Contents

Preface

This lecture note is based on *Linear Algebra, 2nd edition* by K. Hoffman and R. Kunze.

Chapter 1

Vector Space

1.1 Matrix

Remind. A **field** is a *good* algebraic structure, which has the addition and the multiplication. Formally, a field $(F, +, \cdot)$, or simply F, is a pair of a set and two operations which is from $F \times F$ to F satisfying the following:

- $(F, +)$ is an **abelian group**, that is, $+$ is commutative, associative, and there is an additional identity 0 and the inverse element $-a$ of a for all $a \in F$.
- $(F^\times, \cdot)$ is also an abelian group, that is, $\cdot$ is commutative, associative, and there is an multiplicational identity 1 and the inverse element a^{-1} of a for all $a \in F^\times$, where $F^\times = F - \{0\}$.
- $+$ and $\cdot$ are *compatible*, which means $\cdot$ is distributing over $+$.

We simply write $a - b := a + (-b)$ and $a/b = ab^{-1}$.

A **matrix** over a field F is a rectangular arrangement of *scalars*, elements of the field F. The space of m by n matrices is denoted as $\mathfrak{M}_{m,n}(F)$.

Example 1.1.1.

- $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$ are well-known(?) fields.
- $\mathfrak{M}_{m,n}(F) \approx F^{mn}$, without the product.
- Matrices do not form a field.

1.1.1 Transpose and Trace

Definition 1.1.1. For every m by n matrix $A \in \mathfrak{M}_{m,n}(F)$, the **transpose** of A is defined as follows:

$$A^\mathsf{T} = (a_{ji})_{n,m}.$$

For every n by n *square* matrix $A \in \mathfrak{M}_{n,n}(F)$, the **trace** of A is defined as follows:

$$\operatorname{tr} A = \sum_{i=1}^{n} a_{ii}.$$

Proposition 1.1.1 (Linearity of transpose and trace)**.** *For every pair of m by n matrices A and $B \in \mathfrak{M}_{m,n}(F)$ and every pair of scalars $a,\ b \in F$,*

$$(aA + bB)^{\mathsf{T}} = aA^{\mathsf{T}} + bB^{\mathsf{T}}.$$

For every pair of n by n square matrices A and $B \in \mathfrak{M}_{n,n}(F)$ and every pair of scalars $a,\ b \in F$,

$$\operatorname{tr}(aA + bB) = a \operatorname{tr} A + b \operatorname{tr} B.$$

Proof. ㅎㅎ. $\square$

Proposition 1.1.2 (Behaviour of transpose and trace)**.**

$$(AB)^{\mathsf{T}} = B^{\mathsf{T}} A^{\mathsf{T}}, \qquad \operatorname{tr} A = \operatorname{tr} A^{\mathsf{T}}, \qquad \operatorname{tr}(AB) = \operatorname{tr}(BA).$$

Proof. Try it! $\square$

1.1.2 Inverse Matrix

Definition 1.1.2. For a *square* matrix $A \in \mathfrak{M}_{n,n}(F)$, if there is another square matrix $B \in \mathfrak{M}_{n,n}(F)$ such that

$$AB = I = BA,$$

then we call $B = A^{-1}$ the(?) **inverse matrix** of A.

Proposition 1.1.3 (Uniqueness of inverse matrix)**.** *The inverse matrix of a matrix A is unique (if exists). This justifies the occurrence of* ***'the'*** *above.*

Proof. Let those be A^{-1} and $\tilde{A}^{-1}$, then

$$\tilde{A}^{-1} = (A^{-1}A)\tilde{A}^{-1} = A^{-1}(A\tilde{A}^{-1}) = A^{-1}.$$

$\square$

Question. If $AB = I$ for two square matrices A and B, what can we say about the invertibility of them? We will solve this problem using a 'function,' which is from and to some vector spaces, defined below.

1.2 Vector Space

Definition 1.2.1. A **vector space** V over F, or simply an ***F*-vector space** V, is a *good* algebraic structure, which has the addition $+: V \times V \to V$ and the F-scalar multiplication $\mathrm{SM}_F: F \times V \to V$. Formally, a field $(V, F, +, \mathrm{SM}_F)$, or simply V, is a pair of a set, a field and two operations satisfying the following:

- $(V, +)$ is an abelian group.
- SM_F and $\cdot_F$ are compatible: $(ab)v = a(bv)$ for every $a, b \in F$ and $v \in V$, and $1v = v$ for all $v \in V$.
- $+$s and SM_F are compatible: $(a+b)v = av + bv$, $a(v+w) = av + aw$ for every $a, b \in F$ and $v, w \in V$.

We simply write $v - w := v + (-w)$.

Example 1.2.1. The following structures are examples of vector space.

- $\mathbf{\{0\}}$ is a vector space over *arbitrary field*, and is called the **trivial space.**
- $\mathbb{R}^n$ and $\mathbb{C}^n$ are vector spaces. In fact, for any field F, F^n is an F-vector space, trivially.
- Hence, a matrix space $\mathfrak{M}_{m,n}$ of m by n matrices over F is a vector space since is *the same with* F^{mn}, and the polynomial space of n-th degree $\mathbf{P}_n[t] = \{\sum_{i=0}^{n} a_i t^i : a_i \in F\}$ is also a vector space, *the same with* F^{n+1}.
- (Field extension) If there are two fields which one is a subfield of another, namely $E \geq F$, then E is a F-vector space, with its addition and multiplication (as a scalar multiplication.)
- (**dual space!**) A **linear functional** f on V is a *linear* map from V to F, that is,
$$f(av + b) = af(v) + b$$
for every $a, b \in F$ and $v \in V$. The **dual space** of V is the space of linear functionals on V, and it forms a F-vector space, with the following operations:
$$(f+g)(v) = f(v) + g(v), \qquad (af)(v) = af(v).$$
The dual space is one of the most interesting things not only in linear algebra, but in abstract algebra, or even in *any* branches which use the term *dual* (e.g., in projective geometry).

Proposition 1.2.1.

- $0v = 0$ *and* $(-1)v = -v$.

  Proof. $0v = (0+0)v = 0v + 0v$ implies $0v = 0$ and $v + (-v) = 0 = 0v = (1 + (-1))v = 1v + (-1)v = v + (-1)v$ implies $-v = (-1)v$. □

- *Every* ***linear combination*** $\sum_i a_i v_i$ *of vectors* $v_i \in V$ *is in* V.

  Proof. Easy induction on the number of summands(terms). ☐

Definition 1.2.2. A subset $W \subseteq V$ is called a **subspace** of V if it forms a vector space itself with the *inherited* operations from V. We denote it $W \leq V$.

Proposition 1.2.2.

$$W \leq V \quad \Longleftrightarrow \quad \forall c \in F,\ \forall v,\ w \in W, \quad cv + w \in W.$$

Proof. ($\Leftarrow$) The other axioms of vector space is satisfied by the fact that the operations are inherited by V, and hence it suffices to show that the operations are closed. First, let $c = -1$ and $v = w$. Then we have $(-1)w + w = 0 \in W$ by the proposition above. Then, the addition is closed if we let $c = 1$; and so is the scalar multiplication if we let $w = 0$, which is in W, as we proved. Hence W forms a vector space.

($\Rightarrow$) $cv + w$ is a linear combination. ㅋㅋ. ☐

Example 1.2.2.

- Removing a coordinate(**projection**):

$$F^2 = \{(a, b) :\ a, b \in F\} \leq \{(a, b, c) :\ a, b, c \in F\} = F^3.$$

- **Symmetric**, **alternating** matrices over F:

$$\mathrm{Sym}_n(F) = \{A \in \mathfrak{M}_{n,n}(F) :\ A = A^{\mathsf{T}}\} \leq M_{n,n}(F),$$
$$\mathrm{Alt}_n(F) = \{A \in \mathfrak{M}_{n,n}(F) :\ A = -A^{\mathsf{T}}\} \leq M_{n,n}(F).$$

- **Hermitian** matrices:

$$\mathrm{Her}_n = \{A \in \mathfrak{M}_{n,n}(\mathbb{C}) :\ A = \overline{A^{\mathsf{T}}}\} \leq M_{n,n}(\mathbb{C}).$$

- The solution space of a system of homogeneous linear (differential) equations.

1.2.1 Spanned subspace

Definition 1.2.3. For a subset S of a vector space V, the **subspace spanned by *S*** is the following set:

$$\langle S \rangle = \left\{ \sum_{\text{finite}} a_i v_i : \quad a_i \in F,\ v_i \in S \right\},$$

where $\sum_{\text{finite}} a_i v_i$ means $a_i \neq 0$ for *only* some finitely many indices i, and $a_i = 0$ for others; i.e., we only add finitely many vectors.[†]

And we call a element of $\langle S \rangle$ a **linear combination** of S.

[†]We take finitely many vectors since it is not sufficient to define 'convergence' of an infinite series with just axioms of vector space.

Example 1.2.3. Let $\mathbb{R}^\infty$ be the space of all sequences which are **eventually zero**, that is, there are only *finitely many* nonzero terms. Then,

$$\mathcal{E}^\infty = \left\{ \mathbf{e}_i = \left(0,\ \cdots,\ 0,\ \underset{i-\text{th}}{1},\ 0,\ \cdots \right) : \quad i \in \mathbb{N} \right\}$$

spans $\mathbb{R}^\infty$. Hence, for example, a sequence

$$(1,\ 1,\ 1,\ \cdots)$$

is not a linear combination of $\mathcal{E}^\infty$.

Proposition 1.2.3. *For every subset $S \in V$, $\langle S \rangle$ is a subspace of V.*

Proof. Use **Proposition 1.2.2**. □

Proposition 1.2.4. *For every subset $S \in V$, $\langle S \rangle$ is the smallest subspace of V which contains S, namely,*

$$\langle S \rangle = \bigcap_{\substack{S \subseteq W \\ W:\ \textit{vector space}}} W.$$

Proof. ($\subseteq$) Since W is a vector space containing S, it must contain other linear combinations of S also. Therefore $\langle S \rangle \subseteq W$ for every W satisfying the condition whence

$$\langle S \rangle \subseteq \bigcap_{\substack{S \subseteq W \\ W:\ \text{vector space}}} W.$$

($\supseteq$) $\langle S \rangle$ is a vector space containing S. □

1.2.2 Basis

Definition 1.2.4. A subset $S = \{v_i :\ i \in I\}$ is **linearly independent** if every vector of S cannot be represented by a linear combination other vectors; i.e.,

$$\forall i \in I,\ \forall a_j \in F, \quad v_i \neq \sum_{\substack{j \neq i \\ \text{finite}}} a_j v_j.$$

Equivalently, if every linear combination whose coefficients are not all zero is non-zero, the subset is linearly independent. If not, S is **linearly dependent**.

Definition 1.2.5. A **(Hamel) basis** $\mathfrak{B}$ of V is a subset of V which satisfies the followings:

$$\langle \mathfrak{B} \rangle = V$$

and

$$\mathfrak{B} \text{ is linearly independent.}$$

We usually fix the order of elements of $\mathfrak{B}$, which is called an **ordered basis**. Hereafter, *every basis is an ordered basis.* For example, a basis $\{(1,0),\ (0,1)\}$ and $\{(0,1),\ (1,0)\}$ are different bases.

Example 1.2.4.

- (**standard basis**)

$$\mathcal{E} = \left\{ \mathbf{e}_i = \left(0,\ \cdots,\ 0,\ \underset{i-\text{th}}{1},\ 0,\ \cdots,\ 0 \right) : \quad 1 \le i \le n \right\}$$

  is a basis for F^n, for arbitrary field F.

- $\{(1,0),(1,1)\}$ is a basis for $\mathbb{R}^2$ (over the field $\mathbb{R}$).

1.2.3 Dimension

Definition 1.2.6. 'The' **dimension** $\dim V$ of given vector space V is the **cardinality**(the number of elements of given set for finite set) of a basis.

Definition 1.2.7. A **finite dimensional vector space** is a vector space whose bases are all finite.†

We consider *finite dimensional vector spaces only* unless there is an additory description.

Lemma 1.2.1. *Let* $\mathfrak{B} = \{v_i :\ 1 \le i \le n\}$ *be a basis, and* $\mathfrak{C} = \{w_j :\ 1 \le j \le m\}$ *span* V*. Then* $m \ge n$*.*

Proof. If there is a vector v of $\mathfrak{B}$ which is not in $\mathfrak{C}$, without loss of generality, rename it $v_1 \neq w_i$. Then $\mathfrak{C} \cup \{v_1\}$ is linearly dependent: since $\mathfrak{C}$ spans V, there is a linear combination of $\mathfrak{B}$ represents $-v_1$ whence $\mathfrak{C} \cup \{v_1\}$ is linearly dependent. Hence there is at least one vector w_t which is represented by a linear combination of others. Then let $\mathfrak{C}_1 = \mathfrak{C} \cup \{v_1\} - \{w_t\}$.

Repeat this process. It is possible up to n-th stage since $\mathfrak{B}$ is linearly independent: if there is a linear dependence, it must contain a w-vector. Hence we obtain

$$\mathfrak{C}_n = \{v_1,\ \cdots,\ v_n,\ w_{r_1},\ \cdots,\ w_{r_{m-n}}\}$$

and it is linearly dependent. □

Theorem 1.2.1 (uniqueness of dimension)**.** *Let* $\mathfrak{B}$ *and* $\mathfrak{C}$ *be two bases of finite dimensional vector space* V*. Then* $|\mathfrak{B}| = |\mathfrak{C}|$*.*

Proof. $|\mathfrak{B}| \ge |\mathfrak{C}|$ and $|\mathfrak{C}| \ge |\mathfrak{B}|$ by **Lemma 1.2.1**. □

Well... How about the existence? The existence of the dimension needs the existence of the basis of V.

Theorem 1.2.2 (existence of basis)**.** *Every vector space has a basis, if* ***AC****(Axiom of Choice) assumed. In addition, it is equivalent to* ***AC****.*

Proof. $\exists\mathfrak{B} \iff$ **ZL** $\iff$ **AC**. See a set theory textbook. □

† It is the best way for defining finite dimensional vector space since the dimension is *not* well-defined yet.

1.2.4 Basis extension

We can *extend* a basis of smaller space to a larger space.

Theorem 1.2.3 (basis extension)**.** *Let $W \leq V$ be two vector spaces and $\mathfrak{C}$ be a basis for W. Then there is a basis $\mathfrak{B}$ of V which contains $\mathfrak{C}$.*

Proof. Induction on $\boldsymbol{n}-\boldsymbol{m}$, where $n = \dim V$ and $m = \dim W$. If $n - m = 0$, just let $\mathfrak{B} = \mathfrak{C}$. (Why?) Now, assuming there is a vector in $V - W$, take a vector v in $V - W$. Then v is linearly independent with $\mathfrak{B}$ (that is, $\tilde{\mathfrak{B}} = \mathfrak{B} \cup \{v\}$ is linearly independent) and hence $\tilde{W} = \left\langle \tilde{\mathfrak{B}} \right\rangle$ is a vector space which $\tilde{W} \leq V$. Since $n - m$ decreases, the induction proceeds. □

1.2.5 Sum and direct sum

Definition 1.2.8. For a set $\{S_i\}_{i\in I}$ of sets with a common addition, we define the **sum** of $\{S_i\}$ as follows:

$$\begin{aligned}\sum_{i\in I} S_i &= \left\{ \sum_{\text{finite}} s_i : \quad s_i \in S_i \right\} \\ &= \left\{ \sum_{i\in I} s_i : \quad s_i \in S_i \text{ and all } s_i = 0 \text{ but for finitely many } i \right\}.\end{aligned}$$

Definition 1.2.9. For a set $\{W_i\}_{i\in I}$ of *subspaces* of V which is mutually disjoint:

$$W_i \cap W_j = \{0\}, \qquad \text{for } i \neq j,$$

we define the **direct sum** of $\{W_i\}$ just the sum of them:

$$\bigoplus_{i\in I} W_i = \sum_{i\in I} W_i.$$

If the set is not mutually disjoint, even if it is mutually disjoint itself, we *make it be* mutually disjoint: isolate the vectors with giving different coordinates for each vector space. For example, if $V \oplus W \neq \{0\}$,

$$V \oplus W :\approx \{(v,\ w) : \quad v \in V,\ w \in W\}.$$

Trivially the direct sum of some vector spaces is a vector space.

Example 1.2.5.

- $\mathbb{R} \oplus \mathbb{R} \approx \mathbb{R}^2$,
- $\mathbb{R} \oplus \mathbb{R} \oplus \cdots \approx \mathbb{R}^\infty$.

Additional.

- For finite spaces, the direct sum of them is *the same*(isomorphic) with the cartesian product.

- The direct sum can be represented by a commutative diagram:

$$
\begin{tikzcd}
X & & \\
& V \oplus W \arrow[ul, dotted, "f"] & V \arrow[l, "i_1"] \arrow[ull, bend right, "f_1"'] \\
& W \arrow[u, "i_2"'] \arrow[uul, bend left, "f_2"] &
\end{tikzcd}
$$

  where the cartesian product is represented by:

$$
\begin{tikzcd}
X \arrow[dr, dotted, "f"'] \arrow[drr, bend left, "f_1"] \arrow[ddr, bend right, "f_2"'] & & \\
& X \times Y \arrow[r, "p_1"'] \arrow[d, "p_2"] & X \\
& Y &
\end{tikzcd}
$$

1.3 Linear Transformation

Definition 1.3.1. Let V and W be two F-vector spaces, then a map $f: V \to W$ is **linear** if

$$f(cv + dw) = cf(v) + df(w)$$

for every $c,\ d \in F$ and $v,\ w \in V$.

In another viewpoint, a linear map is a **vector space homomorphism** since it *preserves* the operations of vector spaces.

Example 1.3.1.

- A map $f: F \to F, \quad x \mapsto ax$ is a linear map from and to F. It is why maps of this kind are called *linear*.

- Producting a matrix is a linear map. For a matrix $A \in \mathfrak{M}_{m,n}(F)$,

  $$L_A: F^m \to F^n, \quad X \mapsto AX$$

  is a linear map from F^m to F^n. We will show that every linear map (from and to finite dimensional vector spaces) can be represented in this way, i.e., *matrices and linear transformations are the same things.*

- Another familiar linear maps are differentiation and integration. Let $\mathcal{C}^n$ be the space of function from and to $\mathbb{R}$ which is n-th differentiable and has continuous n-th derivative. Then for $n \in \mathbb{N}$, the **differentiation operator**

  $$D: \mathcal{C}^n \to \mathcal{C}^{n-1}, \qquad f \mapsto f'$$

  is a linear transformation since $(cf + dg)' = cf' + dg'$. Similarly, the **integration operator**

  $$J: \mathcal{C}^{n \geq 0} \to \mathcal{C}^{n+1}, \qquad f \mapsto \int_0^x f \ \mathrm{d}x$$

  is linear. (Here, F is not a field but the antiderivative of f.) Similarly, partial differentiation operators are linear.

- Transpose and trace are linear.

Theorem 1.3.1. *If f is linear, values of f at the basis elements determine f. Equivalently, there is an one-to-one correspondence between f and $f(\mathbf{e}_i)$'s.*

Proof.

$$\begin{tikzcd} f(v) \arrow[r, bend left, "v=\mathbf{e}_i"] & f(\mathbf{e}_i) \arrow[l, bend left, "v=\sum_i a_i \mathbf{e}_i"] \end{tikzcd}$$

$\square$

Proposition 1.3.1. *Let denote the space of all linear transformations from V to W as $\mathfrak{L}(V, W)$. Then $\mathfrak{L}(V, W)$ is a vector space dimension of mn.*

Definition 1.3.2. For an F-vector space V, a linear map $f: V \to F$ is called a **linear functional**. And it forms a vector space, which is called the **dual space** of V.

$$V^* = \{f: V \to F \mid f \text{ linear}\}.$$

Since V^* is an F-vector space, we can define the **double dual** V^{**} of V as follows:

$$V^{**} = (V^*)^* = \{\alpha: V^* \to F \mid \alpha \text{ linear}\}.$$

An example of elements of V^{**} is **evaluation**:

$$\alpha_a(f) = f(a), \qquad f \in V^*.$$

Definition 1.3.3. For a linear map $f: V \to W$, the **kernel** or the **null space** of f is

$$\ker f = f^{-1}(0) = \{v \in V: f(v) = 0\},$$

where 0 is the zero vector of W. The **image** of f is just $\operatorname{im} f = f(V)$.

Proposition 1.3.2. *The kernel and the image of a linear map form subspaces of V, respectively.*

Theorem 1.3.2 (dimension theorem)**.** *For a linear map $f: V \to W$,*

$$\dim \ker f + \dim \operatorname{im} f = \dim V.$$

Proof. Proof by basis extension. Let $\mathfrak{B} = \{v_i: 1 \le i \le n\}$ be a basis of $\ker f$. Then there is a basis $\mathfrak{C} = \mathfrak{B} \cup \{w_j: 1 \le j \le m\}$ of V which contains $\mathfrak{B}$, and we will show that $f(\mathfrak{C} - \mathfrak{B})$ is a basis of $\operatorname{im} f$.

For an arbitrary vector $v = \sum_i a_i v_i + \sum_j b_j w_j$ of V,

$$f(v) = f\left(\sum_i a_i v_i + \sum_j b_j w_j\right) = \sum_j b_j f(w_j)$$

since v's are in the kernel. Since v's and w's are linearly independent, so are $f(w)$'s. Hence $f(w)$'s form a basis of $\operatorname{im} f$. □

We call $\dim \ker f$ the **nullity** of f and denote it as $\operatorname{null} f$.

Definition 1.3.4.

- A **monomorphism** is an injective homomorphism.
- An **epimorphism** is an surjective homomorphism.
- An **isomorphism** is an bijective homomorphism.
- An **automorphism** is an bijective homomorphism from and to itself.

Figure 1.1: A pigeon.

Theorem 1.3.3 (vector space version of pigeonhole principle)**.** *Let* $f: V \to W$ *is linear, and suppose* $\dim V = \dim W = n < \infty$. *Then the followings hold:*

- *if* f *is a monomorphism, then it is an isomorphism;*
- *if* f *is an epimorphism, then it is an isomorphism.*

Proof. Let f be a monomorphism; suppose that f is not surjective. Then there is a vector $w \in W$ such that $\forall v \in V,\ w \neq f(v)$. Since $f(0) = 0$, other vectors in V are not mapped to 0 and hence $\ker f = \{0\}$. And we get $\dim \operatorname{im} f = n$ from $\dim \ker f = 0$. Hence $\operatorname{im} f = W$.

Changing im and ker proves the rest part of the theorem. □

Now we can prove the question in **Section 1.2**.

Theorem 1.3.4. *For two square matrices* A *and* $B \in \mathfrak{M}_{n,n}(F)$*, if* $AB = I$*, then* $A = B^{-1}$.

Proof. $AB = I$ implies that B is left-invertible, which is equivalent to that L_B is a monomorphism. Since $L_B: F^n \to F^n$, L_B is an isomorphism whence B is invertible:

$$B^{-1} = \begin{pmatrix} | & & | \\ L_B^{-1}\mathbf{e}_1 & \cdots & L_B^{-1}\mathbf{e}_n \\ | & & | \end{pmatrix}.$$

Multiplying B^{-1} right in the both sides of $AB = I$, we obtain $A = B^{-1}$. Similarly, A is invertible and $A^{-1} = B$. □

1.3.1 Rank

Definition 1.3.5. For a matrix $A \in \mathfrak{M}_{m,n}(F)$, the **row space** is a space which is generated by the row vectors of A. Similarly, the **column space** is a space which is generated by the column vectors of A. Then the **row**(*column*) **rank** is the dimension of the row(*column*) space.

Example 1.3.2. We can know the row rank and the column rank in the (R, C)-REF of the matrix; one can do elementary row operations:

$$\begin{aligned} A &= \begin{pmatrix} 1 & 0 & 1 \\ 5 & 6 & 7 \\ 0 & 3 & 1 \end{pmatrix} \\ &\sim_r \begin{pmatrix} 1 & 0 & 1 \\ 0 & 6 & 2 \\ 0 & 3 & 1 \end{pmatrix} \\ &\sim_r \begin{pmatrix} 1 & 0 & 1 \\ 0 & 6 & 2 \\ 0 & 0 & 0 \end{pmatrix} \\ &\sim_r \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & \frac{1}{3} \\ 0 & 0 & 0 \end{pmatrix}; \end{aligned}$$

hence the row space is $\{(a, b, a + b/3)^\top : \ a, b \in R\}$ whence the row rank is 2; while the column space is $\{(a + c, b + c/3, 0)^\top = (\tilde{a}, \tilde{b}, 0)^\top : \ \tilde{a}, \tilde{b} \in R\}$ whence the column rank is also 2. Otherwise one can do elementary column operations:

$$\begin{aligned} A &= \begin{pmatrix} 1 & 0 & 1 \\ 5 & 6 & 7 \\ 0 & 3 & 1 \end{pmatrix} \\ &\sim_c \begin{pmatrix} 1 & 0 & 0 \\ 5 & 6 & 2 \\ 0 & 3 & 1 \end{pmatrix} \\ &\sim_c \begin{pmatrix} 1 & 0 & 0 \\ 5 & 6 & 0 \\ 0 & 3 & 0 \end{pmatrix} \\ &\sim_c \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -\frac{5}{2} & \frac{1}{2} & 0 \end{pmatrix}; \end{aligned}$$

which makes the same result.

Are the row rank and the column rank the same? The answer is...

Lemma 1.3.1. *For a matrix $A \in \mathfrak{M}_{m,n}(F)$ over an* ordered field *F,*

$$\text{col rk}\, A = \text{col rk}\, A^\top.$$

Proof. It is suffices to show that col rk $A \leq$ col rk $A^\top$, since it implies

$$\text{col rk}\, A^\top \leq \text{col rk}(A^\top)^\top = \text{col rk}\, A$$

which completes the proof of lemma.

We will show that $Av = 0$ if and only if $(A^\mathsf{T}A)v = 0$ whence

$$\text{col rk}\, A = \text{col rk}(A^\mathsf{T}A) \leq \text{col rk}\, A^\mathsf{T}$$

; the last inequality follows because each column of $A^\mathsf{T}A$ is a linear combination of the columns of A^T. First, $Av = 0 \implies A^\mathsf{T}Av = 0$ trivially. Conversely,

$$A^\mathsf{T}Av = 0 \implies v^\mathsf{T}A^\mathsf{T}Av = 0 \implies (Av)^\mathsf{T}Av = 0 \implies Av = 0,$$

by positive-definiteness of the dot product. (More generalized version of proof uses the orthogonal complement, or even **Erdős-Kaplansky Theorem**(?). See http://math.stackexchange.com/questions/2315/is-the-rank-of-a-matrix-the-same-of-its-transpose-if-yes-how-can-i-prove-it) □

Theorem 1.3.5 (rank theorem)**.** *The row rank and the column rank are the same, and we call it the* ***rank*** *of the given matrix.*

Proof 1. Count the number of *leading 1* in RREF. □

Proof 2 assuming that F is an ordered field. It is trivial that the row rank of A equals the column rank of A^T. Since the column rank of A^T is the same with of A, the proof completed. □

Theorem 1.3.6 (rank-nullity theorem)**.** *Let $A \in \mathfrak{M}_{m,n}(F)$ be a matrix, then*

$$\operatorname{rank} A + \dim\ker L_A = m.$$

We call $\dim\ker L_A$ *the* ***nullity*** *of A, and denote* $\operatorname{null} A$. *Hence*

$$\operatorname{rank} A + \operatorname{null} A = m = \dim\operatorname{dom} L_A.$$

Proof. We know that $A\mathbf{e}_i$ is i-th column of A. Hence the column space of A is just the image of $L_A : F^m \to F^n$, hence $\text{col rk}\, A = \dim\operatorname{im} L_A$. By the dimension theorem(**Theorem 1.3.2**),

$$\dim\ker L_A + \text{col rk}\, A = \dim F^m = m.$$

□

Definition 1.3.6. Given an m by n matrix A of rank r, a **rank decomposition** of A is a representation by a product $A = PQ$ of two matrices $P \in \mathfrak{M}_{m,r}(F)$ and $Q \in \mathfrak{M}_{r,n}(F)$.

Theorem 1.3.7. *A rank decomposition of a matrix exists, but not uniquely.*

Proof. □

1.3.2 Matrix representation and similarity

Definition 1.3.7. For a finite dimensional vector space V and a basis $\mathfrak{B} = \{v_i\}_{i\in I}$ of V, every vector v in V can be represented as a linear combination of $\mathfrak{B}$ *uniquely*, namely

$$v = \sum a_i v_i;$$

and we call the row vector

$$[v]_{\mathfrak{B}} = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}$$

the **coordinate** vector.

Theorem 1.3.8. *For a F-vector space homomorphism $f: V \to W$ and the bases $\mathfrak{B}$ and $\mathfrak{C}$ of V and W, respectively, there is a unique matrix $[f]_{\mathfrak{C}}^{\mathfrak{B}} \in \mathfrak{M}_{\dim V,\ \dim W}(F)$ such that*

$$[f]_{\mathfrak{C}}^{\mathfrak{B}}[v]_{\mathfrak{B}} = [fv]_{\mathfrak{C}}.$$

Proof. Let $n = \dim V$, $m = \dim W$, $\mathfrak{B} = \{v_i\}_{i=1}^n$ and $\mathfrak{C} = \{w_j\}_{j=1}^m$, then

$$\begin{aligned}
[f(v)]_{\mathfrak{C}} &= \left[f\left(\sum_i a_i v_i\right)\right]_{\mathfrak{C}} \\
&= \left[\sum_i a_i f(v_i)\right]_{\mathfrak{C}} \\
&= \left[\sum_i a_i \sum_j b_{ij} w_j\right]_{\mathfrak{C}} \\
&= \left[\sum_j \left(\sum_i a_i b_{ij}\right) w_j\right]_{\mathfrak{C}} \\
&= \begin{pmatrix} \sum_i a_i b_{i1} \\ \vdots \\ \sum_i a_i b_{im} \end{pmatrix} \\
&= \begin{pmatrix} b_{11} & \cdots & b_{n1} \\ \vdots & \ddots & \vdots \\ b_{1m} & \cdots & b_{nm} \end{pmatrix} \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} \\
&= [f]_{\mathfrak{C}}^{\mathfrak{B}}[v]_{\mathfrak{B}},
\end{aligned}$$

where $[f(v_i)]_{\mathfrak{C}} = [b_{i1} \ \cdots \ b_{im}]^{\mathsf{T}}$ whence

$$[f]_{\mathfrak{C}}^{\mathfrak{B}} = \begin{pmatrix} [f(v_1)]_{\mathfrak{C}} & \cdots & [f(v_n)]_{\mathfrak{C}} \end{pmatrix}.$$

Uniqueness follows from the uniqueness of the coordinate representation. $\square$

Proposition 1.3.3 (composition and product)**.** *Let $V \xrightarrow{f} W \xrightarrow{g} U$ be two homomorphisms. Then*

$$[g \circ f]^{\mathfrak{B}}_{\mathfrak{D}} = [g]^{\mathfrak{C}}_{\mathfrak{D}}[f]^{\mathfrak{B}}_{\mathfrak{C}}.$$

Proof. Easy. □

Therefore, if the bases are fixed, there is a *one-to-one correspondence* between the space of matrices and the space of linear transformations; where the operations are preserved under the correspondence, as follows:

$$f + g \quad \longleftrightarrow \quad [f] + [g]$$

and

$$f \circ g \quad \longleftrightarrow \quad [f][g].$$

We just say that,

"the matrices are the same thing as the linear transformations."

1.3.3 Basis transition

Theorem 1.3.9. *Let $\mathfrak{B}$ be a basis of F^n and let A is an n by n square matrix. Then $A\mathfrak{B} = \{Av_i : \ v_i \in \mathfrak{B}\}$ is a basis of F^n if and only if A is invertible.*

Proof. Denote $\mathfrak{B} = \{v_i : \ 1 \le i \le n\}$ like a column vector, although F^n is not a field, namely,

$$\mathfrak{B} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}.$$

And define $A\mathfrak{B}$ as componentwise product:

$$A\mathfrak{B} = \begin{pmatrix} Av_1 \\ Av_2 \\ \vdots \\ Av_n \end{pmatrix}.$$

($\Rightarrow$) If Y is a basis, then $\exists C \in \mathfrak{M}_{n,n}(F), \quad C(A\mathfrak{B}) = \mathfrak{B}$ since $A\mathfrak{B}$ spans F^n. $(CA)\mathfrak{B} = \mathfrak{B}$ implies that $(CA)v_i = v_i$ for every basis element $v_i \in \mathfrak{B}$ whence $CA = I$ and it implies that A is invertible.

($\Leftarrow$) Let $v = \sum_i a_i v_i$ and

$$R = \begin{pmatrix} a_1 & \cdots & a_n \end{pmatrix}$$

be the coefficient matrix. Then

$$v = \sum a_i v_i = R\mathfrak{B} = (RA^{-1})(A\mathfrak{B}) = \tilde{R}(A\mathfrak{B})$$

whence $A\mathfrak{B}$ is a basis of F^n. □

Theorem 1.3.10. *If $\mathfrak{B}$ and $\tilde{\mathfrak{B}}$ are bases of V and $\mathfrak{C}$ and $\tilde{\mathfrak{C}}$ are bases of W, then for linear $f: V \to W$,*

$$[\mathrm{id}_W]^{\mathfrak{C}}_{\tilde{\mathfrak{C}}}[f]^{\mathfrak{B}}_{\mathfrak{C}}[\mathrm{id}_V]^{\tilde{\mathfrak{B}}}_{\mathfrak{B}} = [f]^{\tilde{\mathfrak{B}}}_{\tilde{\mathfrak{C}}}.$$

Furthermore, $[\mathrm{id}]^{\bullet}_{\bullet}$ are all invertible.

Proof. For the first assertion, just use **Proposition 1.3.3**. For the second one, since

$$[\mathrm{id}]^{\tilde{\mathfrak{B}}}_{\mathfrak{B}}[\mathrm{id}]^{\mathfrak{B}}_{\tilde{\mathfrak{B}}} = [\mathrm{id}]^{\mathfrak{B}}_{\mathfrak{B}} = I_{\dim \bullet},$$

they are invertible by **Theorem 1.3.4**.

□

Proposition 1.3.4. *For two bases $\mathfrak{B}$ and $\tilde{\mathfrak{B}}$ of V, the* ***transition matrix*** *$[\mathrm{id}_V]^{\mathfrak{B}}_{\tilde{\mathfrak{B}}}$ is invertible. Conversely, for a basis $\mathfrak{B}$ of V and an invertible square matrix U whose the number of row is the dimension of the space V, there is a basis $\tilde{\mathfrak{B}}$ of V such that*

$$U = [\mathrm{id}_V]^{\mathfrak{B}}_{\tilde{\mathfrak{B}}}.$$

Proof. The first assertion was proved in **Theorem 1.3.10**. For the second, let

$$U = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix}$$

and $\mathfrak{B} = \{v_i\}$. We want another basis $\tilde{\mathfrak{B}} = \{w_i\}$ which satisfies

$$U = [\mathrm{id}_V]^{\mathfrak{B}}_{\tilde{\mathfrak{B}}} = \begin{pmatrix} | & & | \\ [v_1]_{\tilde{\mathfrak{B}}} & \cdots & [v_n]_{\tilde{\mathfrak{B}}} \\ | & & | \end{pmatrix},$$

that is,

$$v_i = \sum_j a_{ij} w_j, \qquad \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix} = U \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix}$$

Hence we have

$$\tilde{\mathfrak{B}} = U^{-1}\mathfrak{B}.$$

□

1.3.4 Similarity

Definition 1.3.8 (similarity)**.** For two square matrices A and $\tilde{A}$, we say those are **similar** if there is an invertible matrix U such that

$$\tilde{A} = U^{-1}AU,$$

and denote $A \sim \tilde{A}$.

Proposition 1.3.5. *Similarity relation is an* ***equivalence relation****, that is, satisfies the following three properties:*

- *(Reflexivity)* $A \sim A$,
- *(Symmetricity)* $A \sim B \Longrightarrow B \sim A$,
- *(Transitivity)* $A \sim B \sim C \implies A \sim C$.

Example 1.3.3 (similarity)**.** For two square matrices A and $\tilde{A}$, we say those are **similar** if there is an invertible matrix U such that

$$\tilde{A} = U^{-1}AU,$$

and denote $A \sim \tilde{A}$.

Proposition 1.3.6 (similarity)**.** *For a linear operator* $f : V \to V$ *and the bases* $\mathfrak{B}$, $\tilde{\mathfrak{B}}$ *of* V, *two matrix representations of* f *are similar, that is,*

$$[f]_{\mathfrak{B}}^{\mathfrak{B}} \sim [f]_{\tilde{\mathfrak{B}}}^{\tilde{\mathfrak{B}}}.$$

Proof.

$$[\mathrm{id}_V]_{\tilde{\mathfrak{B}}}^{\mathfrak{B}}[f]_{\mathfrak{B}}^{\mathfrak{B}}[\mathrm{id}_V]_{\mathfrak{B}}^{\tilde{\mathfrak{B}}} = [f]_{\tilde{\mathfrak{B}}}^{\tilde{\mathfrak{B}}}.$$

□

Chapter 2

Matrix Group

2.1 Linear Groups

Definition 2.1.1 (normal subgroup)**.** A **normal subgroup** N of a given group G is a subgroup which left and right cosets gN and Ng are the same: $gN = Ng$, i.e.,

$$gNg^{-1} = N$$

for every $g \in G$. We denote it as $N \triangleleft G$.

Proposition 2.1.1 (why normal?)**.** *The quotient group G/N (read G mod N) is well-defined if and only if N is normal of G.*

Proof. Whatever we take, the equivalence class must be the same. If N is normal and letting $x \sim \tilde{x}$, i.e., $x^{-1}\tilde{x} \in N$,

$$\tilde{x}N \subseteq (xN)N = xN$$

and *vice versa*. If $\bar{x} = \bar{\tilde{x}}$ if $x \sim \tilde{x}$ and $\overline{xy} = \bar{x}\bar{y}$, we have

$$gN = (1g)N = NgN = hNgN = (hg)N$$

for every $h \in N$, hence $N = g^{-1}hgN$ so that N is normal: letting $n = g^{-1}hg\tilde{n}$, we have $g^{-1}hg = n\tilde{n}^{-1} \in N$ for every $h \in N$ whence $g^{-1}Ng \subseteq N$. $\square$

Definition 2.1.2 (general linear group and special linear group)**.** The **general linear group** of a given vector space V is a (multiplicative) group of automorphism on V; that is, $\mathrm{GL}(V) = \mathfrak{L}(V, V)^\times$. And the **special linear group** of V is the subgroup of $\mathrm{GL}(V)$ which is consisted by linear transformations whose determinant are all 1.

We denote $\mathrm{GL}(n, F) = \mathrm{GL}(F^n)$, and $\mathrm{SL}(n, F) = \mathrm{SL}(F^n)$.

Theorem 2.1.1 (first isomorphism theorem)**.** *For any homomorphism $\varphi : G \to H$ for two groups G and H,*

$$G/\ker\varphi \approx \operatorname{im}\varphi.$$

Proof.

$$\begin{tikzcd} G \arrow[rr, "\varphi"] \arrow[dr] & & \operatorname{im}\varphi \\ & G/\ker\varphi \arrow[ur, "\approx"'] & \end{tikzcd}$$

□

Proposition 2.1.2 (GL and SL)**.**

$$\mathrm{GL}(V)/\,\mathrm{SL}(V) \approx F^{\times}.$$

Definition 2.1.3 (center)**.** The **center** $Z(G)$ of a group G is the subgroup of elements which satisfy the 'commutative law', i.e.,

$$Z(G) = \{z \in G : \quad zg = gz \text{ for every } g \in G\}.$$

Z for *zentrum*, which means 'center' in German.

Proposition 2.1.3 (normality of the center)**.**

$$Z(G) \triangleleft G.$$

Proof. Trivially,

$$gZ(G) = \{gz : \ z \in Z(G)\} = \{zg : \ z \in Z(G)\} = Z(G)g.$$

□

Example 2.1.1. What are the centers of (a) $\mathrm{GL}(n, F)$ and (b) $\mathrm{SL}(n, F)$?

Answer: (a) $0 \neq cI$'s, (b) αI's where $\alpha^n = 1$.

Proof. (a) is just all. Let $AZ = ZA$ for all invertible A. Then, especially, for all *elementary matrices*, $EZ = ZE$. Note that multiplying E left is the same with elementary *row* operating, while multiplying right is for elementary *column* operating. (Especially, for E_{i+cj}'s.) Hence we obtain that Z is diagonal. Instead a more detailed explanation, we see an example:

$$\begin{pmatrix} 1 & 3 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a & *_1 \\ *_2 & b \end{pmatrix} = \begin{pmatrix} a + 3*_2 & *_1 + 3b \\ *_2 & b \end{pmatrix},$$

$$\begin{pmatrix} a & *_1 \\ *_2 & b \end{pmatrix} \begin{pmatrix} 1 & 3 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} a & *_1 + 3a \\ *_2 & b + 3*_2 \end{pmatrix},$$

hence $*_1$ and $*_2$ are zero. Similar details says that Z must be diagonal, for bigger matrices.

Now, the proof is done: since $E_{i\leftrightarrow j}$ is an elementary matrix, $Z_{ii} = Z_{jj}$, for every i and j pair. Therefore Z is a 'nonzero'(since Z is invertible!) multiple of I.

Not so surprisingly, the proof works on any *ring* with 1; if we modify 'nonzero' to 'invertible', that is, $c \in R^{\times}$. □

Additional (divide by center?)**.** Dividing by center means to ignore the difference due to the elements of Z. Since

$$Z(G) = \{z \in G : \quad z = gzg^{-1} \text{ for every } g \in G\},$$

we have some 'morphisms' $\varphi_g : \ a \mapsto b = gzg^{-1}$ and *their group*

$$\mathrm{Inn}(G) = \{\varphi_g : \quad g \in G\}.$$

We call this group the **inner automorphism group** of G.

We want to show that $G/Z(G) \approx \mathrm{Inn}(G)$. The idea is easy: use the homomorphism $\varphi_\bullet$ above:

$$\varphi_\bullet : \quad G \to \mathrm{Inn}(G).$$

The kernel of this homomorphism is just the center of G, since the (multiplicational) identity of $\mathrm{Inn}(G)$ is the identity function id_G and, from

$$\varphi_z = z \bullet z^{-1} = \mathrm{id}_G, \qquad \forall g \in G,$$

i.e.,

$$\varphi_z(g) = zgz^{-1} = g, \qquad \forall g \in G,$$

we get $zg = gz$ whence $z \in Z(G)$. Therefore $\ker(\varphi_\bullet) = Z(G)$, and by the first isomorphism theorem, we obtain

$$G/Z(G) \approx \mathrm{Inn}(G).$$

Definition 2.1.4 (PGL and PSL)**.** The **projective general linear group** is defined by

$$\mathrm{PGL}(V) = \mathrm{GL}(V)/Z(\mathrm{GL}(V)).$$

The **projective special linear group** is defined by

$$\mathrm{PSL}(V) = \mathrm{SL}(V)/Z(\mathrm{SL}(V)).$$

Projective geometry is difficult...

2.2 Orthogonal group

Definition 2.2.1 (orthogonal transformation)**.** For an *inner product space* $(V, \langle \bullet, \bullet \rangle)$ (or even just a quadratic space with non-degenerate symmetric bilinear form), an **orthogonal transformation** of V is an invertible linear transformation which preserves the given inner product, that is, such $A \in \mathrm{GL}(V)$:

$$\langle v, \ w \rangle = \langle Av, \ Aw \rangle.$$

The group of such transformations is called the **orthogonal group** $\mathrm{O}(V)$ of V, and also denote $\mathrm{O}(n, F) = \mathrm{O}(F^n)$ and $\mathrm{O}(n) = \mathrm{O}(n, \mathbb{R})$. F^n is considered with *dot product.*

Similarly, $\mathrm{SO}(V) = \{T \in \mathrm{O}(V) : \ \det T = 1\}$, and analogous definitions for $\mathrm{SO}(n, F)$ and $\mathrm{SO}(n)$. Obviously, it is called the **special orthogonal group** of V.

Definition 2.2.2 (unitary group)**.** If we give a *hermitian form* $(V,\ \langle\bullet,\bullet\rangle)$ rather than an inner product, where the given field is 'trivially' the field $\mathbb{C}$ of complex number, we define analogously **unitary group** $\mathrm{U}(n)$ as we defined the orthogonal group:

$$\langle v,\ w\rangle = \langle Av,\ Aw\rangle, \qquad A \in \mathrm{GL}(n,\ \mathbb{C}).$$

We *already know* what is $\mathrm{SU}(n)$ and how to call it :D.

Proposition 2.2.1.

- $\mathrm{SO}(V) \triangleleft \mathrm{O}(V) \triangleleft \mathrm{GL}(V)$;
- $\mathrm{SO}(V) \triangleleft \mathrm{SL}(V) \triangleleft \mathrm{GL}(V)$;
- $\mathrm{O}(n, F) = \{A \in \mathfrak{M}_{n,n}(F)^\times : \ A^{-1} = A^\mathsf{T}\}$, *if the inner product is a standard one, so-called dot product. (Canonically isomorphic!)*

Also the followings hold: for $\mathrm{O}(n, F)$, *every element is a matrix with pairwise orthonormal columns (or rows).*

Good, well, why it is called 'orthogonal'? It is because these preserves the 'angle' of two vectors, especially the *orthogonality.* Then, *what* is orthogonal? Which matrices are orthogonal?

Proposition 2.2.2.
In $\mathbb{R}^2$, $\mathrm{O}(2)$ *consists of rotations and reflections. And the group of rotations is just* $\mathrm{SO}(2)$.

Proof. From

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \qquad AA^\mathsf{T} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} a & c \\ b & d \end{pmatrix} = \begin{pmatrix} a^2+b^2 & ac+bd \\ ac+bd & c^2+d^2 \end{pmatrix} = I,$$

we obtain $a^2+b^2 = 1 = c^2+d^2$ and $ac+bd = 0$. Solutions for the first equality are just sines and cosines, namely:

$$a = \cos x, \quad b = \sin x, \quad c = \cos y, \quad d = \sin y.$$

(Orders of sine and cosine do not have to consider; since there is an inversion $\theta \mapsto \frac{\pi}{2} - \theta$.) Evaluating to another equality,

$$\cos x \cos y + \sin x \sin y = \cos(y-x) = 0, \qquad y - x = \frac{2k-1}{2}\pi.$$

Hence $y = x + \frac{2k-1}{2}\pi$. Substituting it, we get

$$c = -\sin x \sin\left(\frac{2k-1}{2}\pi\right) = \mp \sin x, \qquad d = \cos x \sin\left(\frac{2k-1}{2}\pi\right) = \pm\cos x.$$

Due to a *custom* in math and other sciences, we use $\theta = -x$ and finally get

$$A = \begin{pmatrix} \cos\theta & -\sin\theta \\ \pm\sin\theta & \pm\cos\theta \end{pmatrix}.$$

If the signa of second row are pluses,

$$A_+ = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} = R_\theta,$$

where R_θ is the **rotation matrix** of angle θ. Since $\det R_\theta = 1$, $R_\theta \in \mathrm{SO}(2)$.

If the signa of second row are minuses,

$$A_- = \begin{pmatrix} \cos\theta & -\sin\theta \\ -\sin\theta & -\cos\theta \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} R_\theta = S_{-\theta/2},$$

where S_φ is a **reflection matrix** w.r.t. a line $\theta = \varphi$ in polar coordinate system. (Draw it $\sim$.) Note that $\det S_\varphi = -1$. $\square$

How about 3-dimensional space? We consider a rotation on a line, the *axis*. For example, there are 'basic' three rotations:

$$R_x(\theta) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{pmatrix},$$

$$R_y(\theta) = \begin{pmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{pmatrix},$$

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

Surprisingly, they are almost *all*, i.e., the following holds. (Details are omitted.)

Theorem 2.2.1 (decomposition of rotation)**.** *For every 'rotation' $R \in \mathrm{SO}(3)$,*

$$R = R_z(\alpha)\, R_y(\beta)\, R_x(\gamma)$$

where Tait-Bryan angles of R are α, β, γ, about axes z, y, x respectively.

Following our knowledge, there efinition for *arbitrary rotation* is quite obvious, and only acceptable:

Definition 2.2.3 (rotation)**. Rotation** is an element of SO.

We must figure out the following definitions.

Definition 2.2.4 (PGO, PSO, PGU, PSU)**.** Projective (general) orthogonal group $\mathrm{PGO}(V)$ and projective special orthogonal group $\mathrm{PSO}(V)$. Similarly for U's...

Example 2.2.1. Calculate them! What is $Z(\mathrm{O}(V))$ and $Z(\mathrm{SO}(V))$?

Proof. Same with **Example 1.1.** A difference is that $\det Z = \pm 1$ in $\mathrm{O}(V)$. Another one is for SO: for odd-dimensional V, $Z(\mathrm{SO}(V)) = \{I\}$ (a trivial group) since $\det \pm I = \pm 1$; while $Z(\mathrm{SO}(V)) = \{\pm I\}$ for even-dimensional V since $\det \pm I = 1$. ☐

Corollary 2.2.1. PSO $\approx$ SO *for odd-dimensional vector space* V*.*

Proposition 2.2.3.

$$\mathrm{PSU}(2) \approx \mathrm{SO}(3), \qquad \mathrm{SU}(2) \xrightarrow{double} \mathrm{SO}(3),$$

where $\xrightarrow{double}$ *means that there is a double covering.*

"The shortest path between two truths in the real domain passes through the complex domain." —Jacques Hadamard.

Proof. $Z(\mathrm{SU}(2)) = \{\pm I\}$? Trivial. Then it suffices to show that $\mathrm{PSU}(2) \approx \mathrm{SO}(3)$. A transformation A of PSU(2) satisfies (U) $AA^\dagger = I$ by the definition. Use same method as **Proposition 1.5.**: let

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

then

$$AA^\dagger = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} \bar{a} & \bar{c} \\ \bar{b} & \bar{d} \end{pmatrix} = \begin{pmatrix} a\bar{a} + b\bar{b} & a\bar{c} + b\bar{d} \\ \bar{a}c + \bar{b}d & c\bar{c} + d\bar{d} \end{pmatrix} = I$$

whence

$$|a|^2 + |b|^2 = 1 = |c|^2 + |d|^2, \qquad a\bar{c} + b\bar{d} = 0.$$

The first equality gives us

$$a = e^{i\varphi_1}\cos x, \quad b = e^{i\varphi_2}\sin x, \quad c = e^{i\varphi_3}\cos y, \quad d = e^{i\varphi_4}\sin y,$$

and the second equality gives

$$\cos x \cos y + e^{i(-\varphi_1+\varphi_2+\varphi_3-\varphi_4)}\sin x \sin y = 0,$$

since $\overline{e^{i\theta}} = e^{-i\theta}$ for real θ. If $-\varphi_1 + \varphi_2 + \varphi_3 - \varphi_4 \neq 0$, the equality must not hold unless $b = d = 0$, which leads to a contradiction. Also, since (S) $\det A = 1$,

$$ad - bc = e^{i(\varphi_1+\varphi_4)}(\cos x \sin y - \sin x \cos y) = e^{i(\varphi_1+\varphi_4)}\sin(y - x) = 1$$

whence $\varphi_1 + \varphi_4 = \varphi_2 + \varphi_3 = k\pi$ and $y - x = \frac{2k-1}{2}\pi$. Therefore

$$A = \begin{pmatrix} e^{i\varphi_1}\cos x & e^{i\varphi_2}\sin x \\ \mp e^{-i\varphi_2}\sin x & \pm e^{-i\varphi_1}\cos x \end{pmatrix}.$$

Finally, (P) ignore one signum of them, then we have

$$A = \begin{pmatrix} e^{i\varphi_1}\cos\theta & -e^{i\varphi_2}\sin\theta \\ e^{-i\varphi_2}\sin\theta & e^{-i\varphi_1}\cos\theta \end{pmatrix}.$$

Hence, for example, there is an 'isomorphism'

$$\begin{pmatrix} e^{i\varphi_1}\cos\theta & -e^{i\varphi_2}\sin\theta \\ e^{-i\varphi_2}\sin\theta & e^{-i\varphi_1}\cos\theta \end{pmatrix} \leftrightarrow (\theta, \varphi_1, \varphi_2) \leftrightarrow R_z(\theta)R_y(\varphi_1)R_x(\varphi_2),$$

since $R_\bullet(\alpha+\beta) = R_\bullet(\alpha)R_\bullet(\beta)$. Therefore $\mathrm{PSU}(2) \approx \mathrm{SO}(3)$. □

2.3 SO(1,1)

Definition 2.3.1 (indefinite orthogonal group)**.** Consider the Euclidean space only, i.e., $F = \mathbb{R}$. The **indefinite orthogonal group** $\mathrm{O}(p, q)$ is something like O, but the inner product is not provided while the following bilinear form is given:

$$\langle v, w\rangle = v^\mathsf{T} \operatorname{diag}(\underbrace{1, \cdots, 1}_{p}, \underbrace{-1, \cdots, -1}_{q})w,$$

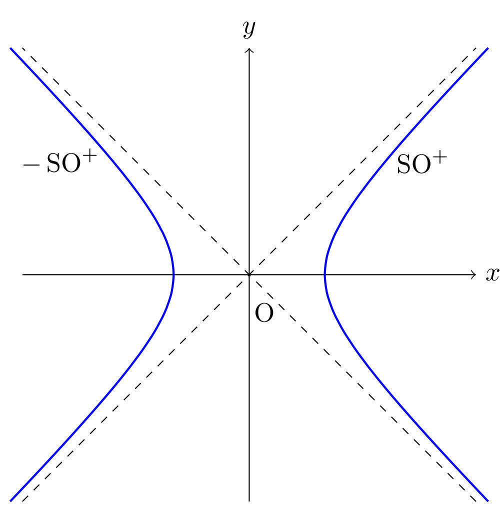
for $(p+q)$-dimensional vectors v and w. For instance, $\mathrm{O}(n) = \mathrm{O}(n, 0) = \mathrm{O}(0, n)$. And $\mathrm{SO}(p, q)$ is ...

We are interested in O(1,1) and O(1,3) in particular.

Proposition 2.3.1. $\mathrm{SO}(1,1)$ *can be represented by a hyperbolae* $x^2 - y^2 = 1$*, hence 2 connected curves.* SO^+ *is the 'connected' component of this group which contains the identity* I*,*

$$\mathrm{SO}^+ = \left\{ \begin{pmatrix} \cosh\theta & \sinh\theta \\ \sinh\theta & \cosh\theta \end{pmatrix} : \quad \theta \in \mathbb{R} \right\}.$$

In fact, we call the connected component of a given 'topological' group that contains the identity element the ***identity component*** *of given group.*



Proof. Completely same process. Note that if $A \in \mathrm{SO}(1,1)$,

$$v^\mathsf{T} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} w = \langle v, w \rangle = \langle Av, Aw \rangle = v^\mathsf{T} A^\mathsf{T} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} Aw,$$

hence

$$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = A^\mathsf{T} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} A.$$

Then we have

$$\mathrm{SO}(1,1) = \left\{ \pm \begin{pmatrix} \cosh\theta & \sinh\theta \\ \sinh\theta & \cosh\theta \end{pmatrix} : \quad \theta \in \mathbb{R} \right\}.$$

We *can(?)* represent it as a parametrized hyperbola:

$$\pm \begin{pmatrix} \cosh\theta & \sinh\theta \\ \sinh\theta & \cosh\theta \end{pmatrix} \longleftrightarrow \pm \begin{pmatrix} \cosh\theta \\ \sinh\theta \end{pmatrix},$$

then SO^+ and $-\mathrm{SO}^+$ are connected components of $\mathrm{SO}(1,1)$. $\square$

What does ‘connected’ means? Detail definition is in *topology*: it cannot separated by some open sets.

We will stop our work here about groups for the time being. If we learn *topology* or *Lie group theory*, it will continue...

Chapter 3

Similarity

3.1 Eigen-*something*

The prefix *eigen-* is adopted from the German word *eigen* for "own-" or "unique to", "peculiar to". We will study about some *unique* something, up to similarity.

Definition 3.1.1 (eigenvalue, eigenvector, eigenspace)**.** For a linear *operator* $T \in \mathfrak{L}(V, V)$, if

$$Tv = \lambda v$$

for a scalar $\lambda \in F$ and a *nonzero* vector $v \in V$, we call λ an **eigenvalue** and v an **eigenvector** of T. The **eigenspace** of λ is a subspace

$$E_\lambda = \{v \in V : \quad Tv = \lambda v\} = \ker(T - \lambda I)$$

of all vectors whose eigenvalue is λ.

Example 3.1.1. Check whether E_λ is a subspace of V.

Proposition 3.1.1. *If $f(t) \in F[t]$ and $v \in E_\lambda$, then $f(T)v = f(\lambda)v$.*

Theorem 3.1.1. *TFAE(the followings are equivalent):*

(a) *λ is an eigenvalue of T.*

(b) *$T - \lambda I$ is singular, i.e., non-invertible.*

(c) $\det(T - \lambda I) = 0$.

Proof. We already(and MUST) know that (b) and (c) are equivalent. If $T - \lambda I$ is invertible,

$$Tv = \lambda v \quad \Longleftrightarrow \quad (T - \lambda I)v = 0 \quad \Longleftrightarrow \quad v = 0$$

and hence λ is not an eigenvalue. And if λ is an eigenvalue, $T - \lambda I$ is not bijective and hence singular. $\square$

Thus, determinant of $\lambda I - T$ is important to decide whether or not λ is an eigenvalue of T. Hence we define a *polynomial*:

Definition 3.1.2 (characteristic polynomial)**.** The **characteristic polynomial** $\phi_T(t)$ is a polynomial defined by

$$\phi_T(t) = \det(tI - T).$$

We will write $\chi\phi$ instead of the term 'characteristic polynomial' since it is so long. XD

Here, t behaves like a *scalar* since it used instead of a scalar λ.

Note that λ is an eigenvalue iff $\phi_T(\lambda) = 0$.

Example 3.1.2. Check if $\chi\phi$ is really a polynomial.

Example 3.1.3. Calculate the $\chi\phi$ of a matrix

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

Find its eigenvalues and eigenspaces, and calculate $\phi_A(A)$. Note that

$$f(T) = \sum a_n T^n$$

for a polynomial $f(t) = \sum a_n t^n \in F[t]$ and a linear operator T.

3.2 Diagonalizability

Why we consider it? A big *raison d'etre* of eigen-something is *diagonalization* of a linear operator. First, from its name, we can define as follows:

Definition 3.2.1 (diagonalization)**.** A **diagonalization** of a linear operator $T \in \mathfrak{L}(V, V)$ is a representation T as a similar operator of a diagonal operator $D = \operatorname{diag}(d_1, \cdots, d_n)$. If there is a diagonalization of T, that is $T \sim D$ for a diagonal operator D, then we call T is **diagonalizable**.

Example 3.2.1. Determine whether the following matrices are diagonalizable, where $F = \mathbb{Q}$:

$$A = \begin{pmatrix} -1 & 3 & -1 \\ -3 & 5 & -1 \\ -3 & 3 & 1 \end{pmatrix}, \qquad B = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

If a matrix is not diagonalizable in given field, consider $F = \mathbb{R}$ and $F = \mathbb{C}$.

If T is diagonalizable, then $[T]_{\mathfrak{B}}^{\mathfrak{B}} = U^{-1}DU$ for a diagonal matrix D, supposing the basis $\mathfrak{B}$ is given for the vector space V; and there is another basis

$\mathfrak{C}$ for V such that $U = [\mathrm{id}_V]^{\mathfrak{C}}_{\mathfrak{B}}$. Evaluating this, we obtain $D = [T]^{\mathfrak{C}}_{\mathfrak{C}}$. Since it is diagonal, we have

$$[D]^i = D\mathbf{e}_i = [T]^{\mathfrak{C}}_{\mathfrak{C}}[w_i]_{\mathfrak{C}} = [Tw_i]_{\mathfrak{C}}$$

and

$$D\mathbf{e}_i = d_i\mathbf{e}_i = [d_i w_i]_{\mathfrak{C}},$$

where $D = \mathrm{diag}(d_1, \cdots, d_n)$ and $\mathfrak{C} = \{w_1, \cdots, w_n\}$. Hence we have $Tw_i = d_i w_i$, i.e., new basis must consist of eigenvectors, and the diagonal matrix contains corresponding eigenvalues. It is equivalent to the original definition. Hence we can re-define diagonalizability of a linear operator without matrices:

Definition 3.2.2 (redefine of diagonalizability)**.** A linear operator $T \in \mathfrak{L}(V, V)$ is diagonalizable if there is a basis for V whose elements are all eigenvectors of V.

Since eigenvectors span V, there are n linearly independent eigenvectors.

Proposition 3.2.1. *If the eigenvalues of T are mutually different, T is diagonalizable.*

Proof. If λ's are different, eigenvectors are linearly independent. □

Proposition 3.2.2. *If H is Hermitian, that is $H = H^\dagger$, then H can be diagonalized by a unitary operator U, i.e., $U^{-1} = U^\dagger$.*

Proof. Exercise. □

Theorem 3.2.1. *Let $T \in \mathfrak{L}(V, V)$ and λ_i's are eigenvalues of T. Then TFAE:*

(a) T *is diagonalizable,*

(b) $\phi_T(t) = \prod(x - \lambda_i)^{e_i}$, $e_i = \dim E_{\lambda_i}$,

(c) $V = \bigoplus E_{\lambda_i}$,

(d) $\dim V = \sum \dim E_{\lambda_i}$.

Proof. **(a)⇒(b)** $\chi\phi$ is invariant under similarity, since

$$tI - T = U^{-1}(tI - D)U,$$

for example. Hence

$$\phi_T(t) = \phi_D(t) = \prod(t - \lambda_i)^{e_i}.$$

A term due to a basis element appears once in the characteristic polynomial, hence the exponent of $t - \lambda_i$ is the (maximum) number of independent vectors in E_{λ_i}, i.e., dimension.

(b)⇒(c)⇒(d)⇒(a) ㅎㅎ. For (b) to (c), use dimension argument. Note that $\sum e_k = n$.

□

3.3 Cayley-Hamilton Theorem and Minimal Polynomial

From **Example 3.1.3**, we can know $\phi_A(A)$ for some matrices. Is it a general result? The answer is YES, and it is called *Cayley-Hamilton theorem*!

Theorem 3.3.1 (Cayley-Hamilton)**.**

$$\phi_T(T) = 0.$$

For $n = 2$, let $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. Then

$$\phi_A(t) = (t-a)(t-d) - bc = t^2 - (a+d)t + (ad-bc)$$

and we get a *familiar*(?) form:

$$T^2 - (a+d)T + (ad-bc) = 0.$$

Proof(?). Evaluating $t = T$,

$$\phi_T(T) = \det(TI - T) = \det(T - T) = 0.$$

(**NOT A PROOF.**) $\lightning$

First, t behaves as a scalar. And also 0 is a zero matrix, rather than a scalar 0, in a formula $\phi_T(T) = 0$.

Then how to prove it? We will consider $t^n \phi_T(t^{-1})$.

Proof. Let $\phi_T(t) = \sum_{i=0}^n c_i t^i$, then

$$t^n \phi_T(t^{-1}) = \sum_{i=0}^n c_i t^{n-i} = t^n \det\left(t^{-1}I - T\right) = \det(I - tT).$$

From

$$\det(A)I = A \cdot \operatorname{adj} A,$$

we get

$$\det(I - tT)I = (I - tT)\operatorname{adj}(I - tT).$$

In order to 'remove' $I - tT$ in the RHS, multiplying $\sum_{i=0}^m (tT)^i$ left,

$$\begin{aligned}
\left(\sum_{i=0}^m (tT)^i\right)\left(\sum_{i=0}^n c_i t^{n-i}\right) &= \left(\sum_{i=0}^m (tT)^i\right)\det(I - tT)I \\
&= \left(\sum_{i=0}^m (tT)^i\right)(I - tT)\operatorname{adj}(I - tT) \\
&= \left(I - (tT)^{m+1}\right)\operatorname{adj}(I - tT).
\end{aligned}$$

By definition of classical adjoint, every entry of this matrix is a polynomial of degree less than n. Hence RHS have terms of degree less than n or greater than or equal to m; for big m, the terms of degree $d \in [n, m)$ in LHS must be vanished. Hence, with m big enough, we obtain that the coefficient of the term of degree n is zero. Now, observing the coefficient of the term of degree n, we get

$$\sum_{i=0}^{n} c_i T^i = 0.$$

Hence $\phi_T(T) = 0$. □

Definition 3.3.1 (annihilating ideal)**.**

$$\mathcal{I}_T = \{p(t) \in F[t] : \quad p(T) = 0\}.$$

A polynomial in $\mathcal{I}_T$ is called an **annihilating polynomial**.

Since $\phi_T(t) \in \mathcal{I}_T$, by Cayley-Hamilton theorem, $\mathcal{I}_T \neq \emptyset$.

Theorem 3.3.2 (minimal polynomial)**.** *There is a monic annihilating polynomial which has the smallest degree. 'Monic' means that the coefficient of the highest order term is 1. We call this polynomial the* ***minimal polynomial*** *$m_T(t)$. And also,*

$$m_T(t) | p(t), \qquad p(t) \in \mathcal{I}_T;$$

especially, $m_T(t) | \phi_T(t)$.

Proof. Since $\deg \mathcal{I}_T$ is a subset of $\mathbb{N}$, there is the minimal degree d. If there is two different monic annihilating polynomial of degree d, denoting m_1 and m_2, we have $m_1 - m_2 \in \mathcal{I}_T$ which leads to a contradiction.

If $m_T(t) \not| \, p(t)$ for every $p(t) \in \mathcal{I}_T$, by division algorithm, we get that the remainder $r(t) = p(t) \bmod m_T(t)$ is also an annihilating polynomial which has the degree less than of $m_T(t)$, a contradiction. □

Proposition 3.3.1. *$\mathcal{I}_T$ is really an ideal. (Of a ring $F[t]$.)*

Proof. Exercise. □

Example 3.3.1. Find the $\chi\phi$ and $m\phi$ of a matrix

$$A = \begin{pmatrix} 5 & -6 & -6 \\ -1 & 4 & 2 \\ 3 & -6 & -4 \end{pmatrix}.$$

Answer.

$$\phi_A(t) = (t-1)(t-2)^2, \qquad m_A(t) = (t-1)(t-2).$$

□

3.4 Invariant and Triangularizability

Definition 3.4.1 (invariant subspace)**.** For $T \in \mathfrak{L}(V, V)$ and $W \leq V$, W is called invariant under T if $TW \leq W$.

Example 3.4.1.

- $F[t]$ is invariant under $D = \frac{\mathrm{d}}{\mathrm{d}t}$.
- Every space is invariant under a projection.
- Suppose there are two linear operator T and S on V, which commute, i.e., $TS = ST$. Let $W = \operatorname{im} S$ and $N = \ker S$, then W and N are invariant under T, since $TW = TSV = STV \leq SV = W$ and $Sn = 0 \implies STn = TSn = 0$.

Let $W \leq V$ be invariant under T, and $\mathfrak{C} = \{w_i\}_{i=1}^m$ be a basis of W. Then,

$$[T]_{\mathfrak{C}}^{\mathfrak{C}} = \begin{pmatrix} [T\restriction_W]_{\mathfrak{C}}^{\mathfrak{C}} & * \\ \mathbf{0} & * \end{pmatrix},$$

since, letting $\mathfrak{B} = \{w_i, v_j\}_{i=1, j=1}^{m,\ n-m} \supseteq \mathfrak{C}$ be a basis of V,

$$Tw_i = \sum a_i w_i + \sum 0 v_j.$$

Theorem 3.4.1. *Let $W \leq V$ be invariant under T, then*

$$\phi_{T\restriction_W} | \phi_T \qquad \textit{and} \qquad m_{T\restriction_W} | m_T.$$

Proof. Let W has a basis $\mathfrak{C}$ and $\mathfrak{B}$ is a basis of V which is extended from $\mathfrak{C}$. Then we have

$$[T]_{\mathfrak{B}}^{\mathfrak{B}} = \begin{pmatrix} [T\restriction_W]_{\mathfrak{C}}^{\mathfrak{C}} & * \\ \mathbf{0} & * \end{pmatrix}.$$

For $\chi\phi$,

$$0 = \phi_T\left([T]_{\mathfrak{B}}^{\mathfrak{B}}\right) = \begin{pmatrix} \phi_T\left([T\restriction_W]_{\mathfrak{C}}^{\mathfrak{C}}\right) & * \\ \mathbf{0} & * \end{pmatrix}.$$

And by above, we have

$$\phi_T\left([T]_{\mathfrak{B}}^{\mathfrak{B}}\right) = 0 \implies \phi_T\left([T\restriction_W]_{\mathfrak{C}}^{\mathfrak{C}}\right) = 0,$$

which completes the remained part of proof. □

Example 3.4.2. Consider a diagonalizable transformation T, and let W_i's be its eigenspaces, then it suits perfectly to above theorem, and it makes the 'sufficient-necessary condition' of diagonalizability clear. But if T is not diagonalizable, it cannot be adopted since we do not know the other components of given block matrix.

We define the following as a generalization of 'annihilator ideal':

Definition 3.4.2 (conductor (ideal))**.** Let W be an *invariant* subspace for T and let v be a vector in V. The **T-conductor of v into W** is the set $S_T(v;\ W)$ which consists of all polynomials $g \in F[t]$ such that $g(T)v \in W$.

If $W = 0$, we denote it as $\mathcal{I}_T(v) = S_T(v;\ 0)$ and call the **T-annihilator of v**. And $\mathcal{I}_T = \bigcap_v \mathcal{I}_T(v)$ is the T-annihilator of V, which annihilates all the vectors of V.

Proposition 3.4.1. *Conductor is an ideal in $F[t]$.*

Definition 3.4.3 (conductor (vector))**.** The monic generator of the ideal $S(v;\ W)$ is also called the **conductor** of v into W.

Analogous proofs of one for uniqueness of minimal polynomial prove also for the conductors, trivially. And, since $\mathcal{I}_T$ is the *strongest* polynomials, the T-conductors divide the minimal polynomial for T.

3.5 Minimal Polynomials and Triangular-/Diagonalizability

Lemma 3.5.1. *Suppose*

$$m_T(t) = \prod (t - c_i)^{r_i}, \qquad c_i \in F,$$

and let $W \lneq V$ be invariant under T. Then there exists a vector $v \notin W$ such that

$$\exists \lambda\text{: eigenvalue of } T: \qquad (T - \lambda I)v \in W,$$

that is, a linear polynomial is a T-conductor for some v.

Proof. Let $w \in V \setminus W$, and g be the T-conductor of w into W. ($g(T)w = 0$.) Then $g | m_T$, and since $w \notin W$, g cannot be a constant. ($g(T)w = kw \in W \implies k = 0 = g$ which is contradict to the fact that g is a generator of an nontrivial ideal.) Therefore

$$g(t) = \prod (t - c_i)^{e_i}; \qquad \sum e_i > 0.$$

Choose j so that $e_j > 0$, then $g = (t - c_j)h$ for some h. Since $v = h(T)w \notin W$ (g is minimal in the sense of degree) and $g(T)w = (T - cI)v \in W$, we just found v! Obviously, c is an eigenvalue. $\square$

We conclude(?) with the following necessary-sufficient condition of diagonalizability and trigonalizability(trivial meaning), in the sense of minimal polynomial:

Theorem 3.5.1. *T is triangularizable iff $m_T = \prod (t - \lambda_i)^{e_i}$, where λ_i's are distinct.*

Proof. ($\Longleftarrow$) Let $W = 0$, then above lemma says $\exists v \exists \lambda (T - \lambda I)v = 0$. Hence it forms an eigenspace, and there is a basis $\mathfrak{B}$ of V extending $\{v\}$; therefore we have

$$[T]_{\mathfrak{B}}^{\mathfrak{B}} = \begin{pmatrix} \lambda & ** \\ \mathbf{0} & * \end{pmatrix}.$$

By an induction on the dimension of square matrix ($*$ for above), we obtain a triangularization of T:

$$[T]_{\mathfrak{B}}^{\mathfrak{B}} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{pmatrix}.$$

($\Longrightarrow$) Calculate it$\sim$. □

Corollary 3.5.1. *Every linear operator is triangularizable if the given field is algebraically closed.*

Another proof of Corollary: using induction. There is an eigenvalue and an eigenvector since the field is algebraically closed. Hence let λ and v the chosen ones, extend v to a basis $\mathfrak{B}$ of V, and denote $\mathfrak{C} = \mathfrak{B} - \{v\}$ and $W = \langle \mathfrak{C} \rangle$. Then:

$$[T]_{\mathfrak{B}}^{\mathfrak{B}} = \begin{pmatrix} \lambda & ** \\ \mathbf{0} & * \end{pmatrix}.$$

And we know that

$$* = [\pi \circ T]_{\mathfrak{C}}^{\mathfrak{C}}$$

where π 'removes' v-component:

$$\pi : V \to W; \qquad av + \sum_{w_i \in \mathfrak{C}} b_i w_i \mapsto \sum_{w_i \in \mathfrak{C}} b_i w_i.$$

$$\begin{tikzcd} W \arrow[r, "T"] \arrow[dr, "\pi\circ T"'] & TW \arrow[d, "\pi"] \\ & W \end{tikzcd}$$

Since $\pi \circ T$ is linear, an induction completes the proof. □

Theorem 3.5.2. *T is diagonalizable iff $m_T = \prod (t - \lambda_i)$, where λ_i's are distinct.*

Proof. ($\Longrightarrow$) Trivial. Think as a linear transformation each of $T - \lambda_i I$'s.

($\Longleftarrow$) Let $W = \bigoplus E_{\lambda_i}$ be the space spanned by all of the eigenvectors of T, and suppose $W \neq V$. By **Lemma 3.5.1**, there is a vector $v \notin W$ and an eigenvalue λ_j such that $w = (T - \lambda_j I)v \in W$. Since $w \in W$, it is represented by a linear combination of eigenvectors uniquely:

$$w = \sum_{w_i \in E_{\lambda_i}} w_i,$$

noting that $Tw = \sum \lambda_i w_i$.

Let $m_T = (t - \lambda c_j)g$ for some polynomial g, and

$$g(t) - g(c_j) = (t - c_j)h(t)$$

for some polynomial h. Then we have

$$g(T)v - g(c_j)v = h(T)(T - c_j I)v = h(T)w \in W$$

and $g(T)v \in W$ whence $g(c_j)v \in W$. Since $v \notin W$, $g(c_j) = 0$. It contradicts the assumption that m_T has distinct roots. □

3.6 Simultaneous Triangular-/Diagonal-ization

We want to find a basis which triangularizes all of the transformations in a family $\mathscr{F}$ *simultaneously.*

The subspace W is **invariant under** $\mathscr{F}$ if W is invariant under each operators.

Since all diagonal matrices commute, if T and S diagonalized simultaneously, then

$$(U^{-1}TU)(U^{-1}SU) = (U^{-1}SU)(U^{-1}TU)$$

and hence $TS = ST$. Therefore we consider only a family whose elements commute mutually, for simultaneous diagonalization.

For simultaneous triangularization, one does not have to satisfy the commutating condition; however it is a *sufficient* condition for simultaneous triangularization, as we will see.

Lemma 3.6.1. *Let $\mathscr{F}$ be a commuting family of triangularizable linear operators on V. Let W be a proper subspace of V which is invariant under $\mathscr{F}$, then there is a vector $v \in V \setminus W$ such that*

$$\forall T \in \mathscr{F}, \quad Tv \in \langle v \rangle \oplus W.$$

Proof. It is too taxing to deal with infinitely many operators; hence we use a basis: let $\{T_1, \cdots, T_r\}$ be 'a'(need not to be unique) maximal linearly independent subset of $\mathscr{F}$; i.e. a basis for $\langle \mathscr{F} \rangle \leq \mathfrak{L}(V, V)$. ($\mathfrak{L}(V, V)$ is a f.d.v.s.) Then it is sufficient to check for these basis elements only.

By **Lemma 3.5.1**, for a single operator, we can find a vector $v_1 \in V \setminus W$ and a scalar λ_1 such that $(T_1 - \lambda_1 I)v_1 \in W$. Since W is invariant under T_1,

$$V_1 = \{v \in V : \quad (T_1 - \lambda_1 I)v \in W\} \gneq W.$$

And V_1 is invariant under $\mathscr{F}$.

Now, in order to use induction, consider V_1 instead of V. Let W be a proper subspace of V_1, and $U_2 = T_2 \restriction_W$ instead of T_1 of above procedure. Since $m_{U_2} | m_{T_2}$, we may apply **Lemma 3.5.1** to new W and U_2 and consider as of T_2. We obtain a vector $v_2 \in V_1 \setminus W$ and a scalar λ_2 such that $(T_2 - \lambda_2 I)v_2 \in W$.

Note that, since $v_2 \in V_1$, both of $(T_1 - \lambda_1 I)v_2$ and $(T_2 - \lambda_1 I)v_2$ belong to W. And let

$$V_2 = \{v \in V_1 : \quad (T_2 - \lambda_2 I)v \in W\},$$

then V_2 is invariant under $\mathscr{F}$.

Continue this process by an induction, then we can find $v = v_r$ as the desired vector. □

Theorem 3.6.1. *Let $\mathscr{F}$ be a commuting family of triangularizable linear operators on V. Then it can be triangularized simultaneously.*

Proof. Induction. Now it is easy. (Same with the proof of **Theorem 3.5.1**.) □

Now, finish with diagonalization.

Theorem 3.6.2. *Let $\mathscr{F}$ be a commuting family of diagonalizable linear operators on V. Then it can be diagonalized simultaneously.*

Proof. Almost same process as for triangularization, at this point, however, it is easier to proceed by induction on $\dim V$.

If $\dim V = 1$, automatically proved. Let $\dim V = n$ and choose any $cI \neq T \in \mathscr{F}$. Let λ_i's be the distinct eigenvalues of T and let $W_i = E_{\lambda_i} = \ker(T - c_i I)$. W_i is invariant under every operator which commutes with T; and each operator in

$$\mathscr{F}_i = \{T \restriction_{W_i} : \quad T \in \mathscr{F}\}$$

is diagonalizable since its minimal polynomial divides the minimal polynomial for the corresponding operator in $\mathscr{F}$. Operators in $\mathscr{F}_i$ can be diagonalized simultaneously since $\dim W_i < \dim V$ by a basis $\mathfrak{B}_i$. Then $\mathfrak{B} = (\mathfrak{B}_i)$ is a desired basis. □

Chapter 4

Decomposition

4.1 Direct Decomposition

First, see some examples of *direct decomposition* of a vector space.

Remind (direct sum)**.** Call W_i's are **independent** and denote

$$\bigoplus W_i = \sum W_i$$

if for every vector in $\sum W_i$ the coordinate representation of it is unique.

It is obvious that W_i's are **independent** iff $\mathfrak{B} = (\mathfrak{B}_i)$ is an ordered basis for $\sum W_i$ where each $\mathfrak{B}_i$ is one for W_i.

Remind (projection, or idempotent)**.** One such that $E^2 = E$.

We have $V = \ker E \oplus \operatorname{im} E$. And E is trivially diagonalizable with

$$[E]_{\mathfrak{B}} = \begin{pmatrix} I & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}$$

where $\mathfrak{B} = (\text{basis for } \operatorname{im} E, \text{ basis for } \ker E)$.

Theorem 4.1.1. *If $V = \bigoplus W_i$, then there exist projections E_i such that:*

- $E_i E_j = 0$ *if* $i \neq j$,
- $I = \sum E_i$,
- $\operatorname{im} E_i = W_i$.

Conversely, if there are projections E_i which satisfy above two from the top and let $\operatorname{im} E_i =: W_i$, then $V = \bigoplus W_i$.

Proof. ($\Longrightarrow$) Take

$$E_j : \bigoplus W_i \xrightarrow[\text{projection}]{\text{canonical}} W_j.$$

($\Longleftarrow$) Obvious. (Find the unique coordinate representation of a vector.) $\square$

Suppose each of W_i is invariant under T, then $T_i = T \restriction_{W_i}$ is a linear operator on W_i, and

$$Tv = \sum T_i v_i$$

if $v = \sum v_i$ is the unique coordinate representation with $v_i \in W_i$. We says that T is the direct sum of T_i's. If the basis is given by $\mathfrak{B} = (\mathfrak{B}_i)$ where each $\mathfrak{B}_i$ is one for W_i, then $[T]_{\mathfrak{B}}^{\mathfrak{B}}$ is a form of block diagonal matrix:

$$[T]_{\mathfrak{B}}^{\mathfrak{B}} = \begin{pmatrix} [T_1]_{\mathfrak{B}_1}^{\mathfrak{B}_1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & [T_2]_{\mathfrak{B}_2}^{\mathfrak{B}_2} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & [T_k]_{\mathfrak{B}_k}^{\mathfrak{B}_k} \end{pmatrix}.$$

Hence for matrices, A is the direct sum of A_i's if $A = \operatorname{diag}(A_1, \cdots, A_k)$ where diag denotes the block diagonal.

Theorem 4.1.2. *Let $V = \bigoplus W_i$ and E_i's be canonical projections. Then W_i's are all invariant under T iff T commutes with each of E_i's.*

Proof. ($\Longrightarrow$) Let $v = \sum v_i$, then

$$E_j T v = E_j \sum T_i v_i = E_j T_j v_j = T_j v_j = T v_j = T E_j v.$$

($\Longleftarrow$)

$$TW_i = TE_i V = E_i T V \leq E_i V = W_i.$$

$\square$

Similar procedure can be adopted to the eigenspace decomposition $V = \bigoplus E_{\lambda_i}$:

Theorem 4.1.3. *Let T be a diagonalizable operator(hence there is the eigenspace decomposition of V w.r.t. T), then there exist projections D_i such that:*

- $T = \sum \lambda_i D_i$,
- $I = \sum D_i$,
- $D_i D_j = 0$ *if* $i \neq j$,
- $\operatorname{im} D_i = E_{\lambda_i}$.

Conversely, if there are distinct scalars λ_i and nonzero operators D_i which satisfy above three from the top, then T is diagonalizable, λ_i's are eigenvalues, and D_i's are projections satisfy $\operatorname{im} D_i = E_{\lambda_i}$.

Proof. TOTALLY SAME PROCEDURE. Omit. $\square$

Hence, if $T = \sum \lambda_i D_i$, then for any polynomial g,

$$g(T) = \sum g(\lambda_i) D_i.$$

And we obtain

$$T^r = \left(\sum \lambda_i D_i\right)^r = \sum \lambda_i^r D_i,$$

since all of heterogeneous terms disappear. From this formulation, we have

$$g(T) = 0 \iff \forall i \ g(\lambda_i) = 0,$$

which means $m_T(t) = \prod(t - \lambda_i)$.

Note that, if $p_j(t) = \prod_{i \neq j} \frac{t - \lambda_i}{\lambda_j - \lambda_i}$, we have $p_j(\lambda_i) = \delta_{ij}$ whence

$$p_j(T) = p_j \left(\sum \lambda_i D_i\right) = \sum \delta_{ij} D_i = D_j.$$

(Hence D_j's not only commute with T but every polynomials in T.)

In fact, we have

$$g(t) = \sum g(\lambda_i) p_j(t).$$

Plugging $g = 1$ and $g = t$,

$$1 = \sum p_i, \qquad t = \sum \lambda_i p_i.$$

(Except $k = 1$. In this case T is trivially diagonalizable.) Evaluating T and using above formulae,

$$I = \sum D_i, \qquad T = \sum \lambda_i D_i.$$

Observe that if $i \neq j$, then $p | p_i p_j$ whence $D_i D_j = 0$. And $p_i(T) \neq 0$ since $\deg p_i < \deg p$. Applying to above theorem, we just proved the sufficient-necessary condition of diagonalizability with another method.

4.2 Primary Decomposition

It is a generalization of what we did above.

Theorem 4.2.1. *Let T be a linear operator on V, and factorize*

$$m_T(t) = \prod_{i=1}^{k} p_i(t)^{r_i},$$

where p_i's are distince irreducible monic polynomials. Let $W_i = \ker p_i(T)^{r_i}$, then:

- $V = \bigoplus W_i$,
- $TW_i \leq W_i$,
- *letting* $T_i = T \restriction_{W_i}$, $m_{T_i}(t) = p_i(t)^{r_i}$.