Honors Mathematics
Introduction to Linear Algebra

Yao Shaoxiong

UM-SJTU Joint Institute

November 20, 2021

Table of contents

Linear System

For a vector space V together with a field $\mathbb{F}$.
A ***linear system*** of m (algebraic) equations in n unknowns $x_1, ..., x_n \in V$ is a set of equations

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

where $b_1, ..., b_m \in V$ and $a_{ij} \in \mathbb{F}$, $i = 1, ..., m, j = 1, ..., n$.
Comment: Note that vector spaces are not limited to $\mathbb{R}^n$ and fields are not limited to $\mathbb{R}$ and $\mathbb{C}$.

Linear System

- If $b_1 = b_2 = ... = b_m = 0$, then Eq. 1 is called a ***homogeneous system***. Otherwise, it is called an ***inhomogeneous system.***
- If $m > n$ we say that the system is ***underdetermined***, if $m > n$ it is called ***overdetermined***.
- A ***solution*** of a linear system 1 is a tuple of elements $(y_1, ..., y_n) \in V^n$ such that all equations are satisfied.
- Three possible situations are
  - unique solution
  - no solution
  - an infinite number of solutions
- A homogeneous system always has the ***trivial solution***.

$$x_1 = x_2 = ... = x_n = 0$$

Two ways to understand linear system(For $V, \mathbb{F} = \mathbb{R}$)

Method 1. Intersection of m "planes" in $\mathbb{R}^n$.
For $n = 2$ and $m = 2$

$$a_{11}x_1 + a_{12}x_2 = b_1$$
$$a_{21}x_1 + a_{22}x_2 = b_2$$

The solution is the intersection of two lines defined by two equations.
Method 2. Span n vectors of in $\mathbb{R}^m$.
We rewrite the equation,

$$\begin{pmatrix} a_{11} \\ a_{21} \end{pmatrix} x_1 + \begin{pmatrix} a_{21} \\ a_{22} \end{pmatrix} x_2 = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$$

Solution is the appropriate coefficients to make two vectors span the third vector.
Comment: This idea refers to Prof. Gilbert Strange.

The Gauss-Jordan Algorithm

The algorithm transforms the linear system

$$\left.\begin{array}{ccc|c} * & * & * & \diamond \\ * & * & * & \diamond \\ * & * & * & \diamond \end{array}\right. \Rightarrow \left.\begin{array}{ccc|c} 1 & * & * & \diamond \\ 0 & 1 & * & \diamond \\ 0 & 0 & 1 & \diamond \end{array}\right. \Rightarrow \left.\begin{array}{ccc|c} 1 & 0 & 0 & \diamond \\ 0 & 1 & 0 & \diamond \\ 0 & 0 & 1 & \diamond \end{array}\right.$$

We can use three basic operations:

1. Interchanging two rows,
2. Multiplying each element in a row with a number,
3. Adding a multiple of one row to another row.

The system in the middle is called ***upper triangular form***; the system is called ***diagonal form***.
The first step is called ***forward elimination***; the second step is called ***backward substitution***.

The Gauss-Jordan Algorithm

There are two situations that there is no unique solution.

1. No solution exists.

$$\begin{array}{ccc|c} 1 & * & * & \diamond \\ 0 & 1 & * & \diamond \\ 0 & 0 & 0 & \diamond \end{array}$$

2. Infinitely many solutions exist.

$$\begin{array}{ccc|c} 1 & * & * & \diamond \\ 0 & 1 & * & \diamond \\ 0 & 0 & 0 & \diamond \end{array}$$

Comments

- Do not forget the operation Interchanging.
- When there are infinitely many solution, we need to use some variables as parameters in the expression of all solutions. For example, we rename $x_3 = \alpha$ and express x_1 and x_2.

The Solution Set

1.1.5. Definition

The ***solution set*** is the set of all *n-tuples* of elements $x_1, ..., x_n$ in vector space V that satisfy 1.

A solution set can have

- one element
- no element
- infinitely many elements

These are all possible situations.

Question:

Why it is impossible to have 2 solutions?
What is the structure of the solution set?

Fundamental Lemma for Homogeneous Equations

1.1.8. Lemma.

The homogeneous system

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n &= 0 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n &= 0 \end{aligned}$$

of m equations in n real ot complex unknowns $x_1, \dots, x_n$ has a non-trivial solution if $n > m$.

Proof.

The point is that the coefficients are not necessarily non-zero.
For $m = 1$, since all coefficients cannot be zeros, otherwise it is not an equation with m variables. Since $n > m$, we can set $x_i = 1$ for $2 \leq i \leq n$, and then

$$x_1 = -\frac{1}{a_{11}}(a_{12} + \dots + a_{1n})$$

Fundemental Lemma for Homogeneous Equations

Proof(Continued).

For $m > 1$, we use induction. If we have proved the case for $m-1$, the case m will have an extra equation.

$$a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n = 0$$

This equation cannot have all zero coefficients, then we assume that the non- zeros coefficient is a_{mi}, then we have

$$x_i = -\frac{1}{a_{mi}}\Big(\sum_{j=1, j\neq i}^{n} a_{mj}x_j\Big)$$

we substitute this back, and we get to the situation $n-1$ and $m-1$. For this case, we have non-trivial solution. Note that if we eliminate extra variables, we can set the variables eliminated equal to 1 and other elements 0.

Linear Independence

1.2.1. Definition

Let V be a real or complex vector space and $v_1, \ldots, v_n \in V$. Then the vectors $v_1, \ldots, v_n$ are said to be ***independent*** if for $\lambda_1, \ldots, \lambda_n \in \mathbb{F}$

$$\sum_{k=1}^{n} \lambda_k v_k \quad \Rightarrow \quad \lambda_1 = \ldots = \lambda_n = 0$$

A finite set $M \subset V$ is called an ***independent set*** if its elements are independent.

Comment:

This is the original definition of linear independence.

Linear Combinations and Span

1.2.4. Definition.

Let $v_1, \ldots, v_n \in V$ and $\lambda_1, \ldots, \lambda_n \in \mathbb{F}$. Then the expression

$$\sum_{k_1}^{n} \lambda_k v_k = \lambda_1 v_1 + \ldots + \lambda_n v_n$$

is called a ***linear combination*** of the vectors $v_1, \ldots, v_n$.
The set

$$span\{v_1, \ldots, v_n\} = \{y \in V : y = \sum_{k=1}^{n} \lambda_k v_k, \lambda_1, \ldots, \lambda_n \in \mathbb{F}\}$$

is called the (***linear***) ***span*** or the ***linear hull*** of the vectors $v_1, \ldots, v_n$.

Linear Combinations and Span

1.2.6. Lemma.

The vectors $v_1, ..., v_n \in V$ if and only if none of them is contained in the span of all the others.

Proof.

We prove the contraposition of the statement that is

The vectors $v_1, ..., v_2 \in V$ are dependent if and only if one of the is in the non-zero linear combination of all other elements.

The statement follows from the definition of the linear independence.

Span of Subsets

If V is a vector space and M is some subset of V, then we can define the ***span of M*** as the linear combination of all elements in M, i.e.,

$$spanM := \{v \in V : \exists_{n \in \mathbb{N}} \exists_{\lambda_1,\ldots,\lambda_n \in \mathbb{F}} \exists_{m_1,\ldots m_n \in M} : v = \sum_{i=1}^{n} \lambda_i m_i\}$$

Note the difference is that we can take M as a infinite set, and this is different from previous operation on finite number of elements.

Basis

1.2.8. Definition.

Let V be a real or complex vector space. An n-tuple $\mathcal{B} = (b_1, \dots, b_n) \in V^n$ is called an ***(ordered and finite) basis*** of V if every vector v has a unique representation

$$v = \sum_{i=1}^{n} \lambda_i b_i, \quad \lambda_i \in \mathbb{F}.$$

The numbers λ_i are called the ***coordinates*** of v with respect to $\mathcal{B}$. Sometimes we are not interested in the order of elements of a basis and write $\mathcal{B} = \{b_1, \dots, b_n\}$, replacing the tuple by a set. This is known as an ***unordered basis***.

Characterization of Bases

1.2.10. Theorem.

Let V be a real or complex vector space. An n-tuple $\mathcal{B} = \{b_1, ..., b_n\} \in V^n$ is a basis of V if and only if

(i) the vectors $b_1, ..., b_n$ are linearly independent, i.e., $\mathcal{B}$ is an independent set, and

(ii) $V = span\mathcal{B}$

Proof.

$\Rightarrow$

1. Since each element can be uniquely expressed as the combination of basis, then we have $V = span\mathcal{B}$.
2. For the zero vector, it has unique representation with 0 coefficients, this is equivalent to state that

$$\sum_{i=1}^{n} \lambda_i b_i = 0 \quad \Rightarrow \quad \lambda_1 = ... = \lambda_n = 0.$$

Characterization of Bases

$\Leftarrow$ If $span\mathcal{B} = V$, then each element in V can be expressed as a linear combination of elements in $\mathcal{B}$. We need to prove this representation is unique.

$$v = \sum_{i=1}^{n} \lambda_i b_i = \sum_{i=1}^{n} \mu_i b_i, \quad \lambda_i, \mu_i \in \mathbb{F}.$$

Then

$$0 = \sum_{i=1}^{n} (\lambda_i - \mu_i) b_i$$

Since the b_i are all independent, this implies

$$\lambda_i - \mu_i = 0, \quad i = 1, ..., n,$$

so the representation is unique.

Finite- and Infinite- Dimensional Spaces

1.2.11. Definition.

Let V be a real or complex vector space. Then V is called ***finite-dimensional*** if either

- $V = \{0\}$ or
- V processes a finite basis.

If V is not finite-dimensional, we say that it is ***infinite-dimensional***.

Length of Bases

1.2.13. Theorem.

Let V be a real or complex finite-dimensional vector space, $V \neq \{0\}$. Then any basis of V has the same length (number of elements) n.

Proof.

In this proof, we will use fundamental lemma of homogeneous system.
Assume two basis $\mathcal{A} = (a_1, ..., a_n)$ and $\mathcal{B} = (b_1, ..., b_m)$ have different length $m > n$. Then for each element in $\mathcal{B}$, we have

$$b_j = \sum_{i=1}^{n} c_{ij} a_i$$

for $j = 1, ..., m$. Then if

$$\sum_{j=1}^{m} \lambda_j b_j = 0$$

Length of Bases

Proof (Continued).

we have

$$(\sum_{j=1}^{m}\lambda_j c_{1j})a_1 + (\sum_{j=1}^{m}\lambda_j c_{2j})a_2 + ... + (\sum_{j=1}^{m}\lambda_j c_{nj})a_n = 0$$

since $\mathcal{A}$ is a basis, elements in it are linearly independent, so

$$\begin{aligned} \lambda_1 c_{11} + \lambda_2 c_{12} + ... + \lambda_m c_{1m} &= 0 \\ &\vdots \\ \lambda_1 c_{n1} + \lambda_2 c_{n2} + ... + \lambda_m c_{nm} &= 0 \end{aligned}$$

there will be a non-trivial solution to this system. This contradicts the assumption that $\mathcal{B}$ is a basis.

Dimension

1.2.14. Definition

Let V be a finite-dimensional real or complex vector space. We define the ***dimension*** of V, denoted $dim V$ as follows:

(i) If $V = \{0\}$, $dim V = 0$.

(ii) If $V \neq \{0\}$, $dim V = n$, where n is the length of any basis of V.

If V is an infinite dimensional vector space, we write $dim V = \infty$.

1.2.16. Remark.

For an n-dimensional vector space V and a set $A \subset V$. A few questions arise naturally:

1. If A has n elements, is it a basis?
2. If A is independent, what is the maximum number of elements A can contain?

Basis Extension Theorem

To answer questions on the previous slide, we will prove a fundamentally important result called the ***basis extension theorem***. First, we need a lemma:

1.2.17. Lemma.

Let $a_1, \ldots, a_{n+1} \in V$ and assume that $a_1, \ldots, a_n$ are independent and that $a_1, \ldots, a_{n+1}$ are dependent. Then a_{n+1} ua a linear combination of $a_1, \ldots, a_n$.

Proof.

Since $a_1, \ldots, a_{n+1}$ are dependent, we will have $\sum_{i=1}^{n+1} \lambda_i a_i = 0$ with not all $\lambda_i = 0$ for $i = 1, \ldots, n+1$.
Here $\lambda_{n+1} \neq 0$ since $a_1, \ldots, a_n$ are independent, so we have

$$a_{n+1} = -\frac{1}{\lambda_{n+1}} \sum_{i=1}^{n} \lambda_i a_i$$

Maximal Subsets

1.2.18. Definition.

Let V be a real or complex vector space and $A \subset V$ a ***finite*** set. An independent subset $F \subset A$ is called ***maximal*** if every $x \in A$ is a linear combination of elements of F.

If A is finite and $F \subset A$, then $spanF = spanA$. There many maximal subsets for a given A.

1.2.20. Theorem.

Let V be a vector space and $A \subset V$ a finite set. Then every independent subset $A' \subset A$ lies in some maximal subset $F \subset A$.

Proof.

We define an algorithm to find this maximal subset that contains A'.

Maimal Subsets

Proof(Continued).

If we cannot find $x \in A \backslash A'$ that is not linear combination of elements in A' then A' is the maximal.
If we find an element that is not linear combination of elements in A', we will add this element in the set A' and define

$$A'' = A' \cup \{x\}$$

Then we continue find element in $A \backslash A''$. Because A is finite, this process will end, and we find A' is the subset of a maximal subset of A.

Basis Extension Theorem

1.2.21. Basis Extension Theorem

Let V be a finite-dimensional vector space and $A' \subset V$ an independent set. Then there exists a basis of V containing A'.

Proof.

Write $A' = \{a_1, ..., a_m\}$ and choose a basis $\mathcal{A} = \{a_{m+1}, ..., a_{m+n}\}$ of V, $dim V = n$. We now define

$$A = \{a_1, ..., a_{m+n}\} A'$$

By Theorem 1.2.20 there exists a maximal independent subset F of A containing A'.
Since $\mathcal{A}$ is a basis, $V = span\mathcal{A} = spanA$. Furthermore, $spanF = spanA$, so $spanF = V$. Thus F is a basis.

Basis Extension Theorem

1.2.22. Corollary

Let V be a n-dimensional vector space, $n \in \mathbb{N}$. Then any independent set A with n elements is a basis of V.

Proof.

By the basis extension theorem there is a basis containing A. Since this basis will have n elements, A itself is this basis.

1.2.23. Corollary.

Let V be an n-dimensional vector space, $n \in \mathbb{N}$. Then an independent set A may have at most n elements.

Proof.

By the basis extension theorem there is a basis containing A. Since this basis will have n elements, A may not have more elements than this.

Sums of Vector Spaces

1.2.24. Definition.

Let V be a real or complex vector space and U, W be sets in V.

(i) We define the ***sum of U and W*** by

$$U + W := \{v \in V : \exists_{u \in U} \exists_{w \in W} : v = u + w\}$$

(ii) If U and W are subspaces of V with $U \cap W = \{0\}$, the sum $U + W$ is called ***direct***, and we denote it by $U \oplus W$.

1.2.27. Lemma.

The sum $U + W$ ov vector spaces U, W is direct if and only if all $x \in U + W$, $x \neq 0$, have a ***unique*** representation $x = u + w$, $u \in U$, $w \in W$.

Proof.

$\Rightarrow$ If we have U and W are direct, we cannot have two representation of x. Otherwise we will have $U \cap W \neq \emptyset$.

$\Leftarrow$ If the representation is unique, and $x \in U \cap W$, then

$$x = x + 0 = \frac{x}{2} + \frac{x}{2}$$

then the representation is not unique.

1.2.28. Theorem.

Let V be a vector space and U, $W \subset V$ be finite-dimensional subspaces of V. Then

$$dim(U + W) + dim(U \cap W) = dimU + dimW.$$

Proof.

We consider a basis of $U \cap W$, $\mathcal{R} = \{r_1, ..., r_t\}$.

This is an independent set in U and W. From basis extension theorem, we can find a basis $\mathcal{A} = \{r_1, ..., r_t, a_1, ..., a_n\}$ in U and a basis $\mathcal{B} = \{r_1, ..., r_t, b_1, ..., b_m\}$ in W.

Proof(Continued).

We first show that $A \cup B = \{r_1, ..., r_t, ..., a_1, ..., a_n, b_1, ..., b_m\}$ is an independent set.
For

$$\sum \lambda_i r_i + \sum \mu_j a_j + \sum \eta_k b_k = 0$$

Since we have originally $\mathcal{A}$ and $\mathcal{B}$ are independent sets, if μ_j are 0, then λ_i and η_k are 0, and it is same for ηk.
If there exists λ_i not equals to 0 and η_k not equals to 0, we rearrange the equation

$$v = \sum \lambda_i r_i + \sum \mu_j a_j = -\sum \eta_k b_k$$

since v equals to an element in U and an element in W, so it is in $U \cap W$. v has an unique representation using $\mathcal{R}$, then

$$\sum \kappa_s r_s = -\sum \eta_k b_k$$

since $\mathcal{B}$ is an independent set, we have $\kappa_s = \eta_k = 0$. We have the same result for μ_j.

Proof(Continued).

$$\sum \lambda_i r_i + \sum \mu_j a_j + \sum \eta_k b_k = 0$$
$$\Rightarrow \lambda_i = \mu_j = \eta_k = 0$$

The set $\mathcal{A} \cup \mathcal{B}$ is independent.
For $x \in U + W$, $x = u + w$, then since $u = \sum \lambda_i r_i + \sum \mu_j a_j$ and $w = \sum \lambda_i' r_i + \sum \eta_k b_k$

$$x = \sum (\lambda_i + \lambda_i') r_i + \sum \mu_j a_j + \sum \eta_k b_k$$

so $span(\mathcal{A} \cup \mathcal{B}) = U + W$. We have a basis in $U + W$ with $t + n + m$ elements, so

$$\begin{aligned} dim(U + W) + dim(U \cap W) &= (t + n + m) + t \\ &= (t + n) + (t + m) = dimU + dimW \end{aligned}$$

Inner Product Spaces

1.3.1. Definition

Let V be a real or complex vector space. Then a map $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{F}$ is called a ***scalar product*** or ***inner product*** if for all $u, v, w \in V$ and all $\lambda \in \mathbb{F}$

(i) $\langle v, v \rangle \geq 0$ and $\langle v, v \rangle = 0$ if and only if $v = 0$,

(ii) $\langle u, v + w \rangle = \langle u, v \rangle + \langle u, w \rangle$,

(iii) $\langle u, \lambda v \rangle = \lambda \langle u, v \rangle$,

(iv) $\langle u, v \rangle = \overline{\langle v, u \rangle}$.

The pair $(V, \langle \cdot, \cdot \rangle)$ is called an ***inner product space***.

Comment:

For $v \in V$, we consider the inner product $\langle v, \cdot \rangle$ as a function from V to $\mathbb{F}$, then this function is linear.

The Induced Norm

1.3.4. Definition.

Let $(V, \langle\cdot,\cdot\rangle)$ be can inner product space. The map

$$||\cdot|| : V \to \mathbb{R}, \qquad ||v|| = \sqrt{\langle v, v\rangle}$$

is called the ***induced norm*** on V.

Comment:

The first and second requirements of the norm are true from definition (i) and (iii).
The triangle is true from Cauchy-Schwarz inequality.

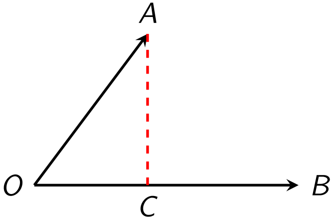
1.3.6. Cauchy-Schwarz Inequality.

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

$$|\langle u, v\rangle| \leq ||u|| \cdot ||v||$$

for all $u, v \in V$, where $||\cdot||$ is the induced norm.

Proof.



Here $u = \overrightarrow{OA}$ and $v = \overrightarrow{OB}$,
We have

$$||\overrightarrow{OC}|| = \frac{|\langle v, u\rangle|}{||v||}$$

then if we let $e := v/||v||$. Then

$$\overrightarrow{OC} = \langle e, u\rangle e$$

So we consider the length of $\overrightarrow{AC}$

$$\begin{aligned} ||\overrightarrow{AC}||^2 &= ||u - \langle e, u\rangle e^2 = \langle u - \langle e, u\rangle e, u - \langle e, u\rangle e\rangle \\ &= ||u||^2 - |\langle e, u\rangle|^2 \end{aligned}$$

Since the square is non-negative,

$$|\langle u, v\rangle|^2 = ||v||^2 \cdot |\langle u, e\rangle|^2 \leq ||u||^2 \cdot ||v||^2.$$

The Induced Norm

1.3.7. Corollary.

The induced norm is actually a norm, i.e., it satisfies

(i) $||v|| \geq 0, ||v|| = 0 \Leftrightarrow v = 0$,

(ii) $||\lambda v|| = |\lambda| \cdot ||v||$,

(iii) $||u + v|| \leq ||u|| + ||v||$

for all $u, v \in V$ and $\lambda \in \mathbb{F}$.

Proof.

We only need to check the third property.

$$\begin{aligned} ||u + v||^2 &= ||u||^2 + ||v||^2 + \langle u, v \rangle + \langle v, u \rangle \\ &= ||u||^2 + ||v||^2 + 2Re\langle u, v \rangle \\ &\leq ||u||^2 + ||v||^2 + 2\langle u, v \rangle \\ &\leq ||u||^2 + ||v||^2 + 2||u||||v|| \\ &= (||u|| + ||v||)^2. \end{aligned}$$

Orthogonality

1.3.11. Definition.

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

(i) Two spaces $u, v \in V$ are called ***orthogonal*** or ***perpendicular*** if $< u, v >= 0$

(ii) We called

$$M^{\perp} := \{v \in V : \forall_{m \in M} \langle m, v\rangle = 0\}$$

the ***orthogonal complement*** of a set $M \subset V$.

For short, we sometimes write $v \perp M$ instead of $v \in M^{\perp}$ or $v \perp M$ for all $m \in M$.

1.3.12. Lemma.

The orthogonal complement $M^{\perp}$ is a subspace of V.

Proof.

This result follows the linearity of inner product in definition.

Orthogonality

Pythagoras's Theorem.

Let $(V, \langle \cdot, \cdot \rangle)$ be an inner product space and M some subset of V. Let $z = x + y$, where $x \in M$ and $y \in M^{\perp}$. Then

$$||z||^2 = ||x||^2 + ||y||^2$$

.

Parallelogram Rule

Let V be a real or complex vector space. Then every norm on V, if it is induced by some inner product, then it satisfies

$$||x + y||^2 + ||x - y||^2 = 2(||x||^2 + ||y||^2)$$

for all $x, y \in V$.

Orthogonality

Exercise

Suppose that $V_1, ..., V_m$ are inner product spaces. Show that the equation

$$\langle (u_1, ..., u_m), (v_1, ..., v_m) \rangle = \langle u_1, v_1 \rangle + \cdots + \langle u_m, v_m \rangle$$

defined an inner product on $V_1 \times \cdots \times V_m$.
Note that for $V_1, ..., V_m$, they may have different definition of inner product.

Orthonorlma Systems

1.3.14. Definition.

Let $(V, \langle \cdot, \cdot \rangle)$ be an inner product space. A tuple of vectors $(v_1, ..., v_r) \subset V$ is called a ***(finite) orthonormal system*** if

$$\langle v_j, v_k \rangle = \delta_{jk} := \begin{cases} 1 & \text{for } j = k \\ 0 & \text{for } j \neq k \end{cases} \qquad j, k = 1, ..., r,$$

i.e.,if $||v_k|| = 1$ and $v_j \perp v_k$ for $j \neq k$.

1.3.16. Lemma.

Let $(V, \langle \cdot, \cdot \rangle)$ be an inner product space and $\mathcal{F} = (v_1, ..., v_r) \subset V$ an orthonormal system. Then the elements of $\mathcal{F}$ are linearly independent.

Proof.

We assume that $\sum \lambda_i v_i = 0$, and take inner product of this equation with v_i. We have $\lambda_1 = \cdots = \lambda_r = 0$.

Orthonormal Bases

1.3.17. Definition.

Let $(V, \langle\cdot,\cdot\rangle)$ be a finite-dimensional inner product vector space and $\mathcal{B} = (e_1, \dots, e_n)$ a basis of V. If $\mathcal{B}$ is also an orthonormal system, we say that $\mathcal{B}$ is an ***orthonormal basis*** (ONB).

1.3.18. Theorem.

Let $(V, \langle\cdot,\cdot\rangle)$ be a finite-dimensional inner product vector space and $\mathcal{B} = (e_1, \dots, e_n)$ an orthonormal basis of V. Then every $v \in V$ has the basis representation

$$v = \sum_{j=1}^{n} \langle e_j, v\rangle e_j.$$

1.3.19. Definition.

The numbers $\langle e_j, v\rangle$ are called ***Fourier coefficients*** of v with respect to the basis $\mathcal{B}$. The vector

$$\pi_{e_i} v := \langle e_i, v\rangle e_i$$

is called the ***projection of v onto e_i***.

Orthonormal Bases

Proof of Theorem 1.3.18.

Since $\mathcal{B}$ is a basis, for every $v \in V$ it has basis representation

$$v = \sum_{j=i}^{n} \lambda_j e_j.$$

Then we take inner product of this equation with e_k

$$\langle e_k, v \rangle = \sum_{j=1}^{n} \lambda_j \langle e_k, e_j \rangle = \lambda_k,$$

this gives the coefficients of the basis representation.

1.3.20. Parseval's Theorem

Let $(V\langle \cdot, \cdot \rangle)$ be a finite-dimensional inner product vector space and $\mathcal{B} = \{e_1, ..., e_n\}$ an orthonormal basis of V.
Then

$$||v||^2 = \sum_{i=1}^{n} |\langle v, e_i \rangle|^2$$

for any $v \in V$.

The Projection Theorem

1.3.21. Projection Theorem.

Let $(V, \langle \cdot, \cdot \rangle)$ be a (possibly infinite-dimensional) inner product vector space and $(e_1, ..., e_r)$, $r \in \mathbb{N}$, be an orthonormal system in V. Denote $U := \operatorname{span}\{e_1, ..., e_r\}$.
Then for every $v \in V$ there exists a unique representation

$$v = u + w \qquad \text{where } u \in U \text{ and } w \in U^\perp$$

and $u = \sum_{i=1}^{r} \langle e_i, v \rangle e_i, \ w := v - u$.

1.3.22. Definition.

The vector

$$\pi_U v := \sum_{i=1}^{r} \langle e_i, v \rangle e_i$$

is called the ***orthogonal projection of*** v ***onto*** U. Note that U may consist different orthonormal system, but this expression will give the same $\pi_U v$ from the uniqueness of the Projection Theorem.

The Projection Theorem

Proof of the Projection Theorem.

We first show the uniqueness of the decomposition: Assume that $v = u + w = u' + w'$. Then by Pythagoras's theorem,

$$0 = ||u - u' + w - w'||^2 = ||u - u'||^2 + ||w - w'||^2$$

so $||u - u'|| = ||w - w'|| = 0$. Thus $u = u'$ and $w = w'$.
Now we need to confirm that $u \in U$ and $w \in U^{\perp}$. It is clear that

$$u = \sum_{i=1}^{r} \langle e_i, v \rangle e_i$$

lies in U. So we need to check that $w \in U^{\perp}$.

The Projection Theorem

Proof of the Projection Theorem(Continued).

For this orthonormal basis $\mathcal{B} = (e_1, ... e_r)$, we have

$$\begin{aligned}
\langle e_i, v - u \rangle &= \langle e_i, v \rangle - \sum_{j=1}^{r} \langle e_i, e_j \rangle \langle e_j, v \rangle \\
&= \langle e_i, v \rangle - \sum_{j=1}^{r} \delta_{ij} \langle e_j, v \rangle \\
&= \langle e_i, v \rangle - \langle e_i, v \rangle \\
&= 0
\end{aligned}$$

so $w = v - u$ is orthogonal to an orthonormal basis in U. For an arbitrary $u' \in U$ with basis representation $u' = \sum \lambda_i' e_i$

$$\langle u', w \rangle = \sum_{i=1}^{r} \overline{\lambda'} \langle e_i, w \rangle = 0.$$

So $w \in U^{\perp}$.

Orthogonal Subspaces

1.3.24. Corollary.

Let $(V, \langle \cdot, \cdot \rangle)$ be a (possibly infinite-dimensional) inner product vector space and let $U \subset V$ be a finite-dimensional subspace. Then

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

If V is finite-dimensional, then

$$\dim V = \dim U + \dim U^{\perp}.$$

Projection theorem states that V has unique representation with $u \in U$ and $w \in U^{\perp}$, so from Lemma 1.2.27. they are orthogonal and from Theorem 1.2.28. the values of dimension has this relation.

Comment:

Note this definition of orthogonality is different from geometrical definition. Two planes in $\mathbb{R}^3$ are not orthogonal.

Bessel's Inequality

1.3.25. Bessel's Inequality.

Let $()V, \langle \cdot, \cdot \rangle)$ be an (possibly infinite-dimensional) inner product space and $(e_1, ..., e_n)$ and orthonormal system in V. Then, for any $v \in V$ and any $r \leq n$,

$$\sum_{i=1}^{r} |\langle e_i, v \rangle|^2 \leq ||v||^2.$$

Proof.

From the Projection Theorem and Pythagoras' Theorem 1.3.13 we have $v - u \perp u$, so $||v - u||^2 + ||u||^2 = ||v||^2$, then

$$0 \leq ||v - u||^2 = ||v||^2 - ||u||^2 = ||v||^2 - \sum_{i=1}^{r} |\langle e_i, v \rangle|^2.$$

Best Approxiamtion

Using finite number of bases, we cannot express an arbitrary $v \in V$, however, it is possible for us to find the expression that minimizes $||v-u||^2$ with $u \in U$, here u has basis representation

$$
\begin{aligned}
||v - \sum_{i=1}^{r} \lambda_i e_i||^2 &= ||v||^2 + \sum_{i=1}^{r} |\lambda_i|^2 - \sum_{i=1}^{r} \lambda_i \langle v, e_i \rangle - \sum_{i=1}^{r} \overline{\lambda_i} \langle e_i, v \rangle \\
&= ||v||^2 + \sum_{i=1}^{r} (\langle e_i, v \rangle - \lambda_i)\overline{(\langle e_i, v \rangle - \lambda_i)} - \sum_{i=1}^{r} |\langle e_i, v \rangle|^2 \\
&= ||v||^2 - \sum_{i=1}^{r} |\langle e_i, v \rangle|^2 + \sum_{i=1}^{r} |\langle e_i, v \rangle - \lambda_i|^2
\end{aligned}
$$

Note the first two terms are constant, and the last term will be 0 if $\lambda_i = \langle e_i, v \rangle$, i.e., the coefficients are just Fourier coefficients. If we extend the number of bases from r to r', then

$$
||v - \sum_{i=1}^{r'} \langle e_i, v \rangle e_i|| \leq ||v - \sum_{i=1}^{r} \langle e_i, v \rangle e_i|| \qquad \text{for any } r' \geq r,
$$

the approximation will be better.

Gram-Schmidt Orthonormalization

Let $(V, \langle \cdot, \cdot \rangle)$ be an inner product vector space and $(v_1, ..., v_n)$ is a system of vectors in V.

Motivation:

Given a vector v, what is the best approximation using this system? Or, what is the best approximate vector in $\text{span}\{v_1, ..., v_n\}$?

Solution:

An orthonormal system in $\text{span}\{v_1, ..., v_n\}$ will give the result by inner product. But an orthonormal system will be arbitrary, what is the most convenient one we can find?

Gram-Schmidt Orthonormalization

We start from $\mathcal{B} = \{v_1\}$, and add another normalized vector that is orthogonal to the span of existing basis.

(i) Normalize v_1

$$w_1 = \frac{v_1}{||v_1||}, \quad \mathcal{B} = \{w_1\}$$

(ii) Add the orthogonal part of v_2 to this system,

$$v_2' = v_2 - \langle w_1, v_2 \rangle w_1 \in \text{span}\{v_1\}^\perp$$

Normalize v_2'

$$w_2 = \frac{v_2'}{||v_2'||}, \quad \mathcal{B} = \{w_1, w_2\}$$

$\vdots$

(iii) Add the orthogonal part of v_n to this system

$$v_n' = v_n - \sum_{i=1}^{n-1} \langle w_i, v_n \rangle w_i \in \text{span}\{w_1, \ldots, w_{n-1}\}^\perp$$

Normalize v_n'

$$w_n = \frac{v_n'}{||v_n'||}, \quad \mathcal{B} = \{w_1, \ldots, w_n\}$$

Linear Maps on Vector Spaces

1.4.1. Definition.

Let $(U, \oplus, \odot)$ and $(V, \boxplus, \boxdot)$ be vector spaces that are either both real or both complex. Then a map $L : U \to V$ is said to be ***linear*** if it is both ***homogeneous*** and ***additive***, i.e.,

$$L(\lambda \odot u) = \lambda \boxdot L(u) \quad \text{and} \quad L(u \oplus v) = L(u) \boxplus L(v)$$

for all $u, v \in U$ and $\lambda \in \mathbb{F}$. The set of all linear maps $L : U \to V$ is denoted by $\mathcal{L}(U, V)$.

1.4.2. Remark.

(i) Note that

$$L(0 \oplus 0) = L(0) \boxplus L(0) = L(0)$$

so $L(0) = 0$. Then we will use same symbol 0 for the zero in U or in V.

(ii) The identity map $\mathrm{id} : V \to V$, $\mathrm{id}(v) = v$, is linear.

Linear Maps are Structure-Preserving

A linear map

$$L : U \to V$$

between vector spaces $(U, \oplus, \odot)$ and $(V, \boxplus, \boxdot)$ is a ***structure-preserving map***.

Note that the composition of linear maps will still be linear, i.e., if $L_1 \in \mathcal{L}(U, V)$ and $L_2 \in \mathcal{L}(V, W)$, then $L_1 \circ L_2 \in \mathcal{L}(U, W)$.

We assume that L is bijective such that L^{-1} exists. Then for a linear map $\sigma : V \to V$, we can define $\pi : U \to U$,

$$\pi = L^{-1} \circ \sigma \circ L$$

since L and σ are linear, the new function π is also a linear map.

$$\begin{tikzcd} U \arrow[r, "L"] \arrow[d, "\pi"'] & V \arrow[d, "\sigma"] \\ U & V \arrow[l, "L^{-1}"'] \end{tikzcd}$$

Comments on Structure-Preserving

Other thoughts....

Here you see a formula that will appear frequently in your further study:

$$z = L^{-1} \circ x \circ L$$

One interpretation is:

The operation on one space is done by another operation on another space.

A similar formula has appeared before, recall that the projection of vector v on a normal vector e is

$$\pi_e = \langle e, v \rangle e$$

You will further understand that a vector is a linear map $e : \mathbb{F} \to \mathbb{F}^n$, and the inner product $\langle e, \cdot \rangle$ is the adjoint.

If $\text{span}\{v\} \neq \text{span}\{e\}$, and a function $f : \text{span}\{v\} \to \text{span}\{e\}$ is linear, such that

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

then f will be a multiple of π_e.

Homomorphisms and Finite-Dimensional Spaces

1.4.4. Theorem.
Let U, V be real or complex spaces. If $(b_1, ..., b_n)$ is a basis in U, then for every $(v_1, ..., v_n) \in V^n$, there exists a unique linear map $L : U \to V$ such that $Lb_k = v_k$, $k = 1, ..., n$.
Note that we assume $\dim U = n$ but V can be infinite-dimensional.

Proof.
(Uniqueness) Assume another linear map $M \in \mathcal{L}(U, V)$ with $Mb_k = v_k$.
For any $u \in U$, $u = \sum \lambda_k b_k$,

$$\begin{aligned} Lu &= L(\sum \lambda_k b_k) = \sum \lambda_k L(b_k) \\ &= \sum \lambda_k v_k = \sum \lambda_k M(b_k) = M(\sum \lambda_k v_k) = Mu. \end{aligned}$$

so $Lu = Mu$ for any $u \in U$, we then have $L = M$.

Proof(Continued).

(Existence) We define the linear map for $u \in U$, $u = \sum \lambda_k b_k$,

$$Lu := \sum \lambda_k v_k$$

We need to check this is a linear map. If $u, u' \in U$ are $u = \sum \lambda_k b_k$ and $u' = \sum \lambda_k b_k$, we have

$$L(u + u') = \sum (\lambda_k + \lambda'_k) v_k = \sum \lambda_k v_k + \sum \lambda'_k v_k = Lu + Lu'$$

Let $\mu \in \mathbb{F}$,

$$L(\mu u) = L(\sum \mu \lambda_k v_k) = \sum \mu \lambda_k v_k = \mu \sum \lambda_k v_k = \mu L(u)$$

so this is a linear map.

Coordinate Map and Dual Space

1.4.6. Examples. Let V a real or complex vector space, and $(b_1, ..., b_n)$ be a basis of V if it is finite-dimensinonal.

(i) If $\dim V = n$, the ***corrdinate map***

$$\varphi : V \to \mathbb{F}^n, \qquad v = \sum \lambda_k b_k \mapsto \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix}$$

is linear (and bijective).

(ii) Then $\mathcal{L}(V, \mathbb{F})$ is known as the ***dual space*** of V and denoted by V^*.
If $\dim V = n$, then we can find a basis of the dual space V^*,

$$b_k^* : V \to \mathbb{F}, \qquad b_k^*(b_j) = \delta_{jk}.$$

Note that if V^* is an inner product space, then

$$b_k^* = \langle b_k, \cdot \rangle$$

can express the basis in V^*.

Range and Kernel

1.4.8. Definition.

Let U, V be real or complex vector spaces and $L \in \mathcal{L}(U, V)$. Then we define the ***range*** of L by

$$\mathrm{ran} L := \{v \in V : \exists_{u \in U} v = Lu\}$$

and the ***kernel*** of L by

$$\ker L := \{u \in U : Lu = 0\}$$

We can notice that $\mathrm{ran} L \subset V$ and $\ker L \subset U$ are subspaces.

1.4.9. Remark.

For $L \in \mathcal{L}(U, V)$, it is injective if and only if $\ker L = \{0\}$.

Proof.

$$\begin{aligned} L \text{ is injective } &\Leftrightarrow Lx = Ly \;\Rightarrow\; x = y, \text{ for any } x, y \in V \\ &\Leftrightarrow L(x-y) = 0 \;\Rightarrow\; x - y = 0, \text{ for any } x, y \in V \\ &\Leftrightarrow Lu = 0 \;\Rightarrow\; u = 0, \text{ for any } u \in V \\ &\Leftrightarrow \ker L = \{0\}. \end{aligned}$$

Range and Kernel

Digression.

Let U, V be vector spaces with $\dim U = n$. For a basis $(b_1, ..., b_n) \in U$, and we define L by a n-tuple $(Lb_1, ..., Lb_n) \in V$, then what are $\operatorname{ran} L$ and $\ker L$?

Solution.

We apply Gram-Schmidt to $(Lb_1, ..., Lb_n)$ and apply the same linear combinations on $(b_1, ..., b_n)$.
Then

$$(Lb_1, ..., Lb_n) \xrightarrow{Gram-Schmidt} (c_1, ..., c_r, 0, ..., 0)$$

$$(b_1, ..., b_n) \xrightarrow{\text{Same Linear Combinations}} (b_1', ..., b_n')$$

here $(c_1, ..., c_r)$ is an orthonormal basis. Then we get

$$\operatorname{ran} L = \operatorname{span}\{c_1, ..., c_r\}$$
$$\ker L = \operatorname{span}\{b_{r+1}', ..., b_n'\}.$$

Nomenclature

Fancy names for linear maps. A homomorphism $L \in \mathcal{L}(U, V)$ is

- an ***isomorphism*** if L is bijective (isos means "equal");
- an ***endomorphism*** if $U = V$;
- an ***automorphism*** if $U = V$ and L is bijective;
- ***epimorph*** if L is surjective;
- ***monomorph*** if L is injective;

Comments:

Possibly only two terms will occur again...
Isomorphism L "relabels" elements from U to V, such that they have the same structure.
Homomorphism L partly preserves the structure from U to V.
Prof. Zach McKenzie likes these two terms very much, you will see them frequently in Discrete Math (if you are ECE major), such as group homomorphism, lattice homomorphism, and graph isomorphism...

Isomorphisms

1.4.11. Theorem.

Let U, V be finite-dimensional vector spaces and $L \in \mathcal{L}(U, V)$. Then L is an isomorphism if and only if every basis $(b_1, \ldots, b_n)$ of U, the tuple $(Lb_1, \ldots, Lb_n)$ is a basis of V.

Proof.

($\Rightarrow$) Assume that L is bijective. For $y \in V$, there will be unique $x = L^{-1}y \in U$ because L is bijective. And $x = \sum \lambda_k b_k$ is the unique basis representation in U. Now

$$y = L(\sum \lambda_k b_k) = \sum \lambda_k \cdot Lb_k$$

so y has unique representation using $(Lb_1, \ldots, Lb_n)$. Then $(Lb_1, \ldots, Lb_n)$ is a basis in V, from the ***original definition***.

Isomorphisms

Proof(Continued).

($\Leftarrow$) Note the the coordinate map will be a bijection, so $\varphi : U \to \mathbb{F}^n$ with basis $(b_1, ..., b_n)$ and $\phi : V \to \mathbb{F}^n$ with basis $(Lb_1, ..., Lb_n)$ are bijections. Then from Theorem 1.4.4. $L = \phi^{-1} \circ \varphi$ are the same map, then L is a bijection from U to V since the composition of bijections will still be a bijection.

1.4.12. Definition.

Two vector spaces U and V are called ***isomorphic***, written $U \cong V$, if there exists an isomorphism $\varphi : U \to V$.

1.4.13. Lemma.

Two finite-dimensional vector spaces U and V are isomorphic if and only if they have the same dimension:

$$U \cong V \qquad \Leftrightarrow \qquad \dim U = \dim V$$

Proof.

L is isomorphism $\Leftrightarrow$ $(b_1, ..., b_n)$ and $(Lb_1, ..., Lb_n)$ are basis.

The Dimension Formula

1.4.14. Dimension Formula.

Let U, V be real or complex vector spaces, $\dim U < \infty$. Let $L \in \mathcal{L}(U, V)$. Then

$$\dim \operatorname{ran} L + \dim \ker L = \dim U.$$

Proof.

Let $\dim U = n$, and $\dim \ker L = r$, $(a_1, ..., a_r)$ is a basis of $\ker L$.
By Basis Extension Theorem, we can find $(a_1, ..., a_r, a_{r+1}, ..., a_n)$ is a basis in U.
We want to show that L is a bijection from $\operatorname{span}\{a_{r+1}, ..., a_n\}$ to $\operatorname{ran} L$. ***(We change the domain of this function)***
(1)Surjection. For $y \in \operatorname{ran} L$, there exists $x \in U$, with $x = \sum \lambda_k a_k$, $Lx = y$. Because

$$y = Lx = L(\lambda_1 a_1 + \cdots + \lambda_n a_n) = \lambda_{r+1} L a_{r+1} + \cdots + \lambda_n L a_n,$$

we have $z = \lambda_{r+1} a_{r+1} + \cdots + \lambda_n a_n \in \operatorname{span}\{a_{r+1}, ..., a_n\}$, with $Lz = y$.

The Dimension Formula

Proof(Continued).

(2)Injection. If for $y \in \operatorname{ran} L$, there are $Lx = Lx' = y$ with $x = \sum_{k=r+1}^{n} \lambda_k a_k$ and $x = \sum_{k=r+1}^{n} \lambda'_k a_k$, then $L(x - x') = 0$ means $x - x' \in \ker L$, so we have

$$(\lambda_{r+1} - \lambda'_{r+1})a_{r+1} + \cdots + (\lambda_n - \lambda'_n)a_n = \eta_1 a_1 + \cdots + \eta_r a_r.$$

We have $(a_1, ..., a_n)$ is a basis, so $\lambda_k = \lambda'_k$ for $k = r + 1, ..., n$.

1.4.15. Corollary.

Let U, V be real or complex finite-dimensional vector spaces with $\dim U = \dim V$. Then a linear map $L \in \mathcal{L}(U, V)$ is injective if and only if it is surjective.

Proof.

We notice that $\ker L = \{0\}$ are $\operatorname{ran} L = V$ are equivalent.

Normed Vector Spaces and Bounded Linear Maps

1.4.16. Definition.

Let $(U, ||\cdot||_U)$ and $(V, ||\cdot||_V)$ be normed vector spaces. Then a linear map $L : U \to V$ is said to be ***bounded*** if there exists some constant $c > 0$ (called a ***bound*** for L) such that

$$||Lu||_V \leq c \cdot ||u||_U \qquad \text{for all } u \in U.$$

1.4.18. Examples.

The map $\int : C([a, b]) \to \mathbb{R}$ definite integral is a bounded map.

$$\int_a^b f \leq |b - a| \cdot \sup_{x \in [a,b]} |f(x)|$$

Note that continuous function is bounded on $[a, b]$, so $\sup_{x \in [a,b]}$ exists. Here the norm of $C((a, b))$ is $\sup_{x \in [a,b]} |\cdot|$.
Question: Will it still be bounded with a different norm?

Bounded Linear Maps

1.4.17. Remark.

If U is finite-dimensional vector space, then any linear map is bounded.

Proof.

Consider $(b_1, ..., b_n)$ is a basis in U, then for $u \in U$, with $u = \sum \lambda_k b_k$,

$$\begin{aligned} \|Lu\|_V &\leq |\lambda_1|\|Lb_1\|_V + \cdots + |\lambda_n|\|Lb_n\|_V \\ &\leq n \cdot \max_{k=1,\ldots,n} |\lambda_k|\|Lb_k\|_V \\ &\leq n \cdot \|u\|_U. \end{aligned}$$

So every liner map is bounded.

The Operator Norm

1.4.19. Definition and Theorem.

Let U, V be normed vector spaces. Then the set of bounded linear maps $\mathcal{L}(U, V)$ is also a vector space and

$$\|L\| := \sup_{u \in U,\, u \neq 0} \frac{\|Lu\|_V}{\|u\|_U} = \sup_{u \in U,\, \|u\|_U = 1} \|Lu\|_V.$$

defines a norm, the so-called ***operator norm*** or ***induced norm*** on $\mathcal{L}(U, V)$.

Proof.

This is a norm since $\|\cdot\|_V$ is a norm, for each $\|u\|_V = 1$,

$$\|Lu\| \geq 0, \text{ and } \|Lu\| = 0 \Rightarrow Lu = 0$$

so $\sup \|Lu\|_V \geq 0$ and $\sup \|Lu\|_V = 0$ iff $Lu = 0$ for any $u \in U$.
We also have

$\sup \|\lambda Lu\|_V = |\lambda| \cdot \sup \|Lu\|_V$, $\sup \|Lu + L'u\| \leq \sup \|Lu\| + \sup \|L'u\|$

The Operator Norm

Theorem.

The operator norm has the additional property that

$$\|L_2 L_1\| \leq \|L_2\| \cdot \|L_1\|, \qquad L_1 \in \mathcal{L}(U, V), \qquad L_2 \in \mathcal{L}(V, W).$$

Proof.

From definition

$$\begin{aligned}
\|L_2 L_1\| &= \sup_{u \in U,\ \|u\|_U = 1} \|L_2 L_1 u\| \\
&\leq \sup_{u \in U,\ \|u\|_U = 1} \left(\sup_{v \in V,\ \|v\|_V = 1} \|L_2 v\| \cdot \|L_1 u\| \right) \\
&= \sup_{v \in V,\ \|v\|_V = 1} \|L_2 v\| \cdot \sup_{u \in U,\ \|u\|_U = 1} \|L_1 u\| \\
&= \|L_2\| \cdot \|L_1\|.
\end{aligned}$$

A Calculus of Linear Maps

The word ***calculus*** means a "scheme of calculating".

Motivation

Recall Theorem 1.4.4., a linear map from U to V can be determined by a basis $(b_1, ..., b_n)$ in U and $(Lb_1, ..., Lb_n)$ in V. We want to find a good notation to "store" these information.

Solution

For finite-dimensional case, we notice that if $\dim U = n$ and $\dim V = m$, we can find isomorphism $\varphi_1 : U \to \mathbb{R}^n$ and $\varphi_2 : V \to \mathbb{R}^m$. For linear maps from $\mathbb{R}^n$ to $\mathbb{R}^m$, we can write them down using ***matrices***.

Comment:

We are supposed to replace $\mathbb{R}$ by $\mathbb{F}$ here since U and V may be complex.

Matrices

1.5.1. Definition.

An matrix is a map

$$a : \{1, ..., m\} \times \{1, ..., n\} \to \mathbb{F}, \qquad (i, j) \mapsto a_{ij}.$$

And it can be represented by a graph

$$A := \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}$$

We denote all $m \times n$ matrix over $\mathbb{F}$ by $\mathrm{Mat}(m \times n; \mathbb{F})$. We often write $\mathbb{C}$ or $\mathbb{R}$ instead of $\mathbb{F}$.

Matrices as Linear Maps

1.5.3. Theorem.

For $A \in \mathrm{Mat}(m \times n; \mathbb{F})$ and $(e_1, ..., e_n)$ is the standard basis in $\mathbb{F}^n$, then we define $j : \mathrm{Mat}(m \times n; \mathbb{F}) \to \mathcal{L}(\mathbb{F}^n, \mathbb{F}^m)$ by

$$(e_1, ..., e_n) \to (Le_1, ..., Le_n) = (a_{\cdot 1}, ..., a_{\cdot n})$$

finding L that satisfies this condition. Then $j(A) = L$ is an isomorphism, and $\mathrm{Mat}(m \times n; \mathbb{F}) \cong \mathcal{L}(\mathbb{F}^n, \mathbb{F}^m)$.

Proof.

From Theorem 1.4.4., L is unique and j is a function.
(1)Injection. If $j(A) = j(A')$, then each column are the same, so $A = A'$.
(2)Surjection. For each $L \in \mathcal{L}(\mathbb{F}^m, \mathbb{F}^n)$, there exists $(Le_1, ..., Le_n)$, so j is a surjection.

Comment:

Be very careful that a matrix is defined on ***standard basis***.
We will replace L by A since j is a bijection.

Matrices as Linear Maps

Calculation Method.

For $x \in \mathbb{F}^n$,

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = x_1 \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \cdots + x_n \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} = \sum_{k=1}^{n} x_k e_k .$$

then Ax is

$$\begin{aligned} Ax &= A \sum_{k=1}^{n} x_k e_k = \sum_{k=1}^{n} x_k A e_k = \sum_{k=1}^{n} x_k a_{.k} \\ &= x_1 \begin{pmatrix} a_{11} \\ \vdots \\ a_{m1} \end{pmatrix} + \cdots + x_n \begin{pmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{pmatrix} = \begin{pmatrix} x_1 a_{11} + \cdots + x_n a_{1n} \\ \cdots \\ x_1 a_{m1} + \cdots + x_m a_{mn} \end{pmatrix} \end{aligned}$$

Comment:

The last expression looks like inner product on $\mathbb{F}^n$ and rows of A are involved.

Composition and Matrix Product

1.5.4. Definition and Theorem.

If $A \in \mathrm{Mat}(l \times m; \mathbb{F})$ and $B \in \mathrm{Mat}(m \times n; \mathbb{F})$, then we define the composition of A and B as the ***product of A and B*** by

$$AB \in \mathrm{Mat}(l \times n; \mathbb{F}), \qquad AB := \left(\sum_{k=1}^{m} a_{ik} b_{kj} \right)_{\substack{i=1,\dots,l \\ j=1,\dots,n}}$$

Proof.

We need to check that AB is the composition of A and B.
The composition of A and B is supposed to have n columns, and ABe_j is the jth column,

$$ABe_j = Ab_j = \sum_{k=1}^{m} a_{\cdot k} b_{kj}$$

so the entry on ith row and jth column is

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

Matrix Product

Comment on matrix product:

From composition of linear maps, the product is associative but not commutative.

Example.

Calculate the matrix product,

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

Question:

What is the result for general case?

Vectors are Matrices

Motivation.

Note that vectors in $\mathbb{F}^n$ is written as a map $u : \{1\} \times \{1, ..., n\} \to \mathbb{F}$, which is exactly a $n \times 1$ matrix. So what is the linear map it represents?

Solution.

From the definition of matrices, we know that $u \in \mathcal{L}(\mathbb{F}, \mathbb{F}^n)$. So it maps 1 to u.

Comment:

So there are three interpretation of a vector in $\mathbb{F}^n$, a vector, a matrix and a linear map. We have three vector spaces are isomorphic,

$$\mathbb{F}^n \cong \mathrm{Mat}(n \times 1, \mathbb{F}) \cong \mathcal{L}(\mathbb{F}, \mathbb{F}^n).$$

For finite-dimensional vector space with $\dim V = n$, we can replace $\mathbb{F}^n$ by V.

Matrices of Inner Product

Motivation.

We have shown that for $v \in V$, $\langle v, \cdot \rangle$ is a linear map in $\mathcal{L}(V, \mathbb{F})$, so what is the matrix of this linear map?

Solution.

Assume that $\dim V = n$ and $(e_1, ..., e_n)$ is a standard basis in V. If v^* is the matrix of $\langle v, \cdot \rangle$, then

$$v^* = (\langle v, e_1 \rangle, ..., \langle v, e_n \rangle) = (\overline{\langle e_1, v \rangle}, ..., \overline{\langle e_n, v \rangle}) = (\overline{v_1}, ..., \overline{v_n}).$$

Note that we transfer a linear map $v \in \mathcal{L}(\mathbb{F}, V)$ to a linear map $v^* \in \mathcal{L}(V, \mathbb{F})$.

Comment:

This result shows that

$$V^* = \mathcal{L}(V, \mathbb{F}) \cong \mathrm{Mat}(1 \times n; \mathbb{F}).$$

The Matrix of the Adjoint of a Linear Map

Motivation.

Recall that 'dual' is means a map that reverses the domain and the codomain, so what is T^* that corresponds to $T \in \mathcal{L}(U, V)$?

$$\begin{tikzcd}
\mathbb{F} \arrow[r, shift left, "u"] \arrow[rr, bend left=60, "v"] & U \arrow[l, shift left, "u^*"] \arrow[r, shift left, "T"] & V \arrow[l, shift left, "T^*"] \arrow[ll, bend left=60, "v^*"]
\end{tikzcd}$$

We can determine T^* from this diagram, if A is the matrix of T,

$$\begin{tikzcd}
u \arrow[r] \arrow[d] & v = Tu \arrow[r] & v \arrow[d] \\
u^* \arrow[r] & v^* = u^* T^* \arrow[r] & v^*
\end{tikzcd}$$

we want to find the matrix of T^* denoted A^*.

The Matrix of the Adjoint of a Linear Map

Solution.
The relation between T and T^* is

$$(Tu)^* = u^* T^*,$$

we transfer T to A, $u \in U$ to $x \in \mathbb{F}^n$ with $x = \sum x_i e_i$.
(1)Left From the additivity of dual,

$$(Ax)^* = \sum_{i=1}^{n} (x_i a_{\cdot i})^* = \sum_{i=1}^{n} \overline{x_i} a_{\cdot i}^*,$$

here $a_{\cdot i}^*$ is the dual of $a_{\cdot i}$ in $\mathcal{L}(\mathbb{F}^n, \mathbb{F})$.
(2)Right Assume that the matrix of T^* is $A' \in \mathrm{Mat}(n \times m; \mathbb{F})$, when we take $e_j \in \mathbb{F}^m$,(Sorry using A^* directly may be misleading...)

$$x^* A' e_j = x^* a'_{\cdot j} = \langle x, a'_{\cdot j} \rangle,$$

The Matrix of the Adjoint of a Linear Map

Solution(Continued).

so

$$x^*A' = (\langle x, a'_{\cdot 1}\rangle, ..., \langle x, a'_{\cdot m}\rangle) = \begin{pmatrix} \overline{x_1}a'_{11} & & \overline{x_1}a'_{1m} \\ + & & + \\ \vdots & , ..., & \vdots \\ + & & + \\ \overline{x_n}a'_{n1} & & \overline{x_n}a'_{nm} \end{pmatrix} = \sum_{i=1}^{n} \overline{x_i}a'_{i\cdot}$$

(3)Left equals right,

$$(Ax)^* = \sum_{i=1}^{n} \overline{x_i}a^*_{\cdot i} = x^*A' = \sum_{i=1}^{n} \overline{x_i}a'_{i\cdot}$$

so we have

$$a^*_{\cdot i} = a'_{i\cdot}$$

we will replace A' by A^* afterwards. This means that

The rows of A^* are the dual of columns of A.

Adjoint of a Linear Map

Definition.

Let U, V be vector spaces and $T \in \mathcal{L}(U, V)$, then the adjoint of T denoted $T^* \in \mathcal{L}(V, U)$ is the unique linear map that satisfies,

$$\langle Tu, v \rangle = \langle u, T^* v \rangle$$

for any $u \in U$, $v \in V$.

Comment:

Note that this definition is the same as the previous result,

$$\langle Tu, v \rangle = (Tu)^* v = u^* T^* v = \langle u, T^* v \rangle.$$

Matrix Transpose

Definition.
For $A = (a_{ij}) \in \mathrm{Mat}(m \times n; \mathbb{F})$ we define the ***transpose*** of A by

$$A^T \in \mathrm{Mat}(n \times m; \mathbb{F}), \; A^T = (a_{ji}).$$

Theorem.
The ***adjoint*** of a matrix is

$$A^* \in \mathrm{Mat}(n \times m; \mathbb{F}), \; A^* = \overline{A}^T = (\overline{a_{ji}}).$$

Matrix Transpose

We stated that adjoint is the "dual" of a linear map. However, the dual map is defined as follows.

Definition and Theorem.
If $T \in \mathcal{L}(U, V)$, then the ***dual map*** of T is the linear map $T' \in \mathcal{L}(V^*, U^*)$ defined by $T'(v^*) = v^* \circ T$ for $v^* \in V$.
If $(b_1, ..., b_n)$ is a basis of U and $(c_1, ..., c_m)$ is a basis of V, with

$$\varphi_{A1}(b_1, ..., b_n) = \varphi_{A2}(b_1^*, ..., b_n^*) = (e_1, ..., e_n)$$
$$\varphi_{B1}(c_1, ..., c_m) = \varphi_{B2}(c_1^*, ..., c_m^*) = (e_1, ..., e_m),$$

we have the matrix of T is A and the matrix of T' is A', then

$$A' = A^T.$$

Matrix Transpose

$$
\begin{tikzcd}
\mathbb{R}^n \arrow[d, "A"'] & U \arrow[l, "\varphi_{A1}"'] \arrow[d, "T"'] \arrow[r, dash] & U^* \arrow[r, "\varphi_{A2}"] & \mathbb{R}^n \\
\mathbb{R}^m & V \arrow[l, "\varphi_{B1}"'] \arrow[r, dash] & V^* \arrow[u, "T'"'] \arrow[r, "\varphi_{B2}"] & \mathbb{R}^m \arrow[u, "A'"']
\end{tikzcd}
$$

Proof.

Each column of A' is

$$
\begin{aligned}
A' e_j &= \varphi_{A2} \circ T' \circ \varphi_{B2}^{-1} e_j \\
&= \varphi_{A2} \circ T'(c_j^*) \\
&= \varphi_{A2} \circ (c_j^* \circ T),
\end{aligned}
$$

we use $T'(v^*) = v^* \circ T$ here.

Matrix Transpose

Proof(Continued).

To calculate $c_j^* \circ T$, we notice that

$$c_j^* \circ T = c_j^* \circ \varphi_{B1}^{-1} \circ A \circ \varphi_{A1}.$$

From the left, $c_j^* \circ \varphi_{B1}^{-1} = e_j^*$ since its product with basis are the same. From previous result we know that,

$$e_j^* A = a_{j\cdot}.$$

Then we have

$$A' e_j = \varphi_{A2} \circ a_{j\cdot} \circ \varphi_{A1}.$$

Matrix Transpose

Proof(Continued).

We consider that

$$\begin{aligned} a_{j\cdot} \circ \varphi_{A1} &= \sum_{i=1}^{n} a_{ji} e_i^* \varphi_{A1} \\ &= \sum_{i=1}^{n} a_{ji} b_i^*. \end{aligned}$$

In the last step we used $e_i^* \varphi_{A1} = b_i^*$ similar to $c_j^* \circ \varphi_{B1}^{-1} = e_j^*$.
Finally we have

$$\begin{aligned} A' e_j &= \varphi_{A2} \sum_{i=1}^{n} a_{ji} b_i^* \\ &= \sum_{i=1}^{n} a_{ji} e_i. \end{aligned}$$

We have $A' = (a_{ji}) = A^T$ as the result.

Properties of Adjoint

Theorem.
Let U, V be vector spaces. If $T \in \mathcal{L}(U, V)$ and $T^* \in \mathcal{L}(V, U)$, then
$$(\operatorname{ran} T)^\perp = \ker T^*$$
$$(\operatorname{ran} T^*)^\perp = \ker T.$$

Proof.
Note that $v \in (\operatorname{ran} T)^\perp$ is
$$(\forall u \in U)\langle Tu, v\rangle = 0 \quad \Leftrightarrow \quad (\forall u \in U)\langle u, T^* v\rangle = 0,$$
the second statement is only possible for $T^* v = 0$, then $v \in \ker T^*$.

Properties of Adjoint

Comment:

Note that this theorem can be expressed as,

$$\operatorname{ran} T \oplus \ker T^* = V,$$
$$\operatorname{ran} T^* \oplus \ker T = U.$$

If U, V are finite-dimensional,

$$\dim \operatorname{ran} T + \dim \ker T^* = \dim V$$
$$\dim \operatorname{ran} T^* + \dim \ker T = \dim U.$$

Properties of Adjoint

Motivation.

For vector spaces U, V, we have $T \in \mathcal{L}(U, V)$. Then we can find $\operatorname{ran} T$ and $\ker T$, so we can determine $\operatorname{ran} T^* = (\ker T)^\perp$ and $\ker T^* = (\operatorname{ran} T)^\perp$. Can we express T^* using basis of these subspaces?

$$\operatorname{ran} T^* = \operatorname{span}\{b_1, ..., b_r\} \qquad \operatorname{ran} T = \operatorname{span}\{c_1, ..., c_r\}$$

$$\ker T = \operatorname{span}\{b_{r+1}, ..., b_n\} \qquad \ker T^* = \operatorname{span}\{c_{r+1}, ..., c_m\}$$

If $T(b_i) = c_i$ for $i = 1, ..., r$, then is $T^*(c_i) = b_i$ the adjoint of T?

Properties of Adjoint

Solution.

We only have this result for suitable basis. If $u = \sum_{i=1}^{n} \lambda_i b_i$ and $v = \sum_{j=1}^{m} \mu_j c_j$, then

$$\langle Tu, v \rangle = \langle T \sum_{i=1}^{n} \lambda_i b_i, \sum_{j=1}^{m} \mu_j c_j \rangle = \langle \sum_{i=1}^{r} \lambda_i c_i, \sum_{j=1}^{m} \mu_j c_j \rangle$$

since we have $\text{span}\{c_1, ..., c_r\} \perp \text{span}\{c_{r+1}, ..., c_n\}$.

$$\begin{aligned}\langle Tu, v \rangle &= \langle \sum_{i=1}^{r} \lambda_i c_i, \sum_{j=1}^{r} \mu_j c_j \rangle + \langle \sum_{i=1}^{r} \lambda_i c_i, \sum_{j=r+1}^{m} \mu_j c_j \rangle \\ &= \langle \sum_{i=1}^{r} \lambda_i c_i, \sum_{j=1}^{r} \mu_j c_j \rangle.\end{aligned}$$

Properties of Adjoint

Solution(Continued).

The same result for adjoint

$$\langle u, T^* v\rangle = \langle \sum_{i=1}^{r} \lambda_i b_i, \sum_{j=1}^{r} \mu_j b_j\rangle,$$

so we have

$$\langle \sum_{i=1}^{r} \lambda_i c_i, \sum_{j=1}^{r} \mu_j c_j\rangle = \langle \sum_{i=1}^{r} \lambda_i b_i, \sum_{j=1}^{r} \mu_j b_j\rangle.$$

Because λ_i and μ_j are arbitrary, we have

$$\langle b_i, b_j\rangle = \langle c_i, c_j\rangle \qquad i, j = 1, ..., r.$$

With basis that satisfies this equation, we can have for $T(b_i) = c_i$ the adjoint is $T^*(c_i) = b_i$.

Second Way to Understand Matrix times Vector

Recall that a matrix times a vector has results look like inner product, and we explain this using adjoint. For $A \in \mathrm{Mat}(m \times n; \mathbb{F})$, $x \in \mathbb{F}^n$,

$$Ax = (x^* A^*)^* = (\langle x, a^*_{\cdot 1} \rangle, \dots, \langle x, a^*_{\cdot m} \rangle)^*,$$

here $a^*_{\cdot i}$ is the ith column of A^*.

$$Ax = \begin{pmatrix} \overline{\langle x, a^*_{\cdot 1} \rangle} \\ \vdots \\ \overline{\langle x, a^*_{\cdot m} \rangle} \end{pmatrix} = \begin{pmatrix} \langle a^*_{\cdot 1}, x \rangle \\ \vdots \\ \langle a^*_{\cdot m}, x \rangle \end{pmatrix} = \begin{pmatrix} a_{1\cdot} x \\ \vdots \\ a_{m\cdot} x \end{pmatrix}$$

We have proved that the rows of A^* are the dual of columns of A, then the rows of A are the dual of the columns of A^*.
In the last step, we use

$$a_{i\cdot} = \langle a^*_{\cdot i}, \cdot \rangle,$$

the final result is matrix product between $a_{i\cdot}$ and x.

Dimension of $U \times V$ and $\mathcal{L}(U, V)$

Theorem.

Let U, V be real or complex finite-dimensional vector spaces. For $\dim U = n$ and $\dim V = m$,

$$\dim(U \times V) = n + m$$

Note that $\mathcal{L}(U, V) = n \times m$ since matrix has $n \times m$ independent entries.

Proof.

For a basis $(b_1, \dots, b_n)$ in U and a basis $(c_1, \dots, c_m)$ in V, we find a basis in $U \times V$ by

$$((b_1, 0), \dots, (b_n, 0), (0, c_1), \dots, (0, c_m))$$

Note that for $u \in U$, $u = \sum \lambda_k b_k$ and for $v \in V$, $v = \sum \mu_j c_j$, so

$$(u, v) = \sum \lambda_k (b_k, 0) + \sum \mu_j (0, c_j)$$

is the unique representation, $\dim(U \times V) = n + m$.

Block Matrix

Motivation.

If we have two linear maps $L_1 \in \mathcal{L}(U, W)$ and $L_2 \in \mathcal{L}(V, W)$, can we define $L \in \mathcal{L}(U \times V, W)$?
In other way, if we have $L_1 \in \mathcal{L}(U, V)$ and $L_2 \in \mathcal{L}(U, W)$, can we define $L \in \mathcal{L}(U, V \times W)$?

Solution.

If we have the matrix of L_1, L_2 are A, B, and the matrix of L is C.
Then in the first situation,

$$C = \begin{pmatrix} A & B \end{pmatrix}.$$

In the second situation,

$$C = \begin{pmatrix} A \\ B \end{pmatrix}.$$

This method gives rise to block matrix.

Block Matrix

Properties.

From definition, we can simplify the result when matrices times from left,

$$T \begin{pmatrix} A & B \end{pmatrix} = \begin{pmatrix} TA & TB \end{pmatrix}.$$

Then from right,

$$\begin{pmatrix} A \\ B \end{pmatrix} T = \begin{pmatrix} AT \\ BT \end{pmatrix}.$$

Elementary Matrix Manipulations

When we apply Gauss-Jordan Algorithm to solve a linear system, we use three types of linear maps in matrix form.

Definition.

For a linear system $Ax = b$ with $b \in \mathbb{F}^m$ and $A \in \mathrm{Mat}(m \times n; \mathbb{F})$, we have three types of elementary matrices:

1. Row switching.
$$a_{i\cdot} \leftrightarrow a_{j\cdot},$$
2. Row multiplications.
$$ka_{i\cdot} \to a_{i\cdot}, \text{ where} k \neq 0,$$
3. Row addition.
$$a_{i\cdot} + ka_{j\cdot} \to a_{i\cdot}.$$

Comment:

Note that $\mathrm{ran}A^*$ does not change, then $\ker A$ which is $(\mathrm{ran}A^*)^\perp$ does not change.

Inverse of a Matrix

If $A \in \text{Mat}(n \times n; \mathbb{F})$, we find the inverse map of A by solving linear system $Ax = b$.

1.5.9. Definition.

A matrix $A \in \text{Mat}(n \times n; \mathbb{F})$ is called ***invertible*** if there exists some $B \in \text{Mat}(n\times)$ such that

$$AB = BA = \text{id} = \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix}.$$

We then write $B = A^{-1}$ and say that A^{-1} is the ***inverse*** of A.

Comment:

If $BA = \text{id}$, then

$$AB = A \cdot \text{id} \cdot B = A(BA)B = (AB)(AB),$$

so $AB = \text{id}$.

Find the Inverse

Gauss-Jordan Algorithm

For $A \in \mathrm{Mat}(m \times n; \mathbb{F})$, we write it in juxtaposition with the identity matrix and apply the Gauss-Jordan Algorithm to them at the same time, then

$$B \begin{pmatrix} A & \mathrm{id} \end{pmatrix} = \begin{pmatrix} \mathrm{id} & \mathrm{B} \end{pmatrix}.$$

So we get B is A^{-1}. Here B is the multiplication of all elementary matrices.

Inverse Maps

Recall that to express a linear map $L \in \mathcal{L}(U, V)$ with $\dim U = n, \dim V = m$, we need two bijections $\varphi_A \in \mathcal{L}(U, \mathbb{F}^n)$ and $\varphi_B \in \mathcal{L}(V, \mathbb{F}^m)$. Then we can determine the matrix $A \in \mathrm{Mat}(m \times n; \mathbb{F})$ such that

$$L = \varphi_B \circ A \circ \varphi_A.$$

Solution.

To express the inverse of isomorphism $L \in \mathcal{L}(U, V)$ with $\dim U = \dim V = n$, we have

$$L^{-1} = \varphi_A^{-1} \circ A^{-1} \circ \varphi_B^{-1},$$

where A^{-1} is the inverse of A.

Inverse Maps

Examples.

For $L : \mathcal{P}_2 \mapsto \mathcal{P}_2$, such that

$$ax^2 + bx + c \mapsto \frac{a+b+c}{3}x^2 + \frac{a+b}{2}x + \frac{a-c}{2}.$$

with

$$\varphi_A : \mathcal{P}_2 \mapsto \mathbb{F}^3, \qquad \varphi_A(1, x, x^2) = (e_1, e_2, e_3)$$
$$\varphi_B : \mathcal{P}_2 \mapsto \mathbb{F}^3, \qquad \varphi_B(1, x, x^2) = (e_1, e_2, e_3)$$

find L^{-1} with $\varphi_A^{-1}, \varphi_B^{-1}$.

Comment:

Different φ_A, φ_B will not change the expression of L^{-1}.

Linear Maps - Active and Passive Points of View

We only consider $2D$ case as in class, take $T(e_1, e_2) = (b_1, b_2)$.

Active.

When we apply T to $x \in \mathbb{R}^2$, we find

$$Tx = T(x_1 e_1 + x_2 e_2) = x_1 b_1 + x_2 b_2.$$

Passive.

For $x \in \mathbb{R}^2$,

$$x = T^{-1} Tx = T^{-1}(x_1 b_1 + x_2 b_2),$$

so T^{-1} acts on it without changing it.

Change of Basis

We need to express x using basis $(e'_1, ..., e'_n)$ if we have known the standard representation.

Active.

We write down the matrix of T such that $T(e_1, ..., e_n) = (e'_1, ..., e'_n)$. We have two representations of x,

$$x = \sum_{i=1}^{n} x_i e_i = \sum_{i=1}^{n} x'_i e'_i,$$

then apply T^{-1} on both sides,

$$T^{-1} \sum_{i=1}^{n} x_i e_i = \sum_{i=1}^{n} x'_i T^{-1} e'_i = \sum_{i=1}^{n} x'_i e_i,$$

so $T^{-1}x$ has coordinate coefficients equal to x'_i.

Comment:

We solve a linear system here, we denote $x' = \sum_{i=1}^{n} x'_i e'_i$, then $x = Tx'$, so $x' = T^{-1}x$.

Change of Basis

We applied linear map and found another vector, but we can find the matrix of basis change directly as well.

Passive.

We consider $\mathrm{id} : \mathbb{R}^2 \to \mathbb{R}^2$, take

$$\varphi_A(e_1, e_2) = (e_1, e_2)$$
$$\varphi_B(e_1', e_2') = (e_1, e_2)$$

Then $x' = Ax$ where $x' = x_1' e_1 + x_2' e_2$ and $x = x_1 e_1 + x_2 e_2$.

$$\begin{tikzcd} \mathbb{R}^2 \arrow[r, "\mathrm{id}"] \arrow[d, "\varphi_A"'] & \mathbb{R}^2 \arrow[d, "\varphi_B"] \\ \mathbb{R}^2 \arrow[r, "A"] & \mathbb{R}^2 \end{tikzcd} \qquad A \circ \varphi_A = \varphi_B \circ \mathrm{id}$$

Note that $\varphi_A = \mathrm{id}, \varphi_B = T^{-1}$, then $A = T^{-1}$.

Reflection in $\mathbb{R}^2$

The reflection matrix with respect to x-axis is

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

Example.

Find the matrix of the linear map $L : \mathbb{R}^2 \to \mathbb{R}^2$ such that it takes reflection of a vector with respect to the line through

$$y = \begin{pmatrix} 1 \\ 2 \end{pmatrix}.$$

Reflection in $\mathbb{R}^2$

Solution.

We consider the basis

$$b_1 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, b_2 = \begin{pmatrix} 2 \\ -1 \end{pmatrix}.$$

If $x = \lambda_1 b_1 + \lambda_2 b_2$, then

$$Lx = \lambda_1 b_1 - \lambda_2 b_2,$$

so we define $T(e_1, e_2) = (b_1, b_2)$, then

$$L = TAT^{-1},$$

we use basis change in previous slide.

Structure of Solution Sets

Motivation.

For linear system $Ax = b$, the solution set S is not a subspace, so how to express S?

Sloution.

Note that if $x_1, x_2 \in S$, then $A(x_2 - x_1) = 0$, and $x_2 - x_1 \in \mathrm{ker}A$.
If $Ax_p = b$, then

$$x_p + \mathrm{ker}A := \{x \in \mathrm{ker}A : x_p + x\}$$

is the solution set S. We have to find two things to solve this system, a particular solution x_p, a basis of $\mathrm{ker}A$.

Introduction to Linear Algebra

Systems of Linear Equations

Finite-Dimensional Vector Spaces

Inner Product Spaces

Linear Maps

Matrices

Theory of Systems of Linear Equations

Determines

Solution Sets

Defitions.

Solution Set For a linear systems of equations $Ax = b$, the ***solution set*** is

$$\text{Sol(A,b)} = \{x \in \mathbb{R}^n : Ax = b\}.$$

Particular Solution If $x_0 \in \mathbb{R}^n$ satisfies

$$Ax_0 = b,$$

then we say that x_0 is a ***particular solution*** of $Ax = b$.

Associated homogeneous solution set

$$\text{Sol}(A, 0) = \{x \in \mathbb{R}^n : Ax = 0\} = \ker A.$$

Structure of Solution Set

A very important, fundamental result states:
The solution set of $Ax = b$ is the sum of the homogeneous solution set and a particular solution.

1.6.1. Lemma.
Let $x_0 \in \mathbb{R}^n$ be a particular solution of $Ax = b$. Then

$$\mathrm{Sol}(A, b) = \{x_0\} + \mathrm{ker} A = \{y \in \mathbb{R}^n : y = x_0 + x, x \in \mathrm{ker} A\}.$$

Proof.
Note that for $x \in \mathbb{R}^n$, we have

$$x \in \mathrm{Sol}(A, b) \Leftrightarrow x - x_0 \in \mathrm{ker} A.$$

Solvability of Systems of Equations

1.6.2. Corollary

If x_0 is a particular solution of $Ax = b$ and $\{v_1, ..., v_r\}$ is a basis of $\ker A$, then

$$\mathrm{Sol}(A, b) = \{x \in \mathbb{R}^n : x = x_0 + \lambda_1 v_1 + \cdots + \lambda_r v_r : \lambda_1, ..., \lambda_r \in \mathbb{R}\}.$$

Here $r = \dim \ker A$.

1.6.3. Corollary.

If we have a solution to the linear system of equations $Ax = b$, this solution is unique if and only if $\ker A = \{0\}$.

Proof. We have

$$\mathrm{Sol}(A, b) = \{x_0\} \Leftrightarrow \ker A = \{0\}.$$

Solvability of Systems of Equations

This gives rise to a further, fundamentally important result:

1.6.4. Fredholm Alternative.
Let $A \in \mathrm{Mat}(n \times n; \mathbb{R}^n)$. Then

- either $Ax = b$ has a unique solution for any $b \in \mathbb{R}^n$
- or $Ax = 0$ has a non-trivial solution.

Proof.
If A is bijective, then $x = A^{-1}b$;
If A is not bijective, it is not injective, then $\ker A \neq \{0\}$.

Comment:
This result is only for $A \in \mathrm{Mat}(n \times n)$. For $A \in \mathrm{Mat}(n \times m)$, A is injective is equivalent to A is surjective.

Matrix Rank

1.6.5. Definition.

Let $A \in \operatorname{Mat}(m \times n; \mathbb{F})$ be a matrix columns $a_{\cdot j}$, $1 \le j \le n$, and rows $a_{i \cdot} \in \mathbb{F}^n$, $1 \le i \le m$. Then we define

- the ***column rank*** of A to be

$$\text{column rank } A := \dim \operatorname{span}\{a_{\cdot 1}, \dots, a_{\cdot n}\}$$

- and the ***row rank*** of A to be

$$\text{row rank } A := \dim \operatorname{span}\{a_{1 \cdot}, \dots, a_{m \cdot}\}.$$

1.6.6. Remarks.

- $\text{row rank } A = \text{column rank } A^T$
- $\text{column rank } A = \dim \operatorname{ran} A$

Matrix Rank

1.6.7. Definition and Theorem.

Let $A \in \mathrm{Mat}(m \times n; \mathbb{F})$. Then the column rank is equal to the row rank and we define the ***rank*** of A by

$$\mathrm{rank} A := \text{column rank} A = \text{row rank} A.$$

Proof.

Recall that $\overline{A}^T = A^*$, we assume that $\dim \mathrm{ran} \overline{A}^T = \dim \mathrm{ran} A^T$ and $\dim \ker \overline{A}^T = \dim \ker A^T$ here, then

$$\dim \mathrm{ran} A + \dim \ker A^T = m$$

$$\dim \mathrm{ran} A^T + \dim \ker A^T = m.$$

We subtract these two equations and have

$$\dim \mathrm{ran} A = \dim \mathrm{ran} A^T,$$

so row rank$A$ = column rank$A$.

Matrix Rank

Questions:

Why $\dim \operatorname{ran} A^T = \dim \operatorname{ran} \overline{A}^T$?

Answer:

If we have $(b_1, ..., b_r)$ is a basis of $\operatorname{ran} A^T$, we state that $(\overline{b_1}, ..., \overline{b_r})$ is basis of $\operatorname{ran} \overline{A}^T$. For $x \in \operatorname{ran} \overline{A}^T$,

$$x = \sum_{i=1}^{m} \lambda_i \overline{a_{i\cdot}}^T,$$

then $\overline{x} \in \operatorname{ran} A^T$ and we have

$$\overline{x} = \sum_{i=1}^{m} \overline{\lambda_i} a_{i\cdot}^T = \sum_{j=1}^{r} \mu_j b_j.$$

Answer(Continued).

Note that we have μ_j are unique since b_j are basis, then

$$x = \sum_{j=1}^{r} \overline{\mu_j}\,\overline{b_j}$$

is the unique expression with $\overline{b_j}$, here we use the fact that $\overline{(\cdot)}$ is a bijection from $\mathbb{C}$ to $\mathbb{C}$. This is true for $x \in \operatorname{ran}\overline{A}^T$, then $(\overline{b_1}, ..., \overline{b_r})$ is a basis.

Comment:

Note that $\overline{(\cdot)}$ is not an isomorphism, but $(\overline{b_1}, ..., \overline{b_r})$ is still a basis.

Existence of Solutions

1.6.8. Theorem.

There exists a solution for $Ax = b$ if and only if $\operatorname{rank} A = \operatorname{ran}(A|b)$, where

$$(A|b) = \begin{pmatrix} a_{11} & \cdots & a_{1n} & b_1 \\ \vdots & & \vdots & \vdots \\ a_{m1} & \cdots & a_{mn} & b_m \end{pmatrix} \in \operatorname{Mat}((n+1) \times m).$$

Proof.

If $Ax = b$ has a solution we have

$$b \in \operatorname{ran} A \Leftrightarrow \operatorname{ran} A = \{b\} + \operatorname{ran} A.$$

Note that $\operatorname{rank} A = \dim \operatorname{ran} A$ and $\operatorname{rank}(A|b) = \dim\left(\{b\} + \operatorname{ran} A\right)$ we have

$$\operatorname{rank} A = \operatorname{rank}(A|b).$$

Introduction to Linear Algebra

The Determinant in $\mathbb{R}^2$

Motivation.

We want to find the area of a parallelogram spanned by a and b,

$$P(a,b) = \left\{ x \in \mathbb{R}^2 : x = \lambda a + \mu b, 0 \leq \lambda \leq 1, 0 \leq \mu \leq 1 \right\}.$$

Solution.

This map is a bilinear map,

$$\det : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$$

with properties,

1. $\det(e_1, e_2) = 1$,
2. $\det(\lambda a, b) = \lambda \det(a, b)$, $\det(a + c, b) = \det(a, b) + \det(c, b)$,
3. $\det(a, a) = 0$.

The Determinant in $\mathbb{R}^2$

Solution(Continued).

The last property is true for vectors in parallel, it can be extended to

$$\det(a, b) = -\det(b, a)$$

So for $a = a_1 e_1 + a_2 e_2$, $b = b_1 e_1 + b_2 e_2$,

$$\begin{aligned}\det(a, b) =& a_1 b_1 \det(e_1, e_1) + a_1 b_2 \det(e_1, e_2) + \\ & a_2 b_1 \det(e_2, e_1) + a_2 b_2 \det(e_2, e_2) \\ =& a_1 b_2 - a_2 b_1.\end{aligned}$$

We then find the area of $P(a, b)$ to be,

$$A(a, b) = |a_1 b_2 - a_2 b_1|.$$

Vector Product in $\mathbb{R}^3$

We then define the vector product in $\mathbb{R}^3$ as,

1. Length, $|a \times b| = A(a, b)$.
2. Direction, $a \times b \perp \text{span}\{a, b\}$.
3. Orientation, $(a, b, a \times b)$ should form a "right-hand system".

We can define a map $\times : \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}^3$. Note that since in $\mathbb{R}^n$ for $n > 3$, $\dim(\text{span}\{a, b\})^\perp \geq 1$, so we cannot define vector product.

The Determinant in $\mathbb{R}^3$

Motivation.

We want to find the area of a parallel epiped spanned by three vectors $a, b, c \in \mathbb{R}^3$.

Solution.

We have the determinant in $\mathbb{R}^3$ as an ***oriented volume***,

$$\det : \mathbb{R}^3 \times \mathbb{R}^3 \times \mathbb{R}^3, \qquad \det(a, b, c) = \langle a \times b, c \rangle.$$

Note that previous three properties are preserved.

The Determinant in $\mathbb{R}^n$

1.7.13. Definition.

A function $f : \mathbb{R}^n \times \cdots \times \mathbb{R}^n \to \mathbb{R}$ is said to be a ***p-multilinear map*** (or ***p-multilinear form***) if f is linear in each entry, i.e.,

$$f(\lambda a_1, a_2, ..., a_p) = \lambda f(a_1, a_2, ..., a_p)$$

and

$$f(a_1 + b, a_2, ..., a_p) = f(a_1, a_2, ..., a_p) + f(b, a_2, ..., a_p)$$

for $b, a_1, ..., a_p \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$ and analogously hold for the other entries.

Definition.

A $p - multilinearform$ is said to be ***alternating*** if $f(a_1, ..., a_p) = 0$ whenever $a_j = a_k$ for any $j \neq k$.

It is said to be ***normed*** if $f(e_1, ..., e_n) = 1$, where $e_1, ..., e_n$ are the standard basis vectors in $\mathbb{R}^n$.

Characterization of Alternating Forms

1.7.14. Lemma.

Let $f : \mathbb{R}^n \times \cdots \times \mathbb{R}^n \to \mathbb{R}$ be a p-multilinear map. Then the following are equivalent:

(i) f is alternating

(ii) Only exchange two entries will change the sign,

$$f(..., a_i, ..., a_j, ...) = -f(..., a_j, ..., a_i, ...)$$

(iii) $f(a_1, ..., a_p) = 0$ if $a_1, ..., a_p$ are linearly dependent.

Comment:

If f is normed, the third one can be

$$f(a_1, ..., a_p) = 0 \text{ if and only if } a_1, ..., a_p \text{ are linearly dependent.}$$

Characterization of Alternating Forms

Proof.

We prove that $(i) \Rightarrow (ii) \Rightarrow (iii) \Rightarrow (i)$,

$(i) \Rightarrow (ii)$ We use similar result as,

$$f(a+b, a+b) = f(a,a) + f(a,b) + f(b,a) + f(b,b)$$

Then

$$0 = f(..., a_i, ..., a_j, ...) + f(..., a_j, ..., a_i, ...).$$

$(ii) \Rightarrow (iii)$ We first show that $(ii) \Rightarrow (i)$.

$$\begin{aligned} & f(..., a_i, ..., a_i, ...) = -f(..., a_i, ..., a_i, ...) \\ \Rightarrow & f(..., a_i, ..., a_i, ...) = 0. \end{aligned}$$

Characterization of Alternating Forms

Proof(Continued).

We can assume that

$$a_p = \sum_{i=1}^{p-1} \lambda_i a_i,$$

then

$$f(a_1, ..., a_p) = \sum_{i=1}^{p-1} \lambda_i f(a_1, ..., a_{p-1}, a_i) = 0.$$

$(iii) \Rightarrow (i)$ Note that $a_i = 1 \cdot a_i$, then

$$f(..., a_i, ..., a_i, ...) = 0$$

since they are dependent.

Expansion of Determinant

For $n \in \mathbb{N}$, $n > 1$, we can expand a alternating n-multilinear form $f : \mathbb{R}^n \times \cdots \times \mathbb{R}^n \cong \mathrm{Mat}(n \times n; \mathbb{R}) \to \mathbb{R}$ as

$$f(a_1, ..., a_n) = \sum_{i=1}^{n^n} a_{1g_i(1)} \cdots a_{ng_i(n)} f(e_{g_i(1)}, ..., e_{g_i(n)}).$$

Here $g_i : \{x_1, ..., x_n\} \to \{x_1, ..., x_n\}$, and there are n^n maps. Since f is alternating, $f(e_{g_i(1)}, ..., e_{g_i(n)}) = 0$ if g_i is not bijective. The entries left are,

$$f(a_1, ..., a_n) = \sum_{i=1}^{n!} a_{1\pi_i(1)} \cdots a_{n\pi_i(n)} f(e_{\pi_i(1)}, ...e_{\pi_i(n)}).$$

Then we introduce the concept permutation.

Permutations as Transpositions

1.7.5. Definition.
The set of all ***permutations of n elemnts***,

$$S_n = \{\pi : \{x_1, ..., x_n\} \to \{x_1, ..., x_n\} : \pi \text{ bijective}\}$$

together with the group operations "composition of maps",
$\pi_1 \circ \pi 2(x) = \pi_1(\pi_2(x))$ is called the ***symmetric group***.

1.7.6. Definition.
A permutation in S_n that leaves exactly $n - 2$ elements invariant is called a ***transposition***.

Comment:
Note that the number of all permutations of n elements is $n!$.

Permuatations as Transpositions

1.7.7. Lemma.

Every permutation $\pi \in S_n$, $n \geq 2$, is a composition of transpositions, $\pi = \tau_1 \circ \cdots \circ \tau_k$.
Note that τ_j and k are ***not*** unique.

Proof.

We use induction here. First for $n = 2$, $\pi \in S_2$ is transposition.
Then assume we have for $\tilde{\pi} \in S_n$, $\tilde{\pi}$ can be decomposed as transpositions.
For $\pi \in S_{n+1}$, we consider τ that interchanges $n+1$ and $\pi(n+1)$, so

$$\tilde{\pi} = \tau \circ \pi \in S_n.$$

Here because $\tilde{\pi}(n+1) = n+1$, we can change its domain to $\{1, ..., n\}$. Note that $\tau \circ \tau = \mathrm{id}$,

$$\pi = \tau \circ \tilde{\pi}.$$

We proved that π is still product of transpositions.

Sign of a Permutation

We have proved that a permutation is a product of transpositions, but the number of transpositions can either be odd or even.

1.7.8. Definition and Theorem.

Let $\pi \in S_n$ be represented as a composition of k transpositions, $\pi = \tau_1 \circ \cdots \circ \tau_k$. Then the ***sign*** of π,

$$\mathrm{sgn}\pi := (-1)^k$$

does not depend on the representation chosen.

We have for normed n-multilinear form f,

$$f(e_{\pi(1)}, ..., e_{\pi(n)}) = \mathrm{sgn}\pi.$$

Group Actions

1.7.9. Definition.

Let $(G, \circ)$ be a group and X a set. Then an ***action (or operation) of G on X from left*** is a map

$$\Phi : G \times X \to X, \qquad (g, x) \mapsto \Phi(g, x).$$

We often write $\Phi(g, x) = gx$. It has properties,

(i) $ex = x$ ($e \in G$ is the unit element),

(ii) $(a \circ b)x = a(bx)$ for $a, b \in G$, $x \in X$.

We say that G acts (operates) on X.

1.7.10. Proposition.

Let X be the set of all maps $f : \mathbb{R}^n \times \cdots \times \mathbb{R}^n \to \mathbb{R}$. Then S_n acts on X via

$$(\pi f)(x_1, \ldots, x_n) = f(x_{\pi(1)}, \ldots, x_{\pi(n)}), \qquad \pi \in S_n.$$

Group Actions

Proof.

This is a group action if we rewrite it as

$$(\pi f) = f \circ \pi,$$

then we have

$$ef = f \circ e = f, \; (\sigma \circ \pi)f = f \circ \sigma \circ \pi = \sigma(\pi f).$$

1.7.11. Lemma.

Denote by $\Delta : \mathbb{R}^n \to \mathbb{R}$ the function

$$\Delta(x_1, \ldots, x_n) = \prod_{i<j}(x_j - x_i).$$

Then

$$\tau\Delta = -\Delta \qquad \text{for any transposition } \tau \in S_n.$$

Group Actions

Proof.

We consider τ that interchanges x_r and x_s, and can assume that $r < s$. For the term that x_r, x_s appear both,

$$\tau(x_r - x_s) = -(x_r - x_s)$$

For other terms that x_r or x_s appear,

- $j < r : (x_r - x_j)(x_s - x_j)$
- $r < j < s : (x_s - x_j)(x_j - x_r)$
- $s < j : (x_j - x_s)(x_j - x_r)$

the sign does not change. The result is $\tau\Delta = -\Delta$.

Sign of a Permutation

1.7.12. Corollary.

For every permutation $\pi = \tau_1 \circ \cdots \circ \tau_k \in S_n$,

$$\pi\Delta = (\tau_1 \circ \cdots \circ \tau_k)\Delta = (-1)^k \Delta.$$

In particular,

$$\mathrm{sgn}\pi = (-1)^k,$$

does not depend on the decomposition of π into transpositions and is therefore well-defined.

Proof.

The right hand can be either Δ or $-\Delta$. This is determined by $\pi\Delta$ and not decomposition, so $\mathrm{sgn}\pi$ is well-defined.

1.7.15. Theorem.

For every $n \in \mathbb{N}$, $n > 1$, there exists a unique, normed, alternating n-multilinear form $\det : \mathbb{R}^n \times \cdots \times \mathbb{R}^n \cong \mathrm{Mat}(n \times n; \mathbb{R}) \to \mathbb{R}$. Furthermore,

$$\det(a_1, ..., a_n) = \det A = \sum_{\pi \in S_n} \operatorname{sgn}\pi \ a_{1\pi(1)} \cdots a_{n\pi(n)}.$$

Proof.

We have proved that we can expand $\det(a_1, ..., a_n)$ by multilinear,

$$\det(a_1, ..., a_n) = \sum_{\pi \in S_n} a_{1\pi(1)} \cdots a_{n\pi(n)} f(e_{\pi(1)}, ..., e_{\pi(n)}).$$

Here we use $\sum_{\pi \in S_n}$ instead of $\sum_{i=1}^{n!}$. We have also shown that $\operatorname{sgn}\pi = f(e_{\pi(1)}, ..., e_{\pi(n)})$ is well-defined. Then we will have

$$\det(a_1, ..., a_n) = \sum_{\pi \in S_n} \operatorname{sgn}\pi \ a_{1\pi(1)} \cdots a_{n\pi(n)}.$$

Determinants of Transposed Matrices

1.7.16. Lemma.

Let $A \in \mathrm{Mat}(n \times n; \mathbb{R})$. Then

$$\det A = \det A^T.$$

Proof.

We notice that for $\pi \in S_n$, we can relabel $a_{1\pi(1)} \cdots a_{n\pi(n)}$ with second term in order,

$$a_{1\pi(1)} \cdots a_{n\pi(n)} = a_{\pi^{-1}(1)1} \cdots a_{\pi^{-1}(n)n}.$$

Note that π is bijective, so its inverse exists. Also since for transpositions $\tau \circ \tau = \mathrm{id}$, the inverse of $\pi = \tau_1 \circ \cdots \circ \tau_n$ is

$$\pi^{-1} = \tau_k \circ \cdots \circ \tau_1,$$

then $\mathrm{sgn}\pi = \mathrm{sgn}\pi^{-1}$.

Determinants of Transposed Matrices

Proof(Continued).

Since S_n is a group, it are the same to sum over π or π^{-1}.

$$\begin{aligned}
\det A &= \sum_{\pi \in S_n} \operatorname{sgn}\pi \; a_{1\pi(1)} \cdots a_{n\pi(n)} \\
&= \sum_{\pi^{-1} \in S_n} \operatorname{sgn}\pi^{-1} \; a_{\pi^{-1}(1)1} \cdots a_{\pi^{-1}(n)n} \\
&= \sum_{\pi \in S^n} \operatorname{sgn}\pi \; a_{\pi(1)1} \cdots a_{\pi(n)n} \\
&= \det A^T.
\end{aligned}$$

Comment:

We prove that two ways of picking elements are equal, each one in each column and each one in each row.

Leibnitz Formula and ELementary Row Operations

1.7.17. Leibnitz Formula.

$$\det A = \sum_{\pi \in S_n} \mathrm{sgn}\pi \; a_{1\pi(1)} \cdots a_{n\pi(n)}.$$

1.7.18. Corollary.

For matrix A, P is the matrix of elementary row manipulation and P^T is the corresponding column manipulation, then

$$\det(PA) = \det(AP^T).$$

Proof.

We first note that column manipulation will change $\det A$ by a constant, that is $\det(\cdot P^T) = \lambda_P \det(\cdot)$. Then

$$\det(PA) = \det(A^T P^T) = \lambda_P \det A^T = \lambda_P \det A = \det(AP^T).$$

Elementary Row Operations

We have proved that elementary row operations have the same effect as column operations, then we have

- Row switching.

$$\det(\cdot) \to -\det(\cdot),$$

- Row multiplications.

$$\det(\cdot) \to k\det(\cdot),$$

- Row addition. Determinant does not change.

Determinants of Triangular Matrices

1.7.19. Proposition.

Let $A \in \mathrm{Mat}(n \times n; \mathbb{R})$ has upper triangular form, i.e.,

$$A = \begin{pmatrix} \lambda_1 & & * \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}$$

for diagonal elements $\lambda_1, \dots, \lambda_n \in \mathbb{R}$ and arbitrary values (denoted by $*$) above the diagonal. Then

$$\det A = \lambda_1 \cdots \lambda_n.$$

Proof.

For all $\pi \in S_n$, if $\pi(1) \neq 1$, $a_{1\pi} = 0$, then we need $\pi(1) = 1$. Because π is bijective and $\pi(1) = 1$, we need $\pi(2) = 2$. By induction we can only choose $\pi = \mathrm{id}$ and $\operatorname{sgn} \mathrm{id} = 1$. Then we have the result as above.

1.7.20. Proposition.

A matrix $A \in \mathrm{Mat}(n \times n; \mathbb{R})$ is invertible if and only if $\det A \neq 0$.

Proof.

($\Leftarrow$) We proved that if $(a_1, ..., a_n)$ is dependent then $\det(a_1, ..., a_n) = 0$, then if $\det A \neq 0$, we have A is invertible.
($\Rightarrow$) We denote the inverse of A as A^{-1}, then

$$\det \mathrm{id} = \det(A^{-1}A) = \sum_{\pi \in S_n} \mathrm{sgn}\pi \; a^{-1}_{1\pi(1)} \cdots a^{-1}_{n\pi(n)} \det(a_1, ..., a_n) \neq 0.$$

So we need to have $\det A \neq 0$.

Determinants and Systems of Equations

1.7.21. Fredholm Alternative.

Let $A \in \mathrm{Mat}(n \times n; \mathbb{R})$. Then either

- $\det A = 0$, in which $Ax = 0$ has a non-zero solution, order
- $\det A \neq 0$, then $Ax = b$ has a unique solution $x = A^{-1}b$ for any $b \in \mathbb{R}^n$.

We just rewrite Fredholm Alternative here.

1.7.22. Cramer's Rule.

Let $A = (a_1, ..., a_n) \in \mathrm{Mat}(n \times n; \mathbb{R})$, $a_1, ..., a_n \in \mathbb{R}^n$, be invertible. Then the system $Ax = b$, $b \in \mathbb{R}^n$, has the solution

$$x_i = \frac{1}{\det A} \det(a_1, ..., a_{i-1}, b, a_{i+1}, ..., a_n), \qquad i = 1, ..., n.$$

Determinants and Systems fo Euqations

Proof.

We note that $Ax = \sum_{k=1}^{n} x_k a_k$ for $A = (a_1, ..., a_n) \in \mathrm{Mat}(n \times n; \mathbb{R})$.

Therefore,

$$\begin{aligned}
&\det(a_1, ..., a_{i-1}, b, a_{i+1}, ..., a_n)\\
=&\det(a_1, ..., a_{i-1}, Ax, a_{i+1}, ..., a_n)\\
=&\det(a_1, ..., a_{i-1}, \sum_{k=1}^{n} x_k a_k, a_{i+1}, ..., a_n)\\
=&\sum_{k=1}^{n} x_k \det(a_1, ..., a_{i-1}, a_k, a_{i+1}, ..., a_n)\\
=&x_i \det(a_1, ..., a_{i-1}, a_i, a_{i+1}, ..., a_n) + 0\\
=&x_i \det A.
\end{aligned}$$

Minors and Cofactors

1.7.23. Definition.

Let $A = (a_{ij}) \in \mathrm{Mat}(n \times n; \mathbb{R})$. Denote the $(n-1) \times (n-1)$ matrix obtained from A by deleting the ith row and jth column by

$$A_{ij} = (a_{kl})_{\substack{1 \le k,l \le n.\\ k \ne i,\, l \ne j}}$$

Then

$$m_{ij} := \det A_{ij}$$

is called the ***(i,j)th minor of A***. The number

$$c_{ij} := (-1)^{i+j} m_{ij} = (-1)^{i+j} \det A_{ij}$$

is called the ***(i,j)th cofactor of A*** and the matrix

$$\mathrm{Cof} A := (c_{ij})_{1 \le i,j \le n}$$

is called the ***cofactor matrix of A***.

1.7.24. Definition.

Let $A = (a_{ij}) \in \mathrm{Mat}(n \times n; \mathbb{R})$. The transpose of the cofactor matrix of A is called the ***adjugate*** of A, denoted by

$$A^\sharp := (\mathrm{Cof}A)^T.$$

1.7.25. Theorem.

Let $A = (a_{ij}) \in \mathrm{Mat}(n \times n; \mathbb{R})$ be invertible. Then

$$A^{-1} = \frac{1}{\det A} A^\sharp.$$

Determinants and Inversion of Matrices

1.7.26. Lemma.

Let $A = (a_1, ..., a_n) \in \mathrm{Mat}(n \times n; \mathbb{R})$ and e_i be the ith standard basis vector in $\mathbb{R}^n$. Then

$$\det(a_1, ..., a_{j-1}, e_i, a_{j+1}, ..., a_n) = (-1)^{i+j} \det A_{ij} = c_{ij}$$

where c_{ij} is the (i, j)th cofactor of A.

Proof.

We need to interchange column $n - j$ times and column $n - i$ times to change,

$$\begin{aligned} &\det(a_1, ..., a_{j-1}, e_i, e_{j+1}, ..., a_n) \\ =&(-1)^{n-i+n-j} \det \begin{pmatrix} A_{ij} & * \\ 0 & 1 \end{pmatrix} \\ =&(-1)^{i+j} \det A_{ij}. \end{aligned}$$

In the last step, the only permutations we can choose are $\pi(n) = n$.

Proof of Therorem 1.7.25.

Let $A^{-1} = (x_1, ..., x_n) = (x_{ij})$ be a matrix of column vectors $x_1, ..., x_n$. The inverse of A satisfies $AA^{-1} = \mathrm{id}$, so the jth column of A^{-1} satisfies $Ax_j = e_j$, $j = 1, ..., n$. By Cramer's rule and Lemma 1.7.26,

$$x_{ij} = \frac{1}{\det A}\det(a_1, ..., a_{i-1}, e_j, a_{i+1}, ..., a_n) = \frac{1}{\det A}(-1)^{i+j}A_{ji}.$$

Laplace Expansion

1.7.27. Laplace Expansion.

For $A \in \mathrm{Mat}(n \times n; \mathbb{R})$ and any $j = 1, \dots, n$ the recursion formula

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

is true.

Comment:

Note that we can expand the determinant of $n \times n$ matrix into n determinants of $(n-1) \times (n-1)$ matrices.
The result is also true for rows,

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

Lapalce Expansion

Proof.

Let $A = (a_1, ..., a_n)$, $a_k \in \mathbb{R}^n$, $k = 1, ..., n$. Then the $j$the column has the representation $a_j = \sum_{i=1}^{n} a_{ij} e_i$ and

$$\begin{aligned}
\det A &= \det(a_1, ..., a_{j-1}, \sum_{i=1}^{n} a_{ij} e_i, a_{j+1}, ..., a_n) \\
&= \sum_{i=1}^{n} a_{ij} \det(a_1, ..., a_{j-1}, e_i, a_{j+1}, ..., a_n) \\
&= \sum_{i=1}^{n} a_{ij} (-1)^{i+j} \det A_{ij},
\end{aligned}$$

where the last inequality follows form Lemma 1.7.26.

Product Rule for Determinants

1.7.28. Proposition.

Let $A, B \in \mathrm{Mat}(n \times n; \mathbb{R})$. Then $\det(AB) = \det A \det B$.

Proof.

Note that the function

$$f(b_1, ..., b_n) = \det(Ab_1, ..., Ab_n)$$

is a n-multilinear map. We can expand it as determinant and have

$$\begin{aligned} f(b_1, ..., b_n) &= f(\sum_{i=1}^{n} a_i b_{1i}, ..., \sum_{i=1}^{n} a_i b_{ni}) \\ &= \sum_{\pi \in S_n} \mathrm{sgn}\pi \ b_{1\pi(1)} \cdots b_{n\pi(n)} \det(a_1, ..., a_n) \\ &= \det B \det A. \end{aligned}$$

Comment: A flaw of the proof in the slides is that $\det A$ can be 0. Note that if A is not bijective, then AB is not bijective, so $\det(AB) = 0$ if $\det A = 0$.