Notes of Linear Algebra

Xie Zejian Zhang Songxin

2021-03-05

Contents

1 Background Knowledge **3**

2 Vector Space **4**

2.1 Linear independence and basis 4

2.2 Free vector space 5

2.3 Linear mappings 6

2.4 Subspace and factor space 7

2.4.1 Subspace and Sum 7

2.4.2 Factor Space 9

2.5 Inner Product spaces 11

2.5.1 Orthogonal 11

2.6 Dimension 12

2.7 Convex sets 14

2.8 Matrix and linear space 14

2.8.1 Projection Matrix 16

2.8.2 Linear transformation 18

3 Linear Mappings 19

3.1 Basic properties 19

3.1.1 Induced Linear Mappings 20

Matrix Analysis 24

4 Eigenvalues 25

4.1 Symmetric matrices and Spectral Decomposition 27

4.2 Eigenprojections 28

4.3 Advanced in eigenvalues 29

4.4 Quadratic form 29

4.5 Nonnegative Definite Matrix 30

5 Singular Value Decomposition 31

Chapter 1

Background Knowledge

Definition 1.1 (Group)**.** A group is a set G with a binary low of composition

$$\mu : G \times G \to G$$

denoting as $\mu(x, y) = xy$.

- $(xy)\, z = x\,(yz)$
- There exists an element e called the identity s.t. $xe = ex = x$
- To each $x \in G$ there is an element x^{-1} s.t. $xx^{-1} = x^{-1}x = e$

Let G and H be two groups, then a mapping $\phi : G \to H$ is called a homomorphism if

$$\phi(xy) = \phi x \phi y \qquad x, y \in G$$

A group is called commutative or abelian if for each $x, y \in G$, $xy = yx$.

Definition 1.2 (field)**.** A field is a set K on which two binary lows of composition s.t.

- K is a commutative group with respect to addition.
- The set $K - \{0\}$ is a commutative group with respect to multiplication.
- Addition and multiplication are connected by the distributive low,

$$(\alpha + \beta)\,\gamma = \alpha\gamma + \beta\gamma$$

Chapter 2

Vector Space

2.1 Linear independence and basis

Definition 2.1 (linear independence)**.** A family of vectors $\{x_i\}_{i\in I}$ is called **linear independent** if the vectors x_i are linearly independent i.e.

$$\sum_{i\in I}\alpha_i x_i = 0 \implies \alpha_i = 0 \quad \text{for each } \ i$$

Definition 2.2 (system of generators)**.** A subset $S \subset E$ is called a system of generators of E if every vector $x \in E$ is a linear combination of vectors in S.

Proposition 2.1. *1. Every finitely generated non-trivial vector space has a finite basis.*

2. Suppose that $S = \{x_1, \dots, x_m\}$ is a finite system of generators of E and that the subset $R \subset S$ by $R = \{x_1, \dots, x_r\}$ $(r \leq m)$ consists of linearly independent vectors. Then there exists a basis T of E s.t. $R \subset T \subset S$.

Proof. Just need to notice that every basis is the system of generators, and it is a minimal one.

$\square$

Theorem 2.1. *Let E be a non-trivial vector space. Suppose S is a system of generators and R is a family of linearly independent vectors in E s.t. $R \subset S$. Then there exists a basis T of E s.t. $R \subset T \subset S$.*

Proof. Consider the partially order defined between R and S, find some $X \subset E$ s.t.

- $R \subset X \subset S$

- the vectors in X are linearly independent.

We note this partially order as $\mathcal{P}(R,S)$.
Notice that for every chain $\{X_\alpha\} \subset \mathcal{P}(R,S)$ has a maximal element $A = \bigcup_\alpha X_\alpha$. It is obvious that $A \in \mathcal{P}(R,S)$ (Notice that $R \subset A \subset S$ and the property of a chain of the set that contains linearly independent vectors.)
So we prove that every chain $\{X_\alpha\} \subset \mathcal{P}(R,S)$ has a upper bound in $\mathcal{P}(R,S)$, so Zorn's Lemma implies that there exists a maximal element $T \in \mathcal{P}(R,S)$ s.t. vectors in T are linearly independent.
Then we just need to show that T generates E. Give $x \in E$, suppose that x is linearly independent to vectors in T. Notice that S generates E, so

$$x = \sum_{i \in I'} \alpha_i x_i \qquad \text{for some } x_i \in S$$

If x is linearly independent to vectors in T then exists some $i \in I'$ s.t. x_i is linearly independent to vectors in T and note this set as $\{x_j\}_{j\in J} \subset S$, consider the set $\{x_j\}_{j\in J} \cup T \supsetneq T$ which leads to a contradiction of the maxmality of T. So T is a basis of E.

□

Corollary 2.1. 1. *Every system of generators of E contains a basis. In particular, every non-trivial vector space has a basis.*
2. *Every family of linearly independent vectors of E can be extended to a basis.*

2.2 Free vector space

Let X be an arbitrary set and consider all maps $f : X \to \mathbb{K}$ s.t. $f(x) \neq 0$ only for finitely many $x \in X$, denoting the set of these maps by $F(X)$, it is easy to show that $F(X)$ is a vector space.

Now give a basis of $F(X)$. For any $a \in X$, let f_a be:

$$f_a(x) = \begin{cases} 1 & x = a \\ 0 & x \neq a \end{cases}$$

Then $\{f_a\}_{a\in X}$ forms a basis of $F(X)$.

$F(X)$ is called the **free vector space over** X.

2.3 Linear mappings

Definition 2.3 (linear mapping)**.** Suppose that E and F are vector spaces, and let $\varphi : E \to F$ be a set mapping s.t.

$$\varphi(x+y) = \varphi(x) + \varphi(y) \text{ for all } x, y \in E$$

and

$$\varphi(\alpha x) = \alpha\varphi(x) \text{ for all } \alpha \in \mathbb{K}, x \in E$$

Then we call the mapping φ satisfying above conditions linear mappings.
Moreover, if $F = \mathbb{K}$, then we called φ a **linear function** on E.

Corollary 2.2. *Linear mappings preserve linear relations.*

Proof. Suppose φ be a linear mappings, and let $u = \alpha x + \beta y \in E$, then

$$\varphi(u) = \varphi(\alpha x + \beta y) = \alpha\varphi(x) + \beta\varphi(y)$$

□

Let $\varphi : E \to F, \psi : F \to G$ be linear mappings, then the composition of them $\psi \circ \varphi : E \to G$ is defined by:

$$(\psi \circ \varphi)(x) = \psi(\varphi(x))$$

It is easy to show that $\psi \circ \varphi$ is still a linear mapping.

Proposition 2.2. *Suppose S is a system of generators of E and $\varphi_0 : S \to F$ where F is also a vector space. Then φ_0 can be extended in at most one way to linear mapping $\varphi : E \to F$.*
And the extension exists iff such an extension is that

$$\sum_i \alpha_i \varphi_0(x_i) = 0$$

whenever $\sum_i \alpha_i x_i = 0$.

Proof.

- $\Longrightarrow$: Suppose φ to be a linear mapping and it is the extension of φ_0, then $\varphi\left(\sum_{i=1}^n \alpha_i x_i\right) = \sum_{i=1}^n \alpha_i \varphi(x_i)$ for each $x_i \in E$.
  And for each $x_i \in S$,

  $$\varphi\left(\sum_{i=1}^n \alpha_i x_i\right) = \sum_{i=1}^n \alpha_i \varphi(x_i) = \sum_{i=1}^n \alpha_i \varphi_0(x_i)$$

  so $\varphi(0) = \varphi_0(0) = 0$.

- $\Longleftarrow$: For any $x \in E$, define there exists some $\{x_i\}_{i \in I} \subset S$ s.t. $x = \sum_{i \in I} \alpha_i x_i$. Define

$$\varphi(x) = \sum_{i \in I} \alpha_i \varphi_0(x_i)$$

It is obvious that φ is that linear mapping.

□

Notice that if S is a basis of E, let φ_0 be a set map from S to E, then φ_0 can be extended in a unique way to a linear mapping $\varphi : E \to F$.

Proposition 2.3. *Let $\varphi : E \to F$ be a linear mapping and $\{x_\alpha\}$ be a basis of E. Then φ is a linear isomorphism iff the vectors $y_\alpha = \varphi(x_\alpha)$ form a basis for F.*

Proof.
- $\Longrightarrow$: As φ is a linear isomorphism, so for any $y \in F$, there exists a unique $x \in E$ s.t. $x = \varphi^{-1}(y)$. Notice that $\{x_\alpha\}$ is a basis, so $x = \sum_\alpha a_\alpha x_\alpha$ for some a_α, so $y = \varphi(x) = \varphi(\sum_\alpha a_\alpha x_\alpha) = \sum_\alpha a_\alpha \varphi(x_\alpha)$. That means $\{\varphi(x_\alpha)\}$ generates F. Then we need to prove the linear independence.
  Let $\sum_\alpha \lambda_\alpha x_\alpha = 0$, then $\lambda_\alpha = 0$ for each α. Then let $\sum_\alpha \gamma_\alpha \varphi(x_\alpha) = 0$, then

$$\sum_\alpha \gamma_\alpha \varphi(x_\alpha) = \varphi\left(\sum_\alpha \gamma_\alpha x_\alpha\right) = 0$$

  so $\sum_\alpha \gamma_\alpha x_\alpha = 0$ which means $\gamma_\alpha = 0$ for each α. So $\{\varphi(x_\alpha)\}$ is a basis of F.

- $\Longleftarrow$: Let $\{y_\alpha = \varphi(x_\alpha)\}$ be a basis of F, then for each $y \in F$, there exists a unique components (λ_α) s.t. $\sum_\alpha \lambda_\alpha y_\alpha = y$. Then we have

$$\sum_\alpha \lambda_\alpha \varphi(x_\alpha) = \varphi\left(\sum_\alpha \lambda_\alpha x_\alpha\right) = \varphi(x)$$

  for some unique $x \in E$.

□

2.4 Subspace and factor space

2.4.1 Subspace and Sum

Definition 2.4 (Subspace)**.** Let X be a vector space and let $A \subset X$ be a subset of X. Then A is called a subspace if A is also a vector space.

Let S be a non-empty subset of X and there exists a set, noting as X_S, is the linear combination of any vectors in S, X_S is truly a subspace which is called **the subspace generated by** S or **linear closure** of S.

Proposition 2.4. *Let A_1, A_2 be two subspaces of the vector space X and suppose that $A_1 \cap A_2 \neq \varnothing$ then $A_1 \cap A_2$ is still a subspace of X.*

Proof. Notice that if $x \in A_1 \cap A_2$, then $x \in A_1$ and $x \in A_2$, and A_1, A_2 are vector space thus provide the linearity of $A_1 \cap A_2$.

□

Definition 2.5 (sum of subspace)**.** Let A_1, A_2 be two subspaces of a vector space X, then $\{x = x_1 + x_2 : x_1 \in A_1, x_2 \in A_2\}$ is called the **sum of** A_1 **and** A_2, denote as $A_1 + A_2$. It is easy to determine that $A_1 + A_2$ is still a subspace of X.

Notice that the decomposition is not determined uniquely.
Let $x = x_1 + x_2 = x_1' + x_2'$, then $x_1 - x_1' = x_2 - x_2' = z \in A_1 \cap A_2$. Only if $A_1 \cap A_2 = \{0\}$, then $x = x_1 + x_2$ is uniquely determined. In this time, we called that sum as **direct sum** of A_1 and A_2, denote as $A_1 \oplus A_2$.

Proposition 2.5.

- *Let A_1, A_2 be subspaces of X and let S_1, S_2 be systems of generators of A_1 and A_2, then $S_1 \cup S_2$ generates $A_1 + A_2$.*
- *Suppose that $A_1 \cap A_2 = \{0\}$ and T_1, T_2 are basis of A_1, A_2, then $T_1 \cup T_2$ is the basis of $A_1 \oplus A_2$.*

Proof. Give any $x \in A_1 + A_2$, then $x = x_1 + x_2$ for some $x_1 \in A_1, x_2 \in A_2$. $x_1 = \sum_\alpha \lambda_\alpha x_\alpha$ for some $x_\alpha \in S_1$ and $x_2 = \sum_\beta \gamma_\beta x_\beta$ for some $x_\beta \in S_2$, so $x = \sum_\alpha \lambda_\alpha x_\alpha + \sum_\beta \gamma_\beta x_\beta$, notice that every $x_\alpha, x_\beta \in S_1 \cup S_2$, so $S_1 \cup S_2$ generates $A_1 + A_2$.

Now we need to prove that $T_1 \cup T_2$ is linearly independent.
Notice that $T_1 \subset A_1, T_2 \subset A_2$, $A_1 \cap A_2 = \{0\}$, so $T_1 \cap T_2 = \{0\}$. So consider $x \in A_1 \oplus A_2$, $x = \sum_\alpha \lambda_\alpha x_\alpha + \sum_\beta \gamma_\beta x_\beta = 0$, then $A_1 \ni x_1 = \sum_\alpha \lambda_\alpha x_\alpha = -\sum_\beta \gamma_\beta x_\beta = x_2 \in A_2$, so $x_1 = x_2 = 0$, then as the property of basis, $\lambda_\alpha = 0$ for all α and $\gamma_\beta = 0$ for all β.

□

Definition 2.6 (complementary subspace)**.** If A_1 is a subspace of X, and there exists a subspace A_2 s.t. $A_1 \oplus A_2 = E$, then A_2 is called the **complementary subspace** for A_1 in X.

Proposition 2.6 (existence of complementary subspace)**.** *If $A_1 \subset X$ is a subspace, then there exists a $A_2 \subset X$ a subspace s.t. $A_1 \oplus A_2 = X$*

Proof. According to the 2.1, suppose that $\{x_\alpha\}$ is a basis of A_1, then it is linearly independent and so can be extended to a basis of X, denote as $\{x_\gamma\}$. Notice that $\{x_\alpha\} \subset \{x_\gamma\}$ and let $\{x_\beta\} = \{x_\gamma\} - \{x_\alpha\}$. Then let A_2 be the subspace generated by $\{x_\beta\}$.

Observe that $\{x_\alpha\}\cup\{x_\beta\}$ generates X, so $A_1+A_2=X$, then let $x\in A_1\cap A_2$, so $x=\sum_\alpha \lambda_\alpha x_\alpha=\sum_\beta \omega_\beta x_\beta$ which means $\sum_\alpha \lambda_\alpha x_\alpha+\sum_\beta(-\omega_\beta)x_\beta=0$. For vectors in $\{x_\alpha\}$ and $\{x_\beta\}$ are linearly independent, so $\lambda_\alpha=0,\omega_\beta=0$ for all α,β, then $A_1\cap A_2=\{0\}$ which means $X=A_1\oplus A_2$.

□

Corollary 2.3. *Let A_1 be a subspace of X and $\varphi_1 : A_1 \to F$ be a linear mapping. Then φ_1 may be extended to a linear mapping $\varphi : X \to F$.*

Proof. According to the above proposition, there exists a subspace $A_2 \subset X$ s.t. $A_1\oplus A_2=X$. Now define $\varphi_2 : A_2 \to F$ be a linear mapping. Then for any $x\in X$, notice that $x=x_1+x_2$ where $x_1\in A_1, x_2\in A_2$, define

$$\varphi(x)=\varphi_1(x_1)+\beta\varphi_2(x_2) \qquad x=x_1+x_2;\beta\in\mathbb{K}$$

It is easy to show that φ is a linear mapping as φ_1,φ_2 are.

□

2.4.2 Factor Space

Definition 2.7 (factor space)**.** Suppose that X is a vector space and A_1 is a subspace of X. Two vectors $x,x'\in X$ is called **equivalent** mod A_1 if $x-x'\in A_1$. Then $x\sim x'$ is a equivalence relation, that is reflexive, symmetric and transitive.

Then we let X/A_1 denote the **set of equivalence classes**, X/A_1 is a vector space too and define a mapping:

$$\pi : X \to X/A_1$$

by letting $\pi x=\overline{x}, x\in X$ where $\overline{x}$ denotes the equivalence class containing x. Clearly, π is a surjective mapping.

Proof. Now prove the equivalent relation:

- let $x\sim x_1, x_1\sim x_2$, which means $x-x_1\in A_1$ and $x_1-x_2\in A_1$ then $x-x_2=(x-x_1)+(x_1-x_2)\in A_1$.
- Notice that $x-x=0\in A_1$ as A_1 is a subspace.
- Observe that $x-x_1=(-1)(x_1-x)$ which means the symmetry.

□

Proposition 2.7. *There exists precisely one linear structure in X/A_1 s.t. π is a linear mapping.*

Proof. Assume that X/A_1 is made into a vector space s.t. π is a linear mapping. Then

$$\pi(x+y) = \pi(x) + \pi(y)$$

and $\pi(\lambda x) = \lambda\pi(x)$. It shows that we can use a linear mapping π to define the linear structure of X/A_1 and the linear structure of X/A_1 is determined by the linear structure of X, thus unique.

Now define the linear structure of X/A_1. Let $\overline{x}, \overline{y} \in X/A_1$ and $\overline{x} \neq \overline{y}$. Then there exists some $x, y \in X$ s.t. $\pi(x) = \overline{x}$ and $\pi(y) = \overline{y}$. Pick an arbitrary x and y, define:

$$\overline{x} + \overline{y} = \pi(x+y)$$

and

$$\lambda\overline{x} = \pi(\lambda x)$$

We only need to show that π is a linear mapping. Suppose that $x_1 - x_2 \in A_1$ and $y_1 - y_2 \in A_1$, notice that $(x_1 + y_1) - (x_2 + y_2) = (x_1 - x_2) + (y_1 - y_2) \in A_1$ as the property of subspace. Since the picking of x_1, x_2, y_1, y_2 is arbitrary, $\pi(x) = \overline{x}$, $\pi(x+y) = \overline{x} + \overline{y}$. Then π is a communicative group as above. Similarly, it is easy to show that $\pi(\lambda x) = \lambda\pi(x)$. Then π is linear, so it determines the linear structure of X/A_1.

□

Remark. The space discussed above like X/A_1 is called the factor space or quotient space and the linear mapping $\pi : X \to X/A_1$ is called the canonical projection of X onto A_1.

Definition 2.8. Let A_1 be a subspace of X, and suppose $\{x_\alpha\}$ is a family of vectors in X. Then x_α is called **linear dependent mod** A_1 if there are scalars λ_α not all zero s.t. $\sum_\alpha \lambda_\alpha x_\alpha \in A_1$.

A family of vectors is called linearly independent mod a subspace A_1 if they are not linearly dependent mod A_1.

Now consider the canonical projection $\pi : X \to X/A_1$, then $\{x_\alpha\}$ is linearly dependent mod A_1 iff the vectors $\pi(x_\alpha)$ are linearly dependent in X/A_1.

Proof.
- $\Longrightarrow$: Suppose $\{x_\alpha\}$ is linear dependent mod A_1, then $\sum_\alpha \lambda_\alpha x_\alpha \in A_1$ for not all zero λ_α, notice that the linearity of π,

$$\sum_\alpha \lambda_\alpha \pi(x_\alpha) = \pi\left(\sum_\alpha \lambda_\alpha x_\alpha\right)$$

  Observe that $\sum_\alpha \lambda_\alpha x_\alpha = x \in A_1$, and only if $x \in A_1$, $\pi(x) = \overline{0}$ in X/A_1.

- $\Longleftarrow$: Omission.

□

Suppose that $\{x_\alpha\} \cup \{x_\beta\}$ is a basis of X and $\{x_\alpha\}$ generates A_1, then according to 2.6 there exists a A_2 generated by $\{x_\beta\}$ s.t. $A_1 \oplus A_2 = X$.

Proposition 2.8 (basis of a factor space)**.** *$\pi(x_\beta)$ for all β form a basis of X/A_1.*

Proof. First, we need to prove that $\pi(x_\beta)$ generates X/A_1.
Let $\overline{x} \in X/A_1$ be an arbitrary element. We only need to find a $x \in \pi^{-1}(\overline{x})$, notice that if $\overline{x}$ is non-trivial i.e. $\overline{x} \neq \overline{0}$, $x \notin A_1$, so there must exist some γ_β s.t. $x = \sum_\beta \gamma_\beta x_\beta$. Then

$$\pi\left(\sum_\beta \gamma_\beta x_\beta\right) = \pi(x) = \overline{x} = \sum_\beta \gamma_\beta \pi(x_\beta)$$

Second, we observe that $\{x_\beta\}$ is linearly independent mod A_1, so $\pi(x_\beta)$ are linearly independent in X/A_1.

□

2.5 Inner Product spaces

Definition 2.9. Let X be a vector space, a function, $\langle \mathbf{x}, \mathbf{y} \rangle$, defined for all $\mathbf{x} \in X$ and $\mathbf{y} \in X$, is an **inner product** if for any $\mathbf{x}, \mathbf{y}, \mathbf{z} \in X$ and any $c \in \mathbb{R}$:

1. $\langle \mathbf{x}, \mathbf{x} \rangle \geq 0$ and equality holds iff $\mathbf{x} = \mathbf{0}$
2. $\langle \mathbf{y}, \mathbf{x} \rangle = \langle \mathbf{x}, \mathbf{y} \rangle$
3. $\langle \mathbf{x} + \mathbf{y}, \mathbf{z} \rangle = \langle \mathbf{x}, \mathbf{z} \rangle + \langle \mathbf{y}, \mathbf{z} \rangle$
4. $\langle c\mathbf{x}, \mathbf{y} \rangle = c\langle \mathbf{x}, \mathbf{y} \rangle$

2.5.1 Orthogonal

Two vectors are said to be **orthogonal** if $\langle \mathbf{x}, \mathbf{y} \rangle = 0$ and denoted as $\mathbf{x} \perp \mathbf{y}$ and $\mathbf{x} \perp X$ if $\mathbf{x} \perp \mathbf{y}$ for all $\mathbf{y} \in X$.

As one can apply Gram–Schmidt orthonormalization for a basis in a vector space equipped inner product, we have

Theorem 2.2. *Every finite dimensional non-trivial vector space has an orthogonal basis.*

Theorem 2.3. *Let $X \subset \mathbb{R}^m$ is a subspace with an orthogonal basis, then each $\mathbf{x} \in \mathbb{R}^m$ can be expressed uniquely as $\mathbf{x} = \mathbf{u} + \mathbf{v}$ where $\mathbf{u} \in X$ and $\mathbf{u} \perp X$*

Such $\mathbf{u}$ is known as the orthogonal projection of $\mathbf{x}$ onto X and such $\mathbf{v}$ is called **component** of $\mathbf{x}$ orthogonal to X. All orthogonal components is also a vector space.

Definition 2.10. Suppose S is a vector subspace of X then it's orthogonal component $S^\perp$ is collection of all vectors $\mathbf{x}$ in X *s.t.* $\mathbf{x} \perp S$.

One can easily check that an orthogonal component is also a vector subspace of X.

Theorem 2.4. $X = S \oplus S^\perp$

2.6 Dimension

Recall 2.1, every system of generators contains a basis, so if the generators of the system is finite, there exists a finite base of the space.

Definition 2.11 (dim)**.** Consider a vector space X whose basis is the family of finite number of vectors i.e. $\{x_1, \ldots, x_n\}$ generates X and $\sum_{i=1}^n \alpha_i x_i = 0$ whenever $\alpha_i = 0$ for every i. Then denotes the **dim of** X as $\dim X = n$.

Proposition 2.9. *Suppose a vector space X has a basis of n vectors. Then every family of $(n+1)$ vectors is linearly dependent. That means n is the maximum number of linearly independent vectors in X and hence every basis of X consists of n vectors.*

Proof. We use mathematical induction to prove this proposition.

1. Let $n = 1$, let x_1 be a basis of X, then $y_1, y_2 \neq 0$ and $y_1, y_2 \in X$. Then $y_1 = \alpha x, y_2 = \beta x$. Now let $\gamma_1 y_1 + \gamma_2 y_2 = 0$, we can let $\gamma_1 = \alpha\beta, \gamma_2 = -\alpha\beta$ which means y_1, y_2 are linearly dependent.
2. Assume that the proposition holds for every vector space having basis of $r \leq n-1$ vectors by the induction.

3. Let X be a vector space and let $\{x_1, \dots, x_n\}$ be the basis of X and $\{y_1, \dots, y_{n+1}\}$ be an arbitrary family of vectors in X.
   Now consider the factor space $X/\operatorname{span} y_{n+1}$ and the canonical projection $\pi : X \to X/\operatorname{span} y_{n+1}$. As $\{x_i : i = 1, \dots, n\}$ generates X and π is surjective, $\{\pi(x_i) : i = 1, \dots, n\}$ generates $X_1 = X/\operatorname{span} y_{n+1}$, so according to 2.1, it contains a basis of X_1 and as $y_{n+1} = \sum_{i=1}^{n} \alpha_i x_i$ for some not all zero α_i, $\{\overline{x_i} = \pi(x_i) : i = 1, \dots, n\}$ is linearly dependent, so $\dim X_1 \le n-1$, then by the hypothesis of induction, $\{\overline{y_i} = \pi(y_i) : i = 1, \dots, n\}$ are linearly independent.
   so there exists:
   $$\sum_{i=1}^{n} \gamma_i \overline{y_i} = 0 \text{ for non-trivial } \{\gamma_i\}$$
   which means $\{y_i : i = 1, \dots, n\}$ are linearly dependent mod $\operatorname{span} y_{n+1}$ which means
   $$\sum_{i=1}^{n} \gamma_i y_i = \lambda y_{n+1}$$
   leads to the consult that $\{y_1, \dots, y_{n+1}\}$ are linearly dependent.

□

Give a vector space X and a subspace $A_1 \subset X$, then there exists a subspace $A_2 \subset X$ s.t. $A_1 \oplus A_2 = X$ by 2.6. Then let $\{x_\alpha\}$ be a basis of A_1 and $\{x_\beta\}$ be a basis of A_2, notice that $\{x_\alpha\} \cap \{x_\beta\} = \varnothing$ and $\{x_\alpha\} \cup \{x_\beta\}$ generates X. So we easily observe that $\dim X = \dim A_1 + \dim A_2$ if $A_1 \oplus A_2 = X$.

Then according to 2.8, let π be the canonical projection, $\{\overline{x_\beta} = \pi(x_\beta)\}$ forms a basis of X/A_1, so $\dim(X/A_1) = \operatorname{card}\{\overline{x_\beta}\} = \operatorname{card}\{x_\beta\} = \dim A_2$. So $\dim X = \dim A + \dim(X/A_1)$.

Proposition 2.10. *Let $A_1, A_2 \subset X$ be arbitrary subspace of X. Then*

$$dim\, A_1 + dim\, A_2 = dim(A_1 + A_2) + dim(A_1 \cap A_2)$$

Proof. Just let $\{x_\alpha\}$ be the basis of $A_1 \cap A_2$ and let $\{y_\beta\}, \{y_\gamma\}$ be the extending tail i.e. they don't intersect $\{x_\alpha\}$ and $\{x_\alpha\} \cup \{y_\beta\}$ is a basis of A_1 and $\{x_\alpha\} \cup \{y_\gamma\}$ is a basis of A_2.

Let $\operatorname{card}\{x_\alpha\} = \alpha, \operatorname{card}\{y_\beta\} = \beta, \operatorname{card}\{y_\gamma\} = \gamma$. Then $\dim A_1 = \alpha + \beta, \dim A_2 = \alpha + \gamma, \dim(A_1 \cap A_2) = \alpha$. Now we only need to show that $\{x_\alpha\} \cup \{y_\beta\} \cup \{y_\gamma\}$ generates $A_1 + A_2$. It is easy to show by the definition of generators of system. And notice that they are independent with each other. Thus $\{x_\alpha\} \cup \{y_\beta\} \cup \{y_\gamma\}$ is a basis of $A_1 + A_2$ which means $\dim(A_1 + A_2) = \operatorname{card}(\{x_\alpha\} + \{y_\beta\} + \{y_\gamma\}) = \alpha + \beta + \gamma$.

□

2.7 Convex sets

Convex set is a special type subset of a vector space.

Definition 2.12. A set $S \subset \mathbb{R}^m$ is said to be **convex** iff for any $\mathbf{x_1}, \mathbf{x_2} \in S$ and $0 < c < 1$, we have

$$c\mathbf{x_1} + (1-c)\mathbf{x_2} \in S$$

Proposition 2.11. *Suppose $S_1, S_2 \subset \mathbb{R}^m$ and convex, then so is $S_1 \cap S_2$ and $S_1 + S_2$.*

For any set S, the smallest convex contains it is called **convex hull** of S and denoted as $C(X)$.

Theorem 2.5. *If S is convex, so is $\overline{S}$ and $S^\circ = \overline{S}^\circ$*

Lemma 2.1. *Let S be a closed convex set of $\mathbb{R}^m$ and $\mathbf{0} \notin S$, then there exists $\mathbf{a} \in \mathbb{R}^m$ s.t. $\mathbf{a'x} > 0$ for all $\mathbf{x} \in S$.*

Definition 2.13. Let $S_1, S_2 \in \mathbb{R}^m$ be convex and $S_1 \cap S_2 = \varnothing$. Then there exists $\mathbf{b} \neq 0 \in \mathbb{R}^m$ which separate S_1 and S_2.

2.8 Matrix and linear space

Definition 2.14. Let $\mathbf{X}$ be matrix in $\mathbb{R}^{m\times n}$. The subspace of $\mathbb{R}^n$ spanned by the m rows of $\mathbf{X}$ is called the **row space** of $\mathbf{X}$ and denoted as $\mathcal{R}(\mathbf{X})$ and that of $\mathbb{R}^m$ is column space and denoted as $\mathcal{C}(\mathbf{X})$

The column(row) space often equipped:

- Inner product: $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x'Ay}$, $\mathbf{A} = \mathbf{I}$ usually.
- Norm: $\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$
- Metric: $d(\mathbf{x}, \mathbf{y}) = \sqrt{\langle \mathbf{x}-\mathbf{y}, \mathbf{x}-\mathbf{y} \rangle}$

The column space of $\mathbf{X}$ is sometimes also referred to as the **range** or **image** of $\mathbf{X}$. Note

$$\mathcal{C}(\mathbf{X}) = \{\mathbf{y} : \mathbf{y} = \mathbf{Xa}, \mathbf{a} \in \mathbb{R}^n\}$$

Clearly, the rank of $\mathbf{X}$ is just the dimension of $\mathcal{C}(\mathbf{X})$ and that agree with $\dim \mathcal{C}(\mathbf{X'})$, *i.e.*, the number of independent columns of $\mathbf{X}$. The null space $\mathcal{N}(\mathbf{X})$ is the orthogonal space of $\mathcal{C}(\mathbf{X'})$.

Proposition 2.12. *Let* $\mathbf{A} \in \mathbb{R}^{m\times m}$*, then:*

1. $rank(\mathbf{AB}) \le rank(\mathbf{A}) \wedge rank(\mathbf{B})$
2. $|rank(\mathbf{A}) - rank(\mathbf{B})| \le rank(\mathbf{A}+\mathbf{B}) \le rank(\mathbf{A}) + rank(\mathbf{B})$
3. $rank(\mathbf{A}) = rank(\mathbf{A}') = rank(\mathbf{AA}') = rank(\mathbf{A}'\mathbf{A})$

Proof. 1. Note $\mathbf{AB}$ can be seen as linear transformation in $\mathcal{C}(X)$ or so in $\mathcal{C}(X')$ and claim follows.

2. Note

$$\mathbf{A}+\mathbf{B} = \begin{bmatrix}\mathbf{A} & \mathbf{B}\end{bmatrix}\begin{bmatrix}\mathbf{I}\\ \mathbf{I}\end{bmatrix}$$

So property 1 applies and conclude:

$$\text{rank}(\mathbf{A}+\mathbf{B}) \le \text{rank}\left(\begin{bmatrix}\mathbf{A} & \mathbf{B}\end{bmatrix}\right) \le \text{rank}(\mathbf{A}) + \text{rank}(\mathbf{B})$$

Replace $\mathbf{A}$ and $\mathbf{B}$ by $\mathbf{A}+\mathbf{B}$ and $-\mathbf{B}$, we have

$$\text{rank}(\mathbf{A}) \le \text{rank}(\mathbf{A}+\mathbf{B}) + \text{rank}(\mathbf{B})$$

And similar result also hold for $\mathbf{B}$ and then claim follows.

3. It's sufficient to show $\text{rank}(\mathbf{A}) = \text{rank}(\mathbf{A}'\mathbf{A})$ and it's enough to show

$$\mathcal{N}(\mathbf{A}) = \mathcal{N}(\mathbf{A}'\mathbf{A})$$

To see that, note $\mathbf{Ax} = \mathbf{0} \implies \mathbf{A}'\mathbf{Ax} = \mathbf{0}$ clearly and if $\mathbf{A}'\mathbf{Ax} = \mathbf{0}$ we have $\mathbf{x}'\mathbf{A}'\mathbf{Ax} = \mathbf{0}$ and thus $\|\mathbf{A}'\mathbf{x}\| = 0$ and there must be $\mathbf{Ax} = \mathbf{0}$.

□

Proposition 2.13. *Let* $\mathbf{A}, \mathbf{B}, \mathbf{C}$ *are any matrices s.t. all the block matrix involved are defined. We have*

1. $rank\left(\begin{bmatrix}\mathbf{A} & \mathbf{B}\end{bmatrix}\right) \ge rank(\mathbf{A}) \vee rank(\mathbf{B})$
2. $rank\left(\begin{bmatrix}\mathbf{A} & \mathbf{0}\\ \mathbf{0} & \mathbf{B}\end{bmatrix}\right) = rank\left(\begin{bmatrix}\mathbf{0} & \mathbf{A}\\ \mathbf{B} & \mathbf{0}\end{bmatrix}\right) = rank(\mathbf{A}) + rank(\mathbf{B})$
3. $rank\left(\begin{bmatrix}\mathbf{A} & \mathbf{0}\\ \mathbf{C} & \mathbf{B}\end{bmatrix}\right) = rank\left(\begin{bmatrix}\mathbf{C} & \mathbf{A}\\ \mathbf{B} & \mathbf{0}\end{bmatrix}\right) = rank\left(\begin{bmatrix}\mathbf{B} & \mathbf{C}\\ \mathbf{0} & \mathbf{A}\end{bmatrix}\right) = rank\left(\begin{bmatrix}\mathbf{0} & \mathbf{A}\\ \mathbf{B} & \mathbf{C}\end{bmatrix}\right) \ge rank(\mathbf{A}) + rank(\mathbf{B})$

Theorem 2.6. *Let* $\mathbf{B}$ *be matrix in* $\mathbb{R}^{m\times n}$ *and* $\mathbf{A}, \mathbf{C}$ *justify the matrix multiplication:*

$$rank(\mathbf{ABC}) \ge rank(\mathbf{AB}) + rank(\mathbf{BC}) - rank(\mathbf{B})$$

Proof. Note by some linear transformation, we have

$$\begin{bmatrix}\mathbf{B} & \mathbf{0}\\ \mathbf{0} & \mathbf{ABC}\end{bmatrix} \to \begin{bmatrix}\mathbf{B} & \mathbf{BC}\\ \mathbf{AB} & \mathbf{0}\end{bmatrix}$$

and claim follows by proposition 2.13.3.

□

Take $\mathbf{B} = \mathbf{I}$, we have

Corollary 2.4. *If $A \in \mathbb{R}^{m\times n}, B \in \mathbb{R}^{n\times p}$*

$$rank(\mathbf{AB}) \geq rank(\mathbf{A}) + rank(\mathbf{B}) - n$$

2.8.1 Projection Matrix

On the space $\mathbb{R}^m$, there exist projection matrix:

Proposition 2.14. *Suppose $\mathbf{Q}$ is orthogonal matrix, then $\mathbf{QQ'}$ is a projection on $\mathcal{C}(\mathbf{Q})$.*

Such matrix is called **projection matrix** for the space S(if $S = \mathcal{C}(\mathbf{Q})$) and denoted as $\mathbf{P_S}$. Note for fixed S, the orthogonal basis $\mathbf{Q}$ can be various, the projection matrix is unique.

Proposition 2.15. *Suppose $\mathbf{Q_1}$ and $\mathbf{Q_2}$ are orthogonal matrices, and $\mathcal{C}(\mathbf{Q_1}) = \mathcal{C}(\mathbf{Q_2})$, then $\mathbf{Q_1Q_1'} = \mathbf{Q_2Q_2'}$*

Recall the Gram-Schmidt orthonormalization apply linear transformation on $\mathbf{X}$ to finally get orthogonal $\mathbf{Q}$, such process can be represented as

$$\mathbf{Q} = \mathbf{XA}$$

Note $\mathbf{I} = \mathbf{Q'Q} = \mathbf{A'X'XA}$ and $\mathbf{A}$ is full rank square matrix, we have $\mathbf{AA'} = \mathbf{(X'X)^{-1}}$. Consequently:

$$\mathbf{P_X} = \mathbf{QQ'} = \mathbf{X(X'X)^{-1}X'}$$

In fact, $\mathbf{A}$ must be upper triangle and $\mathbf{X} = \mathbf{QA^{-1}}$ is the so called QR decomposition.

Note the projection matrix is symmetric and idempotent, we can show that it's precisely characterization of projection matrix:

Proposition 2.16. *If $\mathbf{P}$ is symmetric and idempotent, then there is a vector space X has $\mathbf{P}$ as projection matrix, and $dim X = rank(\mathbf{P})$.*

Proof.

Lemma 2.2. *Let* $\mathbf{A} \in \mathbb{R}^{m\times n}$ *with rank* r*, then there exists full rank* $F \in \mathbb{R}^{m\times r}$ *and* $G \in \mathbb{R}^{r\times n}$ *s.t.* $\mathbf{A} = \mathbf{FG}$.

By above lemma, we have $\mathbf{P} = \mathbf{FG}$, since $\mathbf{P}$ is idempotent then we have

$$\begin{aligned}
\mathbf{FGFG} = \mathbf{FG} &\implies \mathbf{F'FGFGG'} = \mathbf{F'FGG'} \\
&\implies \mathbf{GF} = \mathbf{I} \implies \mathbf{FGF} = \mathbf{F} \\
&\implies (\mathbf{FG})'\mathbf{F} = \mathbf{G'F'F} = \mathbf{F} \\
&\implies \mathbf{G'} = (\mathbf{F'F})^{-1}\mathbf{F} \\
&\implies \mathbf{P} = \mathbf{F}(\mathbf{F'F})^{-1}\mathbf{F'}
\end{aligned}$$

Thus $\mathbf{P}$ be projection on $\mathcal{C}(\mathbf{F})$. This completes the proof.

□

Now we extend orthogonal projection to oblique case, where $X = S \oplus T$ still but $T \neq S^{\perp}$.

Definition 2.15. Suppose $S \oplus T = \mathbb{R}^m$ and $\mathbf{x} = \mathbf{s} + \mathbf{t}$ where $\mathbf{x} \in \mathbb{R}^m, \mathbf{s} \in S, \mathbf{t} \in T$, then $\mathbf{s}$ is called **projection** on S along T while $\mathbf{t}$ is so on T along S.

Suppose $\mathbf{X} = \begin{bmatrix}\mathbf{S} & \mathbf{T}\end{bmatrix}$ is nonsingular where $\mathbf{S} \in \mathbb{R}^{m\times s}, \mathbf{T} \in \mathbb{R}^{m\times t}$, we have

$$\mathbf{X^{-1}S} = \begin{bmatrix}\mathbf{I} \\ \mathbf{0}\end{bmatrix}, \mathbf{X^{-1}T} = \begin{bmatrix}\mathbf{0} \\ \mathbf{I}\end{bmatrix}$$

They are orthogonal. Thus for arbitrary $\mathbf{y} \in \mathbb{R}^m$, it can be unique expressed as $\mathbf{X^{-1}Sa} + \mathbf{X^{-1}Tb}$. To get the oblique projection, for any $\mathbf{x} \in \mathbb{R}^m$, find $\mathbf{Xy} = \mathbf{x}$, then

$$\mathbf{x} = \mathbf{Xy} = \mathbf{X}(\mathbf{X^{-1}Sa} + \mathbf{X^{-1}Tb}) = \mathbf{Sa} + \mathbf{Tb}$$

The oblique projection matrix is something map $\mathbf{x}$ to $\mathbf{Sa}$ and denoted as $\mathbf{P_{S|T}}$. Note we have orthogonal projection matrix $\mathbf{P}$ map $\mathbf{y}$ to $\mathbf{X^{-1}Sa}$, thus

$$\mathbf{P_{S|T}} = \mathbf{XPX^{-1}} = \mathbf{X}\begin{bmatrix}\mathbf{I}_s & \mathbf{0} \\ \mathbf{0} & \mathbf{0}\end{bmatrix}\mathbf{X}$$

Clearly, $\mathbf{P_{S|T}}$ is still idempotent but not symmetric, unless $S \perp T$.

Another generalization of projection is define $x \perp y$ iff $\mathbf{x'Ay} = 0$, where $\mathbf{A}$ is positive definite and so we have some invertible $\mathbf{B}$ *s.t.* $\mathbf{A} = \mathbf{B'B}$.

Definition 2.16. Then for any $\mathbf{x} \in \mathbb{R}^m$, suppose it can be expressed as $\mathbf{x} = \mathbf{s} + \mathbf{t}$ *s.t.* $\mathbf{s} \in S$ and $\mathbf{s'At} = 0$, then such $\mathbf{s}$ is the orthogonal projection onto S relative A.

We will see both generalization agree.

Let $U = \{\mathbf{z} : \mathbf{z} = \mathbf{Bs}, \mathbf{s} \in S\}$, for decomposition $\mathbf{x} = \mathbf{s} + \mathbf{t}$, we have $\mathbf{Bx} = \mathbf{Bs} + \mathbf{Bt}$, where

$$\mathbf{s'B'Bt} = \mathbf{sAt} = \mathbf{0}$$

Thus $\mathbf{Bt} \in U^{\perp}$, by the uniqueness of orthogonal projection, this generalization is also unique. And if $S = \mathcal{C}(X)$, then $U = \mathcal{C}(BX)$, thus the projection onto U is:

$$\mathbf{P} = \mathbf{BX(X'AX)^{-1}X'B'}$$

which map Bx to Bs and that implies the projection onto S relative to $\mathbf{A}$ is:

$$\mathbf{P} = \mathbf{X(X'AX)^{-1}XA}$$

Definition 2.15 and definition 2.16 agree since in definition 2.15 $\mathbf{X} = \begin{bmatrix} \mathbf{S} & \mathbf{T} \end{bmatrix}$ then $\mathbf{X^{-1}S} \perp \mathbf{X^{-1}T}$ and we have $\mathbf{(X^{-1}Sa)'X^{-1}Tb} = \mathbf{a'S'X^{-1\prime}X^{-1}Tb} = \mathbf{s(XX')^{-1}t} = \mathbf{0}$, that relate to definition 2.16 clearly. For the other direction, it's clear as $\mathbf{P_{T|S}} = \mathbf{I} - \mathbf{P}$.

We can see that $\mathbf{s}$ is the nearest with $\mathbf{x}$, since for any $\mathbf{y} \in S$:

$$\begin{aligned}
d(\mathbf{x}, \mathbf{y}) &= d(\mathbf{x} - \mathbf{s}, \mathbf{y} - \mathbf{s}) \\
&= \mathbf{(x - s)'A(x - s)} + \mathbf{(s - y)'A(s - y)} + 2\mathbf{(x - s)'A(s - y)} \\
&= \mathbf{(x - s)'A(x - s)} + \mathbf{(s - y)'A(s - y)} \\
&\geq \mathbf{(x - s)'A(x - s)} = d(\mathbf{x}, \mathbf{s})
\end{aligned}$$

2.8.2 Linear transformation

All linear mappings $\varphi : \mathbb{R}^n \to \mathbb{R}^m$ can be presented as a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ *s.t.* $\varphi(\mathbf{x}) = \mathbf{Ax}$.

Chapter 3

Linear Mappings

3.1 Basic properties

Definition 3.1 (kernel and image)**.** Suppose X, Y are vector spaces and $\varphi : E \to F$ be a linear mapping. Then the **kernel of** φ denoted as $\ker\varphi$ is the subset $K \subset X$ s.t. if $x \in K \implies \varphi(x) = 0$.

The **image space of** φ denoted as $\operatorname{Im}\varphi$ is the subset $I \subset Y$ s.t. $y \in I \implies$ there exists some $x \in X$ s.t. $\varphi(x) = y$.

Proposition 3.1. *1. Let $\varphi : X \to Y$ be a linear mapping, then* $\ker\varphi$ *is a vector space.*

2. The mapping $\varphi : X \to Y$ is injective iff $\ker\varphi = \{0\}$.

Proof. 1. Let $\varphi : X \to Y$ be a linear mapping, let $x_1, x_2 \in \ker\varphi$. Then

- $\varphi(x_1 + x_2) = \varphi(x_1) + \varphi(x_2) = 0$, so $x_1 + x_2 \in \ker\varphi$.
- $\varphi(\alpha x_1) = \alpha\varphi(x_1) = 0$, so $\alpha x_1 \in \ker\varphi$.

2. Let φ be injective that means for each $y \in \operatorname{Im}\varphi$, $\varphi^{-1}(y) = x$ for some unique $x \in X$. So $\varphi^{-1}(0) = 0$ for only $0 \in X$.
   For the converse, let $\ker\varphi = \{0\}$, give an arbitrary $y \in \operatorname{Im}\varphi$, suppose there exists $x_1, x_2 \in X$ s.t. $\varphi(x_1) = \varphi(x_2) = y$, then $\varphi(x_1 - x_2) = \varphi(x_1) - \varphi(x_2) = 0$, if $x_1 \neq x_2$, there leads to a contradiction about $\ker\varphi = \{0\}$. So φ is injective.

$\square$

3.1.1 Induced Linear Mappings

Definition 3.2 (restriction of linear mapping)**.** Suppose $\varphi : X \to Y$ is a linear mapping and $X_1 \subset X$, $Y_1 \subset Y$ are subspace s.t. $\varphi(x) \in Y_1$ when $x \in X_1$.
Then the linear mapping $\varphi_1 : X_1 \to Y_1$ defined by $\varphi_1(x) = \varphi(x), x \in X_1$ is called **the restriction of** φ to X_1.

Now we can find that $\varphi \circ i_{X_1} = i_{Y_1} \circ \varphi_1$ where $i_{X_1} : X_1 \to X$ is canonical injections, same as i_{Y_1}.

Equivalently, the diagram is commutative.

$$\begin{tikzcd} X \arrow[r, "\varphi"] & Y \\ X_1 \arrow[u, "i_X"] \arrow[r, "\varphi_1"] & Y_1 \arrow[u, "i_Y"'] \end{tikzcd}$$

Let $\varphi : X \to Y$ be linear mapping and $\varphi_1 : X_1 \to Y_1$ be its restriction to subspace $X_1 \subset X, Y_1 \subset Y$. Then there exists precisely one linear mapping

$$\overline{\varphi} : X/X_1 \to Y/Y_1$$

s.t.

$$\overline{\varphi} \circ \pi_X = \pi_Y \circ \varphi$$

where π_X, π_Y are canonical projections on X, Y.

Notice that $\pi_Y(\varphi(x_1)) = \pi_Y(\varphi(x_2))$ whenever $\pi_X(x_1) = \pi_X(x_2)$. Because $\pi_X(x_1) = \pi_X(x_2)$ implies $\pi_X(x_1 - x_2) = \overline{0}$ so $x_1 - x_2 \in \ker \pi_X = X_1$. Then

$$\begin{aligned} \pi_Y \circ \varphi(x_2 - x_1) &= \pi_Y \circ \varphi(x) && \text{for } x \in X_1 \\ &= \pi_Y(y) && \text{for } y \in Y_1 \\ &= \overline{0} \end{aligned}$$

as the existence of the restriction φ_1.

Then we can assert that there exists a mapping s.t. $\overline{\varphi}(x)$ has only one value in Y/Y_1, thus a function. Then we need to show its linearity. Now let $\overline{x_1}, \overline{x_2} \in X/X_1$ and $x_1 \in \pi_X^{-1}(\overline{x_1})$ same as x_2.

$$\begin{aligned} \overline{\varphi}(\alpha\overline{x_1} + \beta\overline{x_2}) &= \overline{\varphi} \circ \pi_X(\alpha x_1 + \beta x_2) \\ &= \pi_Y \circ \varphi(\alpha x_1 + \beta x_2) \\ &= \alpha\pi_Y \circ \varphi(x_1) + \beta\pi_Y \circ \varphi(x_2) \\ &= \alpha\overline{\varphi}(\overline{x_1}) + \beta\overline{\varphi}(\overline{x_2}) \end{aligned}$$

which means the linearity.

Remark. The $\overline{\varphi}$ discussed above is called the **induced mapping in factor space** and the relation of $\overline{\varphi}$ is equivalent to the diagram:

$$\begin{tikzcd} X \arrow[r, "\varphi"] \arrow[d, "\pi_X"] & Y \arrow[d, "\pi_Y"] \\ X/X_1 \arrow[r, "\overline{\varphi}"] & Y/Y_1 \end{tikzcd}$$

Notice that this diagram is commutative.

And the relation can be overwritten by $\overline{\varphi x} = \overline{\varphi x}$.

Let $\varphi : X \to Y$ be a linear mapping and $X_1 = \ker \varphi$, $Y_1 = \{0\}$. Since $\varphi(x) = 0$ when $x \in X_1$, a linear mapping is **induced** by φ :

$$\overline{\varphi} : X/\ker \varphi \to Y/\{0\} = Y$$

s.t.

$$\overline{\varphi} \circ \pi = \varphi$$

where $\pi : X \to X/\ker \varphi$ is the canonical projection.

1. This mapping $\overline{\varphi}$ is injective. In fact if $\overline{\varphi} \circ \pi(x) = 0$, then $\varphi(x) = 0$ which means $x \in \ker \varphi$. Then $\pi(x) = \overline{0}$, so $\ker \overline{\varphi} = \{\overline{0}\}$, according to 3.1, $\overline{\varphi}$ is injective.
2. $\overline{\varphi}$ is a linear isomorphism between $X/\ker \varphi$ and $\operatorname{Im} \varphi$, i.e.

$$\overline{\varphi} : X/\ker \varphi \xrightarrow{\simeq} \operatorname{Im} \varphi$$

   Notice that $\overline{\varphi}$ is injective and since $\operatorname{Im} \varphi$ it is surjective, thus one-to-one and onto.

Then every linear mapping $\varphi : X \to Y$ can be written as a composition of a surjective and injective linear mapping:

$$\begin{tikzcd} X \arrow[r, "\varphi"] \arrow[d, "\pi"] & Y \\ X/\ker \varphi \arrow[ur, "\overline{\varphi}"] & \end{tikzcd}$$

Now consider the linear mapping:

$$\varphi' : X_1/(X_1 \cap X_2) \xrightarrow{\simeq} (X_1 + X_2)/X_2$$

We need to show it is a isomorphism.
First we observe the canonical projection:

$$\pi : X_1 + X_2 \to (X_1 + X_2)/X_2$$

and $\pi \mid_{X_1}$ be the restriction on X_1. Notice that for $x \in X_1 + X_2$:

$$x = x_1 + x_2 \qquad x_1 \in X_1, x_2 \in X_2$$

then

$$\pi(x) = \pi(x_1 + x_2) = \pi(x_1) = \pi \mid_{X_1} (x_1)$$

So we find that $\pi \mid_{X_1}$ is surjective.
Define $\varphi = \pi \mid_{X_1}: X_1 \to (X_1 + X_2)/X_2$, then

$$\ker \varphi = \ker \pi \cap X_1 = X_1 \cap X_2$$

With the above discussion, we notice that $\varphi : X_1 \to (X_1 + X_2)/X_2$ and so

$$X_1/\ker \varphi \xrightarrow{\simeq} (X_1 + X_2)/X_2$$

Proposition 3.2. *Suppose that $\varphi : X \to Y$ and $\psi : X \to Z$ are linear mappings s.t. $\ker \varphi \subset \ker \psi$, then there exists a linear mapping $\omega : X \to Z$ s.t. $\omega \circ \varphi = \psi$.*

Proof. Notice that $\psi(x) = 0$ if $x \in \ker \varphi$, consider the induced linear mapping:

$$\overline{\psi} : X/\ker \varphi \to Z$$

s.t. $\overline{\psi} \circ \pi = \psi$ where $\pi : X \to X/\ker \varphi$ is the canonical projection. The existence of $\overline{\psi}$ is determined by the $\psi \mid_{\ker \varphi}: \ker \varphi \to \{0\}$.

Now let

$$\overline{\varphi} : X/\ker \varphi \xrightarrow{\simeq} \operatorname{Im} \varphi$$

be the linear isomorphism determined by φ and define $\overline{\psi}_1 : \operatorname{Im} \varphi \to Z$ by

$$\overline{\psi}_1 = \overline{\psi} \circ \overline{\varphi}^{-1}$$

Then let $\omega : X \to Z$ be a linear mapping which extends $\overline{\psi}_1$.
Notice that

$$\overline{\varphi}^{-1} \circ \varphi = \overline{\varphi}^{-1} \circ \overline{\varphi} \circ \pi = \pi$$

which means:

$$\omega \circ \varphi = \overline{\psi}_1 \circ \varphi = \overline{\psi} \circ \overline{\varphi}^{-1} \circ \varphi = \overline{\psi} \circ \pi = \psi$$

□

Remark. The result can be expressed in commutative diagram:

$$\begin{tikzcd}
X \arrow[r, "\varphi"] \arrow[d, "\psi"] & Y \arrow[dl, "\omega"] \\
Z &
\end{tikzcd}$$

Matrix Analysis

Chapter 4

Eigenvalues

Suppose $\mathbf{A} \in \mathbb{R}^m$, if $\mathbf{Ax} = \lambda\mathbf{x}$, we say λ eigenvalue of $\mathbf{A}$ and $\mathbf{x}$ is eigenvector of $\mathbf{A}$. To find λ, we solve following characteristic equation of $\mathbf{A}$:

$$|\mathbf{A} - \lambda\mathbf{I}| = 0$$

Recall the Fundamental theorem of algebra, there is m eigenvalues and the times of λ repeated is called **algebraic multiplicity**, or multiplicity for short and denoted as $\mu_{\mathbf{A}}(\lambda)$.

Note the eigenvector for a eigenvalue λ is not unique, in fact, all of them formed a vector space.

Theorem 4.1. *If $S_{\mathbf{A}}(\lambda)$ is all eigenvectors of $\mathbf{A}$ corresponding to λ, then $S_{\mathbf{A}}(\lambda)$ is a vector space.*

The dimension of eigenspace of λ is called **geometric multiplicity** of λ and deonted as $\gamma_{\mathbf{A}}(\lambda)$.

Following are frequently using:

Proposition 4.1. *Let $\mathbf{A} \in \mathbb{R}^{m\times m}$, λ is it's eigenvalue, then the following holds:*

1. *The eigenvalues of $\mathbf{A'}$ are the same as that of $\mathbf{A}$.*
2. *$\mathbf{A}$ is singular iff 0 is a eigenvalues.*
3. *The eigenvalues of $\mathbf{BAB^{-1}}$ are the same as $\mathbf{A}$.*
4. *If $\mathbf{A}$ is orthogonal, $|\lambda_i| = 1$.*
5. *$1 \leq \gamma_{\mathbf{A}}(\lambda) \leq \mu_{\mathbf{A}}(\lambda) \leq m$.*
6. *λ^n is an eigenvalue of $\mathbf{A}^n$ and the eigenspace remain the same, where n can be negative when $\mathbf{A}$ is invertible.*
7. *$tr(\mathbf{A}) = \sum_{i=1}^m \lambda_i$, $|\mathbf{A}| = \prod_{i=1}^m \lambda_i$.*
8. *$\sigma_{\mathbf{AB}} = \sigma_{\mathbf{BA}}$ if ignore zero eigenvalues.*

Proof. **7**. Recall the characteristic equation of the form:

$$(-\lambda)^m + \alpha_{m-1}(-\lambda)^{m-1} + \cdots + \alpha_1(-\lambda) + \alpha_0 = 0$$

By the Vieta's formulas,

$$\sum_{i=1}^{m} \lambda_i = \alpha_{m-1}, \prod_{i=1}^{n} \lambda_i = \alpha_0$$

For α_{m-1}, by the definition of determinant, it comes from term $\prod_{i=1}^{m}(a_{ii}-\lambda)$ and thus equal to $\sum_{i=1}^{m} a_{ii} = \operatorname{tr}(\mathbf{A})$. For α_0, let $\lambda = 0$ in above equation and we have $|\mathbf{A}| = \alpha_0$. This completes the proof.

□

Proposition 4.2. *Let* $\mathbf{A} \in \mathbb{R}^{m\times m}$ *and symmetric,* $\mathbf{c}, \mathbf{d} \in \mathbb{R}^m$*, then*

$$|\mathbf{A} + \mathbf{cd'}| = |\mathbf{A}|\,(1 + \mathbf{d'A^{-1}c})$$

Proof.

$$|\mathbf{A} + \mathbf{cd'}| = \left|\mathbf{A}(\mathbf{I} + \mathbf{A^{-1}cd'})\right| = |\mathbf{A}|\left|\mathbf{I} + \mathbf{A^{-1}cd'}\right| = |\mathbf{A}|\,(1 + \mathbf{c'A^{-1'}d}) = |\mathbf{A}|\,(1 + \mathbf{d'A^{-1}c})$$

where we use the truth:

Lemma 4.1. $|\mathbf{I} + \mathbf{bd'}| = 1 + \mathbf{d'b}$

Since for any orthogonal vector $\mathbf{x}$ to $\mathbf{d}$, $(\mathbf{I} + \mathbf{bd'})\mathbf{x} = \mathbf{x}$, they are eigenvectors of 1 and thus $\mu_{\mathbf{A}}(1) \geq \gamma_{\mathbf{A}}(1) = m-1$. Notice $\operatorname{tr}(\mathbf{I} + \mathbf{bd'}) = m + \mathbf{d'b}$ and that implies there are exactly 1 eigenvalues is $1 + \mathbf{d'b}$ and claim follows by compute $\prod \lambda_i$.

□

Proposition 4.3. *Suppose* $\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_r}$ *belong to different* λ_i*, then they are linearly independent.*

Suppose $\operatorname{eig}(\mathbf{A})$ are all distinct, then let

$$\mathbf{X} = \begin{bmatrix} \mathbf{x_1} & \ldots & \mathbf{x_m} \end{bmatrix}$$

where $\mathbf{x}_i$ is an eigenvector corresponding to λ_i. Then $\mathbf{Ax_i} = \lambda_i \mathbf{x_i}$ implies $\mathbf{AX} = \mathbf{X}\operatorname{diag}(\lambda_i)$. That is, $\mathbf{A} = \mathbf{X\Lambda X^{-1}}$ is **diagonalizable**. If $\mathbf{A}$ is diagonalizable, then it's rank is the number of its nonzero eigenvalues, also, in view of proposition 4.1, $\mu_{\mathbf{A}}(\lambda) = \gamma_{\mathbf{A}}(\lambda)$.

The following theorem stats that a matrix satisfy its own characteristic equation.

Theorem 4.2 (Cayley-Hamilton)**.** *Suppose eig*$(\mathbf{A}) = \lambda_1, \ldots, \lambda_m$ *then*

$$\prod_{i=1}^{m} \mathbf{A} - \lambda_i \mathbf{I} = \mathbf{0}$$

4.1 Symmetric matrices and Spectral Decomposition

Symmetric matrices avoid occurrence of complex eigenvalues:

Theorem 4.3. *Let $\mathbf{A} \in \mathbb{R}^{m \times m}$ be symmetric, then all eigenvalues of $\mathbf{A}$ are real.*

Proof. Suppose $\lambda \in \operatorname{eig}(\mathbf{A})$, then

$$(\mathbf{Ax})^* \mathbf{x} = \overline{\lambda} \mathbf{x}^* \mathbf{x}$$

on the other hand

$$(\mathbf{Ax})^* \mathbf{x} = \mathbf{x}^* \mathbf{Ax} = \lambda \mathbf{x}^* \mathbf{x}$$

thus $\overline{\lambda} = \lambda$ and must be real.

$\square$

Remark. The real eigenvalues suggest real eigenvector existence, suppose $\mathbf{x} = \mathbf{a} + i\mathbf{b}$, then

$$\mathbf{Ax} = \mathbf{Aa} + i\mathbf{Ab} = \lambda \mathbf{a} + i\lambda \mathbf{b}$$

thus $\mathbf{a}$ is also eigenvector.

We have see that sets of eigenvectors comes from different eigenvalues are linearly independent. If $\mathbf{A}$ is symmetric, they are even orthogonal. Suppose $\lambda, \gamma \in \sigma_{\mathbf{A}}$ and $\lambda \neq \gamma$, corresponding to eigenvectors $\mathbf{x}$ and $\mathbf{y}$.

$$\begin{aligned} \lambda \mathbf{x}'\mathbf{y} &= (\lambda \mathbf{x})'\mathbf{y} = (\mathbf{Ax})'\mathbf{y} = \mathbf{x}'\mathbf{A}'\mathbf{y} \\ &= \mathbf{x}'\gamma\mathbf{y} = \gamma\mathbf{x}'\mathbf{y} \implies \mathbf{x}'\mathbf{y} = 0 \end{aligned}$$

Thus, if all the m eigenvalues are distinct, Spectral decomposition can be applied. In fact, it's possible even $\mathbf{A}$ has multiple eigenvalues. To see this, we need following theorem.

Lemma 4.2. *Let $\mathbf{A} \in \mathbb{R}^{m \times m}$ be symmetric and $\mathbf{x} \in \mathbb{R}^m$, then there is some $\lambda_i \in \sigma_{\mathbf{A}}$ s.t.*

$$\lambda_i \in \mathit{span}\left(\mathbf{x}, \mathbf{Ax}, \ldots, \mathbf{A^{r-1}x}\right)$$

for some $r \geq 1$

Proof. Let r be the smallest for which $(\mathbf{x}, \mathbf{Ax}, \ldots, \mathbf{A^r x})$ are linearly dependent. Then there exist not all zero α_i *s.t.*:

$$\alpha_0 \mathbf{x} + \alpha_1 \mathbf{Ax} + \cdots + \alpha_r \mathbf{A^r x} = (\alpha_0 \mathbf{I} + \alpha_1 \mathbf{A} + \cdots + \mathbf{A^r})\mathbf{x} = \mathbf{0}$$

where we let $\alpha_r = 0$ WLOG. By Fundamental Algebra Theorem, there exist γ_i *s.t.*

$$\sum_{i=0}^{r} \alpha_i \mathbf{A}^i = \prod_{i=1}^{m} (\mathbf{A} - \gamma_i \mathbf{I})$$

Now let $\mathbf{y} = [\prod_{i=2}^{m}(\mathbf{A} - \gamma_i \mathbf{I})]\,\mathbf{x}$, its nonzero as $\mathbf{x}, \mathbf{A}\mathbf{x}, \ldots, \mathbf{A}^{r-1}\mathbf{x}$ are linearly independent. Thus $\mathbf{y}$ is in span $\left(\mathbf{x}, \mathbf{A}\mathbf{x}, \ldots, \mathbf{A}^{r-1}\mathbf{x}\right)$ and it follows that

$$(\mathbf{A} - \gamma_1 \mathbf{I})\mathbf{y} = \mathbf{0}$$

and then claim follows.

□

Above lemma gives a way to find a new orthogonal eigenvector from existed $\mathbf{x_1, \ldots, x_h}$, select $\mathbf{x}$ orthogonal to all of them then $\mathbf{A^k x}$ remains orthogonal since

$$\mathbf{x_i' A^k x} = (\mathbf{A^k x_i})'\mathbf{x} = \lambda_i^k \mathbf{x_i' x} = 0$$

so the vector $\mathbf{y}$ given by the lemma is desired. Then we can constructed a set of m eigenvectors that are orthonormal.
As we said before, then so called spectral decomposition applied. Let $\mathbf{Q} = (\mathbf{x_1}, \ldots, \mathbf{x_m})$ constructed by the orthonormal set and become an orthogonal matrix, then $\mathbf{A} = \mathbf{Q\Lambda Q'}$ where $\mathbf{\Lambda} = \mathrm{diag}(\lambda_i)$ as before.

Clearly, in this case, geometric multiplicity and algebraic multiplicity coincide and rank is number of nonzero eigenvalues.

4.2 Eigenprojections

A set of orthonormal eigenvectors can be used to find **eigenprojections** of $\mathbf{A}$.

Definition 4.1. Let λ be an eigenvalues of symmetric $\mathbf{A} \in \mathbb{R}^{m \times m}$ with multiplicity $r \geq 1$, $\{\mathbf{x_i}\}_1^r$ be the orthonormal set of eigenvectors, then the **eigenprojections** of $\mathbf{A}$ is

$$\mathbf{P_A}(\lambda) = \sum_{i=1}^{r} \mathbf{x_i x_i'}$$

This is orthogonal projection for eigenspace $S_{\mathbf{A}}(\lambda)$. Let $\{\lambda_i\}$ be the multiset of eigenvalues and $\{\mu_i\}$ be set of them, then

$$\mathbf{A} = \mathbf{Q\Lambda Q'} = \sum_{i=1}^{m} \lambda_i \mathbf{x_i x_i'} = \sum_{i=1}^{k} \mu_i \mathbf{P_A}(\mu_i)$$

The last term is preferred than the second since it's term are unique.

4.3 Advanced in eigenvalues

Theorem 4.4. *Let* $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{m\times m}$ *with eigenvalues* $\lambda_1, \ldots, \lambda_m$ *and* $\gamma_1, \ldots, \gamma_m$. *Define*

$$M = \max_{ij} |a_{ij}| \vee |b_{ij}|$$
$$\delta(\mathbf{A}, \mathbf{B}) = \frac{1}{m} \sum_{ij} |a_{ij} - b_{ij}|$$

then

$$\max_i \min_j |\lambda_i - \gamma_j| \leq (m+2) M^{1-\frac{1}{m}} \delta(\mathbf{A}, \mathbf{B})^{\frac{1}{m}}$$

That implies if $\mathbf{B_n} \to \mathbf{A}$ pointwise, then $\gamma \to \lambda$.

Proposition 4.4. λ_i *is continues function of elements of* $\mathbf{A}$.

Theorem 4.5. *Suppose* $\mathbf{A} \in \mathbb{R}^{m\times m}$ *is symmetric and* $\lambda \in \sigma_{\mathbf{A}}$. *Then* $\mathbf{P_A}(\lambda)$ *is a continues function of* $\mathbf{A}$.

4.4 Quadratic form

The quadratic form is something of the form $\mathbf{x'Ax}$ as a function of $\mathbf{x} \neq \mathbf{0}$, where $\mathbf{A} \in \mathbb{R}^{m\times m}$ is symmetric. To avoid effect of scale, we often use **Rayleigh quotient**:

$$R(x, \mathbf{A}) = \frac{\mathbf{x'Ax}}{\mathbf{x'x}}$$

Theorem 4.6. $R(\mathbf{x}, \mathbf{A})$ *take minimum in* $S_{\mathbf{A}}(\lambda_m)$ *while maximum in* $S_{\mathbf{A}}(\lambda_1)$.

Consequently, we have:

Theorem 4.7 (Courant–Fischer min–max theorem)**.** *Let* $\mathbf{A} \in \mathbb{R}^{m\times m}$ *be symmetric with eigenvalues* $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m$. *For* $1 \leq h \leq m$, *let* $\mathbf{B}_h \in \mathbb{R}^{m\times(h-1)}$ *and* $\mathbf{C}_h \in \mathbb{R}^{m\times(m-h)}$ *which are orthogonal. Then*

$$\lambda_h = \min_{\mathbf{b}_h} \max_{\mathbf{b}'_h\mathbf{x}=\mathbf{0}} \frac{\mathbf{x'ax}}{\mathbf{x'x}} = \max_{\mathbf{C}_h} \min_{\mathbf{C}'_h\mathbf{x}=\mathbf{0}} \frac{\mathbf{x'ax}}{\mathbf{x'x}}$$

Proof. Let $\mathbf{x}_i$ be eigenvectors corresponding to λ_i. The idea is we should specify $\mathbf{B}_h$ and $\mathbf{C}_h$ to avoid $\mathbf{x}_i$ according the larger (and smaller) occur in the $\mathcal{N}(\mathbf{B'}_h)$, so we can hide them in $\mathcal{C}(\mathbf{B}_h)$. That is, let $\mathbf{B}_h$ constructed by $\{\mathbf{x}\}_1^{h-1}$ and so the next maximum is λ_h.

$\square$

4.5 Nonnegative Definite Matrix

Theorem 4.8. *Suppose* $\mathbf{A} \in \mathbb{R}^{m\times m}$ *is symmetric, then*

1. $\mathbf{A}$ *is positive definite iff* $\lambda > 0$ *for all* $\lambda \in \sigma_{\mathbf{A}}$
2. $\mathbf{A}$ *is positive semidefinite iff* $\lambda \geq 0$ *for all* $\lambda \in \sigma_{\mathbf{A}}$ *and* $0 \in \sigma_{\mathbf{A}}$

Proof. By spectral decomposition, the orthogonal matrix $\mathbf{Q}$ span $\mathbb{R}^m$, thus any $\mathbf{x} = \mathbf{Qa}$ for some $\mathbf{a}$, then

$$\mathbf{x}'\mathbf{A}\mathbf{x} = \mathbf{x}'(\mathbf{Q}\mathbf{\Lambda}\mathbf{Q}')\mathbf{x} = \mathbf{a}'\mathbf{\Lambda}\mathbf{a}$$

Then the claim follows easily.

□

Symmetric matrix often obtained by taking $\mathbf{A} = \mathbf{TT}'$ or $\mathbf{TT}'$, in fact, they share positive eigenvalues.

Theorem 4.9. *Let* $\mathbf{T} \in \mathbb{R}^{m\times m}$ *with rank* r*, then positive eigenvalues of* $\mathbf{TT}'$ *are the same with* $\mathbf{T}'\mathbf{T}$*.*

Proof. □

Chapter 5

Singular Value Decomposition

Theorem 5.1. *If* $\mathbf{A} \in \mathbb{R}^{m \times n}$ *of rank* $r > 0$*, there exist orthogonal matrix* $\mathbf{P} \in \mathbb{R}^{m \times m}$ *and* $\mathbf{Q} \in \mathbb{R}^{n \times n}$ *s.t.* $\mathbf{A} = \mathbf{PDQ'}$ *where* $\mathbf{D}$ *is:*

$$\begin{cases} \Sigma & m = n = r \\ \begin{bmatrix} \Sigma & \mathbf{0} \end{bmatrix} & r = m < n \\ \begin{bmatrix} \Sigma \\ \mathbf{0} \end{bmatrix} & r = n < m \\ \begin{bmatrix} \Sigma & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} & r < m, r < n \end{cases}$$

where $\Sigma \in \mathbb{R}^{r \times r}$ *and is diagonal with positive entries, which are* $\sqrt{\lambda_i}$ *where* $\lambda \in \sigma_{\mathbf{A}}$