BASIC LINEAR ALGEBRA - MATH 223 NOTES

基本线性代数——MATH 223 笔记

Author
Wenyou (Tobias) Tian
田文友
University of British Columbia
英属哥伦比亚大学
2023

Contents

1 Sets and Maps, 集合与映射

1.1 Sets, 集合

A ***set*** is a collection of objects that are **definite, and distinct**; these objects are **elements**（元素）of this set.
We use $\varnothing$ to denote an ***empty set***（空集）, in which there is no element.

To denote a set:

1. List all elements of the set, or
2. Write down the condition for which the elements of the set should satisfy.
Such a set should be a **subset** of a ***universal set***
e.g. $A = \{x \in \mathbb{R} : x \geq 4\}$ means a set A that has all real numbers that are greater than or equal to 4

When listing elements, there are **no repetitions** and **no ordered sequence**. Thus, forms of $\{1, 1, 2, 3\}$ is unacceptable, and $\{1, 2, 3\}$ and $\{2, 1, 3\}$ are the same set.

Some relevant notations include quantifiers:

1. Universal quantifier - $\forall$: means "for every", "for each", or "for all"
2. Existential quantifier - $\exists$: means "there exists", "there are some"

1.1.1 Relationships, 集合关系

Definition 1.1. A is a **subset**（子集）of B if
$\forall x \in A, x \in B$

1.1.2 Operations, 集合运算

We define four operations, these operations mostly involve 2 sets (except for the case where it is finding a complement of a given set).

Definition 1.2. The four operations on sets are defined as follows:
$A \cup B$ - A union B（并集）, is a set that satisfies

$$\forall x \in A \cup B, x \in A \vee x \in B$$

$A \cap B$ - A intersect B（交集）, is a set that satisfies

$$\forall x \in A \cap B, x \in A \wedge x \in B$$

$\bar{A}$ - A's complement（补集）, is a set that satisfies

$$\forall x \in \bar{A}, x \in \mathbb{U} \wedge x \notin A$$

$A - B$ or $A \backslash B$ - A minus B, is a set that satisfies

$$\forall x \in A - B, x \in A \wedge x \notin B$$

We further define the following

$$A \times B := \{(a, b) : a \in A, b \in B\}$$

$A \times B$ is called the ***Cartesian product*** of sets A and B, where its elements are element pairs in the form shown above.
By extending this definition, we further have $\mathbb{R}^n := \mathbb{R} \times \cdots \times \mathbb{R}$, and the element of such a set is in the form $(a_1, \ldots, a_n), a_1, \ldots, a_n \in \mathbb{R}$, which is called an ***n-tuple***.
If we denote $\#A$ as the ***cardinality***（势）of the set A, which is the number of elements in A, then we have a trivial conclusion, where

$$\#(A \times B) = \#A \times \#B$$

.

1.1.3 Logic Statements, 逻辑语句

These logic statements significantly reduce the number of words needed to express certain mathematical conditions and provide structures for mathematical proofs.

1. P implies Q（命题）: $P \implies Q$. This statement is true when P is True and Q is True, or P is False. P is the sufficient condition（充分条件）for Q, Q is the necessary condition（必要条件）for P.
2. Negating an implication: $\neg(P \implies Q) \equiv P \wedge \neg Q$
3. Biconditional (if and only if)（当且仅当）: $P \iff Q \equiv (P \implies Q) \wedge (Q \implies P)$. P and Q are each other's sufficient and necessary conditions（充要条件）.
4. Converse（逆命题）: $Q \implies P$
5. Inverse（否命题）: $\neg P \implies \neg Q$
6. Contrapositive（逆否命题）: $\neg Q \implies \neg P$, an argument's contrapositive is equivalent to the argument.

1.2 Maps, 映射

Functions, as discussed in usual calculus classes, commonly operate with real numbers. $f(x) = x + 3$ is such an example. However, maps, in contrast with functions, provide correspondence rules between more general sets than $\mathbb{R}$. Thus, functions are special forms of maps.
To define a map, one needs 2 sets and 1 rule

Definition 1.3. Let X and Y be two general sets,
we can define a map f in the form

$$f : X \to Y, x \mapsto f(x)$$

such that $\forall x \in X, f(x) \in Y$. X is the domain（定义域）of such a map, and Y is the codomain（到达域）.

Formally, a map $f : A \to B$ is a **subset** of $A \times B$ such that $\forall x \in A$ appears **exactly once** as the 1^{st} coordinate of an element of this subset.

1.2.1 Properties of Maps, 映射的性质

Among all maps, there are some maps that are of particular interest to us because they have special properties.

Definition 1.4. There are generally three types of maps, consider $f : A \to B$
f is ***injective***（单射）, if

$$\forall a_1, a_2 \in A, a_1 \neq a_2 \implies f(a_1) \neq f(a_2)$$

$$\forall a_1, a_2 \in A, f(a_1) = f(a_2) \implies a_1 = a_2$$

f is ***surjective***（满射）, if

$$\forall b \in B, \exists a \in A, f(a) = b$$

f is ***bijective***（双射）, if f is both injective and surjective

By such definitions, for example, $f : \mathbb{R} \to \mathbb{R}, x \mapsto x^2$ is neither injective nor surjective.

1.2.2 Mapping sets, 对集合进行映射

Consider a map $f : X \to Y$, where $A \subset X$ and $B \subset Y$, we define two sets

$$f(A) := \{f(x) \in Y : x \in A\}$$

$$f^{-1}(B) := \{x \in X : f(x) \in B\}$$

We call $f(A)$ the ***image set***（像集）of A, and we call $f^{-1}(B)$ the ***pre-image set***（原像集）of B.

Specifically, consider $B = \{y\}$, then f is

surjective, if $\forall y \in Y, f^{-1}(B) \neq \varnothing$
injective, if $\forall y \in Y, \# f^{-1}(B) \leq 1$

1.2.3 Special maps, 特殊的映射

We first discuss the composite of maps（复合函数）. Consider two maps, $f : A \to B$ and $g : B \to C$, then we denote gf or $g \circ f$ to specify a map whose domain is A, whose codomain is C, where

$$\forall a \in A, (g \circ f)(a) = g(f(a)) \in C$$

Naturally, with the same A and B as given above, we have two maps

$$P_1 : A \times B \to A, (a, b) \mapsto a$$

$$P_2 : A \times B \to B, (a, b) \mapsto b$$

called ***projections***（射影、投影）, where P_1 is the projection on the 1^{st} factor, and P_2 is the projection on the 2^{nd} factor.
It is then obvious to notice that if $A, B \neq \varnothing$, P_1 and P_2 are both surjective, but P_1 is only injective if $\# B = 1$, and P_2 is only injective if $\# A = 1$.
Conveniently, we further define the map

$$\mathrm{Id}_A : A \to A, a \mapsto a$$

as the **identity map**（恒等映射）, if $A \neq \varnothing$.
Finally, we discuss a map whose condition of application is fairly restricted, yet very important.

Definition 1.5. We define an ***inverse map***（反映射）, $f^{-1} : B \to A$, for a given map $f : A \to B$, if we have
f is bijective, and
$f \circ f^{-1} = \mathrm{Id}_B$ and $f^{-1} \circ f = \mathrm{Id}_A$

For example, $f : \mathbb{R} \to \mathbb{R}, x \mapsto x^2$ does not have an inverse map f^{-1}. However, by slight adjusting f's domain and codomain, $f : \mathbb{R}^+ \to \mathbb{R}^+, x \mapsto x^2$ does have an inverse map $f^{-1} : \mathbb{R}^+ \to \mathbb{R}^+, y \mapsto \sqrt{y}$.
Notation wise, we have the following distinction:

Let $f : A \to B$, $y \in B$,
$f^{-1}(y)$ is the image of y under the inverse map, it does not need to exist, but if it does, it is a **single** element in A;
$f^{-1}(\{y\})$ always exists, it can be $\varnothing$;
if f^{-1} exists, then $f^{-1}(\{y\}) = \{f^{-1}(y)\}$

To indicate a bijective map, we use the ***isomorphism***（同构）sign $\cong$,

$$f : A \xrightarrow{\cong} B$$

1.2.4 Additional notes on maps, 有关映射的附言

Consider the following diagram

$$
\begin{tikzcd}
X \arrow[r, "f"] \arrow[d, "h"] & Y \arrow[d, "g"] \\
A \arrow[r, "i"] & B
\end{tikzcd}
$$

If in a diagram, all maps between any two sets (including compositions and possibly multiple compositions) agree, then one calls the diagram ***commutative***（交换图表）.
In this case, if $g \circ f = i \circ h$, the diagram is commutative.
Another note on maps is that, if $f : X \to Y$ and $A \subset X$, then

Definition 1.6. We call the map the ***restriction*** of f to A, read as "f restricted to A", to be

$$f|A : A \to Y, a \mapsto f(a)$$

2 Vector Spaces, 向量空间

2.1 Vector Space, 向量空间

A vector space, in short, is a set V whose addition and multiplication are defined to follow several axioms.

Definition 2.1. A triple $(V, +, \cdot)$ consisting of a set V, an addition map (called vector addition（向量加法）),

$$\begin{aligned} + : V \times V &\to V \\ (v_1, v_2) &\mapsto v_1 + v_2 \end{aligned}$$

and a multiplication map (called scalar multiplication（标量乘法）),

$$\begin{aligned} \cdot : \mathbb{F} \times V &\to V \\ (\lambda, v) &\mapsto V \end{aligned}$$

is called a ***vector space*** if the following 8 axioms hold for maps $+$ and $\cdot$:

$$\begin{aligned}
&1. \forall x, y, z \in V, (x + y) + z = x + (y + z) \\
&2. \forall x, y \in V, x + y = y + x \\
&3. \exists 0 \in V, \forall v \in V, v + 0 = v \\
&4. \forall v \in V, \exists - v \in V, v + (-v) = 0 \\
&5. \forall \lambda, \mu \in \mathbb{F}, v \in V, \lambda(\mu x) = (\lambda\mu)x \\
&6. \forall v \in V, 1v = v \\
&7. \forall \lambda \in \mathbb{F}, x, y \in V, \lambda(x + y) = \lambda x + \lambda y \\
&8. \forall \lambda, \mu \in \mathbb{F}, v \in V, (\lambda + \mu)v = \lambda v + \mu v
\end{aligned}$$

Following this definition, we have some useful remarks that we can use directly:

1. There only exists one zero vector in any vector space.
2. For every $v \in V$, there only exists one $-v$.
3. We may write $x + (-y)$ as $x - y$.

2.1.1 Vector Subspaces, 向量子空间

Definition 2.2. U is a ***vector subspace*** of V, if:

$$\begin{aligned}
&1. U \subset V \\
&2. U \neq \varnothing \\
&3. \forall x, y \in U, x + y \in U \\
&4. \forall \lambda \in \mathbb{F}, u \in U, \lambda u \in U
\end{aligned}$$

This means the zero vector, 0, must be in every vector subspace of a given vector space.

2.2 Field, 域

The definition of a field is very similar to a vector space, in fact, a field is a vector space over itself automatically given the following definition.

Definition 2.3. A ***field*** is a triple $(\mathbb{F}, +, \cdot)$ consisting of a set $\mathbb{F}$, and two rules of composition:

$$\begin{aligned} + : \mathbb{F} \times \mathbb{F} &\to \mathbb{F} \\ (\lambda, \mu) &\mapsto \lambda + \mu \end{aligned}$$

and

$$\begin{aligned} \cdot : \mathbb{F} \times \mathbb{F} &\to \mathbb{F} \\ (\lambda, \mu) &\mapsto \lambda\mu \end{aligned}$$

which satisfy the following 9 axioms:

$$\begin{aligned}
&1.\forall \lambda, \mu, \nu \in \mathbb{F}, (\lambda + \mu) + \nu = \lambda + (\mu + \nu) \\
&2.\forall \lambda, \mu \in \mathbb{F}, \lambda + \mu = \mu + \lambda \\
&3.\exists 0 \in \mathbb{F}, \forall \lambda \in \mathbb{F}, \lambda + 0 = \lambda \\
&4.\forall \lambda \in \mathbb{F}, \exists - \lambda \in \mathbb{F}, \lambda + (-\lambda) = 0 \\
&5.\forall \lambda, \mu, \nu \in \mathbb{F}, \lambda(\mu\nu) = (\lambda\mu)\nu \\
&6.\forall \lambda, \mu \in \mathbb{F}, \lambda\mu = \mu\lambda \\
&7.\exists 1 \in \mathbb{F}, \forall \lambda \in \mathbb{F}, (1 \neq 0) \implies 1\lambda = \lambda \\
&8.\forall \lambda, \mu, \nu \in \mathbb{F}, \lambda(\mu + \nu) = \lambda\mu + \lambda\nu \\
&9.\forall \lambda \in \mathbb{F}, (\lambda \neq 0) \implies (\exists \lambda^{-1} \in \mathbb{F}, \lambda^{-1}\lambda = 1)
\end{aligned}$$

Following this definition, we have 0 and 1 to be uniquely determined.
To show a set to be a field, one needs to show the set contains 0 and 1, and contains unique $-x$ and x^{-1}.

2.2.1 Finite fields, 有限域

Let $\mathbb{F}$ be a field and 1 to be its unit of addition, if $n1 := 1 + \cdots + 1 \neq 0$, for $n \in \mathbb{N}^+$, then we say $\mathbb{F}$ is a ***field of characteristic zero***（特征为 0 的域）. Otherwise, the ***characteristic***（特征）of $\mathbb{F}$ is the smallest positive number p such that $p1 = 0$.
A remark that comes from such a definition is that if the characteristic of $\mathbb{F}$, p, is not 0, then p is a prime number.
From this, we can define a finite field:

> Consider a field $\mathbb{F}_p$, where p is a prime number, then $\mathbb{F}_p$ can be a field of $\{0, 1, \ldots, p-1\}$ by defining the sum and the product to be the remainders of the usual sum and product modulo p.

In this case, $\mathbb{F}_p^n$ is naturally a vector space over $\mathbb{F}_p$ and it can be made into a field by defining

.
If *magically*, we have a way to define multiplication for $\mathbb{F}_p^n$, and we call such a field to be $\mathbb{F}_{p^n}$, then a vector space V that is over $\mathbb{F}_p$ now can become a vector space over $\mathbb{F}_{p^n}$.
One example of such manipulation is to first consider a vector over $\mathbb{R}$, then consider the same vector space over $\mathbb{C}$, since $\mathbb{C}$ is technically defined from $\mathbb{R}^2$.

2.3 Complex numbers, 复数

Complex numbers are investigated because we do not have a solution for the equation $x^2+1=0$ in $\mathbb{R}$. We define i to be a solution to this equation, and $-i$ to be the other.
Now, we want to make a number system with i and well-defined $+$ and $\times$.

Definition 2.4. The ***field of complex numbers*** is defined to be a triple $(\mathbb{C},+,\cdot)$, with $\mathbb{C}:=\mathbb{R}^2$, where

$$(x,y)+(a,b)=(x+a,y+b)$$

$$(x,y)\cdot(a,b)=(xa-yb,xb+ya)$$

Consider a complex number $z=a+bi$, it is convenient to define its ***complex conjugate***（共轭复数）$\bar{z}:=a-bi$. It is also convenient to know $z^{-1}=\frac{a}{a^2+b^2}-\frac{b}{a^2+b^2}i$. Similarly, if the absolute value of a real number on a real number line indicates the distance between this number and the origin, then the absolute value of a complex number on a complex plane can also be interpreted as such, where $|z|=\sqrt{a^2+b^2}$.

2.4 Abelian group, 阿贝尔群

An additive abelian group is a set with a defined set S and a map $+$, where the addition map allows the existence of a ”neutral element” 0, and each $x\in S$ has its negative $-x$.
Similarly, a multiplicative abelian group a set with a defined set S and a map $\times$, where the addition map allows the existence of a ”neutral element” 1 different from 0, and each $x\in S$ has its negative x^{-1}.

3 Dimension, 维度

3.1 Linear independence, 线性独立

3.1.1 Linear combination, 线性组合

Definition 3.1. Consider a set of vectors $\{v_1, \dots, v_n\} \subset V$ and $\lambda_1, \dots, \lambda_n \in \mathbb{F}$, then

$$\lambda_1 v_1 + \cdots + \lambda_n v_n$$

is a ***linear combination*** of $\{v_1, \dots, v_n\}$.
Then, we can define a ***linear span/hull***（线性生成空间）to be

$$L(v_1, \dots, v_n) := \{\lambda_1 v_1 + \cdots + \lambda_n v_n : \lambda_i \in \mathbb{F}\}$$

In this case, $L(v_1, \dots, v_n)$ is a subspace of V.
We define $L(\varnothing) := \{0\}$

Definition 3.2. A set of vectors $\{v_1, \dots, v_n\}$ is ***linearly independent*** if none of these vectors is a linear combination of the rest;
that is, if $\lambda_1, \dots, \lambda_n \in \mathbb{F}$

$$\lambda_1 v_1 + \cdots + \lambda_n v_n = 0 \implies \lambda_1 = \cdots = \lambda_n = 0$$

3.1.2 Basis, 基

Definition 3.3. A set of vectors $\{v_1, \dots, v_n\}$ is a basis of V if the set is linearly independent and it spans V, that is

$$\lambda_1 v_1 + \cdots + \lambda_n v_n = 0 \implies \lambda_1 = \cdots = \lambda_n = 0$$

$$L(v_1, \dots, v_n) = V$$

More exactly, $\forall v \in V$, there is exactly one $(\lambda_1, \dots, \lambda_n) \in \mathbb{F}^n$ such that $v = \lambda_1 v_1 + \cdots + \lambda_n v_n$

Theorem 3.1. If V is a vector space over $\mathbb{F}$, and V has a basis of n elements, then every basis of V has n elements.

In general, consider a vector space over $\mathbb{F}$ that has a basis of n elements, then the simplest basis we can form is the ***canonical/standard basis***（标准基）$(e_1, \dots, e_n)$ of $\mathbb{F}^n$, where

$$\begin{aligned} e_1 &:= (1, \dots, 0) \\ &\vdots \\ e_n &:= (0, \dots, 1) \end{aligned}$$

When demonstrating a set of vectors to be a basis for a vector space, one needs to show its linear independence, and it spanning the vector space. Usually, we use the definition to show such things.
More specifically, if we are to demonstrate linear independence for a set of vectors that spans a function vector space, for example, trigonometric functions $\sin x$ and $\cos x$, then we can plug in different x values for the equation

$$\lambda_1 \sin x + \lambda_2 \cos x = 0$$

to find each λ respectively.

Definition 3.4. A set of linearly independent vectors $\{v_1, \ldots, v_n\}$ is ***maximal*** if we cannot add more vectors while maintaining linear independence.

To extend from this definition, we have a useful lemma

Lemma 3.1. A maximal linearly independent set of vectors is a basis of V.

This then brings one theorem and one more lemma,

Theorem 3.2. The Basis Extension Theorem（基底延拓定理）states that:
Suppose V is a vector space over $\mathbb{F}$, we have a set of linearly independent vectors

$$\{v_1, \ldots, v_r\} \subset V$$

and some other set of vectors $\{w_1, \ldots, w_s\} \subset V$ such that

$$L(v_1, \ldots, v_n, w_1, \ldots, w_s) = V$$

then, there exists a subset of $\{w_1, \ldots, w_s\}$ such that together with $\{v_1, \ldots, v_r\}$ forms a basis of V.

Lemma 3.2. The Basis Exchange Lemma（基底交换引理）states that:
If $\{v_1, \ldots, v_r\}$ and $\{w_1, \ldots, w_s\}$ are bases of V, then $\forall v_i, \exists w_j$ such that replacing v_i by w_j in $\{v_1, \ldots, v_r\}$ we still have a basis.
That being said, we then also have $r = s$.

With "BET" and "BEL", we have one more useful theorem to easily determine linear dependence given a certain condition,

Theorem 3.3. If a vector space V has a basis of n elements, then any collection of more than n vectors must be linearly dependent.

3.2 Dimension, 维度

Since every basis of a given vector space V has the same number of vectors, then

Definition 3.5. If the vector space V over $\mathbb{F}$ has a basis $\{1, \ldots, v_n\}$, then n is called the ***dimension*** of V, abbreviated as $\dim V$.

3.2.1 Infinite dimensions, 无限维度

Given the above definition for the dimension of a vector space, then

Definition 3.6. The vector space V is ***infinite-dimensional*** if V does not possess a finite basis $\{1, \ldots, v_n\}$, where $0 \leq n < \infty$. We then write $\dim V = \infty$.

In order to show some vector space to be infinite-dimensional, since we cannot provide a finite basis, we can find a set $W \subset V$, where for any $w_1, \ldots, w_n \in W$, we can demonstrate the linear independence for such a collection of vectors. This demonstrates this collection of vectors cannot be maximal, as any more vectors added to this collection still ensure linear independence. Since we can always find one more vector to satisfy such linear independence, and W is only a subset of V, then V must be infinite-dimensional. This is most useful when demonstrating function vector spaces to be infinite-dimensional.
This brings a useful remark:

If V is finite-dimensional, U is a subspace of V, then U is finite-dimensional;
If U is infinite-dimensional, U is a subspace of V, then V is infinite-dimensional.
If U is a subspace of V, then $\dim U < \dim V$ if and only if $U \neq V$.

3.2.2 Direct sums, 直接和

Definition 3.7. If U_1, U_2 are subspaces of V, then

$$U_1 + U_2 := \{x + y : x \in U_1, y \in U_2\}$$

If we further know $U_1 \cap U_2 = \{0\}$, then U_1 and U_2 are complementary sets, where

$$U_1 \oplus U_2 = V$$

A useful conclusion we can reach from any direct sum is

Theorem 3.4. The Dimension Formula for Subspaces states that:
If U_1 and U_2 are finite-dimensional subspaces of V, then

$$\dim(U_1 + U_2) + \dim(U_1 \cap U_2) = \dim U_1 + \dim U_2$$

4 Linear Maps and Matrices, 线性映射与矩阵

4.1 Linear Maps, 线性映射

Definition 4.1. Let V and W be vector spaces over $\mathbb{F}$, a map:

$$f : V \to W$$

is called a **linear transformation**（线性变换）if it satisfies:

$$\forall x, y \in V, f(x + y) = f(x) + f(y)$$

$$\forall \lambda \in \mathbb{F}, x \in V, f(\lambda x) = \lambda f(x)$$

This linear map is also called a ***homomorphism***（同态）.
The set of homomorphisms between V and W also forms a vector space over $\mathbb{F}$, denoted as:

$$\mathrm{Hom}_{\mathbb{F}}(V, W) := \{f : V \to W\}$$

with addition and multiplication to be defined as

$$+ : (f + g)(v) = f(v) + g(v)$$

$$\cdot : (\lambda f)(v) = \lambda f(v)$$

More on notations of linear maps,

Definition 4.2. A linear map $f : V \to W$ is called:
a ***monomorphism***（单态射）if it is injective;
an ***epimorphism***（满态射）if it is surjective;
an ***isomorphism***（同构）if it is bijective;
an ***endomorphism***（自同态）if $V = W$;
an ***automorphism***（自同构）if it is bijective and $V = W$

More on an isomorphism, it means that if f is an isomorphism, then these necessarily follows:

f is linear.
There exists $g : W \to V$, such that $f \circ g = \mathrm{Id}_W$ and $g \circ f = \mathrm{Id}_V$.
Given $\{v_1, \ldots, v_n\}$ to be the basis of V, $\{f(v_1), \ldots, f(v_n)\}$ is a basis of W, the converse is true as well.

We then call g the inverse of f, and we can denote it as f^{-1}.
Moreover, any two n-dimensional vector space over the same field $\mathbb{F}$ are isomorphic.
Linear maps between two n-dimensional spaces are surjective if and only if they are injective.

We are then interested in some special spaces of these linear maps,

Definition 4.3. Let $f : V \to W$ be a linear map, then:
the ***kernel***（核，零空间）of f is defined as

$$\mathrm{Ker}f := \{v \in V : f(v) = 0\}$$

the ***image***（像）of f is defined as

$$\mathrm{Im}f := \{f(v) : v \in V\}$$

Then, naturally, $\mathrm{Ker}f \subset V$,$\mathrm{Im}f \subset W$.

We then define a feature about the image of f,

Definition 4.4. The ***rank***（秩）of the linear map f is defined to be

$$\mathrm{rk}f := \dim(\mathrm{Im}f)$$

This then will give us an extremely useful theorem,

Theorem 4.1. The ***rank-nullity theorem***（秩——零化度定理）states that:

$$\mathrm{rk}f + \dim(\mathrm{Ker}f) = \dim V$$

If we further consider the image of f, then given $\{v_1, \dots, v_n\}$ to be the basis of V, there exists a unique set of vectors $\{w_1, \dots, w_n\}$ such that for every $i = 1, \dots, n$, $f(v_i) = w_i$.

4.2 Matrices, 矩阵

Consider a general $m \times n$ matrix, where m is the number of rows, and n is the number of columns,

$$A = (a_{ij}) = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix}$$

If we define $M_{m\times n}(\mathbb{F})$ to be the set of all $m \times n$ matrices with entries in $\mathbb{F}$, then naturally, $A \in M_{m\times n}(\mathbb{F})$, and we can build a **bijective** map

$$L : \mathrm{Hom}_{\mathbb{F}}(V, W) \to M_{m\times n}(\mathbb{F})$$

so that each linear map (homomorphism) between V and W can be represented as a $m \times n$ matrix, given that $\dim V = n$ and $\dim W = m$.
If $v \in V$ and $v = \lambda_1 v_1 + \dots + \lambda_n v_n$ given the basis of V to be $\{v_1, \dots, v_n\}$, then by writing

the coefficients in a column, we can represent

$$v = \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix}$$

Then,

$$Av = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n a_{1i}\lambda_i \\ \vdots \\ \sum_{i=1}^n a_{mi}\lambda_i \end{pmatrix}$$

The general rule behind any matrix calculation is **row** $\times$ **column**.
Thus, essentially, a $m \times n$ matrix is map $\mathbb{F}^n \to \mathbb{F}^m$.
By the above arithmetic rule, we can know that

> The columns of a matrix are the images of the **unit vectors**, $e_1, \dots, e_n$.

From above, by choosing a basis in V, not only did we get a coordinate system in V, but also we can write any vector in V as a column with entries in the field, that is choosing an isomorphism $\mathbb{F}^n \to V$, where

Definition 4.5. A ***canonical basis isomorphism*** is defined to be

$$\begin{aligned} \Phi_{\{v_1,\dots,v_n\}} : \mathbb{F}^n &\to V \\ (\lambda_1, \dots, \lambda_n) &\mapsto \lambda_1 v_1 + \dots + \lambda_n v_n \end{aligned}$$

if we have chosen the basis of V to be $\{v_1, \dots, v_n\}$.

Now, since we can write any linear map as a matrix, then if we choose the basis of V to be $\{v_1, \dots, v_n\}$ and the basis of W to be $\{w_1, \dots, w_m\}$, where $f : V \to W$, A represents f, then the entries of the i^{th} column of A is the coordinate of Av_i in terms of $\{w_1, \dots, w_m\}$. That is

$$f(v_i) = Av_i = a_{1i}w_1 + \dots + a_{mi}w_m$$

The following commutative diagram can assist in understanding:

$$\begin{tikzcd} \mathbb{F}^n \arrow[r, "A"] \arrow[d, "\Phi_{\{v_1,\dots,v_n\}}"] & \mathbb{F}^m \arrow[d, "\Phi_{\{w_1,\dots,w_m\}}"] \\ V \arrow[r, "f"] & W \end{tikzcd}$$

The last thing to highlight is that consider the previously mentioned map L, if we know $L(f) = A$, then

Definition 4.6. The ***rank*** of a matrix A is its column rank, that is, the maximal number of linearly independent columns,

$$\mathrm{rk} A = \mathrm{rk} f$$

given the above assumption.
Similarly, the row rank of the same matrix is the maximal number of linearly independent rows.

Theorem 4.2. Given a matrix A, its column rank = its row rank.

5 Matrix Calculus, 矩阵运算

5.1 Matrix Multiplication, 矩阵乘法

First, we define matrix addition and scalar multiplication:

Definition 5.1. Let $(a_{ij}), (b_{ij}) \in M_{m\times n}(\mathbb{F})$ and $\lambda \in \mathbb{F}$, then

$$(a_{ij}) + (b_{ij}) := (a_{ij} + b_{ij}) \in M_{m\times n}(\mathbb{F})$$

$$\lambda(a_{ij}) := (\lambda a_{ij}) \in M_{m\times n}(\mathbb{F})$$

Now, consider a diagram

$$V \xrightarrow{B} W \xrightarrow{A} U$$

Then, $AB : V \to U$, that is $AB := A \circ B$

Definition 5.2. Let $\dim V = n, \dim W = m, \dim U = r$, if $A = (a_{ik}) \in M_{r\times m}(\mathbb{F})$ and $B = (b_{kj}) \in M_{m\times n}(\mathbb{F})$, the ***product*** $AB \in M_{r\times n}(\mathbb{F})$ is defined by

$$AB := (\sum_{k=1}^{m} a_{ik}b_{kj})_{(i=1,\ldots,r),(j=1,\ldots,n)}$$

Matrix multiplication is associative $A(BC) = (AB)C$, and distributive with respect to addition $A(B+C) = AB+AC$ and $(A+B)C = AC+BC$, but they are NOT commutative.

Definition 5.3. A matrix A is ***invertible***（可逆的）if the associated linear map is an isomorphism; the matrix of the inverse map is then called the matrix ***inverse***（逆矩阵）to A and is denoted by A^{-1}.

There are several useful remarks regarding matrix inversion:

1. Each invertible matrix A is square.
2. If A is invertible, then A^{-1} is invertible.
3. If A, B are invertible, then AB is invertible and $(AB)^{-1} = B^{-1}A^{-1}$
4. If A, B are square matrices, then

$$AB = \mathrm{Id} \iff BA = \mathrm{Id} \iff B = A^{-1}$$

5.2 Elementary Transformations

There are three elementary ROW transformations:

Definition 5.4. For a matrix $A \in M_{m\times n}(\mathbb{F})$, we have:
(R1): Interchanging two rows.
(R2): Multiplication of a row by a scalar $\lambda \neq 0, \lambda \in \mathbb{F}$.
(R3): Addition of an arbitrary multiple of one row to another row.

Elementary transformations do not alter the rank of the matrix.
Thus, if after an arbitrary number of transformations, if the first r entries on the diagonal of the operated matrix are distinct from 0, and the rest $m-r$ rows and all the entries below the diagonal are 0, then $\mathrm{rk} A = r$.

6 System of Linear Equations, 线性方程组

6.1 System of Linear Equations

Generally, if $A \in M_{m\times n}(\mathbb{F})$, then we consider

$$Ax = b$$

to represent a system of linear equations where

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m \end{aligned}$$

where

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}, x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, b = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}$$

If $b_i = 0$, then the system is said to be ***homogeneous***（齐次）.
Then we define

Definition 6.1. Given a system of linear equations $Ax = b$, the solution set is

$$A^{-1}(b) := \{x : Ax = b\}$$

If we already know $Ax_0 = b$, then every solution to this linear system is of form $x_0 + k$, where $k \in \mathrm{Ker}A$.
A useful remark for the solvability of a linear system is that, $Ax = b$ is solvable if and only if

$$\mathrm{rk}A = \mathrm{rk}(A|b)$$

6.2 Solving Linear Systems, 求解线性方程组

We apply Gaussian elimination (elementary row transformation) to reduce A to its echelon form, that is for i^{th} row, the leading 1 should appear at or after i^{th} entry of the row, and all entries before and below the leading 1s should vanish to 0.
A row-reduced echelon form follows all the rules above while having 0s above all leading 1s.
If the linear system has a solution, then $b \in \mathrm{Im}A$.
If the linear system has a UNIQUE solution, then $\mathrm{Ker}A = \{0\}$ and A is injective.
If A is a square matrix, then A is surjective, then a solution exists for every b.
To define a vector space using $Ax = 0$, the vector space is the solution set to this linear system, that is, $\mathrm{Ker}A$. If A is not full rank, then we can express the kernel in parametric

form to find bases for the vector space. For example

$$\begin{pmatrix}1 & 2 & 3\\4 & 5 & 6\\7 & 8 & 9\end{pmatrix} \to \begin{pmatrix}1 & 0 & -1\\0 & 1 & 2\\0 & 0 & 0\end{pmatrix}$$

Let $z = t$, then $x = t$, and $y = -2t$, then

$$\mathrm{Ker}A = t\begin{pmatrix}1\\-2\\1\end{pmatrix}$$

where $\begin{pmatrix}1\\-2\\1\end{pmatrix}$ is the basis of this vector space.

Since row operations preserve Ker$A$ and ImA, we already know the dimension of the kernel is 1 from above, then by rank-nullity theorem, dim(ImA). Furthermore, since the columns of A represent the image of the unit vectors, any 2 vectors can span ImA, thus, we have an equation for ImA, which is

$$s\begin{pmatrix}1\\4\\7\end{pmatrix} + t\begin{pmatrix}2\\5\\8\end{pmatrix}$$

After Gaussian elimination, we define ***pivot columns*** to be columns with a leading "1", where ***non-pivot columns*** to be columns without a leading "1". Each of the non-pivot columns contributes a parameter.

6.3 Inverting a Matrix, 求逆矩阵

First, consider an $n \times n$ square matrix A, where $Ax = b$.

1. If rk$A < n$, A^{-1} does not exist.
2. For some b, there will be infinitely many solutions, then these b form a linear subspace of ImA.
3. For some b, there will be no solution, then these b have the property that $b \notin \mathrm{Im}A$.

Since row operations are essentially matrix compositions:

$$\begin{pmatrix}0 & 1\\1 & 0\end{pmatrix} A : \text{swapping rows}$$

$$\begin{pmatrix}\lambda & 0\\0 & 1\end{pmatrix} A : \text{multiplying a row by } \lambda$$

$$\begin{pmatrix}1 & \lambda\\0 & 1\end{pmatrix} A : \text{Adding multiples of one row to another}$$

Thus, if the row-reduced echelon form of A is equivalent to Id, then A^{-1} exists. So $x = A^{-1}b$ from the above case.

To find A^{-1}, the most direct way is to

$$\begin{pmatrix} A & | & \mathrm{Id} \end{pmatrix} \xrightarrow{Gaussian Elimination} \begin{pmatrix} \mathrm{Id} & | & A^{-1} \end{pmatrix}$$

7 Determinants, 行列式

7.1 Definition of determinants, 行列式的定义

Definition 7.1. There exists a unique map:

$$\det : M_{n\times n}(\mathbb{F}) \to \mathbb{F}$$

such that it satisfies:

1. det is linear in each row, that is preserved under $R_i + \lambda R_j$
2. If $A \in M_{n\times n}(\mathbb{F})$, and $\mathrm{rk} A < n$, then $\det A = 0$.
3. $\det \mathrm{Id} = 1$

Consider a general matrix $A = (a_{ij})$,

Definition 7.2. Let A_{ij} denote the matrix where in A, the i^{th} row and j^{th} column is eliminated.

Therefore, we can write out a recursive formula for determining $\det A$, this process is called ***cofactor expansion*** with respect to the i^{th} row (cofactor expansion with respect to a column is also fine):

$$\det A = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \det A_{ij}$$

Specifically, for a 2×2 matrix,

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$$

Regarding to row operations,

1. If a multiply a row of A by λ, $\det A' = \lambda \det A$
2. If two rows are interchanged in A, $\det A' = -\det A$
3. If a row adds λ of another row in A, then $\det A' = \det A$

Consider a 3×3 ***Vandermonde Matrix***（范德蒙矩阵）,

$$\begin{aligned}
\begin{vmatrix} 1 & a & a^2 \\ 1 & b & b^2 \\ 1 & c & c^2 \end{vmatrix} &= \begin{vmatrix} 1 & a & a^2 \\ 0 & b-a & b^2-a^2 \\ 0 & c-a & c^2-a^2 \end{vmatrix} \\
&= 1 \cdot \begin{vmatrix} b-a & b^2-a^2 \\ c-a & c^2-a^2 \end{vmatrix} \\
&= (b-a)(c-a) \begin{vmatrix} 1 & b+a \\ 1 & c+a \end{vmatrix} \\
&= (b-a)(c-a)(c-b)
\end{aligned}$$

Determinant is a useful theoretical tool to give us information about ”volumes” in $\mathbb{R}^n$.

1. $\begin{vmatrix} a \end{vmatrix}$ gives the length of real number a (absolute value)
2. $\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}$ gives the area of a parallelogram in $\mathbb{R}^2$
3. $\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$ gives the volume of parallelopiped in $\mathbb{R}^3$

In multivariable calculus, the ***Jacobian determinant***（雅可比行列式）gives us the change of variable formula:

$$\begin{vmatrix} \frac{\partial f_1}{\partial x_1} & \dots & \frac{\partial f_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial f_n}{\partial x_1} & \dots & \frac{\partial f_n}{\partial x_n} \end{vmatrix}$$

Some common matrices have easy ways to compute determinants,

1. Upper triangular

$$\begin{vmatrix} a_{11} & * & \dots & * \\ 0 & a_{22} & \dots & * \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & a_{nn} \end{vmatrix} = a_{11}a_{22}\dots a_{nn}$$

2. Lower triangular

$$\begin{vmatrix} a_{11} & 0 & \dots & 0 \\ * & a_{22} & \dots & 0 \\ \vdots & \vdots & & \vdots \\ * & * & \dots & a_{nn} \end{vmatrix} = a_{11}a_{22}\dots a_{nn}$$

3. Block upper triangular

$$\begin{vmatrix} A & * & * \\ 0 & B & * \\ 0 & 0 & C \end{vmatrix} = \det A \det B \det C$$

4. Block lower triangular (ditto)
5. Block diagonal (ditto)

7.2 Properties of Determinants, 行列式的性质

Definition 7.3. Consider a general $n \times n$ matrix $A = (a_{ij})$, then then transpose of A is defined to be:

$$A^t = \begin{pmatrix} a_{11} & a_{21} & \dots & a_{n1} \\ a_{12} & a_{22} & \dots & a_{n2} \\ \vdots & \vdots & & \vdots \\ a_{1n} & a_{2n} & \dots & a_{nn} \end{pmatrix}$$

Then, $\det A^t = \det A$

Proposition 7.1. $\det(AB) = \det A \det B$

7.3 Cramer's Rule, 克莱默法则

Consider a general linear system

$$Ax = b$$

where $A = (a_{ij}) \in M_{n\times n}(\mathbb{F})$, and $b = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}$

Then consider a matrix B_i:

$$B_i = \begin{pmatrix} a_{11} & \dots & a_{1i-1} & b_1 & a_{1i+1} & \dots & a_{1n} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ a_{n1} & \dots & a_{ni-1} & b_n & a_{ni+1} & \dots & a_{nn} \end{pmatrix}$$

which is formed by replacing i^{th} column of A with b.
Then, assume $\det A \neq 0$, and if $Ax = b$, has a unique solution

$$\begin{pmatrix} x_1^\circ \\ \vdots \\ x_n^\circ \end{pmatrix}$$

then

$$x_i^\circ = \frac{\det B_i}{\det A}$$

Proof for Cramer's Rule is omitted at this section, but a useful matrix associated with such proof is M_i:

$$M_i = \begin{pmatrix} a_{11} & \dots & a_{1i-1} & x_i^\circ a_{1i} - b_1 & a_{1i+1} & \dots & a_{1n} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ a_{n1} & \dots & a_{ni-1} & x_i^\circ a_{ni} - b_n & a_{ni+1} & \dots & a_{nn} \end{pmatrix}$$

And $\det M_i = 0 = x_i^\circ \det A - \det B_i$
Suppose now we want to find a formula for A^{-1}, since $AA^{-1} = \mathrm{Id}$, then, the k^{th} column of

A^{-1} is the solution to the equation

$$Ax_k = e_k$$

Assume $x_k = \begin{pmatrix} x^\circ_{k1} \\ \vdots \\ x^\circ_{kn} \end{pmatrix}$, let $A^{-1} = (c_{ij})$, then c_{ij} is the j^{th} component of x_j, so by Cramer's Rule, we replace the i^{th} column with e_j:

$$c_{ij} = \frac{1}{\det A} \times \begin{vmatrix} a_{11} & \dots & a_{1i-1} & 0 & a_{1i+1} & \dots & a_{1n} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ a_{j1} & \dots & a_{ji-1} & 1 & a_{ji+1} & \dots & a_{jn} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ a_{n1} & \dots & a_{ni-1} & 0 & a_{ni+1} & \dots & a_{nn} \end{vmatrix}$$

If we then apply cofactor expansion with respect to the i^{th} column, then

$$c_{ij} = \frac{1}{\det A} \times (-1)^{i+j} \times \begin{vmatrix} a_{11} & \dots & a_{1i-1} & a_{1i+1} & \dots & a_{1n} \\ \vdots & & \vdots & \vdots & & \vdots \\ a_{j-11} & \dots & a_{j-1i-1} & a_{j-1i+1} & \dots & a_{j-1n} \\ a_{j+11} & \dots & a_{j+1i-1} & a_{j+1i+1} & \dots & a_{j+1n} \\ \vdots & & \vdots & \vdots & & \vdots \\ a_{n1} & \dots & a_{ni-1} & a_{ni+1} & \dots & a_{nn} \end{vmatrix} = (-1)^{i+j} \times \frac{\det(A_{ji})}{\det A}$$

To highlight, we can have

Proposition 7.2. Let $A^{-1} = (c_{ij})$, then given $A = (a_{ij})$, we have

$$c_{ij} = (-1)^{i+j} \times \frac{\det A_{ji}}{\det A}$$

A short example is a 2×2 matrix

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad - bc} \times \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

8 Euclidean Vector Space, 欧几里得向量空间

8.1 Inner Product, 内积

Let V be a vector space, and we define an extra structure - inner product, a Euclidean space:

Definition 8.1. An inner product on V over $\mathbb{R}$ is a function:

$$(*, *) : V \times V \to \mathbb{R}$$

which satisfies:

1. Bi-linearity

$$(\lambda_1 x_1 + \lambda_2 x_2, y) = \lambda_1(x_1, y) + \lambda_2(x_2, y)$$

$$(x, \lambda_1 y_1 + \lambda_2 y_2) = \lambda_1(x, y_1) + \lambda_2(x, y_2)$$

2. Symmetry

$$(x, y) = (y, x)$$

3. Positive-definite

$$\forall x \in V, (x, x) \geq 0, x = 0 \iff (x, x) = 0$$

For a complex vector space, we can ask for positive-definite, but then we need to modify bi-linearity and symmetry to Hermitian form, for example, if $x, y \in \mathbb{C}^n, x = (z_1, \ldots, z_n), y = (w_1, \ldots, w_n)$, we can have $(x, y) = z_1\bar{w}_1 + \cdots + z_n\bar{w}_n$.
Any finite-dimensional vector space V over $\mathbb{R}$ has ∞ many inner products.
Thus, a ***Euclidean space*** is defined to be a vector space with some inner product on such a vector space.
This is useful to define Euclidean geometry, where it involves distances and angles.

Definition 8.2. A ***Euclidean norm***（欧几里得范数）is defined to be

$$\|x\| := \sqrt{(x, x)}$$

Then,

Definition 8.3. If we consider $x, y \in V$, the angle between them to be α, then by Law of Cosines, we have

$$\cos\alpha = \frac{(x, y)}{\|x\|\|y\|}$$

In a Euclidean space, we define the unit sphere S to be a set of vectors where

$$S := \{x \in V : \|x\| = 1\}$$

For function vector space, for example, all real continuous functions on $[0, 1]$, we can define its inner product to be

$$(f, g) := \int_0^1 f(x)g(x)\mathrm{d}x$$

For complex-value functions, we can use Hermitian form of some

$$(f, g) = \int f(x)\overline{g(x)}\mathrm{d}x$$

8.1.1 Properties of a Norm, 范数的性质

Norms do not have to always come from an inner product, as axiomatically, we only require a norm to satisfy three conditions:

1. $\|x\| = 0 \iff x = 0$, otherwise $\|x\| > 0$
2. $\|\lambda x\| = |\lambda|\|x\|$
3. $\|x + y\| \leq \|x\| + \|y\|$

8.2 Cauchy-Schwarz Inequality, 柯西——施瓦茨不等式

Theorem 8.1. Cauchy-Schwarz Inequality states that

$$|(x, y)| \leq \|x\|\|y\|$$

In continuous functions on $[a, b]$, this applies to be

$$(\int_a^b f(x)g(x)\mathrm{d}x)^2 \leq \int_a^b f(x)^2\mathrm{d}x \int_a^b g(x)^2\mathrm{d}x$$

In real numbers, we then have

$$(\sum_{i=1}^n x_i y_i)^2 \leq (\sum_{i=1}^n x_i)^2(\sum_{i=1}^n y_i)^2$$

8.3 Gram-Schmidt Orthonormalization, 格拉姆——施密特正交化

The convenience of a Euclidean space is that it gives an orthonormal basis which is of many uses.

Definition 8.4. Let V be a Euclidean vector space, and for some vectors $v_1, \ldots, v_n \in V$, if $\forall i, j = 1, \ldots, n, \|v_i\| = 1 \wedge (i \neq j) \implies (v_i, v_j) = 0$, then $v_1, \ldots, v_n$ are orthonormal.

Lemma 8.1. Any collection of orthonormal vectors is linearly independent

Thus, if $\{v_1, \ldots, v_n\}$ is an orthonormal basis for Euclidean vector space V, if we express $v = \lambda_1 v_1 + \cdots + \lambda_n v_n \in V$, then $\lambda_i = (v, v_i)$ From a random basis, to find an orthonormal basis of the same vector space, we follow this process

Theorem 8.2. Given a random basis $\{v_1, \ldots, v_n\}$:
1. Normalize v_1 to $\tilde{v_1}$ where $\|\tilde{v_1}\| = 1$
2. For $k = 1, \ldots, n-1$, we have

$$\tilde{v_{k+1}} = \frac{v_{k+1} - \sum_{i=1}^{k}(v_{k+1}, \tilde{v_i})\tilde{v_i}}{\|v_{k+1} - \sum_{i=1}^{k}(v_{k+1}, \tilde{v_i})\tilde{v_i}\|}$$

Then, $\{\tilde{v_1}, \ldots, \tilde{v_n}\}$ is an orthonormal basis.

8.3.1 Projections, 投影

Let $V = U \oplus W$, then $\forall v \in V, \exists! v = u + w, u \in U \land w \in W$.
Given a linear subspace U, unless it is over a finite field, there are ∞ many projections, the best of which is the orthogonal complement.

Definition 8.5. An orthogonal complement of U is defined to be

$$U^{\perp} := \{u' \in V : \forall u \in U, (u, u') = 0\}$$

We have $U \oplus U^{\perp} = V$

Orthogonal projectors are thus projectors that project U along $U^{\perp}$ to U.

8.3.2 Relation to Matrix, 与矩阵的关系

Consider $A = (a_{ij}) \in M_{m \times n}$, then $A \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ can be interpreted as a dot product.

Let $r_i = \begin{pmatrix} a_{i1} & \ldots & a_{in} \end{pmatrix}$, and $r_i \cdot x$ is the dot product, then

$$x \in \text{Ker}A \iff r_i \cdot x = 0 \iff x \in L(r_1, \ldots, r_m)^{\perp}$$

This tells us, the column of space of A is $\text{Im}A$, and the ***row space***（行空间）of A is $\text{Ker}(A)^{\perp}$.
And given that $V = \text{Ker}(A) \oplus \text{Ker}(A)^{\perp}$, we have

$$\begin{aligned} \dim(\text{Ker}(A)^{\perp}) &= \dim V - \dim(\text{Ker}A) \\ \text{rk}A &= \dim V - \dim(\text{Ker}A) \end{aligned}$$

That is: **column rank = row rank**

8.4 Orthogonal Matrix, 正交矩阵

Definition 8.6. Linear map $A : V \to W$ is an ***isometry***（等距同构）if

$$(Av_1, Av_2) = (v_1, v_2)$$

where V, W are Euclidean vector spaces, and $v_1, v_2 \in V$.
Usually, it means A is surjective.

Such isometries take any orthonormal basis to an orthonormal basis, therefore, if A is an endomorphism, and let c_k denote the k^{th} column of A, then with

$$c_i \cdot c_j = \delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

We further will have $A^t A = \mathrm{Id}$ More generally,

Theorem 8.3. Let $A : V \to W$, $\dim V = \dim W$, be an isometry, then
$\iff$ $(Av_1, Av_2) = (v_1, v_2)$
$\iff$ orthonormal basis of $V \to$ orthonormal basis of W
$\iff$ columns of A is the orthonormal basis of W
$\iff$ $A^t A = \mathrm{Id} \iff A^t = A^{-1} \iff \det A = \pm 1$
$\iff$ $AA^t = \mathrm{Id} \to$ rows of A form an orthonormal basis
$\to \det A_{ij} = \pm a_{ij}$

Orthonormal transformations of V form a group called $O(V)$ or $O_n(\mathbb{R})$, and those special orthonormal transformations' group is denoted as $SO(V)$ or $SO_3(\mathbb{R})$.

9 Spectral Theorem, 谱定理

9.1 Eigenvalues and Eigenvectors, 特征值与特征向量

Definition 9.1. Let A be a matrix with $a_{ij} \in \mathbb{F}$ and $A : \mathbb{F}^n \to \mathbb{F}^n$, then $\lambda \in \mathbb{F}$ is an eigenvalue of A if

$$\exists v \neq 0, v \in \mathbb{F}^n, Av = \lambda v$$

v is the eigenvector corresponding to the eigenvalue λ

To find eigenvalues of a matrix A, we find the ***characteristic polynomial***（特征多项式）:

$$\det(A - \lambda \mathrm{Id}) = 0$$

Then, to find the eigenvectors corresponding to eigenvalue λ, we just have to find a basis for $\mathrm{Ker}(A - \lambda \mathrm{Id})$.

9.1.1 Basis of Eigenvectors, 特征向量的基

In order for a basis of eigenvectors to exist, we need A to at least be an endomorphism $A : V \to V$, then V **might** have a basis of eigenvectors of A.
The matrix A with respect to the eigenvector basis is ***diagonal***（对角的）.
We need eigenvalues to determine whether or not a basis of eigenvectors exists. Therefore, consider the characteristic polynomial

$$\begin{vmatrix} a_{11} - \lambda & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} - \lambda & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} - \lambda \end{vmatrix}$$

which must have the leading term to be some $(-1)^n \lambda^n$.
By **Fundamental Theorem of Algebra**（代数基本定理）, an n-degree polynomial must have n roots in $\mathbb{C}$, counting with ***algebraic multiplicity***（代数重复度）.
Over $\mathbb{C}$, any polynomial can be factored completely; but over $\mathbb{R}$, the complex roots will come in pairs $\{z, \overline{z}\}$, that is, odd-degree polynomials must have odd number of real roots.
In the complex plane, multiplying by $\lambda = \alpha \pm \beta i = |\lambda| e^{i\phi}$, means rotating ϕ, scaling by $|\lambda| = \sqrt{a^2 + b^2}$, where $\sin \phi = \frac{a}{|\lambda|}, \cos \phi = \frac{b}{|\lambda|}$.

9.1.2 Linear Independence of Eigenvectors, 特征向量的线性独立性

Suppose $\{v_i\}$ is a set of eigenvectors corresponding to distinct eigenvalues λ_i, then,

Base: $i = 1$, $\{v_1\}$ is linearly independent,
Inductive Hypothesis: Assume for $i = k$
Inductive Step: we show for $i = k + 1$,
suppose $\alpha_1 v_1 + \dots + \alpha_{k+1} v_{k+1} = 0$
Then

$$A(\alpha_1 v_1 + \dots + \alpha_{k+1} v_{k+1}) = \lambda_1 \alpha_1 v_1 + \dots + \lambda_{k+1} \alpha_{k+1} v_{k+1}$$

Use this minus λ_{k+1} of supposition, we have

$$\alpha_1(\lambda_1 - \lambda_{k+1})v_1 + \cdots + \alpha_k(\lambda_k - \lambda_{k+1})v_k = 0$$

Since $\lambda_1 \neq \ldots \neq \lambda_{k+1}$, then $\alpha_1 = \cdots = \alpha_k = 0$, which further indicates $\alpha_{k+1} = 0$
$\{v_{k+1}\}$ is thus linearly independent.

9.2 Matrix Diagonalization, 矩阵对角化

Over $\mathbb{C}$, consider a linear operator in $\mathbb{C}^n$ with distinct eigenvalues $\lambda_1, \ldots, \lambda_n$, then the matrix of A in this corresponding basis of eigenvectors is **diagonal**.

9.2.1 Change-of-basis Matrix, 基变更矩阵

Consider in $\mathbb{R}^2$, there is a basis $\{v_1, v_2\}$ and the standard basis $\{e_1, e_2\}$.
Let $v_1 = c_{11}e_1 + c_{21}e_2$, and $v_2 = c_{12}e_1 + c_{22}e_2$, then a change-of-basis matrix C, is thus

$$C = \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix}$$

Then, consider a general vector $x \in V$. We then have

$$x_{e_1,e_2} = Cx_{v_1,v_2}$$

$$x_{v_1,v_2} = C^{-1}x_{e_1,e_2}$$

If we have a linear operator A with respect to the basis $\{e_1, e_2\}$, and a linear operator A' with respect to the basis $\{v_1, v_2\}$, then we should have

$$(Ax_{e_1,e_2})_{v_1,v_2} = A'x_{v_1,v_2}$$

Therefore,

$$\begin{aligned} (Ax_{e_1,e_2})_{v_1,v_2} &= A'x_{v_1,v_2} \\ C^{-1}Ax_{e_1,e_2} &= A'C^{-1}x_{e_1,e_2} \\ C^{-1}A &= A'C^{-1} \\ C^{-1}AC &= A' \end{aligned}$$

9.2.2 Complex Eigenvectors, 复特征向量

If A have complex eigenvectors, then over $\mathbb{R}$, we can bring A to the form:

$$A = \begin{pmatrix} X & 0 & 0 \\ 0 & Y & 0 \\ 0 & 0 & Z \end{pmatrix}$$

where each X, Y, Z is a 2×2 matrix, with the form

$$\begin{pmatrix} |\lambda|\cos\phi & -|\lambda|\sin\phi \\ |\lambda|\sin\phi & |\lambda|\cos\phi \end{pmatrix}$$

Recall: $|\lambda|e^{i\phi} = |\lambda|\cos\phi + |\lambda|i\sin\phi$

9.2.3 Non-distinct Eigenvalues, 非独特特征值

Since we know $\det A' = \det\bigl(C^{-1}AC\bigr) = \lambda_1 \ldots \lambda_n$, thus:

> $\det A' \iff \exists i, \lambda_i = 0$
> This means, with 0 to be an eigenvalue, we have $\exists v \neq 0, Av = 0$, that is $\mathrm{Ker} A \neq \{0\}$. Then, A is neither injective, surjective, nor invertible.

Definition 9.2. Define ***geometric multiplicity***（几何重复度）of eigenvalue λ to be

$$\dim(\mathrm{Ker}(A - \lambda \mathrm{Id}))$$

which is less than or equal to its algebraic multiplicity.

When the geometric multiplicity of λ is less than its algebraic multiplicity, then A cannot be diagonalized.
The best we can do is to find C, such that $C^{-1}AC$ is of Jordan normal form.

Definition 9.3. An $n \times n$ Jordan block is of the form

$$J_\lambda = \begin{pmatrix} \lambda & 1 & 0 & \ldots & 0 \\ 0 & \lambda & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \ldots & \lambda \end{pmatrix}$$

One Jordan block represents the existence of 1 eigenvector.
Let

$$J_0 = \begin{pmatrix} 0 & 1 & 0 & \ldots & 0 & 0 \\ 0 & 0 & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & 0 & 1 \\ 0 & 0 & 0 & \ldots & 0 & 0 \end{pmatrix}$$

Then, every $J_\lambda = \lambda \mathrm{Id} + J_0$.
With $J_0^n = 0$, it is a nilpotent matrix.

Lemma 9.1. The binomial formula states that

$$(J_\lambda)^n = (\lambda \mathrm{Id} + J_0)^n$$

For each eigenvalue λ, we can write a Jordan form with Jordan blocks on the diagonal. The number of Jordan blocks (geometric multiplicity) is the number of eigenvectors associated with this eigenvalue.
Why do we care about diagonalization?

We can find powers of A, with D being A's diagonalized matrix.

$$A^m = CD^mC^{-1} = C \begin{pmatrix} \lambda_1^m & 0 & \dots & 0 \\ 0 & \lambda_2^m & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & \lambda_n^m \end{pmatrix} C^{-1}$$

9.3 Self-adjoint Matrix, 自伴随矩阵

Definition 9.4. Let $A : V \to V$ be an endomorphism, where V is a Euclidean vector space,
A is ***self-adjoint***（自伴随）if

$$(Av, w) = (v, Aw)$$

Therefore, when considering with respect to an orthonormal basis, the matrix A is symmetric, where

$$A = A^t, a_{ij} = a_{ji}$$

This is true because (Ae_i, e_j) represents j^{th} component of A's i^{th} column (a_{ji}), and (e_i, Ae_j) represents i^{th} component of A's j^{th} column (a_{ij}).

9.3.1 Spectral Theorem, 谱定理

Theorem 9.1. A self-adjoint linear operator has a basis of eigenvectors, and we can choose an **orthonormal basis** of eigenvectors.
The process of finding the change-of-basis for self-adjoint linear operators is called the ***Principal Axes Transformation***.

This is:

1. Eigenvectors corresponding to different eigenvalues are orthogonal.
2. If there are "repeated eigenvalues", then part of the matrix corresponding to λ must be λId.

For a self-adjoint matrix, its eigenvalues must have the same geometric and algebraic multiplicity.
Thus, to find an orthonormal basis of eigenvectors (Principal Axes Transformation), we:

1. Find the eigenvalue of A.
2. For each λ, find its eigenvector(s).
3. Within each λ, orthonormalize (Gram-Schimdt) the found eigenvectors.

4. Write the orthonormalized eigenvectors in columns, which form the change-of-basis matrix C.
5. By applying C^tAC, we can diagonalize A where on the diagonal is A's eigenvalues.
5. Or, by applying CDC^t, we can find the linear transformation A with respect to the standard basis.

Theorem 9.2. Spectral decomposition of self-adjoint operators:
If $f : V \to V$ is self-adjoint of a finite-dimensional Euclidean vector space, and $\lambda_1, \dots, \lambda_r$ its distinct eigenvalues, and $P_k : V \to V$ the orthogonal projection onto the eigenspace E_{λ_k}, then

$$f = \sum_{k=1}^{r} \lambda_k P_k$$