Linear Algebra II

Felix Chen

Contents

1 Introduction **2**
- 1.1 recap . . . 2

2 Diagonization **3**
- 2.1 Eigen-things . . . 3
- 2.2 Characteristic polynomial . . . 4

3 Canonical forms **8**
- 3.1 Minimal polynomials and Cayley-Hamilton . . . 8
- 3.2 Invariant subspaces . . . 10
- 3.3 Primary Decomposition . . . 12
- 3.4 Cyclic decomposition . . . 13
- 3.5 Rational canonical forms . . . 16
- 3.6 Primary cyclic decomposition and Jordan canonical forms . . . 18
- 3.7 Semisimple transformations . . . 21
- 3.8 Bonus section . . . 23

4 Inner product spaces **24**
- 4.1 Inner product . . . 24
- 4.2 Orthogonality . . . 27
- 4.3 Adjoint maps . . . 29
- 4.4 Orthogonal maps and Unitary maps . . . 31
- 4.5 Normal maps . . . 32

5 Bilinear forms **35**
- 5.1 Positive define matrices . . . 37
- 5.2 Bilinear forms on inner product spaces . . . 38
- 5.3 Spectral decomposition . . . 39
- 5.4 Further on normal maps . . . 42

6 Bilinear forms **45**
- 6.1 Some special bilinear forms . . . 46
- 6.2 Lie algebras . . . 48
- 6.3 Abelian, nilpotent and solvable Lie algebras . . . 52

§1 Introduction

Teacher: An Jinpeng
Homepage: https://www.math.pku.edu.cn/teachers/anjp/algebra

§1.1 recap

Direct sums of vector spaces Given a field F, let $V_1, \dots, V_k$ be vector spaces over F. The set

$$V_1 \times \cdots \times V_k = \{(v_1, \dots, v_k) \mid v_i \in V_i\}$$

forms a vector space by the operations

$$(v_1, \dots, v_k) + (w_1, \dots, w_k) = (v_1 + w_1, \dots, v_k + w_k)$$

and

$$c \cdot (v_1, \dots, v_k) = (cv_1, \dots, cv_k).$$

We call this vector space the **external direct sum** of $V_1, \dots, V_k$, denoted by $\bigoplus_{i=1}^k V_i$.

Obviously $(U \oplus V) \oplus W \simeq U \oplus (V \oplus W)$.

For every i, we have an injective linear map:

$$\begin{aligned} \tau_i : &V_i \to \bigoplus_{j=1}^k V_j \\ &v_i \mapsto (0, \dots, v_i, \dots, 0) \end{aligned}$$

Lemma 1.1.1

If $\mathcal{B}_i$ are the bases of V_i, then $\bigcup_{i=1}^k \tau_i(\mathcal{B}_i)$ is a basis for $\bigoplus_{i=1}^k V_i$.

In paricular,

$$\dim \bigoplus_{i=1}^k V_i = \sum_{i=1}^k \dim V_i.$$

Proof. Spanning part:

For any $(v_1, \dots, v_k) \in \bigoplus_{i=1}^k V_i$,

$$v_i \in V_i = \operatorname{span}(\mathcal{B}_i) \implies \tau_i(v_i) \in \operatorname{span}\left(\tau_i(\mathcal{B})_i\right) \implies (v_1, \dots, v_k) \in \operatorname{span}\left(\bigcup_{i=1}^k \tau_i(\mathcal{B}_i)\right)$$

Linearly independent part:

If $\bigcup_{i=1}^k \tau_i(\mathcal{B}_i)$ is linearly dependent, i.e. exists $e_{ij} \in \mathcal{B}_i$ satisfying $\exists c_{ij} \in F$,

$$\sum_{i,j} c_{ij} \tau_i(e_{ij}) = 0.$$

This expands to

$$\left(\sum_{j=1}^{m_1} c_{1j} e_{1j}, \dots, \sum_{j=1}^{m_k} c_{kj} e_{kj} \right) = 0.$$

but e_{1j} are linear independent, which implies $c_{1j} = 0$. $\square$

Remark 1.1.2 — Let V be a vector space over F, and $V_1, \dots, V_k$ are subspaces of V.

Consider a linear map $\Phi : V_1 \oplus \cdots \oplus V_k \to V$ by $(v_1, \dots, v_k) \mapsto v_1 + \cdots + v_k$.

Then $\operatorname{Im}(\Phi) = V_1 + \cdots + V_k$. If Φ is injective, i.e. $V_1, \dots, V_k$ are independent, we say $V_1 + \cdots + V_k$ the **internal direct sum** of $V_1, \dots, V_k$.

In this case Φ gives an isomorphism of external and internal sums:

$$\Phi : \bigoplus_{i=1}^{k} V_i \xrightarrow{\sim} \sum_{i=1}^{k} V_i.$$

Lemma 1.1.3

The following statements are equivalent:

1. $V_1, \dots, V_k$ are independent;
2. For $v_i \in V_i, (i = 1, \dots, k)$, if $\sum_{i=1}^{k} v_i = 0$, then $v_i = 0$.
3. For any $1 \le i \le k$, $V_i \cap (V_1 + \cdots + V_{i-1}) = \{0\}$.
4. Given arbitary bases $\mathcal{B}_i$ of V_i, they are disjoint and their union is a basis of $\bigoplus_{i=1}^{k} V_i$.
5. If $\dim V < +\infty$, they are also equivalent to:

$$\dim \sum_{i=1}^{k} V_i = \sum_{i=1}^{k} \dim V_i.$$

Proof. It's easy but verbose so I leave it out. □

Example 1.1.4

Let $\operatorname{char} F \neq 2$, $V = F^{n \times n}$, $V_1 = \{A \in V \mid A^t = A\}$, $V_2 = \{A \in V \mid A^t = -A\}$.

Note that $V_1 \cap V_2 = \{0\}$, and $V_1 + V_2 = V$, hence $V_1 \oplus V_2 = V$ is an internal direct sum.

§2 Diagonization

Example: google page rank?

Given a linear map T, it can be represented as different matrices under different bases. Thus a question arises: What's the simpliest matrix representation of a linear map?

Definition 2.0.1 (Diagonizable maps)**.** Let V be a vector space over F, $T \in L(V)$ is a linear map from V to itself. If the matrix of T under a certain basis is diagonal, we say T is **diagonizable**.

In this case the linear map T can be simply described as a diagonal matrix, thus we'll study under what condition is T diagonizable.

§2.1 Eigen-things

Definition 2.1.1 (Eigenvalue)**.** Let $T : V \to V$ be a linear map, for $c \in F$, let

$$V_c = \{v \in V \mid Tv = cv\} = \ker(T - c \cdot \mathrm{id}_V).$$

If $V_c \neq \{0\}$, we call c an **eigenvalue** of T, and V_c the **eigenspace** of T with respect to c. the vectors in V_c are called **eigenvectors**.

All the eigenvalues of T are called the **spectrum** of T, denoted by $\sigma(T)$.

Proposition 2.1.2
Let $\mathcal{B}$ be a basis of V, then $[T]_{\mathcal{B}}$ is diagonizable $\iff$ vectors in $\mathcal{B}$ are all eigenvectors.

Proof. Let $\mathcal{B} = \{e_1, \ldots, e_k\}$, $A = [T]_{\mathcal{B}}$.

$$Te_j = \sum_{i=1}^{k} A_{ij} e_i.$$

So A is diagonal $\iff A_{ij} = 0$ when $i \neq j$,
$\iff \exists c_j \in F, Te_j = c_j e_j$,
$\iff$ all the vectors e_j are eigenvectors. □

Example 2.1.3
Let $V = F^{n\times n}$, then V_{sym} is the eigenspace of 1, and $V_{antisym}$ is the eigenspace of -1.

Lemma 2.1.4
Let T be a linear operator, then

$$\sigma(T) = \{c \in F \mid \det(c \cdot id_V - T) = 0\}.$$

Proof. $V_c = \ker(c \cdot \mathrm{id}_V - T)$,

$$c \in \sigma(T) \iff V_c \neq \{0\} \iff \det(c \cdot \mathrm{id}_V - T) = 0.$$

□

§2.2 Characteristic polynomial

To define the characteristic polynomial rigorously, we need to introduce one more concept:

Definition 2.2.1 (Rational function field)**.** Let F be a field, and $F[x]$ be its polynomial ring. Define the **rational function field**:

$$H := \{(f, g) \mid f, g \in F[x], g \neq 0\} = F[x] \times (F[x]\backslash\{0\}).$$

This process is similar to the extension from $\mathbb{Z}$ to $\mathbb{Q}$: We define a equivalent relation on H:

$$(f_1, g_1) \sim (f_2, g_2) \iff f_1 g_2 = f_2 g_1.$$

Let $F(x)$ be the set of all the equivalence classes.

Next we define the addition and multiplication as the usual way, and check they are well-defined (here it is left out).

Remark 2.2.2 — This process can be adapted to any integral domain R, which gives its fraction field $\operatorname{Frac}(R)$.

In general, we can define $F(x_1, \dots, x_n) = \operatorname{Frac}(F[x_1, \dots, x_n])$.
Let F be a field, and V a finite dimensional vector space over F, T is a linear operator on V.
We want to find the eigenvalues of T, by Lemma 2.1.4, we need to solve the equation

$$\det(c \cdot \mathrm{id}_V - T) = 0.$$

Definition 2.2.3 (Characteristic polynomial)**.** Let $A \in F^{n \times n}$, consider

$$xI - A \in F[x]^{n \times n} \subset F(x)^{n \times n}.$$

So

$$\det(xI - A) =: f_A(x) \in F(x).$$

The polynomial $f_A(x)$ is called the **characteristic polynomial** of A. Observe that its roots are all the eigenvalues of A.

In fact we can write f_A explicitly:

$$f_A(x) = \sum_{i=0}^{n} (-1)^i \sum \det B x^{n-i}$$

where $\sum \det B$ is over all $i \times i$ principal minors of A. In particular, $f_A(0) = (-1)^n \det A$.

Remark 2.2.4 — In fact, the more intrinsic way to define the characteristic polynomial is to define it as $f_T(x) = (x - c_1)(x - c_2) \cdots (x - c_n)$, where c_i's are eigenvalues of a linear operator T. However, this definition requires the theory of Jordan forms, so it's hard to define it beforehand.

It's clear that similar matrices has the same characteristic polynomial since they represent the same linear operator.

Lemma 2.2.5

Let $A : F^r \to F^n$, $B : F^n \to F^r$ be linear maps. Then $f_{AB}(x) = x^{n-r} f_{BA}(x)$.

Proof 1. Note that

$$\begin{pmatrix} xI_n & A \\ B & I_r \end{pmatrix} \begin{pmatrix} I_n & 0 \\ -B & xI_r \end{pmatrix} = \begin{pmatrix} xI_n - AB & xA \\ 0 & xI_r \end{pmatrix}.$$

and

$$\begin{pmatrix} I_n & 0 \\ -B & xI_r \end{pmatrix} \begin{pmatrix} xI_n & A \\ B & I_r \end{pmatrix} = \begin{pmatrix} xI_n & A \\ 0 & xI_r - BA \end{pmatrix}.$$

By taking the determinant of both equations, we get:

$$x^r \det(xI_n - AB) = x^n \det(xI_r - BA).$$

□

Proof 2. By taking a suitable basis, we may assume $A = \begin{pmatrix} I_m & 0 \\ 0 & 0 \end{pmatrix}$. Suppose $B = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}$, where B_{11} is an $m \times m$ matrix.

Compute

$$AB = \begin{pmatrix} B_{11} & B_{12} \\ 0 & 0 \end{pmatrix}, BA = \begin{pmatrix} B_{11} & 0 \\ B_{21} & 0 \end{pmatrix}.$$

we get $f_{AB}(x) = f_{B_{11}}(x)x^{n-m}, f_{BA}(x) = x^{r-m}f_{B_{11}}(x)$. □

If T is diagonalizable, then $f_T(x) = (x - c_1)\cdots(x - c_n)$, where $\{c_1, \ldots, c_n\} = \sigma(T)$.

Example 2.2.6 (How to diagonalize a matrix)

Let $A = \begin{pmatrix} 5 & -6 & -6 \\ -1 & 4 & 2 \\ 3 & -6 & -4 \end{pmatrix}$, we can compute $f_A(x) = (x-1)(x-2)^2$.

Next we compute the eigenspaces of each eigenvalue:

$$V_1 = (3, -1, 3), V_2 = \text{span}\{(2, 1, 0), (2, 0, 1)\}.$$

denote the eigenvectors by v_1, v_2, v_3.

At last we set $P = (v_1, v_2, v_3)$, we know $P^{-1}AP = \text{diag}\{1, 2, 2\}$.

Example 2.2.7

Let $A = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$, $f_A(x) = x^2 - 2\cos\theta x + 1$, which has no real roots.

But if we regard it as a complex matrix, we can get $\sigma(A) = \{e^{i\theta}, e^{-i\theta}\}$, and $P = \begin{pmatrix} 1 & 1 \\ -i & i \end{pmatrix}$.

Example 2.2.8

Let $A = \begin{pmatrix} \lambda & a & b \\ 0 & \lambda & c \\ 0 & 0 & \lambda \end{pmatrix}$, where $\lambda, a, b, c \in \mathbb{R}$.

$f_A = (x - \lambda)^3$, but its eigenspace has dimension less than 3, so A is not diagonalizable.

From the examples we know not all the matrices can be diagonalized

- When f_A cannot decomposite to products of polynomials of degree 1;
- When the dimensions of eigenspaces can't reach $\dim V$.

The first case can be solved by putting it in a larger field; While the second case is intrinsic.

In what follows we'll take a closer look at the diagonalizable matrices, and find some equivalent statement of being diagonalizable.

Proposition 2.2.9

T can be diagonalize $\iff$ V can decomposite to direct sums of one-dimensional fixed subspaces.

Proof. Since there exists a basis consisting of eigenvectors: $\{e_1, \dots, e_n\}$, then $V = \bigoplus_{i=1}^n Fe_i$.

On the other hand, if $V = \bigoplus_{i=1}^n V_i$, where V_i's are 1-dimensional subspaces fixed under T, take $v_i \in V_i$, it's clear that v_i's form a basis of V, and they are all eigenvectors. This implies T is diagonalizable. □

Proposition 2.2.10

The eigenspaces of different eigenvalues are independent. So their sum is acutually internal direct sums.

Proof. Let $\sigma(T) = \{c_1, \dots, c_r\}$, for any $v_i \in V_{c_i}$, if $v_1 + \cdots + v_r = 0$, let

$$S_1 = (T - c_2 \operatorname{id}_V) \cdots (T - c_r \operatorname{id}_V),$$

then $S_1(v_1 + \cdots + v_r) = Cv_1 = 0 \implies v_1 = 0$.

(As $S_1 v_i = (c_i - c_2) \cdots (c_i - c_r) v_i$ for $1 \le i \le r$.)

Similarly $v_i = 0$ for all i. □

Proposition 2.2.11

Suppose

$$f_T(x) = \prod_{c \in \sigma(T)} (x - c)^{m(c, f_T)}.$$

then $\forall c \in \sigma(T)$ we have $1 \le \dim V_c \le m(c, f_T)$.

Here $\dim V_c$ is called the **geometric multiplicy**, and $m(c, f_T)$ is the **algebraic multiplicy** of c.

Proof. Let $d = \dim V_c \ge 1$.

Take a basis $\{e_1, \dots, e_d\}$ of V_c and extend it to a basis of V: $\{e_1, \dots, e_n\}$.

Since $Te_i = ce_i, \forall i \le d$, so

$$[T]_{(e_i)} = \begin{pmatrix} cI_d & * \\ 0 & * \end{pmatrix}.$$

so $f_T(x) = (x - c)^d g(x)$, which means $m(c, f_T) \ge d$. □

Now we come to a conclusion:

Theorem 2.2.12

The followings are equivalent:

1. T is diagonalizable;
2. $V = \bigoplus_{c \in \sigma(T)} V_c$;
3. $\dim V = \sum_{c \in \sigma(T)} \dim V_c$;
4. $f_T(x) = \prod_{c \in \sigma(T)} (x - c)^{\dim V_c}$.

Proof. This follows immediately by previous propositions. □

§3 Canonical forms

It turns out that not all linear operators can be expressed as diagonal matrix. In this section we proceed in another direction: to find the "simpliest" matrix expression for a general operator.

Definition 3.0.1 (Irreducible maps)**.** Let T be a linear operator on V. We say T is **reducible** if V can be decompose to a direct sum of two T-invariant subspaces $W_1 \oplus W_2$. Otherwise we say T is **irreducible**.

In order to study T, we only need to study the "smaller" maps $T\big|_{W_1}$ and $T\big|_{W_2}$. In this case we denote $T = T\big|_{W_1} \oplus T\big|_{W_2}$. By decompose these smaller maps, we'll eventually get a decomposition of T consisting of irreducible maps:

$$T = \bigoplus_{i=1}^{r} T_{W_i}.$$

Then by taking a basis of each W_i, and they form a basis $\mathcal{B}$ of V. It's easy to observe that $[T]_{\mathcal{B}}$ is a block diagonal matrix.

In the special case when the W_i's are all 1-dimensional subspaces, the map T is diagonalizable. The eigenvectors are the elements in the W_i's and the eigenvalues are actually the map T_{W_i}.

§3.1 Minimal polynomials and Cayley-Hamilton

Definition 3.1.1 (Annihilating polynomial)**.** Let $M_T = \{f \in F[x] \mid f(T) = 0\}$, we say the polynomial in M_T are the **annihilating polynomials** of T.

Note that M_T is an *nonzero* ideal of $F[x]$. This is because $\{\mathrm{id}, T, \ldots, T^{n^2}\} \subset \mathrm{Mat}_{n\times n}(F)$ must be linealy dependent.

Proposition 3.1.2

T is diagonalizable $\iff \exists f \in M_T$ s.t. f is the product of different polynomials of degree 1.

Before we prove this proposition, let us take a look at the properties of annihilating polynomials.

Since $F[x]$ is a PID, M_T must be generated by one element, namely p_T, the *minmimal polynomial* of T. Thus we can WLOG assmue $f = p_T$ in the above proposition.

Speaking of polynomials and linear maps, one thing that pops into our mind is the characteristic polynomial f_T. In fact there is strong relations between p_T and f_T:

Theorem 3.1.3 (Cayley-Hamilton)

The characteristic polynomial of a linear operator T is its annihilating polynomial, i.e. $f_T(T) = 0$.

This theorem is also true when T is a matrix on a module. To prove it more generally, we introduce the concept of modules.

Definition 3.1.4 (Modules over commutative rings)**.** Let R be a commutative ring. A set M is called a **module** over R or an **R-module** if:

- There is a binary operation (addition) $M \times M \to M : (\alpha, \beta) \mapsto \alpha + \beta$ such that M becomes a commutative group under this operation.

- There is an operation (scaling) $R \times M \to M : (r, \alpha) \mapsto r\alpha$ with assosiativity and distribution over addition (both left and right). We also require $1_R\alpha = \alpha$ for all $\alpha \in M$.

Example 3.1.5

A commutative group automatically has a structure of $\mathbb{Z}$-module. (view the group operation as addition in definition of modules)

Example 3.1.6

Let $R = F[x]$, T a linear operator on V. Define $R \times V \to V : (f, \alpha) \mapsto f\alpha := f(T)\alpha$. We can check V becomes a module over R.

We can also define matrices on a commutative ring R, with addition and multiplication identical to the usual matrices. So the determinant and characteristic polynomial make sense as well.

Note that each $m \times n$ matrix represents a homomorphism $R^m \to R^n$.

Proof of Theorem 3.1.3. Take a basis $\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\}$ of V. Let $A = [T]_{\mathcal{B}}$. If we view V as a R-module ($R = F[x]$),

$$(\alpha_1, \ldots, \alpha_n)A = (T\alpha_1, \ldots, T\alpha_n) = (x\alpha_1, \ldots, x\alpha_n) = (\alpha_1, \ldots, \alpha_n) \cdot xI_n.$$

This implies $(\alpha_1, \ldots, \alpha_n)(xI_n - A) = (0, \ldots, 0)$.

Claim 3.1.7. If $f \in F[x]$ s.t. $\exists B \in R^{n\times n}$ s.t. $(xI_n - A)B = fI_n$, then $f(T) = 0$.

Proof of the claim.

$$0 = (\alpha_1, \ldots, \alpha_n)(xI_n - A)B = (\alpha_1, \ldots, \alpha_n) \cdot fI_n = (f(T)\alpha_1, \ldots, f(T)\alpha_n).$$

Since $\alpha_1, \ldots, \alpha_n$ is a basis, $f(T)$ must equal to 0. □

Now it's sufficient to prove f_T satisfies the condition in the claim. This follows from letting $B = A^{\text{adj}}$, the adjoint matrix of A. □

Remark 3.1.8 — In fact this proof is derived from the proof of Nakayama's lemma, which is an important result in commutative algebra.

As a corollary, $p_T \mid f_T$.

Proof of Proposition 3.1.2. First we prove a lemma:

Lemma 3.1.9

Let $T_1, \ldots, T_k \in L(V)$, $\dim V < \infty$. Then

$$\dim \ker(T_1T_2 \ldots T_n) \leq \sum_{i=1}^{k} \dim \ker(T_i).$$

Proof of the lemma. By induction we only need to prove the case $k = 2$.

Note that $\ker(T_1T_2) = \ker(T_2) + \ker(T_1|_{\operatorname{im} T_2})$. So

$$\dim\ker(T_1T_2) = \dim\ker(T_2) + \dim\ker(T_1|_{\operatorname{im} T_2}) \le \dim\ker(T_2) + \dim\ker(T_1).$$

$\square$

If T is diagonalizable, suppose the matrix of T is $diag\{c_1, \dots, c_r\}$, then $g = \prod_{i=1}^r (x - c_i)$ is an annihilating polynomial of T.

Conversely, if $\prod_{i=1}^r (T - c_i I) = 0$, by lemma

$$n = \ker\left(\prod_{i=1}^r (T - c_i I)\right) \le \sum_{i=1}^r \ker(T - c_i I) = \sum_{i=1}^r \dim V_{c_i}.$$

This forces $V = \bigoplus_{i=1}^r V_{c_i}$, which completes the proof. $\square$

§3.2 Invariant subspaces

For an invariant subspace $W \subset V$, there may not exist a subspace W' s.t. $W \oplus W' = V$, so we can instead study the quotient space.

Define $T_W = T|_W \in L(W)$, $T_{V/W} \in L(V/W)$: $T_{V/W}(\alpha + W) = T(\alpha) + W$. It's clear that $T_{V/W}$ is well-defined.

However, this decomposition loses some information about T, i.e. they can't determine T completely. For example when $T = \begin{pmatrix} A & B \\ 0 & C \end{pmatrix}$, the matrix B will not be carried to T_W and $T_{V/W}$ as their matrices are A, C respectively.

Since $\det T = \det T_W \det T_{V/W}$, $f_T = f_{T_W} \cdot f_{T_{V/W}}$. The minimal polynomials satisfy

$$\operatorname{lcm}(p_{T_W}, p_{T_{V/W}}) \mid p_T, \quad p_T \mid p_{T_W} p_{T_{V/W}}.$$

This follows by the definition of $T_W, T_{V/W}$, readers can check it manually. Hint: The image of $p_{T_{V/W}}(T)$ is in W. So by Proposition 3.1.2, T is diagonalizable $\iff$ $T_W, T_{V/W}$ are both diagonalizable and their minimal polynomials are coprime.

Definition 3.2.1 (Simultaneous diagonalization)**.** Let $\mathcal{F} \subset L(V)$, if there exists $\mathcal{B}$ s.t. $\forall T \in \mathcal{F}$, $[T]_{\mathcal{B}}$ is diagonal matrix, then we say $\mathcal{F}$ can be simultaneously diagonalized.

Proposition 3.2.2

Let $\mathcal{F} \subset L(V)$, TFAE:

- $\mathcal{F}$ can be simultaneously diagonalized;
- Any element in $\mathcal{F}$ is diagonalizable, and any two elements commute with each other.

Proof. It's obvious the first statement implies the second.

On the other hand, we proceed by induction on the dimension of the space V.

Assume $\dim V = n \ge 2$, WLOG $T \in \mathcal{F}$ is not a scalar matrix.

Let $\sigma(T) = \{c_1, \dots, c_r\}, V = \bigoplus_{i=1}^r V_{c_i}$, where $r \ge 2$, $V_{c_i} \ne V$. Since T commutes with other elements in $\mathcal{F}$, so $V_{c_i} = \ker(T - c_i \operatorname{id}_V)$ is invariant under all the maps in $\mathcal{F}$.

Hence we can restrict $\mathcal{F}$ to V_{c_i} and apply induction hypothesis, i.e. for any $U \in \mathcal{F}$, $U|_{V_{c_i}}$ can be simultaneously diagonalized.

Therefore $\exists \mathcal{B}_i$ s.t. $[U|_{V_{c_i}}]_{\mathcal{B}_i}$ is diagonal $\implies$ $[U]_{\mathcal{B}}$ is diagonal, where $\mathcal{B} = \bigcup \mathcal{B}_i$. $\square$

Definition 3.2.3 (Triangulable matrix)**.** Let $T \in L(V)$. If $[T]_{\mathcal{B}}$ is an upper triangular matrix for some basis $\mathcal{B}$, we say T is **triangulable**.

Proposition 3.2.4

Let $\dim V = n$, for $T \in L(V)$, TFAE:

(1) T is triangulable;

(2) f_T(or p_T) can be decomposed to product of polynomials of degree 1.

(3) There exists a sequence of T-invariant subspaces $\{0\} = W_0 \subset W_1 \subset \cdots \subset W_n = V$. This kind of sequence is called a **flag**. (Flag itself does not require T-invariant)

Remark 3.2.5 — In particular, when the base field is *algebraically closed*, the above statements always holds.

Proof. It's obvious that (1) $\implies$ (2).

For (2) $\implies$ (3): We proceed by induction, for W_1 just take the space spanned by one of the eigenvectors of T.

Assume that we have constructed W_j for $0 \le j \le i$. Instead of finding an invariant subspace of dimension $i+1$, we'll find an invariant subspace of dimension 1 in V/W_i.

Let Q denote the quotient map $V \to V/W_i$. Consider the map $T_{V/W_i} : \alpha + W_i \mapsto T(\alpha) + W_i$. We have

$$T_{V/W_i} \circ Q = Q \circ T.$$

$$\begin{tikzcd} V \arrow[r, "T"] \arrow[d, "Q"] & V \arrow[d, "Q"] \\ V/W_i \arrow[r, "T_{V/W_i}"] & V/W_i \end{tikzcd}$$

Since $p_{T_{V/W_i}} \mid p_T \implies p_{T_{V/W_i}}$ is product of polynomials of degree 1, T_{V/W_i} must have an eigenvector. Let L denote the subspace spanned by this vector, and $W_{i+1} = Q^{-1}(L)$.

Clearly $\dim W_{i+1} = 1 + \dim W_i = i + 1$. It suffices to check that W_{i+1} is T-invariant:

$$T(W_{i+1}) = T(Q^{-1}(L)) = Q^{-1}(T_{V/W_i}(L)) \subset Q^{-1}(L) = W_{i+1}.$$

Now for the last part (3) $\implies$ (1):

Let $\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\}$, such that $\operatorname{span}\{\alpha_1, \ldots, \alpha_i\} = W_i$. The matrix of T under $\mathcal{B}$ is clearly an upper triangular matrix. $\square$

Proposition 3.2.6

Let F be an algebraically closed field. Suppose the elements of $\mathcal{F} \subset L(V)$ are pairwise commutative, then $\mathcal{F}$ is simultaneously triangulable.

Remark 3.2.7 — The inverse of this proposition is not true: Just let $\mathcal{F}$ be the set consisting of all the upper triangular matrices.

To prove this, we need a lemma:

Lemma 3.2.8

There's a common eigenvector of $\mathcal{F}$.

Proof of lemma. WLOG $\mathcal{F}$ is finite. (In fact, $\operatorname{span}\mathcal{F} \subset L(V)$ is a finite dimensional vector space, so we can take a basis $\mathcal{F}_0$.)

Now by induction, if $T_1, \ldots, T_{k-1}$ have common eigenvector α, let $T_i\alpha = c_i\alpha$. Then

$$W = \bigcap_{i=1}^{k-1} \ker(T_i - c_i \operatorname{id}_V) \neq \{0\}$$

is a T_k-invariant space.

So any eigenvector α' of $T_k\big|_W$ is the common eigenvector. □

Proof of the proposition. It suffices to prove that there exists an $\mathcal{F}$-invariant flag. By the lemma, the proof is nearly identical as the proof of previous proposition. □

§3.3 Primary Decomposition

In this section we mainly study how a linear map is decomposed into irreducible maps and the structure of irreducible maps.

Recall that every vector space V is an $F[x]$-module given a linear operator T. If a subspace $W \subset V$ is a T-invariant space, then W is a submodule of V.

Hence it leads to decompose V into direct sums of submodules.

Definition 3.3.1. Let V, W be isomorphic vector spaces. $T \in L(V)$, $T' \in L(W)$. If there exists an isomorphism $\Phi : V \to W$ s.t. $\Phi \circ T = T' \circ \Phi$, we say T and T' are **equivalent**.

Definition 3.3.2 (Primary maps)**.** Let $T \in L(V)$ be a linear map. We say T is **primary** if p_T is a power of prime polynomials.

Theorem 3.3.3 (Primary decomposition)

Let $T \in L(V)$, $p_T = \prod_{i=1}^k p_i^{r_i}$, where p_i are different monic prime polynomials of degree 1.

We have

$$V = \bigoplus_{i=1}^k W_i, \quad W_i = \ker\left(p_i^{r_i}(T)\right),$$

with $W_i \neq \{0\}$ and $T\big|_{W_i}$ primary.

Proof. Let $f_i = \prod_{j \neq i} p_j^{r_j}$, f_i and p_i are coprime.

Note that $f_i(T) \neq 0$ and $f_i(T)p_i^{r_i}(T) = p_T(T) = 0$, thus $p_i^{r_i}(T)$ is not inversible, which implies $W_i \neq \{0\}$.

W_i independent : If there exists $\alpha_j \in W_j$ s.t. $\sum_{j=1}^k \alpha_j = 0$, applying f_i we get $f_i(\alpha_i) = 0$. But $p_i^{r_i}(\alpha_i) = 0 \implies \alpha_i = 0, \forall i$.

To prove $V = \sum_{i=1}^k W_i$, observe that

$$\gcd(f_1, \dots, f_k) = 1 \implies \exists g_1, \dots, g_k \quad s.t. \quad 1 = \sum_{i=1}^k g_i f_i \implies \alpha = \sum_{i=1}^k g_i(f_i \alpha), \quad \forall \alpha \in V.$$

Since $f_i\alpha \in W_i$, W_i is T-invariant $\implies g_i f_i \alpha \in W_i$.

Lastly, we'll prove that the minimal polynomial q_i of $T\big|_{W_i}$ is $p_i^{r_i}$.

Clearly $p_i^{r_i}(T\big|_{W_i}) = 0$, so $q_i \mid p_i^{r_i}$.

On the other hand, $q_1 q_2 \dots q_k$ is an annihilating polynomial of T, hence

$$\prod_{i=1}^k p_i^{r_i} \mid \prod_{i=1}^k q_i \implies q_i = p_i^{r_i}, \quad \forall i.$$

□

§3.4 Cyclic decomposition

In the following contents we'll assmue $R = F[x]$ if it's not specified.

Definition 3.4.1 (Cyclic maps)**.** Let V be a finite dimensional vector space and $T \in L(V)$. For $\alpha \in V$, $R\alpha = \{f\alpha \mid f \in R\} = \operatorname{span}\{\alpha, T\alpha, \dots\}$ is the smallest T-invariant subspace containing α.

We say T is **cyclic** if $\exists \alpha$ s.t. $V = R\alpha$. In this case α is called a **cyclic vector**.

Here $R\alpha$ is called the cyclic subspace spanned by α.

Remark 3.4.2 — The word "cyclic" comes from the theory of modules.

Note that $\dim R\alpha = 1 \iff \alpha$ is an eigenvector.

Example 3.4.3

Let $A = E_{21} \in F^{2\times 2}$. Then A is cyclic because $A\varepsilon_1 = \varepsilon_2$, $A\varepsilon_2 = 0$. This means ε_1 is a cyclic vector of A,

Now there's a natural question: When is T cyclic and how to find its cyclic vectors?

For a given vector α, let $M_\alpha = \{f \in R \mid f\alpha = 0\}$ is an ideal of R.

Note that $M_T \subset M_\alpha$ as $f \in M_T \implies f(T)\alpha = 0$, so M_α is nonempty, it has a generating element p_α, called the **annihilator** of α.

Proposition 3.4.4

Let $d = \deg p_\alpha$, then $\{\alpha, T\alpha, \dots, T^{d-1}\alpha\}$ is a basis of $R\alpha$. In particular, $\dim R\alpha = \deg p_\alpha$.

Proof. Linear independence:

If $\sum_{i=0}^{d-1} c_i T^i \alpha = 0$, let $g = \sum_{i=0}^{d-1} c_i x^1$.

$$g\alpha = 0 \implies g \in M_\alpha \implies p_\alpha \mid g.$$

But $\deg g \le d - 1 < d = \deg p_\alpha \implies g = 0$.

Spanning:

Clearly $T^i\alpha \in R\alpha$. $\forall f \in R$, let $f = qp_\alpha + r$ with $\deg r < \deg p_\alpha$. Hence $f\alpha = r\alpha \in \operatorname{span}\{\alpha, T\alpha, \dots, T^{d-1}\alpha\}$. □

Since α is a cyclic vector $\iff \dim R\alpha = \dim V$, and $\deg p_\alpha \le \deg p_T \le \deg f_T = \dim V$, so we care whether these two inequalities can attain the equality.

Proposition 3.4.5

There exists $\alpha \in V$ s.t. $p_\alpha = p_T$.

Proof. Let $p_T = \prod_{i=1}^k p_i^{r_i}$.

$$W_i = \ker(p_i^{r_i}(T)) \implies V = \bigoplus_{i=1}^k W_i.$$

We claim that $\ker(p_i^{r_i-1}) \subsetneq W_i$ as $p_{T_{W_i}} = p_i^{r_i}$.

Take a vector $\alpha_i \in W_i \backslash \ker(p_i^{r_i-1}(T))$. By definition $p_{\alpha_i} \mid p_i^{r_i}, p_{\alpha_i} \nmid p_i^{r_i-1} \implies p_\alpha = p_i^{r_i}$.

Let $\alpha = \sum_{i=1}^k \alpha_i$. If $f\alpha = 0$, then $f\alpha_i = 0$ for $i = 1, \dots, k$ as $f\alpha_i \in W_i$.

$$f\alpha_i = 0 \implies p_{\alpha_i} \mid f \implies p_T \mid f.$$

This means we must have $p_\alpha = p_T$. $\square$

Now we come to a conclusion:

Corollary 3.4.6

T is cyclic $\iff \deg p_T = \dim V \iff p_T = f_T$.

In this case, α is a cyclic vector $\iff p_\alpha = p_T$.

Let $n = \dim V$, T be a cyclic map, α be a cyclic vector. By previous proposition, $\{\alpha, T\alpha, \dots, T^{n-1}\alpha\}$ is a basis of V. Denote the basis by $\mathcal{B}$.

Observe that $[T]_\mathcal{B}$ is equal to

$$A = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & -c_0 \\ 1 & 0 & 0 & \cdots & 0 & -c_1 \\ 0 & 1 & 0 & \cdots & 0 & -c_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & -c_{n-2} \\ 0 & 0 & 0 & \cdots & 1 & -c_{n-1} \end{pmatrix}$$

where c_i are the coefficients of $p_\alpha = p_T = f_T = \sum_{i=0}^n c_i x^i$. For a monic polynomial f, define C_f to be the matrix as above, called the **companion matrix** of f.

Proposition 3.4.7

If exists a basis $\mathcal{B}$ s.t. $[T]_\mathcal{B} = C_f$ for some monic polynomial f, then T is cyclic and $p_T = f$.

Proof. Let $\mathcal{B} = \{\alpha_1, \dots, \alpha_n\}$, we have $T^i\alpha_1 = \alpha_{i+1} \implies R\alpha_1 = V$ and $p_{\alpha_1} = f$. $\square$

Remark 3.4.8 — In fact we can check directly that f is the characteristic polynomial of C_f.

This gives another proof of Cayley-Hamilton theorem:

Proof. For any $\alpha \in V$, consider $T_{R\alpha}$:

$$f_{T_{R\alpha}} = f_{C_{p_\alpha}} = p_\alpha \mid f_T$$

This implies that f_T is an annihilating polynomial of α, which means $f_T(\alpha) = 0, \forall \alpha \in V$, i.e. $f_T(T) = 0$. □

Theorem 3.4.9 (Cyclic decomposition)

Let $T \in L(V)$, $\dim V = n$. There exists $\alpha_1, \ldots, \alpha_r \in V$ s.t. $V = \bigoplus_{i=1}^r R\alpha_i$.

Furthermore, $p_{\alpha_r} \mid \cdots \mid p_{\alpha_1} = p_T$, $f_T = \prod_{i=1}^r p_{\alpha_i}$.

Here p_{α_i}'s are called the **invariant factors** of T. The invariant factors are *totally determined* by T.

First we prove a lemma:

Lemma 3.4.10

Let $\alpha \in V$ with $p_\alpha = p_T$, $\forall L \in V/R\alpha$, exists $\beta \in L$ s.t. $p_\beta = p_L$.

Here $f \cdot L := f(T_{V/R\alpha})L$, so $fL = 0 \iff f(T)\beta \in R\alpha$, $\forall \beta \in L$.

Proof. For all $\beta \in L$, we must have $p_\beta L = 0$, since $L = \beta + R\alpha, T(R\alpha) = R\alpha$.

If $p_L\beta \neq 0$, since $p_L\beta \in R\alpha$, thus $p_L\beta = f\alpha$ for some $f \in R$.

Because $p_L \mid p_\beta \mid p_\alpha = p_T$,

$$\left(\frac{p_\alpha}{p_L}\right) f\alpha = p_\alpha\beta = 0.$$

We have $\frac{p_\alpha}{p_L} f$ is an annihilator of α, hence it's a multiple of p_α, i.e. $p_L \mid f$.

Let $f = p_L h$, $\beta_0 = \beta - h\alpha$, we have $p_L\beta_0 = f\alpha - p_L h\alpha = 0 \implies p_{\beta_0} = p_L$. □

Returning to our original theorem, we'll prove by induction on n.

Take $\alpha_1 \in V$ s.t. $p_{\alpha_1} = p_T$. Consider $V/R\alpha_1$, its dimension is strictly lesser than n. By induction hypo, $\exists L_2, L_3, \ldots, L_r \in V/R\alpha_1$, such that

$$V/R\alpha_1 = \bigoplus_{i=1}^r RL_i, \quad p_{L_r} \mid \cdots \mid p_{L_2}.$$

Take $\alpha_i \in L_i$ s.t. $p_{\alpha_i} = p_{L_i}$, we must have $p_{\alpha_r} \mid \cdots \mid p_{\alpha_1} = p_T$.

If there exists $g_i\alpha_i \in R\alpha_i$ s.t. $\sum_{i=1}^r g_i\alpha_i = 0$, then

$$\sum_{i=2}^r g_i L_i = 0 \implies g_i L_i = 0 \implies g_i\alpha_i = 0.$$

For any $\gamma \in V$, since $\gamma \in \gamma + R\alpha_1$, by induction hypo, $\gamma + R\alpha_1 = \sum_{i=2}^r h_i L_i$.

This means $\gamma - \sum_{i=2}^r h_i\alpha_i \in R\alpha_1$, this completes the existence part of the theorem.

As for the uniqueness part, note that $p_T = \operatorname{lcm}(p_1, \ldots, p_r) = p_1$ and $f_T = p_1 \cdots p_r$, suppose $q_1, \ldots, q_s$ are also invariant factors of T, we must have $p_1 = q_1 = p_T$ and $\prod p_i = \prod q_i$.

Assume for contradiction that $\exists 2 \leq t \leq \min\{r, s\}$ s.t. $p_t \neq q_t$, but $p_i = q_i$ for all $i < t$.

Multiplying p_t on both sides of $\bigoplus_{i=1}^r R\alpha_i = \bigoplus_{i=1}^s R\beta_i$ we get:

$$\bigoplus_{i=1}^{t-1} Rp_t\alpha_i = p_tV = \bigoplus_{i=1}^{t-1} Rp_t\beta_i \oplus \bigoplus_{i=t}^{s} Rp_t\beta_i.$$

Now observe that

- For monic polynomial f, g, if $p_\alpha = fg$, then $p_{f\alpha} = g$ as $h(f\alpha) = 0 \iff (fh)\alpha = 0$.

  Hence
  $$\dim Rp_t\alpha_i = \deg p_{p_t\alpha_i} = \deg \frac{p_i}{p_t} = \deg \frac{q_i}{p_t} = \deg Rp_t\beta_i.$$

This implies $\bigoplus_{i=t}^s Rp_t\beta_i = \{0\}$, in particular $p_t\beta_t = 0 \implies p_t \mid q_t$. Similarly $q_t \mid p_t \implies p_t = q_t$, contradiction!

Theorem 3.4.11

Let G be a finite abelian group, then $\exists g_1, \ldots, g_r \in G\backslash\{0\}$, such that $G = \bigoplus_{i=1}^r \mathbb{Z}g_i$ and $|\mathbb{Z}g_r| \mid \cdots \mid |\mathbb{Z}g_1|$.

Remark 3.4.12 — The proof is identical to the proof above.

§3.5 Rational canonical forms

Let $d_i = \deg p_i = \dim R\alpha_i$, $\mathcal{B}_i = \{\alpha_i, \ldots, T^{d_i-1}\alpha_i\}$ is a basis of $R\alpha_i$. Then $[T_{R\alpha_i}]_{\mathcal{B}_i}$ is the companian matrix C_{p_i}, hence T can be represented as a blocked diagonal matrix with each block is C_{p_i} for invariant factors p_i. This is called the **rational canonical form** of T.

Definition 3.5.1. We say $A \in F^{n\times n}$ is **rational** if exists monic $p_1, \ldots, p_r \in F[x]$, such that $p_r \mid \cdots \mid p_1$ and $A = \operatorname{diag}(C_{p_1}, \ldots, C_{p_r})$.

Theorem 3.5.2

Let $T \in L(V)$, then T has a unique rational canonical form.

Proof. If $[T]_{\mathcal{B}'} = \operatorname{diag}(C_{q_1}, \ldots, C_{q_r})$ is another rational canonical form, let $\mathcal{B}' = (\mathcal{B}'_1, \ldots, \mathcal{B}'_r)$.

It's easy to observe that $\operatorname{span}\mathcal{B}'_i = R\beta_i$, where β_i is the first element in $\mathcal{B}_i$, so $V = \bigoplus_{i=1}^r R\beta_i$ is a cyclic decomposition of V, by the previous theorem we deduce the canonical form is unique. □

So far we've proved that $A \sim B \iff A, B$ have the same rational canonical form. Note that this canonical form does not require any extra properties of the base field F.

Next we'll see some applications of it. Different from Jordan canonical forms, rational canonical forms focus more on theory than computaion.

Proposition 3.5.3 (Rational canonical forms don't depend on fields)

Let $A \in F^{n\times n}$ has rational canonical form A', and the invariant factors are $p_1, \ldots, p_r \in F[x]$.

If $K \subset F$ is a smaller field s.t. $A \in K^{n\times n}$, then A' is still the rational canonical form of A in K. i.e. $A' \in K^{n\times n}$, and $\exists P \in K^{n\times n}, A' = PAP^{-1}$.

Proof. Let A'' be the rational form of A on K. By the uniqueness of rational canonical forms, we must have $A' = A''$, since they are both the rational form of A on F. $\square$

Proposition 3.5.4 (Similarity in larger fields implies similarity in smaller fields)
Let A, B be matrices on F, and $A \sim B$ in F. If $A, B \in K^{n\times n}$, where K is a subfield of F, then $A \sim B$ in K as well.

Proof. Let C be the rational canonical form of A, B, since $A, B \in K^{n\times n}$, by the previous proposition, $C \in K^{n\times n}$ and $A \sim C \sim B$ in K. $\square$

Proposition 3.5.5
$\forall A \in F^{n\times n}$, $A \sim A^t$.

Proof. Firstly when $A = C_f$ for some $f \in F[x]$, A has only one invariant factor f. Note that $f_{A^t} = p_{A^t} = f_A = p_A = f$, so the invariant factor of A^t is also f, by rational canonical forms we're done.

Next for generic matrix A, just take the rational canonical form B. By above we have

$$A \sim B \implies A \sim B \sim B^t \sim A^t.$$

$\square$

Example 3.5.6 (How to compute the rational canonical forms (in low dimensions))
Let $A = \begin{pmatrix} 5 & -6 & -6 \\ -1 & 4 & 2 \\ 3 & -6 & -4 \end{pmatrix} \in \mathbb{Q}^{3\times 3}$. First observe that $f_A = (x-1)(x-2)^2$.

Since $(x-1)(x-2)$ is the minimal polynomial of A, so the invariant factors are $p_1 = (x-1)(x-2), p_2 = (x-2)$. Hence the rational canonical form of A is

$$\begin{pmatrix} 0 & -2 & 0 \\ 1 & 3 & 0 \\ 0 & 0 & 2 \end{pmatrix}$$

Next we'll find vectors α_1, α_2 s.t. $p_{\alpha_i} = p_i$. So $P = (\alpha_1, A\alpha_1, \alpha_2)$ will be the transition matrix.

Proposition 3.5.7
Let T be a diagonalizable map, $\sigma(T) = \{c_1, \dots, c_k\}$. Let $V_1, \dots, V_k$ be the primary decomposition of V,

- Let $\alpha = \sum_{i=1}^k \beta_i, \beta_i \in V_i$, then $R\alpha = \operatorname{span}\{\beta_1, \dots, \beta_k\}$, $p_\alpha = \prod_{\beta_i \neq 0}(x - c_i)$.
- Let $d_i = \dim V_i$, then $p_j = \prod_{d_i \geq j}(x - c_i)$.

Proof. Trivial but need some work to check it. $\square$

§3.6 Primary cyclic decomposition and Jordan canonical forms

Theorem 3.6.1

For $T \in L(V)$, T irreducible $\iff$ T is primary and cyclic.

Proof. If T is irreducible, then both the primary and cyclic decomposition have only one term, i.e. T is primary and cyclic.

Conversely, if $V = V_1 \oplus V_2$ is a nontrivial decomposition. Since T is cyclic and primary, assume $f_T = p_T = p^r$, where p is a irreducible polynomial.

Suppose $f_{T_1} = p^s, f_{T_2} = p^t$, then $s + t = r$, $s, t < r$. Since $p_{T_1} \mid p^s, p_{T_2} \mid p^t$,

$$p_T = \operatorname{lcm}(p_{T_1}, p_{T_2}) \mid p^{\max\{s,t\}},$$

contradiction! □

Theorem 3.6.2 (Primary cyclic decomposition)

Let $T \in L(V)$.

- There exists a decomposition $V = \bigoplus_{i=1}^{s} V_i$, each V_i is T-invariant, T_{V_i} primary and cyclic. Let $q_i = p_{T_{V_i}}$.
- $q_1, \ldots, q_s$ are uniquely determined by T (ignoring the permutation). They are called the **elementary divisors** of T.

Proof. Existence follows immediately from the previous theorem.

Uniqueness: Let $V = \bigoplus_{i=1}^{t} W_i$ s.t. T_{W_i} is primary and cyclic. Let $\{u_1, \ldots, u_k\}$ be the set of all the monic prime factors of the minimal polynomials of $T_{W_1}, \ldots, T_{W_t}$.

We can group W_i's by u_i, and each group can be placed in a row in descending order wrt the degree of $p_{T_{W_i}}$.

Let Z_j be the direct sum of the j-th column, note that Z_j is a cyclic decomposition of T.

Now since the cyclic decomposition and primary decomposition are unique, $p_{T_{W_i}}$'s must be unique as well. □

Remark 3.6.3 — The elementary factors depend on the base field.

Since the invariant subspaces of primary subspace are primary, and invariant subspaces of cyclic subspace are cyclic, we can apply both decomposition (in any order) to get the primary cyclic decomposition of any operators.

For a primary cyclic map T, if we choose the base field to be *algebraically closed* (e.g. $\mathbb{C}$), we can write $f_T = p_T = (x - c)^n$. Let $N = T - c\,\mathrm{id}_V$, then $f_T = p_T = x^n$, from rational canonical form we know that N is similar to $\begin{pmatrix} 0 & 0 \\ I_{n-1} & 0 \end{pmatrix}$. Hence T is similar to

$$J_n(c) := \begin{pmatrix} c & & & & \\ 1 & c & & & \\ & 1 & \ddots & & \\ & & \ddots & c & \\ & & & 1 & c \end{pmatrix},$$

such matrix is called a **Jordan block**. Jordan matrices are the blocked diagonal matrices with each block being a Jordan block.

Theorem 3.6.4 (Jordan canonical forms)
If f_T can be decompose to product of polynomials of degree 1, then

- $\exists \mathcal{B}$ s.t. $[T]_\mathcal{B}$ is a Jordan matrix, this is called the **Jordan canonical form** of T.
- The canonical form is unique under permutations of each Jordan blocks.

Proof. This follows immediately from the primary cyclic decomposition of T. ☐

Let's look at the subspaces V_i. We know that T_{V_i} is primary and cyclic, thus $f_i = p_i = (x-c_i)^{r_i}$. Let $N_i = T_{V_i} - \mathrm{id}_{V_i}$, $f_{N_i} = p_{N_i} = x^{r_i}$. Let $\mathcal{B}_i = \{\alpha_i, N_i\alpha_i, \dots, N_i^{r_i-1}\alpha_i\}$, then $[N_i]_{\mathcal{B}_i} = C_{x^{r_i}} = J_{r_i}(0)$.

We can compute the Jordan canonical forms by computing the invariant factors first, and apply the primary decomposition to each factor to get the elementary divisors.

Example 3.6.5
Let $A = \begin{pmatrix} 2 & & \\ a & 2 & \\ b & c & -1 \end{pmatrix} \in \mathbb{C}^{3\times 3}$.

First note that $f_A = (x-2)^2(x+1)$, then $p_A = (x-2)^2(x+1)$ or $(x-2)(x+1)$.

- If $p_A = (x-2)^2(x+1)$, then $p_1 = (x-2)^2(x+1)$, $q_{11} = (x-2)^2, q_{12} = (x+1)$.

  Hence $A \sim \begin{pmatrix} 2 & & \\ 1 & 2 & \\ & & -1 \end{pmatrix}$.

- $p_A = (x-2)(x-1)$, then $p_1 = (x-2)(x+1)$, $p_2 = (x-2)$. The elementary divisors are $x-2, x-2$ and $x+1$.

  Hence $A \sim \begin{pmatrix} 2 & & \\ & 2 & \\ & & -1 \end{pmatrix}$.

Since $p_A = (x-2)(x+1) \iff (A-2I)(A+I) = 0$, i.e. $3a = ac = 0 \iff a = 0$.

Remark 3.6.6 — For generic matrix A, the Jordan canonical form can be derived from the *Smith canonical form* of $xI_n - A$.

The diagonal of Jordan canonical forms are the eigen values of T with *algebraic multiplicy*, and f_T, p_T can be easily written down from it. The number of Jordan blocks with eigenvalue c is equal to $\dim\ker(T - c\,\mathrm{id})$, i.e. the *geometric multiplicy* of c.

Example 3.6.7

We'll compute the Jordan canonical form of $J_n(0)^2$. Since its characteristic polynomial is x^n, and $\dim\ker J_n(0)^2 = 2$, so it has two Jordan block with eigenvalue 0.

But note that $(J_n(0)^2)^m = 0$ iff $m \geq \frac{n}{2}$, thus the minimal polynomial is x^m, the sizes of the Jordan blocks are $\left\lfloor \frac{n}{2} \right\rfloor, \left\lceil \frac{n}{2} \right\rceil$.

Proposition 3.6.8

Let $n = \dim V$, TFAE:

(1) T is nilpotent;

(2) p_T is a power of x ;

(3) $f_T = x^n$;

(4) $T^n = 0$.

Proof. Trivial. □

The nilpotent matrices and diagonalizable matrices are somehow "independent": If A is both nilpotent and diagonalizable, then $A = 0$.

In light of this idea, we present the following theorem:

Theorem 3.6.9 (Jordan decomposition)

Let $T \in L(V)$, $n = \dim V$, and F is algebraically closed. There exists unique $D, N \in L(V)$ s.t. $T = D + N$, where D diagonalizable and N nilpotent, and $DN = ND$.

Moreover there exists $f, g \in F[x]$ s.t. $D = f(T), N = g(T)$.

Proof. For $A \in F^{n\times n}$, $\exists P \in \mathrm{GL}_n(F)$ s.t. $P^{-1}AP = J$, where J is a Jordan matrix.

It's clear that we can find $J_1 + J_2 = J$ with J_1 diagonal, J_2 nilpotent (just exactly as what you think), and we can check $J_1J_2 = J_2J_1$.

Hence $A = PJ_1P^{-1} + PJ_2P^{-1}$ has the desired properties. But now it's hard to prove the uniqueness, so we'll use another approach.

Let $p_T = \prod_{i=1}^{k}(x - c_i)^{r_i}$, and the elementary divisors $q_i = (x - c_i)^{r_i}$. Let $V_i = \ker(q_i(T))$, so $V = \bigoplus_{i=1}^{k} V_i$ is the primary decomposition of T.

> **Claim.** $\exists f \in F[x]$ s.t. $f \equiv c_i (\mathrm{mod}\, q_i)$, $i = 1, 2, \ldots, k$.

(This follows from Chinese Remainder Theorem)

Observe that $f(T)\big|_{V_i} = c_i\,\mathrm{id}_{V_i}$ in this case, thus $f(T)$ is diagonalizable. Since $(T - f(T))\big|_{V_i}$ is nilpotent, so $N = T - f(T)$ is nilpotent. This proves the existence part and the polynomial part.

Now it's easy to prove the uniqueness: If $T = D + N = D' + N'$, since D, N are polynomials of T, D and D' is commutative, hence can be simutaneously diagonalized.

Note that $D - D' = N - N'$ is both diagonalizable and nilpotent, thus it must be 0. (N, N' is commutative, so $(N + N')^{m+m'} = 0$, here $N^m = N'^{m'} = 0$) □

Since this theorem requires the field to be algebraically closed, if T is in a smaller field, we wonder whether D and N is in that field.

Let $A \in \mathbb{R}^{n\times n}$, and $A = D + N$ be its Jordan decomposition. We'll prove that $D, N \in \mathbb{R}^{n\times n}$. By taking conjugates,

$$A = D + N \implies A = \overline{D} + \overline{N}.$$

It's clear that $\overline{D} + \overline{N}$ is also a Jordan decomposition of A, so we must have $D = \overline{D}$, which means $D \in \mathbb{R}^{n\times n}$.

In fact when $\mathbb{R}$ is replaced by any perfect field F, this property still holds. To prove this we need to introduce the semisimple maps.

§3.7 Semisimple transformations

As we've already seen, the "diagonalizable" property depends on the base fields, thus next we'll generalize the concepts of "diagonalizable".

Definition 3.7.1. Let $T \in L(V)$,

- We say T is **simple**(or irreducible) if V has no nontrivial T-invariant subspaces.
- We say T is **semisimple**(or totally reducible) if each T-invariant subspace $W \subset V$ there exists T-invariant subspace Z, s.t. $V = W \oplus Z$.

Obviously simple maps are always semisimple.

Proposition 3.7.2

Let T be a simple linear operator, then $\forall \alpha \in V \backslash \{0\}$, α is a cyclic vector of T.

Lemma 3.7.3

Let $T \in L(V)$.

- If T is semisimple, $V' \subset V$ is T-invariant, then $T_{V'}$ is semisimple.
- If $V = \bigoplus_{i=1}^k V_i$ s.t. T_{V_i} semisimple, then T is semisimple as well.

Proof. Suppose $W \subset V'$ is a T-invariant subspace. Since T is semisimple, $\exists Z \subset V$ s.t. $V = W \oplus Z$, and Z is T-invariant.

Let $Z' = Z \cap V'$, we claim that $V' = Z' \oplus W$.

Clearly $W \cap Z' = \{0\}$ and $W + Z' \subset V'$. For all $v \in V'$, $\exists w \in W, z \in Z$ s.t. $v = w + z$, since $v, w \in V'$, $z = v - w \in V'$ as well, which means $z \in Z'$.

For the second part, (We can assmue $k = 2$, but here we won't use it).

Let $W \subset V$ be a T-invariant subspace. Since T_{V_i} is semisimple, $\exists Z_i \subset V_i$ s.t.

$$V_i = \left(\left(W + \sum_{j=1}^{i-1} V_j \right) \cap V_i \right) \oplus Z_i.$$

Let $Z = \bigoplus_{i=1}^k Z_i$, we claim that $Z \oplus W = V$. If $w \in W \cap Z$, then $w = z_1 + \cdots + z_k$,

$$z_k = w - z_1 - \cdots - z_{k-1} \in Z_k \cap (W + V_1 + \cdots + V_{k-1}) = \{0\}.$$

Thus $z_k = 0$, similarly $z_{k-1} = \cdots = z_1 = 0 = w$.

Note that $W + \sum_{i=1}^j V_i \subset W \oplus \sum_{i=1}^j Z_i$ for all $j = 1, \ldots, k$, so $V = W \oplus Z$. $\square$

Corollary 3.7.4

Let $T \in L(V)$, T is semisimple $\iff$ there exists a T-invariant decomposition $V = \bigoplus_{i=1}^k V_i$ s.t. each T_{V_i} is simple.

Theorem 3.7.5

Let $T \in L(V)$.

- T simple $\iff$ f_T is a prime polynomial;
- T semisimple $\iff$ p_T has no multiple factors.

Proof. T simple $\implies$ T cyclic $\implies$ $f_T = p_T$, so we only need to prove p_T is a prime.

Otherwise $p_T = gh$,

$$0 = p_T(T) = g(T)h(T),$$

So either $g(T)$ or $h(T)$ is not inversible. Thus $\ker(g(T)) \neq \{0\} \implies \ker(g(T)) = V \implies g(T) = 0$, contradiction!

If T is not simple, $\exists W \subset V$, W is T-invariant nontrivial subspace, so $f_T = f_{T_W} \cdot f_{T_{V/W}}$ is not a prime.

T semisimple $\implies$ $\exists V_i$, $V = \bigoplus_{i=1}^k V_i$, such that T_{V_i} is simple $\implies$ $p_{T_{V_i}}$ is prime.

$$p_T = \operatorname{lcm}(p_{T_{V_1}}, \dots, p_{T_{V_k}})$$

has no multiple factors.

Conversely if p_T has no multiple factors, consider the primary cyclic decomposition of T :

$$V = \bigoplus_i W_i, \quad f_{T_{W_i}} \text{ primary.}$$

Since p has no multiple factors, $f_{T_{W_i}} = p_{T_{W_i}}$ is prime polynomial.

Hence T_{W_i} simple $\implies$ T semisimple. □

Corollary 3.7.6

When F is an algebraically closed field:

- T simple $\iff$ $\dim V = 1$.
- T semisimple $\iff$ T is diagonalizable.

This corollary means that "semisimple" is indeed the equivalent description of "diagonalizable" in the algebraic closure.

Note that whether p_T has multiple factors or not does not change under *perfect* field extensions. So "semisimple" is a more general property (it stays the same under more transformations).

Recall that:

Definition 3.7.7 (Perfect fields)**.** If for all prime polynomials $p \in F[x]$, p has no multiple roots in $\overline{F}$, we say F is a **perfect field**.

Finite fields, fields with charcter 0 and algebraically closed fields are always perfect fields.

We can check that when F is perfect, $f \in F[x]$ has no multiple factors iff f has no multiple factors in $\overline{F}[x]$.

Now we can generalize the Jordan decomposition:

Theorem 3.7.8 (Jordan decomposition)
Let $T \in L(V)$, $n = \dim V$, and F is perfect. There exists unique $S, N \in L(V)$ s.t. $T = S + N$, where S semisimple and N nilpotent, and $SN = NS$.

Moreover there exists $f, g \in F[x]$ s.t. $S = f(T), N = g(T)$.

To prove this generalized version, we need the following observation:

Proposition 3.7.9
Let F be a perfect field, $A \in F^{n\times n}$ is semisimple iff A is diagonalizable in $\overline{F}^{n\times n}$.

Proof. A semisimple $\iff$ p_A has no multiple factors in $F[x]$
$\iff$ p_A has no multiple roots in $\overline{F}[x]$
$\iff$ p_A is the product of different monic polynomials of degree 1
$\iff$ A is diagonalizable in $\overline{F}^{n\times n}$. □

Proposition 3.7.10
Let F be a perfect field, $a \in \overline{F}$. Then $a \notin F \iff$ exists an automorphism σ s.t. $\sigma\big|_F = \mathrm{id}_F$, i.e. $\sigma \in \mathrm{Gal}(\overline{F}/F)$ but $\sigma(a) \neq a$.

Remark 3.7.11 — This proof is beyond the scope of this class, but the idea is similar to the conjugate operation on $\mathbb{C}/\mathbb{R}$.

Now we prove the Jordan decomposition:

Proof. Let $A = S + N$ is the Jordan decomposition on $\overline{F}^{n\times n}$. Then by applying σ on this equation,

$$A = \sigma(S) + \sigma(N)$$

holds for all $\sigma \in \mathrm{Gal}(\overline{F}/F)$. Since $\sigma(S)$ is also diagonalizable, $\sigma(N)$ is nilpotent, as σ is an automorphism. So by the uniqueness of Jordan decomposition, $\sigma(S) = S, \sigma(N) = N$.

This implies $S, N \in F^{n\times n}$. □

§3.8 Bonus section

Starting from Galois groups mentioned above, let

$$\mathrm{Aut}(E/F) := \{\sigma \in \mathrm{Aut}(E) \mid \sigma|_F = \mathrm{id}_F\}$$

be the automorphism group of field extension E/F.

Example 3.8.1

Let $F = \mathbb{Q}$, $E = \mathbb{Q}(\sqrt{2})$, then $\sigma : a + b\sqrt{2} \mapsto a - b\sqrt{2}$ is in $\operatorname{Aut}(E/F)$.

If $E = \mathbb{Q}(\sqrt[3]{2})$, if $\sigma \in \operatorname{Aut}(E/F)$, then $\sigma(\sqrt[3]{2})$ is a root of $x^3 - 2 \implies \sigma = \mathrm{id}$. Thus E/F is not a *Galois extension.*

When E/F is a Galois extension, we write $\operatorname{Gal}(E/F) = \operatorname{Aut}(E/F)$.

In the history, this concept is used to solve polynomial equations.

Let $f \in \mathbb{Q}[x]$, let $x_1, \dots, x_n$ be all roots of f. Consider $E = \mathbb{Q}(x_1, \dots, x_n)$, and define $\operatorname{Gal}(f) = \operatorname{Gal}(E/\mathbb{Q})$. Back in the times of Galois, the concept of field haven't been developed yet, so what he did is to consider the bijections between the roots of f.

Galois discovered that f has radical solutions if and only if the group $\operatorname{Gal}(f)$ has a property, and he named it "solvable". Since all the subgroups of S_4 are solvable, thus if $\deg f \leq 4$, f always has radical solutions, but $A_5 < S_5$ is not solvable, so polynomials of degree greater than 4 may not have radical solutions.

One of the ultimate goal of modern algebra is to comprehend the group $\operatorname{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$.

A tool developed for this goal is *group representation.* A representation of a group G is a homomorphism $\varphi : G \to \mathrm{GL}(V)$. Since $\mathrm{GL}(V)$ is something people knows very well, so when the elements of an abstract group G is viewed as linear maps, it's easier to discover more properties of G.

When $G = \operatorname{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$, the representation is called a *Galois representation.* Even one dimensional Galois representations are very nontrivial.

Midterm exam QAQ

§4 Inner product spaces

In this section we always assume the base field to be $\mathbb{R}$ or $\mathbb{C}$.

§4.1 Inner product

Definition 4.1.1 (Inner product)**.** Let V be a vector space, an **inner product** on V is a function $\langle \cdot, \cdot \rangle : V \times V \to F$, $(\alpha, \beta) \mapsto \langle \alpha, \beta \rangle$ such that:

- $\langle \alpha + \beta, \gamma \rangle = \langle \alpha, \gamma \rangle + \langle \beta, \gamma \rangle$, $\langle c\alpha, \beta \rangle = c \langle \alpha, \beta \rangle$, i.e. the linearity of the first entry.
- $\langle \alpha, \beta \rangle = \overline{\langle \beta, \alpha \rangle}$. This implies the *conjugate linearity* of the second entry.
- $\alpha \neq 0 \implies \langle \alpha, \alpha \rangle > 0$.

Remark 4.1.2 — The reason why we require the conjugate property is that we want to make the inner product positive definite: otherwise $\langle i\alpha, i\alpha \rangle = i^2 \langle \alpha, \alpha \rangle$.

The finite dimensional real inner product space is called **Euclid space**, and finite dimensional complex inner product space is called **unitary space**.

In fact the definition of inner space is related to the order in real numbers, so this is not a pure algebraic structure.

Example 4.1.3

Let $V = F^{n\times 1}$. Let $\alpha = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \beta = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$, define $\langle \alpha, \beta \rangle = \sum_{j=1}^n x_j \overline{y_j} = \alpha^t \overline{\beta}$ to be the **standard inner product**.

Denote $\beta^* = \overline{\beta^t}$, then $\langle \alpha, \beta \rangle = \beta^* \alpha$.

Similarly when $V = F^{m\times n}$, $\langle A, B \rangle = \sum_{j,k} A_{jk} \overline{B_j k} = \operatorname{tr}(B^*A) = \operatorname{tr}(AB^*)$.

Definition 4.1.4 (Hermite matrices)**.** Let $A \in F^{n\times n}$, we say A is **Hermite** if $A^* = A$, and **anti-Hermite** if $A^* = -A$.

When $F = \mathbb{R}$, Hermite matrices are symmetrical matrices.

If we also have $\forall X \in F^{n\times 1} \backslash \{0\}$, $X^*AX > 0$, then we say A is **positive definite**.

Example 4.1.5

For all $Q \in \mathrm{GL}_n(F)$, $A = Q^*Q$ is positive definite.

Proposition 4.1.6

Let V be an n dimensional vector space, let $\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\}$ be a basis. For $\alpha, \beta \in V$, let $X = [\alpha]_{\mathcal{B}}$, $Y = [\beta]_{\mathcal{B}}$.

- If $A \in F^{n\times n}$ is positive definite, then

$$\langle \alpha, \beta \rangle = Y^*AX = \sum_{j,k=1}^n A_{kj} x_j \overline{y_k}$$

  is an inner product.

- For any inner product $\langle \cdot, \cdot \rangle$, there exists a unique positive definite matrix A such that the above relations holds.

Proof. It's clear that Y^*AX is an inner product. (just check the definition)

For the latter part, let $A_{kj} = \langle \alpha_j, \alpha_k \rangle$, so A must be unique. By the conjugate linearity of inner product, so A constructed above indeed satisfies desired condition:

$$\langle \alpha, \beta \rangle = \left\langle \sum_{j=1}^n x_j \alpha_j, \sum_{k=1}^n y_k \alpha_k \right\rangle = \sum_{j,k=1}^n x_j \overline{y_k} \langle \alpha_j, \alpha_k \rangle$$

$\square$

Let $T : V \to W$ be an injective linear map, and $\langle \cdot, \cdot \rangle_0$ is an inner product on W. Then T induces an inner product on V:

$$\langle \alpha, \beta \rangle = \langle T\alpha, T\beta \rangle_0, \quad \alpha, \beta \in V.$$

Since T injective, so T actually realizes V as a subspace of W, this inner product is just the original one restricted on the subspace.

Example 4.1.7
Let $V = W = F^{n\times 1}$, $\langle \cdot, \cdot \rangle_0$ is the standard inner product, $Q \in \mathrm{GL}_n(F)$. Then

$$\langle \alpha, \beta \rangle = \langle Q\alpha, Q\beta \rangle_0 = \beta^*(Q^*Q)\alpha.$$

With an inner product, we can assign a "length" to each vector: $\|\alpha\| = \sqrt{\langle \alpha, \alpha \rangle}$. It's clear that:

$$\|c\alpha\| = |c|\|\alpha\|, \quad \|\alpha\| > 0, \forall \alpha \neq 0.$$

Proposition 4.1.8 (Polarization identity)
When $F = \mathbb{R}$,

$$\langle \alpha, \beta \rangle = \frac{1}{4}\left(\|\alpha + \beta\|^2 - \|\alpha - \beta\|^2\right).$$

When $F = \mathbb{C}$,

$$\langle \alpha, \beta \rangle = \frac{1}{4}\sum_{k=1}^{4} i^k \|\alpha + i^k\beta\|^2.$$

Remark 4.1.9 — This means, *inner product is totally determined by length function.*

Proposition 4.1.10 (Cauchy-Schwarz inequality)

$$|\langle \alpha, \beta \rangle| \leq \|\alpha\|\|\beta\|.$$

The equality holds iff α, β linearly dependent.

Proof. WLOG $\alpha, \beta \neq 0$. Let $\gamma = \beta - \frac{\langle \beta, \alpha \rangle}{\|\alpha\|^2}\alpha$ be the orthogonal projection of β on $\alpha^\perp$.

We can check that $\langle \alpha, \gamma \rangle = 0$, so

$$0 \leq \|\gamma\|^2 = \langle \gamma, \beta \rangle = \|\beta\|^2 - \frac{\langle \alpha, \beta \rangle^2}{\|\alpha\|^2},$$

which gives the desired inequality, equality iff $\gamma = 0$ iff α, β linearly dependent. □

Proposition 4.1.11 (Triangle inequality)

$$\|\alpha + \beta\| \leq \|\alpha\| + \|\beta\|.$$

Proof. Square both sides and use Cauchy-Schwarz. □

This means our "length" function is in fact a **norm**.

§4.2 Orthogonality

Definition 4.2.1 (Orthogonality)**.** Let $\alpha, \beta \in V$, we say $\alpha \perp \beta$ if $\langle \alpha, \beta \rangle = 0$.

We can introduce "angles" as well:

Definition 4.2.2 (Angles)**.** When $F = \mathbb{R}$, for $\alpha, \beta \in V \backslash \{0\}$, define

$$\angle(\alpha, \beta) = \arccos \frac{\langle \alpha, \beta \rangle}{\|\alpha\| \|\beta\|} \in [0, \pi].$$

We can see that $\alpha \perp \beta \iff \angle(\alpha, \beta) = \frac{\pi}{2}$.

When $F = \mathbb{C}$, the angle above can be complex, which doesn't make sense, so we won't talk about the angle in $\mathbb{C}$.

Definition 4.2.3 (Orthonormal basis)**.** Let V be an inner product space, let $S \subset V$ be a subset,

- If the vectors in S are pairwise orthogonal, we say S is an **orthogonal set**. Futhermore, if $\|\alpha\| = 1$ for all $\alpha \in S$, we say S is **orthonormal**.
- If S is a basis as well, then S is called an **orthogonal basis** or **orthonormal basis**, respectively.

Note that an orthogonal set can contain the zero vector.

Proposition 4.2.4

If S is an orthogonal set, and $0 \notin S$, then S is linearly independent.

Proof. Let $S = \{\alpha_1, \dots, \alpha_n\}$, if

$$\sum_{j=1}^{n} c_j \alpha_j = 0,$$

take the inner product with α_j for $j = 1, \dots, n$ we get $c_j = 0$, $\forall j$. □

Proposition 4.2.5

If $S = \{\alpha_1, \dots, \alpha_m\}$ is an orthogonal set, then:

$$\left\| \sum_{j=1}^{m} \alpha_j \right\|^2 = \sum_{j=1}^{m} \|\alpha\|^2, \quad \left\langle \sum_{j=1}^{m} x_j \alpha_j, \sum_{j=1}^{m} y_j \alpha_j \right\rangle = \sum_{j=1}^{m} x_j \overline{y_j} \|\alpha_j\|^2.$$

Now we will prove the existence of orthogonal basis, We'll start from a basis $\{\beta_1, \beta_n\}$ to construct an orthogonal basis, and this process is called *Schmidt orthogonalization.*

Theorem 4.2.6 (Schmidt orthogonalization)

Let V be an n-dimensional inner product space, $\{\beta_1, \dots, \beta_n\}$ is a basis of V. Then there exists a unique orthogonal basis $\{\alpha_1, \dots, \alpha_n\}$, such that

$$(\beta_1, \dots, \beta_n) = (\alpha_1, \dots, \alpha_n) N,$$

where N is an upper triangular matrix with diagonal entries equal to 1.

Proof. The idea is to "project" β_j to the subspace spanned by $\beta_1, \dots, \beta_{j-1}$, and let α_j be the orthogonal part.

By induction, let $\beta_1 = \alpha_1$.

$$\alpha_j = \beta_j - \sum_{k=1}^{j-1} \frac{\langle \beta_j, \alpha_k \rangle}{\|\alpha_k\|^2} \alpha_k.$$

It's obvious that $\alpha_j \perp \alpha_k, \forall k = 1, \dots, j-1$, and $\operatorname{span}\{\alpha_1, \dots, \alpha_j\} = \operatorname{span}\{\beta_1, \dots, \beta_j\}$.

Thus $\{\alpha_1, \dots, \alpha_n\}$ is the desired orthogonal basis.

As for the uniqueness, actually α_j can be solved from β_j's: clearly $\alpha_1 = \beta_1$, and

$$\langle \beta_j, \alpha_k \rangle = N_{jk} \langle \alpha_k, \alpha_k \rangle \implies N_{jk} = \frac{\langle \beta_j, \alpha_k \rangle}{\|\alpha_k\|^2} \implies \alpha_j = \beta_j - \sum_{k=1}^{j-1} \frac{\langle \beta_j, \alpha_k \rangle}{\|\alpha_k\|^2} \alpha_k.$$

So α_j is uniquely determined by β_j's. □

Remark 4.2.7 — The above orthogonal basis can be converted to an orthonormal basis $\{\alpha_1', \dots, \alpha_n'\}$ s.t. N' is an upper triangular matrix with positive diagonal entries.

Corollary 4.2.8

Let $S \subset V \backslash \{0\}$ be orthogonal(-normal), then S can be extended to an orthogonal(-normal) basis.

Proposition 4.2.9

Let $S = \{\alpha_1, \dots, \alpha_m\} \subset V \backslash \{0\}$ be an orthogonal set, then for all $\beta \in \operatorname{span} S$ we have:

$$\beta = \sum_{k=1}^{m} \frac{\langle \beta, \alpha_k \rangle}{\|\alpha_k\|^2} \alpha_k.$$

Proposition 4.2.10 (Bessel's inequality)

Conditions as above, then $\forall \beta \in V$,

$$\sum_{k=1}^{m} \frac{|\langle \beta, \alpha_k \rangle|^2}{\|\alpha_k\|^2} \leq \|\beta\|^2.$$

Equality iff $\beta \in \operatorname{span} S$.

Proof. Complete S to an orthogonal basis, by previous propositions, the rest is trivial. □

Let $S \subset V$, define $S^\perp := \{\alpha \in V \mid \alpha \perp \beta, \forall \beta \in S\}$, $S^\perp$ is a vector space and $S^\perp = \operatorname{span}(S)^\perp$.

Proposition 4.2.11

Let V be a finite dimensional inner product space, $W \subset V$ is a subspace, we have $\dim W + \dim W^\perp = \dim V$.

Proof. Take an orthogonal basis B_1 of W, and complete it to an orthogonal basis B of V, then we claim that $B_2 := B\backslash B_1$ is a basis of $W^\perp$. Hence the conclusion follows. $\square$

This means we always have $W \oplus W^\perp = V$.

The orthogonal completion is similar to the annihiltor we studied last semester, in fact, when we view $\langle \cdot, \beta\rangle$ as a function $f_\beta \in V^*$, $f_\beta \in S^0 \iff \beta \in S^\perp$. (Note that the inner product is linear with respect to only the first entry)

This process induces a map $\phi : V \to V^*$ by $\beta \mapsto f_\beta$. It's clear that ϕ is conjugate-linear. So ϕ is a linear map between *real* vector space $V \to V^*$, i.e. $\phi \in \mathrm{Hom}_{\mathbb{R}}(V, V^*)$. thus $\ker\phi = \{0\}$ implies ϕ is an isomorphism on $\mathbb{R}$, so ϕ is a bijection, $\phi(S^\perp) = S^0$.

For $E \subset V^*$, then $E^0 \subset V$, this corresponds to $\phi(S)^0 = S^\perp$. Indeed, $\alpha \in \phi(S)^0 \iff \forall \beta \in S, \langle\alpha, \beta\rangle = 0 \iff \alpha \in S^\perp$. Hence

$$\dim_{\mathbb{C}} W^\perp = 2\dim_{\mathbb{R}} \phi(W^\perp) = 2\dim_{\mathbb{R}} W^0 = \dim_{\mathbb{C}} W^0.$$

The above proposition can be derived directly by $\dim W + \dim W^0 = \dim V$.

We can also get $W = (W^0)^0 = \phi(W^\perp)^0 = (W^\perp)^\perp$.

Definition 4.2.12 (Orthogonal projection)**.** Since $V = W \oplus W^\perp$, for all $\alpha \in V$, there exists unique $\beta \in W, \gamma \in W^\perp$ s.t. $\alpha = \beta + \gamma$. Let $p_W : V \to W$ be the map $\alpha \mapsto \beta$, this is called the **orthogonal projection** from V to W.

§4.3 Adjoint maps

Let $\{\alpha_1, \dots, \alpha_m\}$ be an orthonormal basis of W, then $p_W(\beta) = \sum_{j=1}^m \langle\beta, \alpha_j\rangle\,\alpha_j$. So p_W is a linear map. Moreover $p_W + p_{W^\perp} = \mathrm{id}_V$, $p_W^2 = p_W$. By our geometry intuition, $p_W\beta = \arg\min_\alpha\|\alpha - \beta\|$, this fact is useful in funtional analysis.

Recall that for $T \in L(V)$, $T^t \in L(V^*)$, then what's the map $\phi^{-1}\circ T^t \circ \phi$? Unluckily it's not T, but another map denoted by T^*, the **adjoint map** of T. Keep in mind that T^* depends on the inner product.

$$\begin{array}{ccc} V^* & \xrightarrow{T^t} & V^* \\ \phi\uparrow & & \phi\uparrow \\ V & \xrightarrow{T^*} & V \end{array}$$

Since $T^t\circ\phi = \phi\circ T^* \iff \langle T\alpha, \beta\rangle = \langle\alpha, T^*\beta\rangle$, $\forall \alpha, \beta \in V$, so T^* can be described as the map satisfying this relation.

Proposition 4.3.1

When $\mathcal{B}$ is an orthonormal basis, we have $[T^*]_{\mathcal{B}} = [T]^*_{\mathcal{B}}$.

Proof. Let $\mathcal{B} = \{\alpha_1, \dots, \alpha_n\}$, then $\phi(\mathcal{B})$ is the dual basis of $\mathcal{B}$. i.e. $\phi(\alpha_j)(\alpha_k) = \delta_{jk}$.

Hence $[T^t]_{\phi(\mathcal{B})} = [T]^t_{\mathcal{B}}$. Let $[T^*]_{\mathcal{B}} = A$, then

$$T^*\alpha_k = \sum_{j=1}^n A_{jk}\alpha_j \implies \phi(T^*\alpha_k) = \sum_{j=1}^n \overline{A_{jk}}\phi(\alpha_j).$$

So $[T^t]_{\phi(\mathcal{B})} = \overline{A}$, which completes the proof. $\square$

Proposition 4.3.2

$\ker(T^*) = \mathrm{Im}(T)^\perp$, $\mathrm{Im}(T^*) = \ker(T)^\perp$. $(cT+U)^* = \overline{c}T^* + U^*$, $(TU)^* = U^*T^*$, $T^{**} = T$.

This means the map $T \mapsto T^*$ is a conjugate anti-automorphism of $L(V)$, and it's an involution.

If $T^* = T$, then we say T is **self-adjoint**, and if $T^* = -T$, we say T is **anti self-adjoint**.

Let $F = \mathbb{C}$, T is self-adjoint iff iT is anti self-adjoint. Like a function can be written as a sum of odd and even functions, $\forall T \in L(V)$, there exists unique self-adjoint T_1, T_2 s.t. $T = T_1 + iT_2$. In fact, $T_1 = \frac{T+T^*}{2}, T_2 = \frac{T-T^*}{2i}$.

Let $\mathcal{B}$ be an orthonormal basis, obviously T self-adjoint $\iff$ $[T]_\mathcal{B}$ Hermite.

Example 4.3.3

Let $W \subset V$, p_W be the orthogonal projection. then p_W is self-adjoint as we can choose an orthonormal basis $\mathcal{B}$, such that $[p_W]_\mathcal{B} = \mathrm{diag}\{I_k, 0\}$, where $k = \dim W$.

Let V, W be inner product spaces, we'll study the linear maps $T : V \to W$ which preserves the inner products, i.e.

$$\langle \alpha, \beta \rangle_V = \langle T\alpha, T\beta \rangle_W .$$

If T is an isomorphism, then we say T is the isomorphism between inner product spaces.

Proposition 4.3.4

T preserves inner product $\iff$ T is an isometry, i.e. preserves length.

In particular, isometry is always injective implies that inner product presering maps are always injective.

Proof. Trivial by polarization identity. □

Proposition 4.3.5

Let V, W be finite dimensional inner product spaces, $\dim V = \dim W$, $T \in \mathrm{Hom}(V, W)$, the followings are equivalent:

(1) T preserves inner product;

(2) T is an isomorphism between inner product spaces;

(3) T maps all the orthonormal bases in V to orthonormal bases in W ;

(4) T maps *one* orthonormal basis in V to a orthonormal basis in W.

Proof. It's clear that (1) $\implies$ (2) $\implies$ (3) $\implies$ (4), since T injective $\implies$ T is an isomorphism of vector space.

As for (4) $\implies$ (1), just expand everything using this orthonormal basis. □

Corollary 4.3.6
Inner product spaces with same dimensions are always isomorphic as inner product spaces.

Recall the positive definite matrices we defined earlier, we can also define *positive definite maps*: Let T be a *self-adjoint map*, if

$$\forall \alpha \in V \backslash \{0\}, \quad \langle T\alpha, \alpha \rangle > 0,$$

then we say T is positive definite.

The reason why we require T self-adjoint is that,

$$\langle T\alpha, \alpha \rangle = \langle \alpha, T\alpha \rangle = \overline{\langle T\alpha, \alpha \rangle} \implies \langle T\alpha, \alpha \rangle \in \mathbb{R}.$$

so we can talk about "positive" safely.

§4.4 Orthogonal maps and Unitary maps

Definition 4.4.1 (Orthogonal maps)**.** Let V be a real inner product space, the automorphisms of V (as inner product space) are called **orthogonal maps**, denoted the set by $\mathrm{O}(V)$.

When V is a complex inner product space, we use **unitary maps** and $\mathrm{U}(V)$ instead.

Proposition 4.4.2
Let V be an inner product space,

$$T \in \mathrm{O}(V) \iff T^* = T^{-1}.$$

Proof.

$$T \in \mathrm{O}(V) \iff \langle \alpha, \beta \rangle = \langle T\alpha, T\beta \rangle = \langle \alpha, T^*T\beta \rangle, \quad \forall \alpha, \beta \in V.$$

This also holds for $\mathrm{U}(V)$. □

Proposition 4.4.3
Let $A \in \mathbb{R}^{n\times n}$, TFAE:

- $A^tA = I_n$;
- The column (row) vectors of A form an orthonormal basis of $\mathbb{R}^n$.

Proof. Since A maps the standard basis to the column vectors of A, so the conclusion follows immediately (use A^t to get the row vectors). □

Let $\mathrm{O}(n) = \{A \in \mathbb{R}^{n\times n} \mid A^tA = I_n\}$, and $\mathrm{U}(n) = \{A \in \mathbb{C}^{n\times n} \mid A^*A = I_n\}$. We can see that $A^tA = I_n \implies \det(A) = \pm 1$, and $A^*A = I_n \implies |\det(A)| = 1$.

Let $\mathrm{SO}(n) = \{A \in \mathrm{O}(n) \mid \det A = 1\}$, and $\mathrm{SU}(n) = \{A \in \mathrm{U}(n) \mid \det A = 1\}$. In the language of groups, $\mathrm{SO}(n)$ has only 2 coset in $\mathrm{O}(n)$, while the structure of the cosets of $\mathrm{SU}(n)$ in $\mathrm{U}(n)$ look like S^1.

Example 4.4.4

Let's look at some low dimensional orthogonal groups. $\mathrm{O}(1)=\{1,-1\}$, $\mathrm{SO}(1)=\{1\}=\mathrm{SU}(1)$, $\mathrm{U}(1)=\{z \mid |z|=1\}$.

The group $\mathrm{SO}(2)$ is the rotations of $\mathbb{R}^2$:

$$\mathrm{SO}(2)=\left\{\begin{pmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{pmatrix} \mid \theta\in[0,2\pi)\right\}.$$

While $\mathrm{O}(2)$ also consisting of reflections.

$$\mathrm{SU}(2)=\left\{\begin{pmatrix}z & -\overline{w}\\ w & \overline{z}\end{pmatrix} \mid z,w\in\mathbb{C},|z|^2+|w|^2=1\right\}.$$

In fact these groups are *lie groups*, which means they have the structure of differential manifolds. It's clear that $\mathrm{U}(1)\simeq\mathrm{SO}(2)\simeq S^1$, and we can see $\mathrm{SU}(2)\simeq S^3$.

Theorem 4.4.5 (QR-decomposition)

Any invertible matrix A can be uniquely decomposed to $Q\cdot R$, where $Q\in\mathrm{O}(n)$, R is an upper triangular matrix with positive diagonal entries. When $F=\mathbb{C}$, $\mathrm{O}(n)$ is replaced by $\mathrm{U}(n)$.

Proof. This is essentially Schmidt orthogonalozation. □

Corollary 4.4.6 (Ivasawa decomposition, KAN decomposition)

For all $A\in\mathrm{GL}_n(\mathbb{R})$, it has a unique decomposition $A=A_kA_aA_n$, $A_k\in\mathrm{O}(n)$, A_a is diagonal, A_n is upper triangular matrix with diagonal entries 1. It also holds for $\mathbb{C}$.

Let $\mathcal{B},\mathcal{B}'$ be orthonormal bases of V, $T\in L(V)$. We know that $[T]_{\mathcal{B}'}=P^{-1}[T]_{\mathcal{B}}P$ for some $P\in\mathrm{GL}(V)$. By orthogonality, P must be an orthogonal matrix, wich means $P^t=P^{-1}$.

Definition 4.4.7. Let $A,B\in\mathbb{R}^{n\times n}$, we say they are **orthogonally similar** if $A=P^{-1}BP$ for some $P\in\mathrm{O}(n)$. The name is changed to **unitarily similar** for complex matrices.

Theorem 4.4.8 (Schur triangularization theorem)

Let $F=\mathbb{C}, T\in L(V)$. There exists an orthonormal basis $\mathcal{B}$, such that $[T]_{\mathcal{B}}$ is upper triangular.

Proof. Recall that T is triangulable (which is always true in $\mathbb{C}$) iff there exists a T-invariant flag $\{0\}=W_0\subset W_1\subset\cdots\subset W_n=V$. We can take an orthonormal basis s.t. $W_k=\mathrm{span}\{\alpha_1,\ldots,\alpha_k\}$.

Obviously T is upper triangular under this basis. □

§4.5 Normal maps

Recall that we say two matrices A,B are orthogonally similar, if there exist $P\in\mathrm{O}(n)$ s.t. $B=P^{-1}AP$. Again, we want to find the "simpliest" matrix in each orthogonal equivalent class.

Let $T\in L(V)$ be a linear map, if there exists an orthonomal basis of V s.t. $[T]_{\mathcal{B}}$ is diagonal, then we say T is orthogonally (or unitarily) diagonalizable.

Definition 4.5.1 (Normal maps)**.** Let V be an inner product space, $T \in L(V)$. If $TT^* = T^*T$, then we say T is a **nomal map**.

It turns out that these concepts has close relations:

Theorem 4.5.2

Let V be a finite dimentional inner product space,

- If $F = \mathbb{R}$, then T orthogonally diagonalizable $\iff$ T self-adjoint;
- If $F = \mathbb{C}$, then T unitarily diagonalizable $\iff$ T normal.

Lemma 4.5.3

Let $F = \mathbb{C}$, then T normal $\iff$ there exists self-adjoint commutative maps T_1, T_2 s.t. $T = T_1 + iT_2$.

Proof. If $T = T_1 + iT_2$, then $T^* = T_1 - iT_2$, so $T^*T = TT^*$ since T_1, T_2 commutative.

On the other hand, let $T_1 = \frac{T+T^*}{2}$, $T_2 = \frac{T-T^*}{2i}$. We can check that T_1, T_2 self-adjoint and are commutative. □

Proof of Theorem 4.5.2. For the " $\implies$ " part, let $\mathcal{B}$ be an orthonomal basis such that $[T]_{\mathcal{B}} = \operatorname{diag}\{c_1, \ldots, c_n\}$. Then we have

$$[T^*]_{\mathcal{B}} = [T]_{\mathcal{B}}^* = \operatorname{diag}\{\bar{c}_1, \ldots, \bar{c}_n\}.$$

If $F = \mathbb{R}$, then $T^* = T$, i.e. T self-adjoint.

If $F = \mathbb{C}$, clearly $TT^* = T^*T$, so T is normal.

As for the other part, we need a lemma first.

Lemma 4.5.4

Let V be a f.d. inner product space, $T \in L(V)$. If $W \subset V$ is a T-invariant space, then $W^\perp$ is T^*-invariant.

Proof of the lemma. For all $\alpha \in W^\perp$,

$$0 = \langle \alpha, T\beta \rangle = \langle T^*\alpha, \beta \rangle, \quad \forall \beta \in W.$$

Thus $T^*\alpha \in W^\perp$. □

Corollary 4.5.5

If T is self-adjoint, $W \subset V$ is T-invariant will imply $W^\perp$ is also T-invariant, so T is semisimple.

Lemma 4.5.6

Let V be a f.d. inner product space, $T \in L(V)$ is self-adjoint. We must have $f_T \in \mathbb{R}[x]$, and it can be decomposed to products of polynomials of degree 1.

In particular, $\sigma(T) \subset \mathbb{R}$.

Proof. Let $f_T = \prod_{j=1}^n (x - c_j), c_j \in \mathbb{C}$.

Let $\mathcal{B}$ be an orthonomal basis of V, then $A := [T]_\mathcal{B}$ is Hermite. Let X be a nonzero vector s.t. $AX = c_j X$, then

$$c_j X^* X = X^* A X = (AX)^* X = \bar{c}_j X^* X.$$

So $c_j \in \mathbb{R}$, and we're done. □

Lemma 4.5.7

If T is a self-adjoint map, then all the eigenspaces of T are pairwise orthogonal.

Proof. Let $c_1, c_2 \in \mathbb{R}$ be two eigenvalues of T. Let $\alpha \in V_{c_1}, \beta \in V_{c_2}$.

$$c_1 \langle \alpha, \beta \rangle = \langle c_1 \alpha, \beta \rangle = \langle T\alpha, \beta \rangle = \langle \alpha, T\beta \rangle = \bar{c}_2 \langle \alpha, \beta \rangle = c_2 \langle \alpha, \beta \rangle .$$

Since $c_1 \neq c_2$, we must have $\alpha \perp \beta$, as desired. □

Returning back to Theorem 4.5.2, when T is self-adjoint, let $\sigma(T) = \{c_1, \dots, c_r\}$.

Claim 4.5.8. $V = \bigoplus_{i=1}^r V_{c_i}$, i.e. T is diagonalizable.

Let $W = \bigoplus_{i=1}^r V_{c_i}$, if $W^\perp \neq \{0\}$, then $W^\perp$ is T-invariant.

When $F = \mathbb{C}$, then $T_{W^\perp}$ has eigenvectors; when $F = \mathbb{R}$, then $T_{W^\perp}$ is self-adjoint, so it must have a eigenvector (by lemma).

Since V_{c_i} are pairwise orthogonal, so we can actually take an orthonomal basis of V_{c_i} to get an orthonomal basis of V. Hence T is orthogonally diagonalizable.

Now for the case when T is normal, let T_1, T_2 be self-adjoint maps s.t. $T = T_1 + iT_2$. Since T_1, T_2 commute, the proof is nearly identical to the simutaneously diagonalizable property.

Let $V = \bigoplus_{i=1}^r V_{c_i}$ be the eigenspace decomposition of T_1. Note that V_{c_i} are also T_2-invariant.

Since $(T_2)_{V_{c_i}}$ self-adjoint, $(T_2)_{V_{c_i}}$ is unitarily diagonalizable. Therefore we can concatenate those basis to get a basis of V, and T_1, T_2 are both diagonal under this basis. □

There's another proof of "$\Longleftarrow$" part of the theorem:

Proposition 4.5.9

Let V be an inner product space, $T \in L(V)$ normal. Let $W \subset V$ be a T-invariant space, then $W^\perp$ is T-invariant, and W is T^*-invariant.

Proof. Take an orthonomal basis of $W, W^\perp$, so $A := [T]_\mathcal{B}$ normal.

Since W is T-invariant, $A = \begin{pmatrix} B & C \\ 0 & D \end{pmatrix}$. Note that:

$$AA^* = \begin{pmatrix} BB^* + CC^* & * \\ * & * \end{pmatrix}, \quad A^*A = \begin{pmatrix} B^*B & * \\ * & * \end{pmatrix}.$$

As A normal, $BB^* + CC^* = B^*B$, by looking at the trace of both sides, we get $\operatorname{tr}(CC^*) = 0 \implies C = 0$, the conclusion follows. □

Corollary 4.5.10

Let $A \in \mathbb{C}^{n\times n}$ be an upper triangular martix, then A normal $\iff$ A diagonal.

Proposition 4.5.11

Let T be a normal map, then the eigenspaces of T are pairwise orthogonal.

Proof. Let $\alpha \in V_{c_1}, \beta \in V_{c_2}$, since $\operatorname{span}\{\beta\}$ is a T-invariant space, so $T^*\beta \in \operatorname{span}\{\beta\}$,

$$\langle T^*\beta, \beta\rangle = \langle \beta, T\beta\rangle = \overline{c}_2 \langle \beta, \beta\rangle .$$

Thus $T^*\beta = \overline{c}_2\beta$.

$$c_1 \langle \alpha, \beta\rangle = \langle T\alpha, \beta\rangle = \langle \alpha, T^*\beta\rangle = c_2 \langle \alpha, \beta\rangle .$$

But $c_1 \neq c_2$, we have $\alpha \perp \beta$. □

When $F = \mathbb{C}$: Let $W = \bigoplus_{i=1}^r V_{c_i}$. Since $W^\perp$ is T-invariant, so when $W \perp\neq \{0\}$, T must have eigenvalues in $W^\perp$, contradiction!

Now we've proved that V_{c_i} are pairwise orthogonal, so T is unitarily diagonalizable.

Proposition 4.5.12

Let V be a complex inner product space, $T \in L(V)$ normal,

- T self-adjoint $\iff$ $\sigma(T) \subset \mathbb{R}$;
- T anti self-adjoint $\iff$ $\sigma(T) \subset i\mathbb{R}$;
- T unitary $\iff$ $\sigma(T) \subset \{z : |z| = 1\}$.

Proof. Take an orthonomal basis s.t. $[T]_{\mathcal{B}}$ diagonal. The rest is trivial. □

§5 Bilinear forms

Let V be a finite dimensional vector space, $\dim V = n$.

Definition 5.0.1. Let $F = \mathbb{C}$, we say a function $f : V \times V \to V$ is a **semi bilinear form** if:

- $f(c_1\alpha + \beta, \gamma) = c_1 f(\alpha, \gamma) + f(\beta, \gamma)$;
- $f(\alpha, c_1\beta + \gamma) = \overline{c}_1 f(\alpha, \beta) + f(\alpha, \gamma)$.

Let $\operatorname{Form}(V)$ denote the (semi) bilinear forms on (complex) real vector space V.

For $f \in \operatorname{Form}(V)$, fix a basis $\mathcal{B} = \{\alpha_1, \ldots, \alpha_n\}$ of V, let $[f]_{\mathcal{B}} \in F^{n\times n}$ be the matrix

$$([f]_{\mathcal{B}})_{jk} = f(\alpha_k, \alpha_j).$$

which is called **the matrix of f under $\mathcal{B}$**.

For $\alpha = \sum_{k=1}^n x_k\alpha_k, \beta = \sum_{j=1}^n y_j\alpha_j \in V$. It's clear that

$$f(\alpha, \beta) = \sum_{j,k=1}^n x_k\overline{y}_j f(\alpha_k, \alpha_j) = \sum_{j,k=1}^n x_k\overline{y}_j([f]_{\mathcal{B}})_{jk} = [\beta]^*_{\mathcal{B}}[f]_{\mathcal{B}}[\alpha]_{\mathcal{B}}.$$

From this we know that the map $\mathrm{Form}(V) \to F^{n\times n}, f \mapsto [f]_{\mathcal{B}}$ is a linear isomorphism. Since if $[f]_{\mathcal{B}} = 0$, then $f(\alpha, \beta) = 0$ for all $\alpha, \beta \in V$. Thus it's injective. Obviously it's surjective and linear, so

$$\dim \mathrm{Form}(V) = n^2$$

Example 5.0.2

Let $A \in F^{n\times n}$. Let $f \in \mathrm{Form}(F^{n\times 1})$ be

$$f(X, Y) = Y^* AX, \quad \forall X, Y \in F^{n\times 1}.$$

Let $\mathcal{B}$ be the standard basis of F, it's clear that $[f]_{\mathcal{B}} = A$.

Proposition 5.0.3

Let $\mathcal{B}' = \{\alpha'_1, \ldots, \alpha'_n\}$ be another basis of V, $P \in \mathrm{GL}_n(F)$ satisfies

$$(\alpha'_1, \ldots, \alpha'_n) = (\alpha_1, \ldots, \alpha_n)P.$$

Then $[f]_{\mathcal{B}'} = P^*[f]_{\mathcal{B}}P$.

Proof. Since $[\alpha]_{\mathcal{B}} = P[\alpha]_{\mathcal{B}'}$, just plug this into the definition of $[f]_{\mathcal{B}}$, the rest is trivial. □

Definition 5.0.4. Let $f \in \mathrm{Form}(V)$.

- When $F = \mathbb{R}$, if $\forall \alpha, \beta \in V$ we have $f(\alpha, \beta) = f(\beta, \alpha)$, then we say f is symmetrical (also called Hermite);
- When $F = \mathbb{C}$, if $\forall \alpha, \beta \in V$ we have $f(\alpha, \beta) = \overline{f(\beta, \alpha)}$, we say f is Hermite.

Proposition 5.0.5

When $F = \mathbb{C}$, f Hermite $\iff f(\alpha, \alpha) \in \mathbb{R}, \forall \alpha \in V$.

Proof. For the " $\Longleftarrow$ " direction, consider $f(\alpha + \beta, \alpha + \beta) \in \mathbb{R}$. Expanding we'll get $f(\alpha, \beta) + f(\beta, \alpha) \in \mathbb{R}$, i.e.

$$f(\alpha, \beta) + f(\beta, \alpha) = \overline{f(\alpha, \beta)} + \overline{f(\beta, \alpha)}.$$

Replace α with $i\alpha$, we get

$$f(\alpha, \beta) - f(\beta, \alpha) = -\overline{f(\alpha, \beta)} + \overline{f(\beta, \alpha)}.$$

Combining two equations we get the conclusion. □

Definition 5.0.6. Let $f \in \mathrm{Form}(V)$ be an Hermite form. If $\forall \alpha \in V \backslash \{0\}$, $f(\alpha, \alpha) > 0$, we say f is **positive definite**.

Similarly we can define negative definite and semi positive definite.

Note that a positive definite Hermite form is nothing but an inner product.

§5.1 Positive define matrices

In this section we'll dig deeper into properties of positive definite matrices.

It's clear that if a matrix A is positive definite, then A is inversible, and P^*AP is also positive definite. In particular, P^*P is positive definite.

Theorem 5.1.1 (Cholesky decomposition)
Let $A \in F^{n\times n}$ be a positive definite matrix, there exists a unique upper triangular matrix R with positive diagonal entries s.t. $A = R^*R$.

Proof. Consider the inner product $f(X,Y) = Y^*AX$. Let the standard inner product on V be $f_0(X,Y) = Y^*X$.

Since inner product spaces with same dimensions are isomorphic, so there exists a matrix $R \in \mathrm{GL}_n(F)$, such that
$$R : (F^{n\times 1}, f) \to (F^{n\times 1}, f_0), \quad X \mapsto RX$$
is an isomorphism of inner product space, i.e. $f_0(RX, RY) = f(X,Y)$. This is equivalent to $A = R^*R$.

For any $P \in GL_n(F)$, P is also an isomorphism of $(F^{n\times 1}, f) \to (F^{n\times 1}, f_0)$ iff RP^{-1} preserves the inner product f_0, iff $RP^{-1} \in \mathrm{O}(n)$ or $\mathrm{U}(n)$.

By QR decomposition, $R = RP^{-1} \cdot P$, so there must be a unique P s.t. P upper triangular with positive diagonal entries. □

Corollary 5.1.2
A positive definite $\implies \det A > 0$.

Definition 5.1.3. Let $A \in F^{n\times n}$, for $1 \le k \le n$, define
$$\Delta_k(A) := \det(A_{1\le i\le k}^{1\le j\le k})$$
be the **leading principal minor**.

Theorem 5.1.4
Let $A \in F^{n\times n}$ be an Hermite matrix. Then A positive definite $\iff \Delta_k(A) > 0, k = 1, \dots, n$.

Lemma 5.1.5 (LU decomposition)
Let F be *any field*. For $A \in \mathrm{GL}_n(F)$, the followings are equivalent:

- $\Delta_k(A) \neq 0, k = 1, \dots, n$;
- $A = LU$, where L lower triangular, and U upper triangular with diagonal entries 1.

Proof. On one hand, Let L_k, U_k be the top-left $k \times k$ submatrix of L, U, since L, U inversible, L_k, U_k inversible. By the triangular property, $\Delta_k(A) = \det(L_kU_k) \neq 0$.

On the other hand, it's sufficient to prove:

$$\exists N \text{ strictly upper triangular}, A(N+I_n) \text{ lower triangular}$$

Let A_k be the k-th leading principal submatrix of A, and $\alpha_{k+1}, \beta_{k+1} \in F^{n\times 1}$ the $(k+1)$-th column of A, N.

Now compute the first k rows of the $(k+1)$-th column of $A(N+I)$, which is equal to $A_k\beta'_{k+1} + \alpha'_{k+1}$, where $\alpha'_{k+1}, \beta'_{k+1}$ is the first k entries of $\alpha_{k+1}, \beta_{k+1}$.

Since A_k inversible, $\exists \beta'_{k+1}$ s.t. $A_k\beta'_{k+1} + \alpha'_{k+1} = 0$.

Hence these β'_{k+1} forms a strictly upper triangular matrix N, as desired. □

Proof of the theorem. Let A be an Hermite matrix, if A positive definite, then $\det A \geq 0$.

Let A_k be the upper left $k \times k$ submatrix of A. For $X \in F^{k\times 1}\backslash\{0\}$, we have

$$X^* A_k X = \begin{pmatrix} X \\ 0 \end{pmatrix}^* A \begin{pmatrix} X \\ 0 \end{pmatrix} > 0.$$

Hence A_k positive definite, $\det A_k = \Delta_k(A) \geq 0$.

Conversely, by our lemma let $A = LU$, let $D = (U^*)^{-1}L$, $A = U^*DU$.

Hence A Hermite $\implies$ D Hermite. Moreover D is lower triangular, so D is diagonal.

Some computation yields that $A_k = U_k^* D_k U_k$. Therefore

$$\Delta_k(A) \geq 0 \implies \det(U_k^* D_k U_k) \geq 0 \implies \det D_k \geq 0.$$

From this we deduce that all the diagonal entries of D are positive, so D positive definite $\implies$ A positive definite. □

§5.2 Bilinear forms on inner product spaces

Let V be an inner product space, given a basis of V, recall that there are two linear isomorphism:

$$\mathrm{Form}(V) \to F^{n\times n}, f \mapsto [f]_{\mathcal{B}} \quad L(V) \to F^{n\times n}, T \mapsto [T]_{\mathcal{B}}$$

Hence we can define a map $\mathrm{Form}(V) \to L(V)$ by composing these two isomorphism. Denote this map by $f \mapsto T_f$. It seems like this map also depends on the choice of the basis, but in fact it's independent as long as $\mathcal{B}$ is orthonormal!

Let $\mathcal{B}'$ be another orthonormal basis, then $[T_f]_{\mathcal{B}'} = P^{-1}[T_f]_{\mathcal{B}}P$, while $[f]_{\mathcal{B}'} = P^*[f]_{\mathcal{B}}P$, but P is orthogonal (or unitary), so $P^{-1} = P^*$, i.e. T_f doesn't change under the new basis.

Since T_f do not depend on the basis, thus we wonder whether we can define this map intrinsically.

Proposition 5.2.1

For all $T \in L(V)$, T induces a (semi) bilinear form $f(\alpha, \beta) = \langle T\alpha, \beta\rangle$. We claim that this map $\mathcal{F}$ gives an isomorphism of $L(V)$ and $\mathrm{Form}(V)$.

Proof. Clearly $\mathcal{F}$ is injective:

$$\langle T\alpha, \beta\rangle = 0, \forall \beta \implies T\alpha = 0,$$

thus $\ker \mathcal{F} = \{0\}$.

By dimenional reasons $\mathcal{F}$ must be an isomorphism. □

By considering $\mathcal{F}^{-1}$, we get an one-to-one map $f \mapsto T_f$ s.t.

$$f(\alpha, \beta) = \langle T_f \alpha, \beta \rangle .$$

We'll see that this definition coincide with the initial one. In fact it's sufficient to prove $[T_f]_{\mathcal{B}} = [f]_{\mathcal{B}}$, which is just a bunch of computation ;)

Remark 5.2.2 — We can construct T_f explicitly from f:
The inner product gives a conjugate linear isomorphism

$$\Phi : V \to V^*, \quad \Phi(\alpha)(\beta) = \langle \beta, \alpha \rangle = \overline{\langle \alpha, \beta \rangle}.$$

Similarly, $f \in \mathrm{Form}(V)$ gives a conjugate linear map

$$\Phi_f : V \to V^*, \quad \Phi_f(\alpha)(\beta) = \overline{f(\alpha, \beta)}.$$

Then $T = \Phi^{-1} \circ \Phi_f$ is the desired linear map:

$$\langle T\alpha, \beta \rangle = \overline{\Phi(T\alpha)(\beta)} = \overline{\Phi_f(\alpha)(\beta)} = f(\alpha, \beta).$$

Hence all the properties of linear maps can be carried over to the forms, and vice versa (using the matrix representation).

Corollary 5.2.3
Let $F = \mathbb{C}$, $T \in L(V)$, T self-adjoint iff $\langle T\alpha, \alpha \rangle \in \mathbb{R}, \forall \alpha \in V$.

Proof. T self-adjoint iff f Hermite iff $f(\alpha, \alpha) \in \mathbb{R}$. □

Corollary 5.2.4
Let $f \in \mathrm{Form}(V)$.

- If f Hermite, there exists an orthonormal basis of V s.t. $[f]_{\mathcal{B}}$ is real diagonal.
- If $F = \mathbb{C}$, there exists an orthonormal basis such that $[f]_{\mathcal{B}}$ upper triangular.

§5.3 Spectral decomposition

Theorem 5.3.1 (Spectral decomposition of normal maps)
Let $T \in L(V)$ be a self-adjoint map (or normal map in complex field), let $\sigma(T) = \{c_1, \ldots, c_k\}$, $P_i \in L(V)$ are the projection onto V_{c_i}. Then for any $f \in F[x]$, we have

$$f(T) = \sum_{i=1}^{k} f(c_i) P_i.$$

In particular, $T = \sum_{i=1}^{k} c_i P_i$.

Proof. Consider the orthogonal direct sum

$$V = \bigoplus_{i=1}^{k} V_{c_i},$$

since previously we've proven that T is orthogonally diagonalizable (or unitarily diagonalizable).

Using this decomposition, the conclusion is somewhat trivial. □

Corollary 5.3.2

Each P_i is a polynomial of T.

Proof. Take $f_i \in F[x]$ s.t. $f_i(c_i) = \delta_{ij}$. Then $f_i(T) = \sum_{j=1}^{k} f_i(c_j)P_j = P_i$. □

Using simialr technique, we can consider functions other than polynomials of T, defined by $\phi(T) = \sum_{i=1}^{k} \phi(c_i)T$. By Lagrange interpolation, we can always find a polynomial p s.t. $p(c_i) = \phi(c_i)$ for all $c_i \in \sigma(T)$.

Example 5.3.3

If T semi positive definite normal matrix, $\sigma(T) \subset [0, +\infty)$, so we can define $\sqrt{T} = \sum_{i=1}^{k} \sqrt{c_i} P_i$.

Proposition 5.3.4

T self-adjoint (normal) $\implies$ $\phi(T)$ self-adjoint (normal); $\sigma(\phi(T)) = \phi(\sigma(T))$.

Proof. Note that T and $\phi(T)$ are diagonal matrices under orthonormal basis of V_{c_i}. □

Theorem 5.3.5

Let $T \in L(V)$ be semi positive definite.

- $\sqrt{T}$ semi positive definite, and $\sqrt{T}^2 = T$.
- T positive definite $\iff$ $\sqrt{T}$ positive definite.
- If $S \in L(V)$ semi positive definite and $S^2 = T$, then $S = \sqrt{T}$.

Proof. Since $[\sqrt{T}]_{\mathcal{B}} = \operatorname{diag}(\sqrt{c_1} I_{d_1}, \dots, \sqrt{c_k} I_{d_k})$, the first two statements are trivial.

Let $\sigma(S) = \{s_1, \dots, s_r\}$, $V_i = \ker(S - s_i \,\mathrm{id})$. Since S self-adjoint, $V = \bigoplus_{i=1}^{r} V_i$.

For any $\alpha \in V_i, T\alpha = S^2\alpha = s_i^2\alpha$, thus $V_i \subset \ker(T - s_i^2 \,\mathrm{id})$. Since $s_i \geq 0$, $\sqrt{T} = S$. □

Note that T^*T is always positive definite, so we can consider $\sqrt{T^*T}$. We call the eigen-values of $\sqrt{T^*T}$ **singular values** of T.

In some sense, $\sqrt{T^*T}$ is a semi positive approximation of T:

Lemma 5.3.6
For any $\alpha \in V$, $\|T\alpha\| = \|\sqrt{T^*T}\alpha\|$. In particular, $\ker T = \ker \sqrt{T^*T}$.

Proof. Let $N = \sqrt{T^*T}$,

$$\|N\alpha\|^2 = \langle N\alpha, N\alpha\rangle = \left\langle N^2\alpha, \alpha\right\rangle = \langle T^*T\alpha, \alpha\rangle = \langle T\alpha, T\alpha\rangle = \|T\alpha\|^2.$$

□

Theorem 5.3.7 (Polar decomposition)
Let $T \in L(V)$,

(1) There exists $U \in L(V)$ orthogonal or unitary, $N \in L(V)$ semi positive definite, $T = UN$.

(2) We must have $N = \sqrt{T^*T}$.

(3) T invertible iff N positive definite, in this case U is unique.

Remark 5.3.8 — This is a generalization of $z = re^{i\theta}$ in $\mathbb{C}$. That's where the name comes from.

Proof. If (1) holds, then

$$T^* = NU^* \implies T^*T = NU^*UN = N^2 \implies N = \sqrt{T^*T}.$$

Since T, N are semi positive definite, T invertible iff T positive definite. Now we must have $U = TN^{-1}$, which is unique.

To prove (1), when T invertible, let N, U as above, by our lemma,

$$\|U\alpha\| = \|TN^{-1}\alpha\| = \|\alpha\|$$

Thus U is orthogonal or unitary.

When T is not invertible, $\ker T = \ker N$, thus $\exists U_1 : \mathrm{Im}(N) \to \mathrm{Im}(T)$ s.t. $T = U_1N$. (Just take $N\alpha \mapsto T\alpha$)

Moreover U_1 is an isomorphism of inner product space: $\|U_1N\alpha\| = \|T\alpha\| = \|N\alpha\|$. So U_1 preserves inner product and hence injective. By dimension reasons, U_1 must be an isomorphism.

Now we can take an arbitary isomorphism $U_2 : \mathrm{Im}(N)^\perp \to \mathrm{Im}(T)^\perp$, $U := U_1 \oplus U_2$ is the desired map. □

Looking back at the singular values, consider the image of unit sphere $S \subset V$ under T, $N(S)$ is an ellipsoid:

$$N(S) = \left\{\sum_{i=1}^n c_i x_i \alpha_i : \sum_{i=1}^n x_i^2 = 1\right\}.$$

Since $T = UN$, U is a rotation, so $T(S)$ is also an ellipsoid, whose axes lengths are $2c_i$, where c_i are singular values of T.

Corollary 5.3.9 (Singular value decomposition, SVD)

Let $A \in F^{n\times n}$, then there exists decomposition $A = U_1 D U_2$, where D is a diagonal matrix with non-negative entries, U_1, U_2 are orthogonal or unitary matrices.

Proof. Consider the polar decomposition $A = UN$, let $N = PDP^{-1}$, where P orthogonal or unitary, D non-negative diagonal. Thus we can take $U_1 = UP, U_2 = P^{-1}$.

Note that the diagonal entries of D is precisely the singular value of A. □

Corollary 5.3.10

Let $T \in L(V)$, then T map *some* orthogonal basis to another orthogonal basis.

Proof. Let $T = UN$ be the polar decomposition. Let $\alpha_1, \dots, \alpha_n$ be an orthonormal basis s.t. N diagonal, then

$$T\alpha_i = UN\alpha_i = c_i U\alpha_i,$$

obviously $c_i U\alpha_i$ consititude an orthogonal basis. □

§5.4 Further on normal maps

For $\theta \in \mathbb{R}$, let Q_θ be the rotation of angle θ. The main goal of this section is to prove:

Theorem 5.4.1

Let V be a finite dimensional real inner product space, $T \in L(V)$ normal. There exists an orthonormal basis $\mathcal{B}$ s.t.

$$[T]_{\mathcal{B}} = \operatorname{diag}(a_1, \dots, a_l, r_1 Q_{\theta_1}, \dots, r_m Q_{\theta_m}),$$

where $a_i \in \mathbb{R}, r_j > 0, \theta_j \in (0, \pi)$.

Let's look at a corollary of this theorem first:

Corollary 5.4.2

If T orthogonal, then

$$[T]_{\mathcal{B}} = \operatorname{diag}(I_{l_1}, -I_{l_2}, Q_{\theta_1}, \dots, Q_{\theta_m}).$$

Proof. Applying the theorem, since each block is orthogonal, $a_i = \pm 1$, $r_j = 1$. □

This gives us a comprehension of rotations in higher dimensional spaces.

Here we'll present multiple proofs to emphasize some intermediate result.

Proposition 5.4.3

Let T be a normal map, if $W \subset V$ is T-invariant, then T_W is also normal.

Proof. First note that $W, W^\perp$ are T^*-invariant. For $\alpha, \beta \in W$, we have

$$\langle (T_W)^*\alpha, \beta\rangle = \langle \alpha, T_W\beta\rangle = \langle \alpha, T\beta\rangle = \langle T^*\alpha, \beta\rangle .$$

Thus $(T_W)^* = T^*_W$. The conclusion follows. $\square$

Proposition 5.4.4

Let T be a normal map, there exists an orthogonal decomposition $V = \bigoplus_{i=1}^k V_i$, such that each V_i is T-invariant, and T_{V_i} simple.

Proof. Note that if W is T-invariant, then $W^\perp$ is also T-invariant. By induction and the previous proposition this is trivial. $\square$

Therefore to prove Theorem 5.4.1, we only need to prove the case when T is simple.

Proof of Theorem 5.4.1. WLOG $\dim V > 1$.

Since T simple $\implies f_T \in \mathbb{R}[x]$ prime, thus $\deg f_T = 2$, $\dim V = 2$ and $f_T = (x-c)(x-\bar{c})$.

Take any orthonormal basis $\mathcal{B} = \{\alpha_1, \alpha_2\}$, let $r = |c|$, $A = r^{-1}[T]_{\mathcal{B}}$. Clearly A normal and $\sigma(A) = \{r^{-1}c, r^{-1}\bar{c}\}$, so A is unitarily similar to $\mathrm{diag}(r^{-1}c, r^{-1}\bar{c})$, A is unitary.

Moreover A is a real matrix so A orthogonal, and $\det A = 1$, thus $A = Q_\theta, \theta \in [0, 2\pi]$.

At last by T has no eigenvector, and we can change α_2 to $-\alpha_2$, so we can require $\theta \in (0, \pi)$. $\square$

Proposition 5.4.5

Let $T \in L(V)$, then $\ker(T)^\perp = \mathrm{im}(T^*), \mathrm{im}(T)^\perp = \ker(T^*)$.

Proof. Trivial, just some computation. $\square$

Proposition 5.4.6

Let $T \in L(V)$, $\sigma(T^*) = \overline{\sigma(T)}$,

$$\forall c \in \sigma(T), \quad \dim\ker(T - cI) = \dim\ker(T^* - \bar{c}I).$$

Proof. By the previous proposition,

$$\dim\ker(T - cI) = n - \dim\mathrm{im}(T^* - \bar{c}I) = \dim\ker(T^* - \bar{c}I)$$

which also implies $\sigma(T) = \overline{\sigma(T^*)}$. $\square$

Proposition 5.4.7

If T normal, then $\ker(T - cI) = \ker(T^* - \bar{c}I)$.

Proof. Let $W = \ker(T - cI)$, T^*_W is just $(c\,\mathrm{id}_W)^* = \bar{c}\,\mathrm{id}_W$. Thus $W \subset \ker(T^* 0 \bar{c}I)$, by dimensional reasons they must be equal. $\square$

Proposition 5.4.8

Let T be a normal map, $f, g \in F[x]$ coprime $\implies \ker(f(T)) \perp \ker(g(T))$.

Proof. Since $g(T)^* = \overline{g}(T^*)$, $g(T)$ is normal, thus $\ker(g(T))^\perp = \operatorname{im}(g(T))$.

Let $W = \ker(f(T))$, let $a, b \in F[x]$ s.t. $af + bg = 1$, so $a(T)f(T) + b(T)g(T) = \mathrm{id}_V$. Restrict this equation to W, we get $b(T)_W g(T)_W = \mathrm{id}_W$, hence $W \subset \operatorname{im}(g(T))$. □

Proposition 5.4.9

Let T be a normal map,

- The primary decomposition of T are orthogonal decomposition;
- The cyclic decomposition of T can be orthogonal.

Proof. The first one is trivial by previous proposition.

For cyclic decomposition, we proceed by induction on $\dim V$.

Let $\alpha_1 \in V$ s.t. $p_{\alpha_1} = p_r$, then $(R\alpha_1)^\perp$ are T-invariant, use induction hypo on it and we're done. □

Remark 5.4.10 — This means the primary cyclic decomposition of T can also be orthogonal.

This gives the second proof of Theorem 5.4.1:

Proof. WLOG T normal and primary cyclic, then p_T is primary, and T normal $\implies$ T semisimple, so p_T has no multiple factors, thus p_T prime, which proves the result. □

Next we present the third proof:

Proposition 5.4.11

If $A, B \in \mathbb{R}^{n\times n}$ are unitarily similar, then they are orthogonally similar.

Lemma 5.4.12 (QS decomposition)

For any unitary matrix U, $U = QS$ where Q real orthogonal, S unitary and symmetrical. Moreover $\exists f \in \mathbb{C}[x]$ s.t. $S = f(U^tU)$.

Proof. Let $\sigma(U^tU) = \{c_1, \dots, c_k\}$. We can take a polynomial $f \in \mathbb{C}[x]$ s.t. $f(c_i)^2 = c_i$.

Since U is unitary, $|c_i| = 1 \implies |f(c_i)| = 1$.

Let $S = f(U^tU)$, we claim that S unitary and $S^2 = U^tU$.

Let $U^tU = P\operatorname{diag}(c_1, \dots, c_k)P^{-1}$, where P is unitary, then $S = P\operatorname{diag}(f(c_1), \dots, f(c_k))P^{-1}$ is unitary, and clearly $S^2 = U^tU$.

Let $Q = US^{-1}$, then Q unitary. Since S symmetrical, $S^{-1} = S^* \implies \overline{S^{-1}} = S^t = S$,

$$\overline{Q}Q^{-1} = \overline{U}SSU^{-1} = \overline{U}U^tUU^{-1} = I_n.$$

Hence $\overline{Q} = Q$, Q is real orthogonal. □

Return to the original proposition. Let A, B be real matrices unitarily similar, let $B = UAU^{-1}$, taking the conjuate we get

$$UAU^{-1} = \overline{U}AU^t \implies U^tUA = AU^tU.$$

Let $U = QS$, then $AS = SA$. We have

$$B = UAU^{-1} = QSAS^{-1}Q^{-1} = QAQ^{-1}.$$

Therefore A, B are orthogonally similar. □

Corollary 5.4.13

Let A, B be normal matrices, TFAE:

(1) A, B are unitarily similar (or orthogonally similar);

(2) A, B are similar;

(3) $f_A = f_B$.

Proof. We only need to prove (3) $\implies$ (1).

When $F = \mathbb{C}$, A, B are unitarily similar to diagonal matrices D_1, D_2. Since $f_A = f_B$, D_1, D_2 only differ by a permutation, hence unitarily similar.

When $F = \mathbb{R}$, by the previous proposition and proof for $\mathbb{C}$, we get the result. □

The third proof of Theorem 5.4.1 is to factorize $f_T \in \mathbb{R}[x]$ and use the above corollary.

At last we prove another property of normal maps:

Proposition 5.4.14

Let A be a normal matrix, then A^* is a complex polynomial of A.

Proof. Use the spectral decomposition. □

§6 Bilinear forms

In this section we study the bilinear forms on generic fields. Let $M^2(V)$ denote all the bilinear forms on V.

For $f \in M^2(V)$, Let $(f(\alpha_i, \alpha_j))_{ij}$ be the matrix of f under basis $\{\alpha_i\}$. (Note that this differs by a transpose with previous section)

Obviously $M^2(V) \to F^{n\times n}$ by $f \mapsto [f]_{\mathcal{B}}$ is a linear isomorphism.

Proposition 6.0.1

Let $\mathcal{B}, \mathcal{B}'$ be two basis, P is the transformation matrix between them, for all $f \in M^2(V)$ we have $[f]_{\mathcal{B}'} = P^t[f]_{\mathcal{B}}P$.

Proof. Trivial. □

If $A = P^t BP$ for some $P \in \mathrm{GL}(V)$, we say A, B are **congruent**.
A bilinear form will induce two linear maps $V \to V^*$, namely L_f, R_f:

$$L_f(\alpha)(\beta) = R_f(\beta)(\alpha) = f(\alpha, \beta).$$

Proposition 6.0.2
For any basis $\mathcal{B}$, we have $\operatorname{rank} L_f = \operatorname{rank} R_f = \operatorname{rank}[f]_{\mathcal{B}}$. This number is called the rank of f, denoted by $\operatorname{rank} f$.

If $\operatorname{rank} f = n$, we say f is non-degenrate, this is equivalent to L_f invertible or R_f invertible.

§6.1 Some special bilinear forms

Definition 6.1.1. For $f \in M^2(V)$,

- If $f(\alpha, \beta) = f(\beta, \alpha), \forall \alpha, \beta \in V$, then we say f is **symmetrical**.
- If $f(\alpha, \beta) = -f(\beta, \alpha), \forall \alpha, \beta \in V$, we say f is **anti-symmetrical**.
- If $f(\alpha, \alpha) = 0, \forall \alpha \in V$, we say f is **alternating**.

We denote the above functions by $S^2(V), A^2(V), \Lambda^2(V)$.

We can see that $\Lambda^2(V) \subset A^2(V)$, and they are all subspaces of $M^2(V)$.

Proposition 6.1.2
If $\operatorname{char} F \neq 2$, then $A^2(V) = \Lambda^2(V)$, and $M^2(V) = A^2(V) \oplus S^2(V)$.

Proof. Already proved in last semester. □

Proposition 6.1.3
Let $\mathcal{B}$ be any basis of V,

- f symmetrical $\iff$ $[f]_{\mathcal{B}}$ symmetrical;
- f anti-symmetrical $\iff$ $[f]_{\mathcal{B}}$ anti-symmetrical;
- f alternating $\iff$ $[f]_{\mathcal{B}}$ anti-symmetrical and the diagonal entries are all zero.

Definition 6.1.4 (Quadratic forms)**.** Let $q : V \to F$ be a function, we say q is a **quadratic form** if there exists $f \in M^2(V)$ s.t.

$$q(\alpha) = f(\alpha, \alpha), \quad \forall \alpha \in V.$$

When $V = F^n$, quadratic forms are just a homogenous quadratic polynomial with n variables, i.e.

$$q(X) = X^t AX, \quad A \in F^{n \times n}, X \in F^n.$$

Let $Q(V)$ denote all the quadratic forms on V, it's an F-vector space. By definition there's a surjective linear map $M^2(V) \to Q(V)$ by $\Phi(f)(\alpha) = f(\alpha, \alpha)$.

Proposition 6.1.5

Let char $F \neq 2$,

- The map $\Phi : S^2(V) \to Q(V)$ is an isomorphism.
- Let $q \in Q(V)$, if $f \in S^2(V)$ and $\Phi(f) = q$, then

$$f(\alpha, \beta) = \frac{1}{4}(q(\alpha + \beta) - q(\alpha - \beta)).$$

Proof. The first one can be proved by $\ker(\Phi) = \Lambda^2(V)$ and $M^2(V) = S^2(V) \oplus \Lambda^2(V)$.
The second one is trivial by direct computation. □

From this we can define the matrix of a quadratic form q to be the matrix of the symmetrical bilinear form $\Phi^{-1}(q)$, thus $[q]_{\mathcal{B}}$ is always symmetrical.

Theorem 6.1.6

Let $f \in M^2(V)$,

- If char $F \neq 2$, then $f \in S^2(V) \iff \exists \mathcal{B}$, s.t. $[f]_{\mathcal{B}}$ diagonal;
- $f \in \Lambda^2(V) \iff \exists \mathcal{B}$ s.t. $[f]_{\mathcal{B}}$ is block diagonal with each block being 0 or $\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$.

To prove this theorem, it's sufficient to prove:

Lemma 6.1.7

Let $f \in S^2(V) \cup A^2(V)$, $W \subset V$ is a subspace, let

$$W^\perp = \{\beta \in V \mid f(\alpha, \beta) = 0, \forall \alpha \in W\}.$$

If $f\big|_W$ is non-degenerate, then $V = W \oplus W^\perp$. In this case, let $\mathcal{B}_1, \mathcal{B}_2$ be basis of W, $W^\perp$, and $\mathcal{B} = (\mathcal{B}_1, \mathcal{B}_2)$, we have

$$[f]_{\mathcal{B}} = \operatorname{diag}([f|_W]_{\mathcal{B}_1}, [f|_{W^\perp}]_{\mathcal{B}_2}).$$

Proof. Since $f\big|_W$ non-degenerate, $W \cap W^\perp = 0$. Note that

$$W^\perp = \bigcap_{\alpha \in W} \ker(L_f(\alpha)) = L_f(W)^0.$$

Thus $\dim W^\perp = n - \dim L_f(W) \geq n - \dim W$. This implies that $V = W \oplus W^\perp$.
For the second part, since $f(\alpha, \beta) = 0 \implies f(\beta, \alpha) = 0$, thus the matrix must obey the conclusion. □

Now by induction it's trivial when $n = 1$,

- When $f \in S^2(V)$, WLOG $f \neq 0$, $\exists \alpha$ s.t. $f(\alpha, \alpha) \neq 0$. Let $W = \operatorname{span}\{\alpha\}$, by lemma and induction hypo we're done.

- When $f \in A^2(V)$, there exists α, β s.t. $f(\alpha, \beta) = 1$. Let $W = \text{span}\{\alpha, \beta\}$, similarly by lemma and induction hypo, we're done.

Corollary 6.1.8
For any $q \in Q(V)$, there exists a basis of V s.t. $[q]_{\mathcal{B}}$ diagonal.

The non-degenerate alternating bilinear forms are called **symplectic forms**.

Corollary 6.1.9
If there exists symplectic form f on V, then $\dim V = 2m$ and

$$[f]_{\mathcal{B}} = \begin{pmatrix} 0 & I_m \\ -I_m & 0 \end{pmatrix}$$

for some basis $\mathcal{B}$.

Theorem 6.1.10
Let F be an algebraically closed field, and $\operatorname{char} F \neq 2$. Let $f \in S^2(V)$, there exists a basis $\mathcal{B}$, s.t. $[f]_{\mathcal{B}}$ diagonal and the diagonal entries can only be 0 or 1.

Proof. Use the previous result and multiply some scalars (the root of $x^2 = c$). □

When $F = \mathbb{R}$, using similar technique we can prove the diagonal entries can only be $0, 1$ or -1.

§6.2 Lie algebras

There's a class I missed, so the notes may not be complete.

Definition 6.2.1 (Lie algebra)**.** Let L be a vector space over a field F. Suppose an operation (called **Lie bracket**)

$$L \times L \to L, \quad (x, y) \mapsto [x, y]$$

is given and satisfies:

- (Bilinearity)
$$\begin{cases} [ax + by, z] = a[x, z] + b[y, z], \\ [x, ay + bz] = a[x, y] + b[x, z], \end{cases} \quad \forall x, y, z \in L, a, b \in F;$$
- (Alternativity)
$$[x, x] = 0, \quad \forall x \in L;$$
- (Jacobi identity)
$$[[x, y], z] + [[y, z], x] + [[z, x], y] = 0, \quad \forall x, y, z \in L.$$

Then L is called a **Lie algebra** over F.

The Lie algebra can be viewed as a vectorization of Lie groups, where Lie bracket is the commutator in Lie groups.

Example 6.2.2

On any F-vector space L, one can define a trivial Lie bracket by

$$[x,y]=0, \quad \forall x,y\in L$$

Then L becomes a Lie algebra, called an **abelian Lie algebra**.

We can also define homomorphisms by $\phi([x,y])=[\phi(x),\phi(y)]$.

Definition 6.2.3 (Representation)**.** Let L be a Lie algebra over F. A **representation** of L is a homomorphism $\phi: L\to \mathfrak{gl}(V)$, where V is some finite-dimensional F-vector space.

Example 6.2.4 (Adjoint representation)

Let L be a Lie algebra over F. Define a linear map $\mathrm{ad}: L\to\mathfrak{gl}(L)$ by

$$\mathrm{ad}(x)(y)=[x,y], \quad \forall x,y\in L.$$

We claim that it is a representation, called the **adjoint representation** of L. In fact, it follows from the Jacobi identity that for any $x,y,z\in L$,

$$\begin{aligned}\mathrm{ad}([x,y])(z)&=[[x,y],z]\\&=[x,[y,z]]-[y,[x,z]]\\&=\mathrm{ad}(x)([y,z])-\mathrm{ad}(y)([x,z])\\&=[\mathrm{ad}(x),\mathrm{ad}(y)](z).\end{aligned}$$

Definition 6.2.5 (Subalgebra, ideal, quotient algebra)**.** Let L be a Lie algebra over F.

- If $S,T\subset L$ are subspaces, write
$$[S,T]:=\mathrm{span}\{[x,y]: x\in S, y\in T\}.$$
- A subspace $K\subset L$ is called a **subalgebra** if $[K,K]\subset K$, denoted $K<L$.
- A subspace $I\subset L$ is an **ideal** if $[I,L]\subset I$, denoted $I\lhd L$.
- Let $I\lhd L$. On the quotient space L/I, we introduce the Lie bracket
$$[x+I,y+I]:=[x,y]+I, \quad \forall x,y\in L.$$
Then L/I becomes a Lie algebra, called the **quotient algebra** of L by I.

Example 6.2.6

Let $\phi: L\to L'$ be a homomorphism. Then

$$\ker\phi\lhd L, \quad \mathrm{im}(\phi)\lhd L', \quad \mathrm{im}(\phi)\cong L/\ker\phi.$$

The **center** of L is defined as

$$Z(L):=\{x\in L: [x,y]=0, \forall y\in L\}.$$

We have $Z(L)\lhd L$ and $Z(L)=\ker\mathrm{ad}$.

Definition 6.2.7 (Direct sum)**.** Let $L_1, \dots L_r$ be Lie algebras over F. On the (external) vector space Direct sum $L_1 \oplus \cdots \oplus L_r$ we introduce the Lie bracket

$$[(x_1, \dots, x_r), (y_1, \dots, y_r)] = ([x_1, y_1], \dots, [x_r, y_r])$$

This makes $L_1 \oplus \cdots \oplus L_r$ a Lie algebra, called the **(external) Lie algebra direct sum** of $L_1, \dots, L_r$.

Definition 6.2.8 (Linear Lie algebra)**.** Subalgebras of $\mathfrak{gl}_n(F)$ and $\mathfrak{gl}(V)$ are called **linear Lie algebras**.

We have the following deep result:

Theorem 6.2.9 (Ado-Iwasawa)
All finite-dimensional Lie algebras over F are isomorphic to linear Lie algebras.

Let us introduce some important linear Lie algebras.

Example 6.2.10 (Special linear Lie algebra)
Let
$$\mathfrak{sl}_n(F) = \{x \in \mathfrak{gl}_n(F) : \operatorname{tr}(x) = 0\}, \mathfrak{sl}(V) = \{x \in \mathfrak{sl}(V) : \operatorname{tr}(V) = 0\},$$
where V is a vector space over F. We have $\mathfrak{sl}(V) \lhd \mathfrak{gl}(V)$.

Example 6.2.11 (The Lie algebra $L(V, f)$)
Let V be a finite-dimensional F-vector space, and $f : V \times V \to F$ be a bilinear form. For $x \in \mathfrak{gl}(V)$, we say that f is **invariant under x (in the infinitesimal snese)** if

$$f(xv, w) + f(v, xw) = 0, \quad \forall v, w \in V.$$

This comes from the derivative of Lie groups: Let $L \in \mathrm{GL}(V)$, $g(0) = \mathrm{id}_V$. By taking derivatives at $t = 0$ on
$$f(g(t)v, g(t)w) = f(v, w),$$
we get $f(g'(0)v, w) + f(v, g'(0)w) = 0$.

Let $L(V, f) \subset \mathfrak{gl}(V)$ be the subspace of all $x \in \mathfrak{gl}(V)$ that leave f invariant, we claim that $L(V, f) < \mathfrak{gl}(V)$.

Example 6.2.12

Let's consider 2 special cases of $L(V, f)$:

- Let $V = F^n$, and f be the symmetrical form given by

$$f(v, w) = v^t w, \quad \forall v, w \in F^n.$$

Then $\mathfrak{o}_n(F) := L(F^n, f)$ is called the **orthogonal Lie algebra**. Under the identification $\mathfrak{gl}(F^n) \cong \mathfrak{gl}_n(F)$, we have $\mathfrak{o}_n(F) = \{x \in \mathfrak{gl}_n(F) : x^t + x = 0\}$.

- Let $V = F^{2n}$, and f be the symplectic form given by

$$f(v, w) = v^t \begin{pmatrix} 0 & I_n \\ -I_n & 0 \end{pmatrix} w, \quad \forall v, w \in V.$$

Then $\mathfrak{sp}_{2n}(F) := L(F^{2n}, f)$ is called the **symplectic Lie algebra**.

Suppose $I \lhd L$, and we understand I and L/I, then we understand L (in the rough sense). This motivates the following:

Definition 6.2.13 (Simple Lie algebra, semisimple Lie algebra)**.** Let L be a finite-dimensional Lie algebra over F.

- L is **simple** if it's nonabelian and has no nontrivial ideals.
- L is **semisimple** if it's nonzero and has no nonzero abelian ideal.

Clearly, a simple Lie algebra is semisimple. One of our main purposes is to explain the proof of the following classification theorem:

Theorem 6.2.14

Let L be a finite-dimensional Lie algebra over $\mathbb{C}$.

(1) L is semisimple iff it's isomorphic to the direct sum of finitely many simple Lie algebras.

(2) L is simple iff it's isomorphic to one of the following Lie algebras:

- $\mathfrak{sl}_n(\mathbb{C}), n \geq 2$;
- $\mathfrak{o}_n(\mathbb{C}), n \geq 7$;
- $\mathfrak{sp}_{2n}(\mathcal{C}), n \geq 2$;
- one of the 5 exceptional complex simple Lie algebras, denoted by $\mathfrak{e}_6, \mathfrak{e}_7, \mathfrak{e}_8, \mathfrak{f}_4, \mathfrak{g}_2$ respectively.

Remark 6.2.15 — It can be shown that

$$\mathfrak{o}_2(\mathbb{C}) \cong \mathbb{C}, \quad \mathfrak{o}_3(\mathbb{C}) \cong \mathfrak{sl}_2(\mathbb{C}) = \mathfrak{sp}_2(\mathbb{C}),$$

$$\mathfrak{o}_4(\mathbb{C}) \cong \mathfrak{sl}_2(\mathbb{C}) \oplus \mathfrak{sl}_2(\mathbb{C}), \quad \mathfrak{o}_5(\mathbb{C}) \cong \mathfrak{sp}_4(\mathbb{C}), \quad \mathfrak{o}_6(\mathbb{C}) \cong \mathfrak{sl}_4(\mathbb{C}).$$

§6.3 Abelian, nilpotent and solvable Lie algebras

From now on, let us make the convention that L always denotes a finite-dimensional complex Lie algebra, and V always denoted a complex vector space.

Recall that for $x \in \mathfrak{gl}(V)$, x is said to be semisimple if it's diagonalizable; and nilpotent if $x^r = 0$ for some $r \geq 1$.

Definition 6.3.1 (ad-semisimple and ad-nilpotent)**.** x is **ad-semisimple** if $\mathrm{ad}(x) \in \mathfrak{gl}(V)$ is semisimple. Similarly define ad-nilpotent.

Proposition 6.3.2

Let $L < \mathfrak{gl}(V), x \in L$. If x is semisimple, then it's ad-semisimple. If x is nilpotent, then it's ad-nilpotent.

Remark 6.3.3 — If L is semisimple, then the converse of the proposition holds.

Theorem 6.3.4

A Lie algebra L is abelian iff it consists only of ad-semisimple elements.

For a Lie algebra L, we define two sequences of ideals

$$L = L^0 \supset L^1 \supset \cdots, \quad L = L^{(0)} \supset L^{(1)} \supset \cdots$$

by

$$L^k = [L, L^{k-1}], \quad L^{(k)} = [L^{(k-1)}, L^{(k-1)}].$$

Definition 6.3.5. L is said to be **nilpotent** if $L^k = 0$ for some k. L is said to be **solvable** if $L^{(k)} = 0$ for some k.

It's easy to see $L^k \supset L^{(k)}$, thus nilpotent Lie algebras must be solvable.

Proposition 6.3.6

Let L be a finite-dimensional Lie algebra, TFAE:

- L is semisimple;
- L has no nonzero nilpotent subalgebras;
- L has no nonzero solvable subalgebras.

Theorem 6.3.7 (Engel)

Let $L < \mathfrak{gl}(V)$ be a linear Lie algebra consisting of nilpotent transformations, then the following statement holds:

- There exists $v \in V$ s.t. $Lv = 0$.
- There exists a basis of V s.t. elements in L are all upper triangular.

Remark 6.3.8 — This implies that L is nilpotent.