.

Example 5.2.2. Consider $V = \ell(\mathbb{C}) \supseteq W := \{(a_n) | a_n = 0 \text{ for all but finitely many } n\} \rightsquigarrow W$ is a subspace. However, W is not closed, since

$$v_i = (1, 1/2, ..., 1/2^{i-1}, 0, 0, ...) \in W \text{ but } \lim_{n\to\infty} v_n = (1, 1/2, 1/4, ..., 1/2^{n-1}, 1/2^n, ...) \notin W$$

Theorem 5.2.1 (Hilbert projective theorem)**.** Let V be a Hilbert space and W be closed subspace of V. Then

(1) $\forall v \in V$, $\exists w \in W$ s.t. $\|v - w\|$ is minimal among all $w \in W$.

(2) $W^\perp$ is closed.

(3) $V = W \oplus W^\perp$

(4) $(W^\perp)^\perp = W$. In fact, this is equivalent to W is closed.

(5) There exists orthogonal projection $T : V \to V$ from V to W.

Proof:

(1) For all $v \in V$, $d(W, v) := \inf_{w\in W} \|v - w\|$ exists and ≥ 0. Let $d = d(W, v)$, then pick $w_n \in W$ s.t. $\|w_n - v\| < d + \frac{1}{n}$. By parallelogram law,

$$\begin{aligned} 2(\|w_m - v\|^2 + \|w_n - v\|^2) &= 4\|v - \underbrace{\frac{1}{2}(w_m - w_n)}_{\in W}\|^2 + \|w_m - w_n\|^2 \\ &\geq 4d^2 + \|w_m - w_n\| \end{aligned}$$

$$\begin{aligned} \implies \|w_m - w_n\|^2 &\leq 2\left(\|w_m - v\|^2 + \|w_n - v\|^2 - 2d^2\right) \\ &< 2\left(\frac{2d}{m} + \frac{2d}{n} + \frac{1}{n^2} + \frac{1}{m^2}\right) \longrightarrow 0 \text{ as } m, n \to \infty. \end{aligned}$$

So $\{w_n\}_{n=1}^\infty$ is Cauchy sequence.
Let $w = \lim_{n\to\infty} w_n \in W$ and $\|v - w\| = d$.
Claim: w is unique vector s.t. $\|v - w\|$ is minimal.
pf. Suppose $\|v - w\| = \|v - w'\|$, then by parallelogram law, we have

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

and thus $w = w'$.

(2) Let $\{w_n\}_{n=1}^\infty \subseteq W^\perp$ be Cauchy. Then $\lim_{n\to\infty} w_n = v$ for some $v \in V$.

For any $w \in W$, we have

$$\langle v, w\rangle = \langle v - w_n, w\rangle \leq \|v - w_n\| \cdot \|w\| \to 0 \text{ as } n \to \infty$$

$\implies \langle v, w\rangle = 0\ \forall w \in W$ i.e. $v \in W^\perp$.

(3) By (1), $\forall v \in V$, $\exists! w \in W$ s.t. $\|v-w\|$ is minimal. Then set $v = w + (v-w)$.
Claim: $v - w \in W^{\perp}$.
pf. For any $t \in \mathbb{C}$, $u \in W$, we have $\|v-w\|^2 \leq \|v - \underbrace{w - tu}_{\in W}\|^2$, then

$$|t|^2\|u\|^2 \geq 2\mathrm{Re}(\langle v-w, tu\rangle) \text{ for all } t \in \mathbb{C}$$

Suppose $0 \neq \langle v-w, u\rangle = re^{i\theta}$ for some $r > 0$, $\theta \in R$. Let $t = \varepsilon \cdot e^{i\theta}$,

$$\rightsquigarrow \varepsilon^2 \cdot \|u\|^2 \geq \mathrm{Re}(\bar{t}\langle v-w, u\rangle) = 2\varepsilon r \implies \varepsilon \geq \frac{2r}{\|u\|^2}$$

which is contradict to ε can be arbitrarily small. Hence, $\langle v-w, u\rangle = 0 \implies v - w \in W^{\perp}$.

(4) First, we have $W \subseteq (W^{\perp})^{\perp}$. Also, $V = W \oplus W^{\perp} = (W^{\perp}) \oplus (W^{\perp})^{\perp}$ by (2), (3) and thus $(W^{\perp})^{\perp} = W$.

(5) From (3), define by $T(v) = w$ whenever $v = w + w'$ for $w \in W, w' \in W^{\perp}$

□

5.3 Fourier Analysis

Consider the vector space $PC([0,2\pi]) = \{f : [0,2\pi] \to \mathbb{C} : f(0) = f(2\pi), f : \text{piecewise continuous}\}$ $\subseteq L^2([0,2\pi])$: which is a Hilbert space. Where $L^2(0,2\pi)$ is collecting all $f : [0,2\pi] \to \mathbb{C}$ such that

$$\int_0^{2\pi} |f(x)|^2 \, dx : \text{converges}$$

Definition 5.3.1. $f : [0,2\pi] \to R$ is **piecewise continuous** (**sectionally continuous**) if there exists a partition $P : a = t_0 < t_1 < \cdots < t_n = b$ such that $f(x)$ is continuous on (t_{i-1}, t_i) and $\lim\limits_{x \to t_i^-} f(x) = f(t_i^-)$, $\lim\limits_{x \to t_i^+} f(x) = f(t_i^+)$ both exists with with $f(t_i) = \frac{1}{2}(f(t_i^-) + f(t_i^+))$ for all i.

Define the inner product on $PC([0,2\pi])$ by

$$\langle f, g\rangle = \int_0^{2\pi} f(x)\overline{g(x)} \, dx$$

Fact 5.3.1. $\{e^{ikx}\}_{k \in \mathbb{Z}}$ is orthogonal subset of $PC([0,2\pi])$.

Fact 5.3.2. $\{1, \sin(kx), \cos(kx) : k \in \mathbb{N}\}$ is orthogonal subset of $PC([0,2\pi])$.

Definition 5.3.2 (Fourier coefficients). For a given $f(x)$, let

$$c_k = \frac{1}{2\pi}\int_0^{2\pi} f(t)\overline{e^{ikt}} \, dt = \frac{1}{2\pi}\int_0^{2\pi} f(t)e^{-ikt} \, dt \; \forall k \in \mathbb{Z}$$

and we can associate its **Fourier series** $\sum\limits_{k \in \mathbb{Z}} c_k e^{ikx}$.
In other hand, consider the Fourier series of f from the second orthogonal subset :

$$\frac{1}{2}a_0 + \sum_{k=1}^{\infty} a_k \cos(kx) + b_k \sin(kx)$$

where

$$a_0 = \frac{1}{\pi}\int_0^{2\pi} f(x)\ dx \qquad a_k = \frac{1}{\pi}\int_0^{2\pi} f(x)\cos(kx)\ dx \qquad b_k = \frac{1}{\pi}\int_0^{2\pi} f(x)\sin(kx)\ dx$$

Remark 5.3.1. $\begin{cases} \frac{1}{2}a_0 = c_0 \\ a_k = c_k + c_{-k} \\ b_k = (c_k - c_{-k})i \end{cases}$

Theorem 5.3.1 (Bessel's inequality)**.** For any $f \in PC([0,2\pi])$, we have

$$\sum_{k\in\mathbb{Z}} |c_k|^2 \leq \frac{1}{2\pi}\int_0^{2\pi} |f(x)|^2 dx \quad \text{or} \quad \frac{1}{2}|a_0|^2 + \sum_{k=1}^{\infty}\left(|a_k|^2+|b_k|^2\right) \leq \frac{1}{\pi}\int_0^{2\pi} |f(x)|^2 dx$$

Proof: By best approximate f using $\sum\limits_{k=-n}^{n} c_k e^{ikx}$. □

Property 5.3.1. Let $W = \text{span}_{\mathbb{C}}\{e^{ikx} : k \in \mathbb{Z}\} = \text{span}_{\mathbb{C}}\{1, \sin(kx), \cos(kx) : k \in \mathbb{N}\}$. Then $W^{\perp} = \{0\}$ in $PC([0,2\pi])$. Moreover, since $x \in PC[(0,2\pi)] \setminus W$, we conclude that $PC([0,2\pi]) \neq W \oplus W^{\perp}$ and $PC[(0,2\pi)] = (W^{\perp})^{\perp} \supsetneq W$.

Proof: Claim: Suppose f is piecewise continuous, $f(x) = \frac{1}{2}(f(x^+)+f(x^-))$. If $\langle f, 1\rangle = \langle f, \cos(kx)\rangle = \langle f, \sin(kx)\rangle = 0$ for all k, then $f = 0$.

pf. Suppose $f \neq 0$, WLOG there exists $f(z) > 0$ for some $z \in [0,2\pi]$. Note that f is bounded. If $M = \sup\limits_{x\in[0,2\pi]} f(x) \geq 0$. Find δ small enough such that $f(x) \geq M/2$ for all $x \in (x_0-\delta, x_0+\delta_0)$. Then consider $t(x) = 1 + \cos(x-x_0) - \cos\delta$. Then

$$\begin{cases} |t(x)| \leq 1 & \text{for all } x \text{ outside the interval } (x_0-\delta, x_0+\delta) \\ t(x) > 1 & \text{on the interval } (x_0-\delta, x_0+\delta) \text{ and } t(x) \geq \theta > 1 \text{ on } (x_0-\delta/2, x_0+\delta/2) \end{cases}$$

Since $\displaystyle\int_0^{2\pi} f(x)e^{ikx} = 0$, we have $\displaystyle\int_0^{2\pi} f(x)t(x)^N = 0$ for all $N \in \mathbb{N}$, since $\cos^n x$ can be write as the linear combination of $\{1, \sin(kx), \cos(kx) : k \in \mathbb{N}\}$. Notice that

$$\int_{x_0-\delta}^{x_0+\delta} \underbrace{f(x)t(x)^N}_{\geq 0} dx \geq \int_{x_0-\delta/2}^{x_0+\delta/2} f(x)t(x)^N dx \geq \frac{1}{2}M\theta^N\delta \to \infty \text{ as } N \to \infty$$

$$\left|\int_{\text{outside}} f(x)t(x)^N dx\right| \leq \int_{\text{outside}} |f(x)| \cdot 1^N dx \leq 2\pi M$$

which is contradict . Hence, $M = 0$. □

Corollary 5.3.1. $\{e^{ikx} : k \in \mathbb{Z}\}$ is a maximal orthogonal subset in $PC([0,2\pi])$.

Corollary 5.3.2. If f is PC of period 2π, then f has unique Fourier expansion in the sense.

5.4 Hilbert space (II)

In this section, we will discuss Riesz representation theorem and existance of adjoint operator on the Hilbert space.

Definition 5.4.1. Let $\ell : V \to F$ be a linear functional on an inner product space. We say ℓ is **bounded** if $\exists M > 0$ s.t. $|\ell(x)| \leq M \cdot \|x\|$ for all $x \in V$.
Equivalently, $\sup\limits_{\substack{x\in V \\ \|x\|=1}} |\ell(x)| < \infty$.

Theorem 5.4.1 (Riesz representation theorem)**.** Let $(V, \langle\cdot,\cdot\rangle)$ be a Hilbert space and $\ell : V \to F$ be a bounded linear functional. Then, there exists unique $y \in V$ s.t. $\ell(x) = \langle x, y\rangle$ $\forall x \in V$.

Proof: Write $N = \ker \ell = \{v \in V : \ell(v) = 0\}$

- **Existence:**

  Claim: ℓ is bounded $\implies$ N is a closed subspace.

  pf. Given a Cauchy sequence $\{z_n\}$ in N, write $z = \lim\limits_{n\to\infty} z_n \in V$. Since ℓ is bounded, $\exists M > 0$ s.t. $|\ell(x)| < M \cdot \|x\|$ for all $x \in V$. Then

$$|\ell(z)| = |\ell(z) - \ell(z_n)| \leq< M \cdot \|z - z_n\| \to 0 \text{ as } n \to \infty$$

  Thus, $\ell(z) = 0$ i.e. $z \in N$. □

  Now, since N is closed, we have $V = N \oplus N^{\perp}$ (Hilbert projection theorem)

  - Case 1 : If $N = V$, we just pick $y = 0$.
  - Case 2 : If $N \neq V$, pick $v \in N^{\perp}$ with $v \neq 0$. Then $\ell(v) \neq 0$ ($v \notin N$). For any $x \in V$, $\exists \alpha = \dfrac{\ell(x)}{\ell(v)} \in F$ such that $\ell(x) = \alpha \cdot \ell(v)$. Then

$$\begin{aligned}
&\implies \ell(x - \alpha v) = 0 \implies x - \alpha v \in N \\
&\implies \langle x - \alpha v, v\rangle = 0 \qquad (\text{Since } x - \alpha v \in N, v \in N^{\perp}) \\
&\implies \langle x, v\rangle = \alpha\langle v, v\rangle = \frac{\ell(x)}{\ell(v)} \cdot \|v\|^2 \\
&\implies \ell(x) = \frac{\ell(v)}{\|v\|^2}\langle x, v\rangle = \left\langle x, \frac{\overline{\ell(v)}}{\|v\|^2} v \right\rangle
\end{aligned}$$

- **Uniqueness:** Since $\langle\cdot,\cdot\rangle$ is non-degeneracy.

□

Next, we discuss the existence of adjoint operator on Hilbert space.

Definition 5.4.2. We say a linear transformation $T : V \to W$ is **bounded** if $\exists M > 0$ s.t. $\|T(x)\|_W \leq M \cdot \|x\|_V$ for all $x \in V$.
Equivalently, $\sup\limits_{\substack{x\in V \\ \|x\|=1}} \|T(x)\|_W < \infty$.

Theorem 5.4.2. If V, W are two Hilbert space, $T : V \to W$ is bound, then the adjoint of T, $T^* : W \to V$ exists.

Proof: For any $x \in W$, consider the linear functional

$$\begin{aligned}
\ell_x : \ V &\longrightarrow F \\
y &\longmapsto \langle T(y), x\rangle_W
\end{aligned}$$

Then $|\ell_x(y)| = |\langle T(y), x\rangle_W| \le \|T(y)\|_W \cdot \|x\|_W \le M \cdot \|y\|_V \cdot \|x\|_W$. Thus, ℓ_x is bounded (bound by $M \cdot \|x\|_W$). By Riesz representation theorem, $\exists! z_x \in V$ such that

$$\ell_x(y) = \langle y, z\rangle_V \implies \langle T(y), x\rangle_W = \langle y, z_x\rangle_V$$

Then define $T^*(x) = z_x$. Then $\langle T(y), x\rangle_W = \langle y, T^*(x)\rangle$. Finally, we check that T^* is linear : Let $x_1, x_2 \in W$, then $\forall y \in V$

$$\langle y, T^*(x_1 + x_2)\rangle_V = \langle T(y), x_1 + x_2\rangle_W = \langle T(y), x_1\rangle_W + \langle T(y), x_2\rangle_W = \langle y, T^*(x_1) + T^*(x_2)\rangle$$

Thus, $T^*(x_1 + x_2) = T^*(x_1) + T^*(x_2)$. $\square$

Chapter 6

Homework and bonus

We only include interesting or useful problems. Absolutely, we will not put any calculation questions. Some notation will be defined in homework and will not be defined again in class.

6.1

Problem 6.1.1. Let $T : V \to V$ be a linear operator on V. Check the following sets are ideals of $F[x]$:

(1) the set $\{f(x) \in F[x] : f(T) = 0\}$.

(2) the set $\{f(x) \in F[x] : f(T)(v) = 0\}$, where $v \in V$ is a fixed given vector.

(3) the set $I_{v,W} := \{f(x) \in F[x] : f(T)(v) \in W\}$, where W is a T-invariant subspace of V and $v \in V$ is a given vector.

(4) In part (3), if W is only a subspace of V but not T-invariant, does the statement still hold? Prove it or disprove it by giving a counterexample.

Remark 6.1.1. If V is finite-dimensional, then we know that the first set

$$\{f(x) \in F[x] : f(T) = 0\} = (m_T(x))$$

is a principal ideal generated by the minimal polynomial of T.

Problem 6.1.2. Prove also that if W is a T-invariant subspace of V and $v_1 - v_2 \in W$, then $I_{v_1,W} = I_{v_2,W}$.

Problem 6.1.3. Let $T : V \to V$ be a linear operator on a finite-dimensional vector space V and let v be a non-zero vector in V. The set

$$\{f(x) \in F[x] : f(T)(v) = 0\} = (g(x))$$

is a principal ideal generated by a monic polynomial $g(x) \in F[x]$.

(1) If U is the T-cyclic subspace generated by v, show that $g(x)$ is the minimal polynomial of $T|_U$, and $\dim(U)$ equals the degree of $g(x)$.

(2) Show that the degree of $g(x)$ is 1 if and only if v is an eigenvector of T.

Problem 6.1.4. Let $T : V \to V$ be a linear operator on a finite-dimensional vector space V, let W_1 be a T-invariant subspace of V, and let v be a non-zero vector in V. The set

$$I_{v,W_1} = \{f(x) \in F[x] : f(T)(v) \in W_1\} = (g_1(x))$$

is a principal ideal generated by a monic polynomial $g_1(x) \in F[x]$.

(1) Show that $g_1(x)$ divides the minimal and the characteristic polynomials of T.

(2) Let W_2 be a T-invariant subspace of V such that $W_2 \subseteq W_1$ and let $g_2(x)$ be a monic polynomial such that the set

$$I_{v,W_2} = \{f(x) \in F[x] : f(T)(v) \in W_2\} = (g_2(x))$$

is a principal ideal generated by $g_2(x)$. Show that $g_1(x)$ divides $g_2(x)$.

6.2

Problem 6.2.1. Let T be a diagonalizable linear operator on a finite-dimensional vector space V. Prove that V is a T-cycle subspace if and only if each of the eigenspaces of T is one-dimensional.

Problem 6.2.2. Let T be a linear operator on a finite-dimensional vector space V.

(1) Let λ be an eigenvalue of T. Prove that if $\operatorname{rank}((T - \lambda I)^m) = \operatorname{rank}\left((T - \lambda I)^{m+1}\right)$ for some positive integer m, then $K_\lambda = \ker((T - \lambda I)^m)$.

(2) (Second Test for Diagonalizability.) Suppose that the characteristic polynomial $\operatorname{ch}_T(x)$ splits, and let $\lambda_1, \lambda_2, \ldots, \lambda_k$ be all the distinct eigenvalues of T. Show that T is diagonalizable if and only if $\operatorname{rank}(T - \lambda_i I) = \operatorname{rank}\left((T - \lambda_i I)^2\right)$ for $1 \le i \le k$.

Bonus 1. Let $T : V \to V$ be a linear transformation on a finite-dimensional vector space V over F, where F is an infinite field. Show that V is T-cycle if and only if V has only finitely many T-invariant subspace.

6.3

Problem 6.3.1. Let A be an $n \times n$ matrix whose characteristic polynomial splits. Prove that A and A^t have the same Jordan canonical form, and conclude that A and A^t are similar.

6.4

Problem 6.4.1. Let T be a linear operator on a finite-dimensional vector space V whose characteristic polynomial splits, and let J be the Jordan canonical form of T. Let D be the diagonal matrix whose diagonal entries are the diagonal entries of J, and let $M = J - D$.

(1) Show that M is nilpotent, i.e., $M^k = O$ for some integer k.

(2) Show that $MD = DM$.

(3) If J is given by

$$J=\begin{pmatrix}\lambda & 1 & 0 & \cdots & 0\\ 0 & \lambda & 1 & \cdots & 0\\ 0 & 0 & \lambda & \cdots & 0\\ \vdots & \vdots & \vdots & & \vdots\\ 0 & 0 & 0 & \cdots & 1\\ 0 & 0 & 0 & \cdots & \lambda\end{pmatrix}\in M_m(\mathbb{R}).$$

Show that $(J-\lambda I_m)^m=0$ and that if $r\geq m$, then

$$J^r=\begin{pmatrix}\lambda^r & r\lambda^{r-1} & \frac{r(r-1)}{2!}\lambda^{r-2} & \cdots & \frac{r(r-1)\cdots(r-m+2)}{(m-1)!}\lambda^{r-m+1}\\ 0 & \lambda^r & r\lambda^{r-1} & \cdots & \frac{r(r-1)\cdots(r-m+3)}{(m-2)!}\lambda^{r-m+2}\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & 0 & \cdots & \lambda^r\end{pmatrix}.$$

Moreover, if $f(x)\in\mathbb{R}[x]$, show that

$$f(J)=\begin{pmatrix}f(\lambda) & \frac{1}{1!}f'(\lambda) & \frac{1}{2!}f''(\lambda) & \cdots & \frac{1}{(m-1)!}f^{(m-1)}(\lambda)\\ 0 & f(\lambda) & \frac{1}{1!}f'(\lambda) & \cdots & \frac{1}{(m-2)!}f^{(m-2)}(\lambda)\\ 0 & 0 & f(\lambda) & \cdots & \frac{1}{(m-3)!}f^{(m-3)}(\lambda)\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & 0 & \cdots & f(\lambda)\end{pmatrix}.$$

Problem 6.4.2. Let T be a linear operator on a finite-dimensional vector space, and suppose that $\phi(t)$ is a monic factor of the characteristic polynomial of T. Suppose that x and y are two vectors such that the following ideals are the same and generated by $\phi(t)$:

$$I_T(x):=\{f(t)\in F[t]:f(T)(x)=0\}=(\phi(t))=\{f(t)\in F[t]:f(T)(y)=0\}=:I_T(y).$$

Prove that $x\in Z(y;T)$ if and only if $Z(x;T)=Z(y;T)$, where $Z(v;T)$ is the T-cyclic subspace of V generated by the vector v.

Problem 6.4.3. Let T be a linear operator on a finite-dimensional vector space V. Suppose that $v_1,v_2\in V$ are two vectors such that the following ideals are generated by $\phi_1(x)$ and $\phi_2(x)$.

$$\begin{aligned}I_T(v_1)&:=\{f(x)\in F[x]:f(T)(v_1)=0\}=(\phi_1(x))\\ I_T(v_2)&:=\{f(x)\in F[x]:f(T)(v_2)=0\}=(\phi_2(x)).\end{aligned}$$

Suppose further that $\phi_1(x)$ and $\phi_2(x)$ are coprime monic polynomials.

(1) Show that $Z(v_1;T)+Z(v_2;T)$ is a direct sum.

(2) Show that there is a vector $v_3\in V$ such that the following ideal is generated by $\phi_1(x)\cdot\phi_2(x)$:

$$I_T(v_3):=\{f(x)\in F[x]:f(T)(v_3)=0\}=(\phi_1(x)\cdot\phi_2(x)).$$

(3) Show that there exists $v_3\in V$ such that

$$Z(v_1;T)\oplus Z(v_2;T)=Z(v_3;T).$$

Remark 6.4.1. This exercise allows us to combine two T-cyclic subspaces into a single T-cyclic subspace provided that the minimal polynomials of this two subspaces are coprime.

Bonus 2. Prove the **Jordan-Chevalley decomposition theorem**: Let $A\in M_n(F)$ be a matrix whose characteristic polynomial splits. Then, there exist two unique matrices $S,N\in M_n(F)$ satisfying the conditions: $A=S+N$, S is diagonalizable, N is nilpotent, and $SN=NS$.

6.5

Problem 6.5.1. Let T be a linear operator on a finite-dimensional vector space V over F. Show that V is itself a T-cyclic subspace if and only if $\mathrm{ch}_T(x) = m_T(x)$.

Problem 6.5.2. For any $A \in M_{n\times n}(\mathbb{C})$, we write $A = (a_{ij})_{1\le i,j\le n}$ and define

$$||A|| = \max\{|a_{ij}| : 1 \le i, j \le n\}.$$

(1) Prove that for any $A, B \in M_{n\times n}(\mathbb{C})$, $||AB|| \le n||A|| \cdot ||B||$.

(2) Prove that e^A exists for every $A \in M_{n\times n}(\mathbb{C})$.

Problem 6.5.3. Suppose that J is a single Jordan block:

$$J = \begin{pmatrix} \lambda & 1 & 0 & \cdots & 0 \\ 0 & \lambda & 1 & \cdots & 0 \\ 0 & 0 & \lambda & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & \lambda \end{pmatrix} \in M_m(\mathbb{C}).$$

Show that

$$e^{Jz} = e^{\lambda z} \begin{pmatrix} 1 & \frac{z}{1!} & \frac{z^2}{2!} & \cdots & \cdots & \frac{z^{m-1}}{(m-1)!} \\ 0 & 1 & \frac{z}{1!} & \cdots & \cdots & \frac{z^{m-2}}{(m-2)!} \\ 0 & 0 & 1 & \ddots & \cdots & \frac{z^{m-3}}{(m-3)!} \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \ddots & \frac{z}{1!} \\ 0 & 0 & 0 & \cdots & \cdots & 1 \end{pmatrix},$$

where z is a variable.

Problem 6.5.4. Let A be a square matrix with complex entries. Write

$$S = \{z \in \mathbb{C} \mid |z| < 1 \text{ or } z = 1\}$$

to be a subset of $\mathbb{C}$. Show that $\lim\limits_{m\to\infty} A^m$ exists if and only if both of the following conditions hold.

(1) Every eigenvalue of A is contained in S.

(2) If 1 is an eigenvalue of A, then the dimension of the eigenspace corresponding to 1 equals the multiplicity of 1 as an eigenvalue of A.

Bonus 3. Let $S, T : V \to V$ be two linear operators on a finite-dimensional vector space V. Suppose that any linear operator U on V with $US = SU$ has the property $UT = TU$. Show that T is a polynomial of S, that is, there is an $f(x) \in F[x]$ such that $T = f(S)$.

6.6

Problem 6.6.1. Let $S, T : V \to V$ be two linear operators on a finite-dimensional vector space V whose characteristic polynomials split in $F[x]$. Let $\lambda_1, \ldots, \lambda_k$ be eigenvevlues of S. Then, V has decomposition into generalized eigenspaces with respect to S:

$$V = K_{\lambda_1} \oplus K_{\lambda_2} \oplus \cdots \oplus K_{\lambda_k}.$$

(a) Suppose that $ST = TS$. Show that each generalized eigenspace K_{λ_i} with respect to S is a T-invariant subspace.

(b) However, use counterexample to show that "simultaneously Jordan form" is impossible even if $TS = ST$.

Problem 6.6.2. Let $A, B \in M_{n\times n}(\mathbb{C})$. Suppose that the eigenvalues of A and B are all non-negative real numbers and that $\dim\ker A = \dim\ker A^2$ and $\dim\ker B = \dim\ker B^2$. If $A^2 = B^2$, show that $A = B$.

Problem 6.6.3. Let $A \in M_n(F)$ and $P(x) = \operatorname{adj}(xI - A)$, where adj is the classical adjoint. Show that every entry in the matrix $x^k P(x) - P(x)A^k$ is divisible by $\operatorname{ch}_A(x)$ for $k \in \mathbb{N}$.

Problem 6.6.4. Let T be a linear operator on a finite dimensional vector space over F. Show that every T-invariant subspace W has a T-invariant direct summand W' $(W \oplus W' = V)$ if and only if $m_T(x)$ is a product of distinct irreducible factors.

Remark 6.6.1. This problem generalizes the notion of diagonalizable when the characteristic polynomial splits.

Problem 6.6.5. Let $A \in M_n(F)$. Show that the following are equivalent:

(a) F^n is A-cyclic.

(b) $\operatorname{ch}_A(x) = m_A(x)$.

(c) The subspace $\{X \in M_n(F) \mid AX = XA\} \subseteq M_n(F)$ has dimension n.

(d) For any $B \in M_n(F)$ with $AB = BA$, B is a polynomial in A.

(e) For any column vector $(x_1, x_2, \ldots, x_n)^T \in F^n$, there exist column vectors P and Q in F^n such that $x_k = Q^T A^k P$ for all $1 \le k \le n$.

Problem 6.6.6. In this problem, the solution is regarded as an $n \times n$ matrix.

(a) Let $A, B, C \in M_n(\mathbb{C})$. Show that $X(z) = e^{A(z-z_0)} C e^{B(z-z_0)}$ is the unique solution to the differential equation

$$\frac{d}{dz} X(z) = AX(z) + X(z)B$$

with the initial condition $X(z_0) = C$.

(b) Let $A(z)$ be an $n \times n$ matrix whose entries are smooth functions in z. Show that if $X(z)$ is a solution to

$$\frac{d}{dz} X(z) = A(z)X(z),$$

then

$$\det X(z) = \det X(z_0) \exp\left(\int_{z_0}^{z} \operatorname{tr} A(s)\, ds\right).$$

6.7

Recall. A vector space V over F ($F = \mathbb{R}$ or $\mathbb{C}$) is an **inner product space** if there is a function $\langle \cdot, \cdot \rangle : V \times V \to F$ satisfying conditions

(1) $\langle x + z, y \rangle = \langle x, y \rangle + \langle z, y \rangle$,

(2) $\langle cx, y \rangle = c \langle x, y \rangle$,

(3) $\overline{\langle x, y \rangle} = \langle y, x \rangle$,

(4) $\langle x, x \rangle > 0$ if $x \neq 0$,

for all $x, y, z \in V$ and $c \in F$. Such function $\langle \cdot, \cdot \rangle$ is called an **inner product** on V.

Problem 6.7.1. Let V be an inner product space. Use the above definition, for any $x, y \in V$ and $c \in F$, show the following properties.

(1) $\langle x, y + z \rangle = \langle x, y \rangle + \langle x, z \rangle$.

(2) $\langle x, cy \rangle = \overline{c} \langle x, y \rangle$.

(3) $\langle x, 0 \rangle = \langle 0, x \rangle = 0$.

(4) $\langle x, x \rangle = 0$ if and only if $x = 0$.

(5) If $\langle z, x \rangle = \langle z, y \rangle$ for all $z \in V$, then $x = y$.

Remark 6.7.1. From now on, please feel free to use all the above properties.

Problem 6.7.2. Let $\{v_1, v_2, \ldots, v_k\}$ be an orthogonal set in an inner product space V, and let $a_1, a_2, \ldots, a_k \in F$ be scalars. Prove that

$$\left\| \sum_{i=1}^{k} a_i v_i \right\|^2 = \sum_{i=1}^{k} |a_i|^2 \|v_i\|^2 .$$

Problem 6.7.3. Let V be an inner product space over $\mathbb{C}$. For any $x, y \in V$, prove the following identity.

(a) (Parallelogram law) $\|x + y\|^2 + \|x - y\|^2 = 2\|x\|^2 + 2\|y\|^2$.

(b) (Polar identities) $\langle x, y \rangle = \dfrac{1}{4} \displaystyle\sum_{k=1}^{4} i^k \left\| x + i^k y \right\|^2$, where $i^2 = -1$.

Problem 6.7.4. Let $A, B \in M_{n \times n}(\mathbb{C})$. Show that

$$|\operatorname{tr}(AB^*)| \leq (\operatorname{tr}(AA^*) \operatorname{tr}(BB^*))^{1/2} \leq \frac{1}{2} (\operatorname{tr}(AA^*) + \operatorname{tr}(BB^*)) .$$

Problem 6.7.5. Let $V = \mathbb{R}[x]$ be the real vector space of polynomials in x.

(1) Show that

$$\langle f(x), g(x) \rangle = \int_0^\infty f(x) g(x) e^{-x} dx$$

defines an inner product on V.

(2) Find $a, b, c \in \mathbb{R}$ such that $\{a, bx + c\}$ form an orthonormal subset of V.

Bonus 4. A vector space V over F ($F = \mathbb{R}$ or $\mathbb{C}$) is a **normed space** if there is a function $\|\cdot\| : V \to \mathbb{R}$ satisfying the conditions:

(1) $\|x\| \geq 0$ for all $x \in V$. Also, $\|x\| = 0$ if and only if $x = 0$.

(2) $\|ax\| = |a| \cdot \|x\|$ for all $x \in V$ and $a \in F$.

(3) $\|x + y\| \leq \|x\| + \|y\|$ for all $x, y \in V$.

Such function $\|\cdot\|$ is called a **norm** on V.

(a) Let V be a normed space with norm $\|\cdot\|$. Show that there exists an inner product $\langle \cdot, \cdot \rangle$ on V such that $\|x\|^2 = \langle x, x \rangle$ for all $x \in V$ if and only if the norm satisfies the parallelogram law:
$$\|x + y\|^2 + \|x - y\|^2 = 2\|x\|^2 + 2\|y\|^2$$
for all $x, y \in V$.

(b) Consider V to be the set of real-valued continuous functions defined on $[0, 1]$. Show that the norm
$$\|f\| = \max_{x \in [0,1]} |f(x)|$$
on V is never induced by any inner product.

6.8

Problem 6.8.1. Let W be a finite-dimensional subspace of an inner product space V. From the lecture, we know that $V = W \oplus W^{\perp}$. Define the map $T : V \to V$ by
$$T(v) = w$$
whenever $v = w + u$ for some unique $w \in W$ and $u \in W^{\perp}$.

(1) Prove that T is a linear operator on V and that $T^2 = T$.

(2) Prove that $\operatorname{Im}(T) = W$ and $\ker(T) = W^{\perp}$.

(3) Prove that $\|T(x)\| \leq \|x\|$ for all $x \in V$.

(4) Prove that $T = T^*$.

Remark 6.8.1. Such T is called the **orthogonal projection** from V to W.

Problem 6.8.2. Let V be an inner product space, S and S_0 be subsets of V, and W be a finite-dimensional subspace of V. Prove the following results.

(1) $S_0 \subseteq S$ implies that $S^{\perp} \subseteq S_0^{\perp}$.

(2) $S \subseteq \left(S^{\perp}\right)^{\perp}$; so $\operatorname{span}(S) \subseteq \left(S^{\perp}\right)^{\perp}$.

(3) $W = \left(W^{\perp}\right)^{\perp}$.

Problem 6.8.3. Let W_1 and W_2 be subspaces of a finite-dimensional inner product space. Prove that $(W_1 + W_2)^{\perp} = W_1^{\perp} \cap W_2^{\perp}$ and $(W_1 \cap W_2)^{\perp} = W_1^{\perp} + W_2^{\perp}$.

Problem 6.8.4. Let A be an $n \times n$ matrix. Prove that $\det(A^*) = \overline{\det(A)}$.

Definition. Let $(V, \langle \cdot, \cdot \rangle_V)$ and $(W, \langle \cdot, \cdot \rangle_W)$ be two finite-dimensional inner product spaces with inner products $\langle \cdot, \cdot \rangle_V$ and $\langle \cdot, \cdot \rangle_W$, respectively. Let $T : V \to W$ be a linear transformation. A function $T^* : W \to V$ is called an **adjoint** of T if $\langle T(x), y \rangle_W = \langle x, T^*(y) \rangle_V$ for all $x \in V$ and $y \in W$.

Problem 6.8.5. In this exercise, please use the above definition of adjoint.

(1) Prove that there is a unique adjoint T^* of T, and T^* is linear.

(2) If β and γ are orthonormal bases for V and W, respectively, prove that $[T^*]_\gamma^\beta = \left([T]_\beta^\gamma\right)^*$.

(3) Prove that $\langle T^*(x), y \rangle_V = \langle x, T(y) \rangle_W$ for all $x \in W$ and $y \in V$.

(4) Prove that for all $x \in V$, $T^*T(x) = 0$ if and only if $T(x) = 0$.

(5) Prove that $\operatorname{Im}(T^*)^\perp = \ker(T)$.

Problem 6.8.6. Let V be an inner product space, and let $y, z \in V$. Define $T : V \to V$ by $T(x) = \langle x, y \rangle z$ for all $x \in V$. Prove that T is linear, show that T^* exists, and find an explicit expression for T^*.

Problem 6.8.7. Let V be the vector space over $\mathbb{C}$, consisting of all the complex sequences $(a_1, a_2, a_3, \cdots)$ with only finitely many nonzero entries.[1] Then, we define the inner product[2] on V by

$$\langle (a_1, a_2, a_3, \cdots), (b_1, b_2, b_3, \cdots) \rangle = \sum_{i=1}^{\infty} a_i \overline{b_i}.$$

Let e_i be the sequence with only one nonzero entry being 1 at i-th position.

(1) Let $\sigma_n = e_1 + e_n$ and $W = \operatorname{span}_{\mathbb{C}}(\{\sigma_n : n \geq 2\})$.

  (a) Prove that $e_1 \notin W$, so $W \neq V$.

  (b) Prove that $W^\perp = \{0\}$, and conclude that $W \neq \left(W^\perp\right)^\perp$.

(2) Define linear operator $T : V \to V$ by $T\left((a_m)_{m=1}^\infty\right) = (b_n)_{n=1}^\infty$, where

$$b_k = \sum_{i=k}^{\infty} a_i$$

for every positive integer k.[3]

  (a) Prove that for any positive integer n, $T(e_n) = \sum_{i=1}^{n} e_i$.

  (b) Prove that T has no adjoint.

[1] V is a vector space over $\mathbb{C}$ endowed with usual addition and scalar multiplication.

[2] Notice that the infinite series in the definition of inner product converges because $a_n \neq 0$ for only finitely many n.

[3] Notice that the infinite series in the definition of T converges because $a_n \neq 0$ for only finitely many n.

Bonus 5 (Legendre polynomials)**.** Let $V = \mathbb{R}[x]$ be the space of polynomials with coefficients in $\mathbb{R}$. Let $b > a$ be real numbers. Define the inner product on V by

$$\langle f, g\rangle = \int_a^b f(x)\, g(x)\, dx.$$

For each positive integer n, define

$$q_{2n}(x) = (x-a)^n (x-b)^n .$$
$$p_n(x) = \frac{d^n}{dx^n}\left(q_{2n}(x)\right).$$

(1) Show that

$$\frac{d^{i-1} q_{2n}}{dx^{i-1}}(a) = \frac{d^{i-1} q_{2n}}{dx^{i-1}}(b) = 0$$

for all $i = 1, 2, \ldots, n$.

(2) Show that p_n has degree n.

(3) Show that $p_1, p_2, \ldots, p_n$ are orthogonal to each other.

Remark 6.8.2. After normalizing the polynomials $\{p_n\}_{n=1}^{\infty}$, we get the Legendre polynomials on $[a, b]$.

Bonus 6. Let U, V, W be three finite-dimensional inner product spaces. Let

$$\begin{tikzcd} U \arrow[r, bend left, "S"] & V \arrow[l, bend left, "S^*"] \arrow[r, bend left, "T"] & W \arrow[l, bend left, "T^*"] \end{tikzcd}$$

be a sequence of linear transformations such that $TS = 0$ and T^*, S^* are adjoint of T, S, respectively. Let $\Delta : V \to V$ equal $SS^* + T^*T$ and let $H = \ker(\Delta)$. Show that

$$H = \ker(T) \cap \ker(S^*)$$

and V has a natural orthogonal decomposition as

$$V = H \oplus \operatorname{Im}(S) \oplus \operatorname{Im}(T^*)$$

with orthogonal decompositions

$$\ker(T) = H \oplus \operatorname{Im}(S)$$

and

$$\ker(S^*) = H \oplus \operatorname{Im}(T^*).$$

Bonus 7. Let $(V, \langle \cdot, \cdot \rangle)$ be an inner product space. Suppose that $T : V \to V$ is a linear operator satisfying properties:

(1) T is an abstract projection, i.e, $T^2 = T$.

(2) $\|T(x)\| \leq \|x\|$ for all $x \in V$.

Show that T is an orthogonal projection from V to some subspace W.
(To find out this subspace W is also a part of this exercise.)

6.9

Problem 6.9.1. Let $T : V \to V$ be a linear operator on an inner product space V, and let $W \subseteq V$ be a subspace. Suppose that the adjoint T^* exists. Prove the following results.

(1) If W is T-invariant, then $W^\perp$ is T^*-invariant. Similarly, if W is T^*-invariant, then $W^\perp$ is T-invariant.

(2) If W is both T- and T^*-invariant, then $\left(T|_W\right)^* = (T^*)|_W$.

(3) If W is both T- and T^*-invariant and T is normal, then the restriction of T on W, $T|_W$, is normal.

Definition. A linear operator T on a finite-dimensional inner product space is called **positive definite** [**positive semidefinite**] if T is self-adjoint and $\langle T(x), x\rangle > 0$ [$\langle T(x), x\rangle \geq 0$] for all $x \neq 0$.
An $n \times n$ matrix A with entries from $\mathbb{R}$ or $\mathbb{C}$ is called **positive definite** [**positive semidefinite**] if L_A is positive definite [positive semidefinite].

Problem 6.9.2. Let T be a self-adjoint linear operators on an n-dimensional inner product space V, and let $A = [T]_\beta$, where β is an orthonormal basis for V. Prove the following results.

(1) T is positive definite [semidefinite] if and only if all of its eigenvalues are positive [nonnegative].

(2) T is positive definite if and only if

$$\sum_{i,j=1}^{n} A_{ij} a_j \overline{a_i} > 0$$

for all nonzero n-tuples $(a_1, a_2, \ldots, a_n) \in \mathbb{C}^n$. That is, $x^* A x > 0$ for all nonzero $x \in \mathbb{C}^n$.

(3) T is positive semidefinite if and only if $A = B^* B$ for some square matrix $B \in M_n(\mathbb{C})$.

Problem 6.9.3. Let T be a normal operator on a finite-dimensional complex inner product space V, and let W be a subspace of V. Prove that if W is T-invariant, then W is also T^*-invariant.

Problem 6.9.4. Let T be a normal operator on a finite-dimensional inner product space V. Prove that $\ker(T) = \ker(T^*)$ and $\operatorname{Im}(T) = \operatorname{Im}(T^*)$.

Find $P \in M_4(\mathbb{R})$ such that $PP^T = I_4$ and $P^{-1}AP$ is a diagonal matrix.

Problem 6.9.5. Assume that T is a linear operator on a complex (not necessarily finite-dimensional) inner product space V with an adjoint T^*. Prove the following results.

(1) If T is self-adjoint, then $\langle T(x), x\rangle$ is real for all $x \in V$.

(2) If T satisfies $\langle T(x), x\rangle = 0$ for all $x \in V$, then $T = 0$. (Hint: Replace x by $x + y$ and then by $x + iy$.)

(3) If $\langle T(x), x\rangle$ is real for all $x \in V$, then $T = T^*$.

(4) Does the result (2)(3) above hold if we assume that V is a real inner product space? Prove or give a counterexample.

Problem 6.9.6. Let T be a linear operator on a finite-dimensional inner product space $(V, \langle \cdot, \cdot \rangle)$. Define a new pairing $\langle x, y \rangle_T := \langle T(x), y \rangle$ for $x, y \in V$.

(1) Show that $\langle \cdot, \cdot \rangle_T$ defines an inner product on V if and only if T is a positive definite linear operator on V with respect to $\langle \cdot, \cdot \rangle$.

(2) Let $\langle \cdot, \cdot \rangle'$ be any inner product on V, show that there exists a unique linear operator T on V such that $\langle x, y \rangle' = \langle x, y \rangle_T$ for all $x, y \in V$.

(3) Show that the operator T of (b) is positive definite with respect to both inner products $\langle \cdot, \cdot \rangle'$ and $\langle \cdot, \cdot \rangle$.

(4) If S and T are two positive definite linear operator on V, show that ST is positive definite if and only if S and T commute, i.e. $ST = TS$.

Problem 6.9.7. Let V be a finite-dimensional inner product space.

(1) Let U and T be self-adjoint linear operators on V such that $UT = TU$. Prove that there exists an orthonormal basis for V consisting of vectors that are eigenvectors of both U and T.

(2) Let U and T be self-adjoint operators on V such that T is positive definite. Prove that both TU and UT are diagonalizable linear operators that have only real eigenvalues.

(3) Let U be a diagonalizable linear operator on V such that all of the eigenvalues of U are real. Prove that there exist positive definite linear operators T_1 and T_1' and self-adjoint linear operators T_2 and T_2' such that $U = T_2 T_1 = T_1' T_2'$.

6.10

Problem 6.10.1. Let T be a linear operator on a finite-dimensional complex inner product space V. Show that V has an orthonormal basis of eigenvectors of T with corresponding eigenvalues of absolute value 1 if and only if T is unitary.

Problem 6.10.2. For the following matrix A, find a unitary matrix P and a diagonal matrix D such that $P^* AP = D$.

$$A = \begin{pmatrix} 2 & 3-3i \\ 3+3i & 5 \end{pmatrix}.$$

Problem 6.10.3. Prove that if T is a unitary operator on a finite-dimensional inner product space V, then T has a unitary square root; that is, there exists a unitary operator U such that $T = U^2$.

Problem 6.10.4. Let W be a finite-dimensional subspace of an inner product space V. From the lecture, we have $V = W \oplus W^\perp$. Define $U : V \to V$ by

$$U(v_1 + v_2) = v_1 - v_2,$$

where $v_1 \in W$ and $v_2 \in W^\perp$. Prove that U is a self-adjoint unitary operator.

Remark 6.10.1. Such U is the reflection of V about the subspace W.

Problem 6.10.5. Let T be a linear operator on a finite-dimensional inner product space V. Suppose that T is a projection such that $\|T(x)\| \leq \|x\|$ for all $x \in V$. Prove that T is an orthogonal projection.

Problem 6.10.6. Let T be a normal operator on a finite-dimensional complex inner product space V. Use the spectral decomposition $\lambda_1 T_1 + \lambda_2 T_2 + \cdots + \lambda_k T_k$ of T to prove the following results.

(1) If $T^n = 0$ for some $n \geq 1$, then $T = 0$.

(2) T is a projection if and only if every eigenvalue of T is 1 or 0.

(3) $T = -T^*$ if and only if every λ_i is an pure imaginary number.

Bonus 8. (Gaussian integral) Let A be a positive definite $n \times n$ matrix. Prove that

$$\int_{\mathbb{R}^n} \exp\left(-x^T A x\right) dx_1 \cdots dx_n = \frac{\left(\sqrt{\pi}\right)^n}{\sqrt{\det(A)}},$$

where $x = (x_1, \cdots, x_n)^T$ is a column vector.

Chapter 7

Advance Problem

Problems proposed by Ping-Hsun Chuang (TA).

7.1

Problem 7.1.1. Suppose that $A \in M_n(F)$ and the characteristic polynomial $ch_A(x)$ splits in F. Let $\lambda_1, ..., \lambda_m$ be all distinct eigenvalues of A. Show that

$$n(m-1) \leq \sum_{j=1}^{m} \text{rank}(A - \lambda_j I)$$

Problem 7.1.2. Show that the eigenvalues of the tridiagonal matrix

$$A = \begin{pmatrix} a_1 & -b_1 & 0 & \cdots & 0 & 0 & 0 \\ -c_1 & a_2 & -b_2 & \cdots & 0 & 0 & 0 \\ 0 & -c_2 & a_3 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & a_{n-2} & -b_{n-2} & 0 \\ 0 & 0 & 0 & \cdots & -c_{n-2} & a_{n-1} & -b_{n-1} \\ 0 & 0 & 0 & \cdots & 0 & -c_{n-1} & a_n \end{pmatrix} \in M_n(\mathbb{R}),$$

where $b_i c_i > 0$ for all i, are all real and of multiplicity one.

Problem 7.1.3. Let $A \in M_n(\mathbb{R})$. Show that rank $A =$ rank A^2 if and only if $\lim\limits_{\lambda \to 0} (A + \lambda I)^{-1} A$ exists.

7.2

Problem 7.2.1. Suppose that $A \in M_n(F)$ and the characteristic polynomial $ch_A(x)$ splits in F. Let $\lambda_1, ..., \lambda_m$ be all distinct eigenvalues of A. Also, let $a_1(\lambda_i), ..., a_{r_i}(\lambda_i)$ be all size of Jordan blocks corresponding to the eigenvalue λ_i (counting multiplicity) for each i. Show that the subspace

$$\{X \in M_n(F) | XA = AX\} \subseteq M_n(F)$$

has dimension

$$\sum_{i=1}^{m} \sum_{1 \leq i,k \leq r_i} \min\{a_j(\lambda_i), a_k(\lambda_i)\}$$

Problem 7.2.2. Show that a matrix A can be represented as the product of two involutions (A matrix B is an involution if $B^2 = I$) if and only if the matrices A and A^{-1} are similar.

Problem 7.2.3. Suppose that $T : V \to V$ is a linear operator on a vector space V and that $\{v_1, ..., v_n\}$ is a basis for V that us a single Jordan chain (in other words, a cycle of generalized eigenvectors) for T. Determine a Jordan canonical basis for T^2.

7.3

Problem 7.3.1. Let A be an $n \times n$ diagonalizable matrix with all eigenvalues being $\lambda_1, ..., \lambda_n$. Show that there are right eigenvectors (column vector) $x_1, ..., x_n$ and left eigenvectors (row vectors) $y_1, ..., y_n$ such that

$$A = \sum_{i=1}^{n} \lambda_i x_i y_i$$

Problem 7.3.2. Let $A, B \in M_{n\times n}(\mathbb{C})$. Show that the following are equivalent :

(a) A and B are similar.

(b) Either

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

are similar or

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

are similar.

Problem 7.3.3. Let

$$A = \begin{pmatrix} 0 & 0 & \cdots & 0 & 0 & 0 & -a_0 \\ 1 & 0 & \cdots & 0 & 0 & 0 & -a_1 \\ 0 & 1 & \cdots & 0 & 0 & 0 & -a_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 & 0 & -a_{n-3} \\ 0 & 0 & \cdots & 0 & 1 & 0 & -a_{n-2} \\ 0 & 0 & \cdots & 0 & 0 & 1 & -a_{n-1} \end{pmatrix}$$

Show that there exists an invertible symmetric matrix S such that $A = SA^T S^{-1}$.

7.4

Problem 7.4.1. Let $A = I + xy^T$, where x and y are nonzero n real column vectors. Show that $\det(A) = 1 + x^T y$. Also, determine the Jordan canonical form of the matrix A.

Problem 7.4.2. Let T be a linear transformation on a finite-dimensional vector space over F. Show that V is T-cyclic if and only if any linear operator S on V commuting with T is a polynomial in T.

Problem 7.4.3. Let $A \in M_{n\times n}(\mathbb{C})$. Define

$$\exp(A) = \sum_{k=0}^{\infty} \frac{1}{k!} A^k$$
$$\sin(A) = \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k+1)!} A^{2k+1}$$
$$\cos(A) = \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k)!} A^{2k}$$
$$\log(I_n + A) = \sum_{k=1}^{\infty} \frac{(-1)^k}{k} A^k$$

(a) For any $A \in M_{n\times n}(\mathbb{C})$, show that

$$\sin^2(A) + \cos^2(A) + I_n$$

(b) For any $A \in M_{n\times n}(\mathbb{C})$ with all eigenvalue $|\lambda_i| < 1$, show that

$$\exp(\log(I_n + A)) = I_n + A$$

7.5

Problem 7.5.1. Let $\|\cdot\|$ be the norm induced by the standard inner product on $\mathbb{C}^n$. Show that for any non-zero vectors $x, y \in \mathbb{C}^n$, we have

$$\|x - y\| \geq \frac{1}{2}(\|x\| + \|y\|)\left\|\frac{x}{\|x\|} - \frac{y}{\|y\|}\right\|$$

Problem 7.5.2. Let $A \in M_n(\mathbb{C})$. We equip the vector space $\mathbb{C}^n$ with standard inner product and induced norm $\|\cdot\|$. Define the **numerical range** $W(A)$ by

$$W(A) := \{x^* A x : x \in \mathbb{C}^n \text{ with } \|x\| = 1\} \subseteq \mathbb{C}.$$

(a) Show that $W(A)$ is a convex compact subset in $\mathbb{C}$.

(b) Show that $\frac{1}{n}\operatorname{tr}(A)$ is contained in $W(A)$.

Problem 7.5.3. Let V be an inner product space. Let $e_1, e_2, \cdots, e_n$ be an orthogonal basis of V and $d_1, d_2, \cdots, d_n$ be the lengths of the vectors $e_1, e_2, \cdots, e_n$. Show that an m-dimensional subspace $W \subseteq V$ such that the orthogonal projections of $e_1, e_2, \cdots, e_n$ on W are of equal length exists if and only if

$$d_i^2\left(\frac{1}{d_1^2} + \frac{1}{d_2^2} + \cdots + \frac{1}{d_n^2}\right) \geq m$$

for all $1 \leq i \leq n$.

7.6

Problem 7.6.1. Suppose $f(x)$ is continuous on $\mathbb{R}$ and $f(x+2\pi) = f(x)$. Find the constants c, a_k, b_k (in terms of $f(x)$) such that

$$\int_{-\pi}^{\pi} \left[f(x) - c - \sum_{k=1}^{n} a_k \cos kx + b_k \sin kx \right]^2 dx$$

is minimal.

Problem 7.6.2.

(a) Let $a_k(x)$ be functions with $\left| \sum\limits_{k=1}^{n} a_k(x) \right| \leq M$ for some constant M and all n. Let $p_k \in \mathbb{R}$ such that $\lim\limits_{k\to\infty} p_k = 0$ and $\sum\limits_{k=1}^{\infty} |p_k - p_{k+1}|$ converges. Show that $\sum\limits_{k=1}^{\infty} p_k a_k(x)$ converges uniformly.

(b) Fix $\alpha > 0$ and $1 < \varepsilon < 1$. Show that the series $\displaystyle\sum_{k\geq 1} \frac{\cos kx}{k^\alpha}$ converges uniformly for $\varepsilon \leq x \leq 2\pi - \varepsilon$.

(c) Fix $0 < \varepsilon < 1$. Prove that the series $\displaystyle\sum_{k\geq 1} \frac{\cos kx}{k}$ converges uniformly to $-\log(2|\sin \frac{x}{2}|)$ for $\varepsilon \leq x \leq 2\pi - \varepsilon$.

Problem 7.6.3. Let $A \in M_n(\mathbb{C})$ be nonzero matrix such that $A^* = A$. Show that

$$\operatorname{rank}(A) \geq \frac{(\operatorname{tr}(A))^2}{\operatorname{tr}(A^2)}$$

7.7

Problem 7.7.1. Let $A \in M_n(\mathbb{C})$. Show that the following are equivalent :

(a) A is Hermitian, that is, $A = A^*$.

(b) There is a unitary matrix U such that U^*AU is a real diagonal matrix.

(c) x^*Ax is real for all $x \in \mathbb{C}^n$.

(d) $A^2 = A^*A$.

(e) $A^2 = AA^*$.

(f) $\operatorname{tr}(A^2) = \operatorname{tr}(A^*A)$.

(g) $\operatorname{tr}(A^2) = \operatorname{tr}(AA^*)$.

Problem 7.7.2. Let A be an $n \times n$ Hermitian matrix. Show that the following are equivalent :

(a) A is positive semidefinite, that is, $x^*Ax \geq 0$ for all $x \in \mathbb{C}^n$.

(b) All eigenvalues of A are nonnegative.

(c) $U^*AU = \text{diag}(\lambda_1, ..., \lambda_n)$, for some unitary matrix U and all λ_i are all nonnegative.

(d) $A = B^*B$ for some matrix B.

(e) $A = T^*T$ for some $r \times n$ matrix T with rank $r = \text{rank}(T) = \text{rank}(A)$.

(f) All principal minors of A are nonnegative, that is, the matrices $M_k = (m_{ij})_{1\le i,j\le k}$ defined by $m_{ij} = a_{ij}$ have nonegative determinant for all $1 \le k \le n$.

(g) $\text{tr}(AX) \ge 0$ for all positive semidefinite matrix X.

(h) $X^*AX \ge 0$ for all $n \times m$ matrix X.

Problem 7.7.3. Let $A \in M_n(\mathbb{C})$ and have eigenvalues $\lambda_1, ..., \lambda_n$. Show that the following are equivalent :

(a) A is normal, that is, $AA^* = A^*A$.

(b) $I - A$ is normal.

(c) There exists a unitary matrix U such that $U^*AU = \text{diag}(\lambda_1, ..., \lambda_n)$

(d) There is a set of the unit eigenvectors of A that form an orthonormal basis of $\mathbb{C}^n$.

(e) Every eigenvector of A is an eigenvector of A^*.

(f) $A^* = AU$ for some unitary U.

(g) $A^* = VA$ for some unitary V.

(h) $\text{tr}(A^*A) = \sum_{1\le i,j\le n} |a_{ij}|^2 = \sum_{i=1}^{n} |\lambda_i|^2$.

(i) $\text{tr}(A^*A)^2 = \text{tr}\left((A^*)^2A^2\right)$.

(j) $\|Ax\| = \|A^*x\|$ for all $x \in \mathbb{C}^n$.

(k) $A + A^*$ and $A - A^*$ are commute.

(l) $A^*A - AA^*$ is positive semidefinite.

(m) A commute with A^*A.

(n) A commute with $AA^* - A^*A$.

7.8

Problem 7.8.1 (Converse Spectral Theorem)**.** Let V be a finite-dimensional inner product space. Suppose that V has direct sum decomposition into subspaces

$$V = V_1 \oplus V_2 \oplus \cdots \oplus V_k$$

with $V_i \ne 0$ for all $i = 1, 2, ..., k$. Let $P_i : V \to V_i$ be the orthogonal projection and define

$$T = P_1 + P_2 + \cdots + P_k.$$

Show that $0 < \det T \le 1$. Also, $\det T = 1$ if and only if $V_i \oplus V_j$ is an orthogonal direct sum for any $i \ne j$.

Problem 7.8.2. Let $A \in M_N(\mathbb{C})$. Show that A is a product of two Hermitian matrices if and only if A is similar to A^*.

Problem 7.8.3. Let $A, B \in M_n(\mathbb{C})$. Write $\sigma_1 \geq \cdots \geq \sigma_n$ to be the singular values of A and $\tau_1 \geq \cdots \geq \tau_n$ to be the singular values of $\mathcal{B}$. Show that

$$|\operatorname{tr}(AB)| \leq \sum_{i=1}^{n} \sigma_i \tau_i$$